REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

PREDICTION OF SOLAR ENERGY CONTROL SYSTEM BASED ON MACHINE LEARNING METHOD

Hasanain Alaa Mohammed ALMINSHID

Master's Thesis

Supervisor

Assoc. Prof. Dr. Sefer KURNAZ

Istanbul, 2024

PREDICTION OF SOLAR ENERGY CONTROL SYSTEM BASED ON MACHINE LEARNING METHOD

Hasanain Alaa Mohammed ALMINSHID

Electrical and Computer Engineering

Master’s Thesis

ALTINBAŞ UNIVERSITY

2024

The thesis titled C PREDICTION OF SOLAR ENERGY CONTROL SYSTEM BASED ON MACHINE LEARNING METHOD prepared by HASANAIN ALAA MOHAMMED AL-MINSHID and submitted on 07/06/2024 has been **accepted unanimously** for the degree of Master of Eelecrical and Compeuter Engeneering Department.

____

Assoc. Prof. Dr. Sefer KURNAZ

Supervisor

Thesis Defense Committee Members:

Assoc. Prof. Dr. Sefer KURNAZ	Department of Computer Engineering, Altınbaş University	__________________
Asst. Prof. Dr. Abdullahi Abdu IBRAHIM	Department of Computer Engineering, Altınbaş University	__________________
Asst. Prof. Dr. Serdar KARGIN	Department of Biomedical Engineering, İstanbul Arel University	__________________

I hereby declare that this thesis meets all format and submission requirements of a Master's thesis.

I hereby state that all data and material offered in this capstone project were acquired in complete line with ethical standards and scholastic regulations. I further affirm that, in accordance with the aforementioned standards of academic integrity, all borrowed ideas, arguments, its findings, as well as all the sources indicated in the Reference List, have been correctly cited in the text.

Hasanain Alaa Mohammed ALMINSHID

Signature

DEDICATION

I would like to dedicate this thesis to my supervisor Assoc. Prof. Dr. sefer KURNAZ and also dedicate this thesis to my father, mother and for my family and relatives that standing beside me and encouraging me to obtain a master's degree. I also thank all my friends and colleagues who supported me.

ABSTRACT

PREDICTION OF SOLAR ENERGY CONTROL SYSTEM BASED ON MACHINE LEARNING METHOD

ALMINSHID, Hasanain Alaa Mohammed

M.Sc., Electrical and Computer Engineering, Altınbaş University

Supervisor: Assoc. Prof. Dr. Sefer KURNAZ

Date: 06/2024

Pages: 75

Accurate solar energy generation prediction is critical for optimizing energy management and promoting the shift to renewable energy sources. In this research, we use real-world datasets from solar power plants to train several machine learning models to predict solar energy generation. Solar power generating data, weather sensor data, and auxiliary information are among the databases. We anticipate solar energy generation for various time intervals using machine learning methods such as Linear Regression (LR), Random Forest (RF), Decision Tree (DT), Gradient Boosting Regressor (GBR), AdaBoost Regressor (ADBR), and K-Nearest Neighbours (KNN). Our approach entails training these models on historical data and assessing their performance on test datasets. The findings of this study suggest that machine learning models have promising prediction skills. The R-squared (R2) values for the training and testing datasets demonstrate a good capacity to explain variation in solar energy generation. The DT model, for example, achieved a flawless R2 score of 100% for training datasets and 99.999% for testing datasets. The GBR model came in second, with R2 values of 99.997% for both training and testing. Furthermore, the Root Mean Squared Error (RMSE) values for the GBR model ranged from 1.87 to 109.53 for the KNN model, demonstrating varying degrees of prediction accuracy. In comparison to previous research investigations, our suggested machine learning models regularly show superior prediction accuracy, with R2 scores reaching 99.99% or higher, compared to the best score of 94.50% in previous studies. This study not only advances solar energy prediction using machine learning, but it also highlights the potential of these models in optimizing energy management systems.

Keywords: Solar Energy, Machine Learning, Regression Methods, Control System, Solar Energy Generation.

ÖZET

GÜNEŞ ENERJİSİ KONTROL SİSTEMİNİN MAKİNE ÖĞRENME YÖNTEMİNE GÖRE TAHMİN EDİLMESİ

ALMINSHID, Hasanain Alaa Mohammed

Yüksek Lisans, Elektrik ve Bilgisayar Mühendisliği, Altınbaş Üniversitesi

Danışman: Doç. Dr. Sefer KURNAZ

Tarih: 06/2024

Sayfalar: 75

Doğru güneş enerjisi üretimi tahmini, enerji yönetimini optimize etmek ve yenilenebilir enerji kaynaklarına geçişi teşvik etmek için kritik öneme sahiptir. Bu araştırmada, güneş enerjisi üretimini tahmin etmek amacıyla çeşitli makine öğrenimi modellerini eğitmek için güneş enerjisi santrallerinden gerçek dünya veri kümelerini kullanıyoruz. Veritabanları arasında güneş enerjisi üreten veriler, hava durumu sensörü verileri ve yardımcı bilgiler yer almaktadır. Doğrusal Regresyon (LR), Rastgele Orman (RF), Karar Ağacı (DT), Gradient Boosting Regressor (GBR), AdaBoost Regressor (ADBR) ve K-Nearest gibi makine öğrenme yöntemlerini kullanarak çeşitli zaman aralıklarında güneş enerjisi üretimini öngörüyoruz. Komşular (KNN). Yaklaşımımız, bu modellerin geçmiş veriler üzerinde eğitilmesini ve performanslarının test veri kümeleri üzerinde değerlendirilmesini gerektirir. Bu çalışmanın bulguları, makine öğrenimi modellerinin umut verici tahmin becerilerine sahip olduğunu göstermektedir. Eğitim ve test veri kümelerinin R-kare (R2) değerleri, güneş enerjisi üretimindeki değişimi açıklamak için iyi bir kapasite göstermektedir. Örneğin DT modeli, eğitim veri kümeleri için %100 ve veri kümelerini test etmek için %99,999'luk kusursuz bir R2 puanı elde etti. GBR modeli, hem eğitim hem de test için %99,997'lik R2 değerleriyle ikinci sırada yer aldı. Ayrıca, GBR modeli için Ortalama Karekök Hata (RMSE) değerleri, KNN modeli için 1,87 ile 109,53 arasında değişiyordu ve bu da tahmin doğruluğunun değişen derecelerini gösteriyordu. Önceki araştırma araştırmalarıyla karşılaştırıldığında, önerilen makine öğrenimi modellerimiz düzenli olarak üstün tahmin doğruluğu göstermektedir; önceki çalışmalardaki en iyi puan olan %94,50'ye kıyasla R2

puanları %99,99 veya daha yükseğe ulaşmaktadır. Bu çalışma yalnızca makine öğrenimini kullanarak güneş enerjisi tahminini geliştirmekle kalmıyor, aynı zamanda bu modellerin enerji yönetimi sistemlerini optimize etme potansiyelini de vurguluyor.

Anahtar Kelimeler: Güneş Enerjisi, Makine Öğrenmesi, Regresyon Yöntemleri, Kontrol Sistemi, Güneş Enerjisi Üretimi.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

ABBREVIATIONS

AC	:	Alternative Current
DC	:	Direct Current
KNN	:	K-Nearest Neighbour
ML	:	Machine Learning
RF	:	Random Forest
RMSE	:	Root Mean Squared Error
SL	:	Solar Energy
SL	:	Supervised Learning
USL	:	Unsupervised Learning

1. INTRODUCTION

1.1 BACKGROUND

Rapid expansion of renewable energy sources, such as power plants, has been investigated for the past decade [1]. The expansion of our capacity to generate clean, inexpensive energy is anticipated to stimulate the economy. As a result, there has been a lot of focus on the problems associated with producing solar energy recently. Finding and pinpointing aberrant patterns in the solar systems is a top priority. Such irregularities on photovoltaic (PV) components can be detected and avoided with the help of big data and data-driven approaches. Convolutional neural networks, a key component of deep learning systems, have been shown to be effective in the implementation of machine intelligence [2][3].

Solar photovoltaic PV systems require cutting-edge monitoring tools to track the dynamic evolution of system parameters and notify decision-makers of any unexpected changes. In order to help plant operators better manage their plants and ensure economic integration into smart grids, online monitoring of PV systems is technically useful [4]. Reduced output and increased risk of fire can result from a failure to detect catastrophic failures in PV arrays [5]. If panel owners are aware of anomalies as soon as they arise on the surface of solar panels, they can eradicate them to prevent a further decrease in output [6]. Consequently, enhancing the performance, reliability, and safety of PV plants necessitates the use of technologies for anomaly detection that are both fast and accurate.

Various sorts of abnormalities generally lead to insufficient operation of PV schemes. These irregularities might originate from within or outside the body [7]. The PV system experiences zero production during the day due to an internal fault. Component failure, isolation, inverter shutdown, shade, and inverter maximum power point are frequent issues [8]. The PV's power generation is hampered by factors outside of its control. External abnormalities such as shading, humidity, dust, and temperature are thought to have a considerable impact on PV system output.

The initial cost and Return on Investment are substantial challenges for large-scale power generation with PV systems, mostly due to the unpredictability of the supply of solar energy due to changing weather patterns. The amount of energy received from the Sun per unit of area as electromagnetic radiation within the wavelength range of the measuring device is known as solar irradiance [9]. Since the amount of energy generated from the sun fluctuates with the seasons and other environmental factors, solar irradiation plays a crucial role in solar power generation. To meet the needs of its customers, the power provider must contract with a third party, a firm that generates electricity through the usage of fossil fuels. In light of this intermittent scarcity, it is crucial to comprehend and foresee the output fluctuation [10].

Machine Learning (ML) methods can be used to predict solar power output in addition to the more standard time series model [11]. Prediction using ML is becoming increasingly popular as it can be applied to a wide variety of fields, such as computational statistics, predictive analytics, the healthcare industry, bioinformatics, and behaviourism. When it comes to forecasting sun irradiation and, hence, solar power generation, how do ML methods fare? The relationship between input and output data can be better understood with the use of ML. This feature empowers the machine learning model to address issues of prediction, data mining, spam filtering, and categorization [12]. Once the training data is optimised, it's simple for decision makers to get helpful projections from the trained ML models. ML relies heavily on the quality of the data it uses, which is why pre-processing is so important [13].

ML algorithms, with the aid of training data, sometimes called sample data, construct a numerical model that aids in making predictions or decisions without being explicitly customised [14]. To create predictive models, machine learning combines software engineering with data science. Machine learning constructs or employs the calculations discovered by analysing the data samples. The better the model performs, the more information we feed it.

1.1.1 Global Interest for Solar Power Generation and Forecasting

More than a century ago, the discovery and use of the fuel in home and industrial machinery greatly improved people's standard of living by taking numerous previously unattainable leaps forward. In the short term, consumers reaped the benefits of faster transportation, more convenient electrical and electronic equipment, and a higher industrial output to fulfil rising demand. Rapidly depleting fuel reserves and an exponential increase in fuel demand have resulted in a severe shortage of the product in recent years. This is where we first ran into trouble [15].

The rising demand for fuel meant that fossil fuel reservoirs couldn't keep up, and fuel search is time-consuming and expensive because oil companies need to do extensive surveys, and if signs of the fuel are identified, extensive drilling must be done deep below or in the ocean [16]. It will take several years, not months. Fuel prices have increased due to the imbalance between supply and demand. This has prompted a worldwide search for simple, quickly implemented, and affordably priced alternatives, such as wind, solar, solar thermal, biogas, etc [17].

The situation is becoming increasingly dire as fuel reserves dwindle and the global population continues to grow. Adopting renewable supplies at a lower level is the only feasible answer to both challenges, as it would both fulfil rising energy demands and help save the environment from carbon pollution, protecting the ozone layer in the long term [18][19].

1.2 PROBLEM DEFINITION

When we have a lot of data and that data follows a consistent pattern over a lengthy period of time, we can make very precise predictions. However, if the data format does not exhibit a certain pattern and varies continuously, it becomes impossible to create highly accurate forecast. The primary challenge of solar prediction is that the power output is highly sensitive to prevailing atmospheric conditions at the time in question [20]. Solar energy, for instance, follows a regular cycle. Production of solar energy peaks in the summer and dips in the winter. At the same time, it is possible to make more precise predictions about solar energy production by include weather forecast data.

Predictions made by machine learning models inherently rely on the learned data structure from previously collected data. Accurate estimation requires a high degree of similarity between the features of the historical data and the data to be estimated. However, solar power generation is highly variable due to its reliance on external factors such as weather, seasonal fluctuations, location, time of day, and panel arrangement [21]. This means that forecasting tools cannot provide equally effective outcomes across all locations.

The amount of solar energy produced is rising steadily as more solar power plants are put into operation. Solar power generation may fluctuate with the seasons, but overall, it is on the rise. This is why variations appear during the same time frame each year. Current models do not account for the rising trend of solar power output, thus even if the data structure is determined seasonally, it will be disrupted and high-accuracy predictions will not be possible [22].

The report can be stated as an answer to the problem statement below based on the information presented in the introduction.

In what ways can we use Machine Learning to forecast the short-term production of solar panels?

1.3 PROBLEM SOLUTION

In light of the fact that solar generation exhibits typical behaviour patterns with time-varying, seasonal, and trend patterns, and that the strength of existing models lies in working on stable data, it is essential to shape existing data to a stable form rather than train the models with the rising trend of the data. Since the possible generation may be predicted using mathematical and statistical models based on historical data. For this purpose, data that is prone to fluctuations can be broken down into its trend and seasonal components. As a result, we can derive two stable data from one set of uncertain data. The future value of trend data can be predicted with great accuracy using various approaches appropriate for linear estimate because trend data does not contain any seasonal changes. Since seasonal data tends to be consistent throughout time, its future value can be accurately predicted utilising nonlinear estimate techniques and a variety of machine learning algorithms. The combination of these two forecasts yields a highly reliable final prediction.

1.4 MOTIVATION & OBJECTIVES

Numerous research have investigated various approaches to generation prediction for solar photovoltaic power plants in an effort to identify those most suitable for use in this field. Various machine learning algorithms, including ANN [23], SVR [24], Random Forest [25], KNN [26], and LGBM [27], have been employed to predict solar energy in previous studies and reviews of the relevant literature. Different types of data have been investigated using preexisting algorithms and hybrid models in these works [28]. In any case, these algorithms use the data in its current form or attempt to augment it with additional properties, but they fail to adequately account for the data's underlying trend.

This issue is a time series problem since data on solar power generation is a time series. There are two types of time series data: linear and nonlinear.

When dealing with linear data, linear estimate techniques tend to produce more accurate results, while non-linear techniques tend to be less effective [29]. However, linear approaches cannot reliably forecast complicated data. A review of the relevant literature reveals the need for more accurate solar power forecasting models [30]. The trend problem is addressed and an answer is sought in this thesis. To solve this issue, we suggest various strategies for embedding data trend aspects into machine learning algorithms. In order to circumvent this issue, trend data and current seasonal data are kept in distinct databases. The goal is to use linear estimate techniques to make seasonal forecasts and nonlinear techniques to make forecasts of trend data. In this way, the best features of both approaches can be utilized in the respective data structures.

Improving the accuracy of a forecast for the next day requires resolving the trend problem. As a result, consumers and manufacturers will be able to employ more precise forecasts for day-ahead generation and future decision-making. For instance, small producers may plan to generate or purchase energy from a variety of sources the following day depending on whether or not they anticipate there will be enough generation available. In addition, large producers can anticipate their storage needs for the following day and plan accordingly. Accurate predictions are also required for monitoring, maintenance, dispatching, and scheduling of energy in the solar power community. They can either prepare for future supply by comparing expected and actual solar power generation, or they can troubleshoot existing issues by doing so. Without over- or underestimating the generation from distributed sources, they can be ready for load demand.

1.5 THESIS OUTLINE

The thesis is organised to investigate and evaluate how machine learning methods might be used to estimate solar energy output. In Chapter 2, we begin our trip by doing a thorough Literature Review of the existing literature on the topic of solar energy prediction, machine learning models, and their use in energy forecasting. This part lays the groundwork by exposing the area's lacks, difficulties, and potentials.

The methodology used to foretell solar energy production is laid out in detail in Chapter 3. It describes the data set that was analysed, which included information from solar panels, weather stations, and other sources. This chapter discusses various machine learning models, including Linear Regression (LR), Random Forest (RF), Decision Tree (DT), Gradient Boosting Regressor (GBR), AdaBoost Regressor (ADBR), and K-Nearest Neighbours (KNN), along with their respective configurations. It provides in-depth explanations of data pre-treatment, feature selection, and model training, offering readers a consistent roadmap to understand the study's approach.

The core of the work is cited in the fourth chapter, which deals with Performance Examination. In this chapter, there is given an exhaustive analysis of the results obtained through the application of machine learning models to predict solar energy. A list of metrics, for the assessment of models, including the R-squared (R^2) scores, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE) is compiled. Among other, the highlight of this chapter is the models' prediction abilities showcasing their performance across various scenarios.

The last part of the thesis, Chapter 5 is the conclusion of the study in which the main findings are summarized and the future directions are discussed. This part includes a summary of the findings and their practical application in the renewable energy industry. The study shows that the findings provide useful information on the quality of machine learning if it comes to solar energy production. Furthermore, the chapter suggests directions for further study and propounds on the possibility of enhancing the proposed approach.

2. LITERATURE REVIEW

2.1 INTRODUCTION

Introducing in this chapter, we develop the theoretical framework toward solar forecasting and indulge in the past research of the field. It starts with a definition and overview, with its next part dealing with defect solar prediction and its importance. The chapter goes on to evaluate varying methods of solar power prediction. Also, there is a breakdown of some of the notable studies that examine different approaches to the problem. To come up with conclusions, we prefer to make comparative and contrasting studies of them so that we can notice the pros of each one with some problems or possible changes.

This comparative analysis is expected to be the stepping stone for inventory of research problems and determination of new areas in the subject of solar fault predictions. These are the gaps that are most in the trend of these day like for example, little or no previous research in that area, current method lacked may be some of the variables, or some new technologies that have not been investigated yet. The research will make these gaps visible, and the aim is to make the contribution to the existing body of knowledge and set the way for future studies. One of the sections of the book that talks about is the theoretical foundation and through a wide-ranging literature review we can see the latest developments in the field.

2.2 OVERVIEW OF SOLAR ENERGY

The solar power market has gained significant momentum globally, showing annual growth at a rate of about 30%. As the top producer, China became a holder of the largest share of photovoltaic modules. In 2010 overall, China exported solar panels with a cumulative electrical capacity of 23,000 MW. In 2010, a Beijing-based research firm made a finding that said up to 75% of global solar deals were dealt with by China [31]. The high demand felt by photovoltaic modules has attracted multifarious domestic and international enterprises every year to set up business. With the complicated intervention of the market competition, the prices of photovoltaic modules are slipping [32].

Recent surveys and research indicate a significant decrease in the average per watt cost of photovoltaic (PV) modules. Over a period of four years, the cost has dropped from $1.61 to $0.80. This reduction in cost contrasts sharply with the concurrent increase in power costs within the country [33].

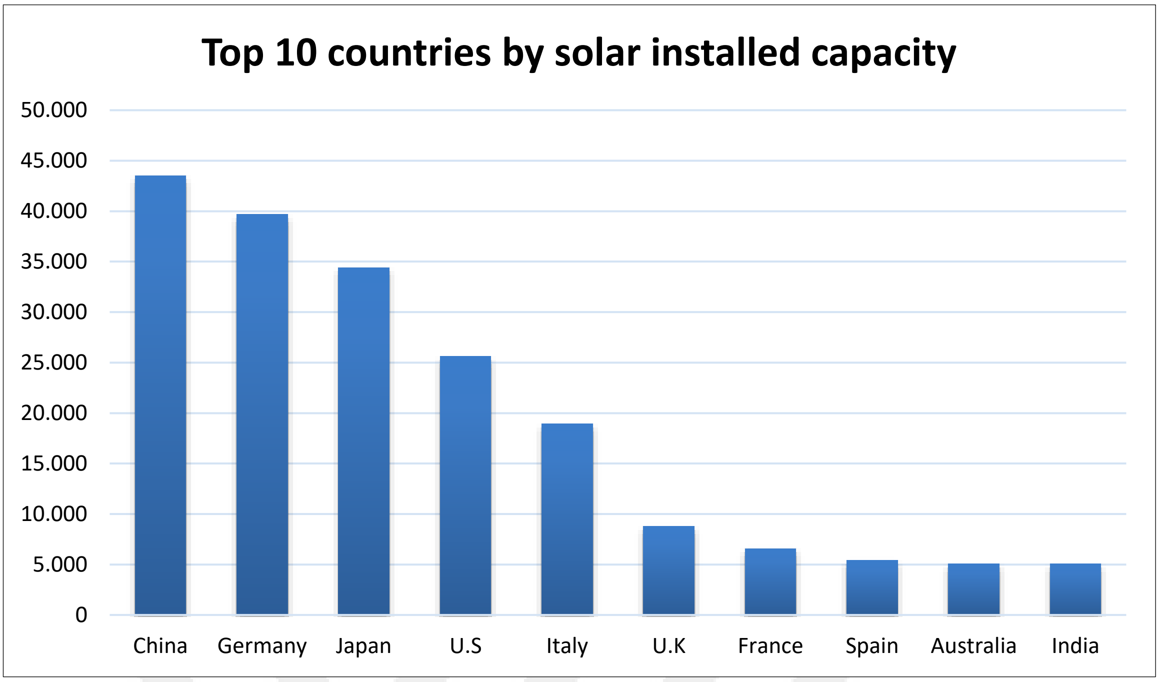

Figure 2.1: Top Ten Solar Installed Capacity Countries [34].

Solar energy is the radiant light and heat captured from the Sun, a medium-sized star at the center of our solar system. It has been radiating for approximately 4.6 billion years and is expected to continue for another 5 billion years [35]. The source of this solar energy is a process known as nuclear fusion, in which hydrogen atoms fuse under great pressure and temperature, producing helium and releasing enormous quantities of energy. Every second, the Sun produces an astounding 3.8 x 1026 Joules of energy [36]. This energy is emitted as electromagnetic radiation in all directions, with a little portion reaching our globe. This solar energy powers the Earth's climate system, accelerates the water cycle, and keeps all life alive. Adoption of solar energy as a renewable resource has taken paramount relevance in today's energy landscape. Traditional fossil fuels such as coal, oil, and natural gas are not only finite, but also cause serious environmental issues such as air pollution and global warming. Furthermore, the mining, transportation, and use of these fuels can cause significant environmental damage, such as habitat destruction and oil spills [37]. Solar energy, on the other hand, is a pure, sustainable, and essentially limitless power source. Solar energy, unlike fossil fuels, produces no hazardous pollutants. Furthermore, capturing solar energy prevents disruptive mining operations as well as the dangers of catastrophic spills [38]. Given rising energy demand and growing concerns about climate change, a transition towards renewable energy sources, particularly solar energy, emerges as a worldwide necessity.

2.2.1 Solar Power Development

The primary focus of solar power research and development is on developing photovoltaic technology to transform the sun's energy into electricity. Solar panels, also known as photovoltaic cells, are composed of semiconductor materials that directly transform solar energy into electricity. These panels can be used for large-scale power generation because they can be mounted on rooftops, fields, solar farms, and even integrated into building materials. Electricity generated by solar power plants can be fed into the grid, helping to reduce reliance on fossil fuels [41].

Numerous initiatives have been launched to facilitate and expand the use of solar energy in residential and commercial settings [39]. About nine hours of daylight each day is typical. The 1981 installation of a solar power system in Mimnniala was an early example of such an initiative [34]. One of the solar power development examples is presented here. In Pakistan, Khukhera near Lasbela (Baluchistan), Malmari in district Thatta, Ghakkar in district Attock, and Dital Khan Legari in district Mirpur Khas all received new solar photovoltaic systems after the initial pilot project was completed successfully [40]. The Government of Punjab has announced the allocation of 5,000 acres of land near the district of Bahawalpur for the construction of the Quaid-e-Azam solar power park. About 400 megawatts (MW) worth of solar power plants have been built and linked to the grid thus far. In 2020, this project is expected to have a total capacity of 1000MW [35].

2.2.1.1 Solar panels types

The photovoltaic effect describes the method through which solar panels, commonly known as PV panels, convert sunlight into electricity. They are essential in the process of converting solar radiation into usable electricity, which can then be used to run everything from individual houses and companies to the entire electrical grid.

Solar modules are constructed using SiO2 (Silicon), a material that is both cheap and widely available. The raw sand must be filtered to a level of 99.9%, or 1 ppm. Purifying sand to the required standards comes at a hefty price [42]. The majority of solar panels fall into one of three categories [43].

2.2.1.2 Mono crystalline solar panels

All solar panels follow the same standard procedure and are constructed from silicon. The silicon used to make monocrystalline solar panels is a single crystal. Because they are grown from a single crystal on silicon, mono crystalline solar panels do not have these grain boundaries [44]. In terms of efficiency, mono crystalline solar panels are at the top of the food chain. The most modern solar panels boast an efficiency of approximately 22% in the lab. Monocrystalline solar panels are more expensive than other types of solar panels due to their higher output efficiency.

Figure 2.2: Photovoltaic Mono Crystalline Module [45].

2.2.1.3 Monocrystalline photovoltaic module

Polycrystalline solar panels, also known as multi-crystalline solar panels, differ in structure from monocrystalline solar panels since they are manufactured using numerous crystals of silicon. As a result of lower standards for purity, their silicon crystals are more affordable. Because of their impure nature, they are not very effective. Poly crystalline solar panels, according to the most recent research [42], have an efficiency of 17%. These modules are designed to withstand high temperatures and work well even in extremely hot climates. Because of this, poly crystalline modules are commonly utilized throughout the Middle East, Africa, and Australia.

Figure 2.3: Photovoltaic Polycrystalline Module [46].

2.2.1.4 Thin film solar panels

Thin-film solar panels are fundamentally different from their bulkier predecessors. Typically, a glass substrate supports a single or many layers of cadmium telluride. Thin-film based solar panels include the more versatile flexible kind. CIGS (Copper Indium Gallium Selenide) is placed onto a plastic sheet to make flexible solar panels functional [48]. While the first generations of thin film solar panels had efficiencies of only 8-10%, the latest generation of these panels, which is presently on the market, boasts an efficiency of 21% [49].

Figure 2.4: Solar Thin Film Module [47].

Thin film solar panels possess several different advantages compared to traditional solar panels [42].

a. Demonstrate exceptional performance in environments characterized by limited solar illumination.
b. It is possible to install the system at any orientation that allows for optimal exposure to sunlight.
c. The temperature coefficient of conventional solar panels is higher in comparison to that of other types.
d. Owing to their low temperature coefficients, these devices exhibit excellent performance in high temperature circumstances.

2.2.2 Applications of Solar Energy

Currently, solar-powered products have gained popularity, encompassing a wide range of applications, ranging from mobile phone chargers to large-scale power plants. Solar energy is increasingly being harnessed and utilized in various sectors. Solar energy is a suitable option for mid-sized residential facilities or small-sized commercial facilities, as indicated by previous research [49]. There exist three primary classifications of solar system configurations.

2.2.2.1 Off-Grid solar systems

An off-grid solar system refers to a self-sustaining energy system that relies solely on solar power as its major source of energy.

Within this particular system, solar energy serves as the predominant source of electricity. Any surplus power generated is afterwards stored within a battery bank, so enabling its utilization during nighttime hours. The off-grid solar system is composed of three main components: the solar inverter, battery bank, and solar panels. The user has provided a numerical reference, indicating the presence of a citation or source [50].

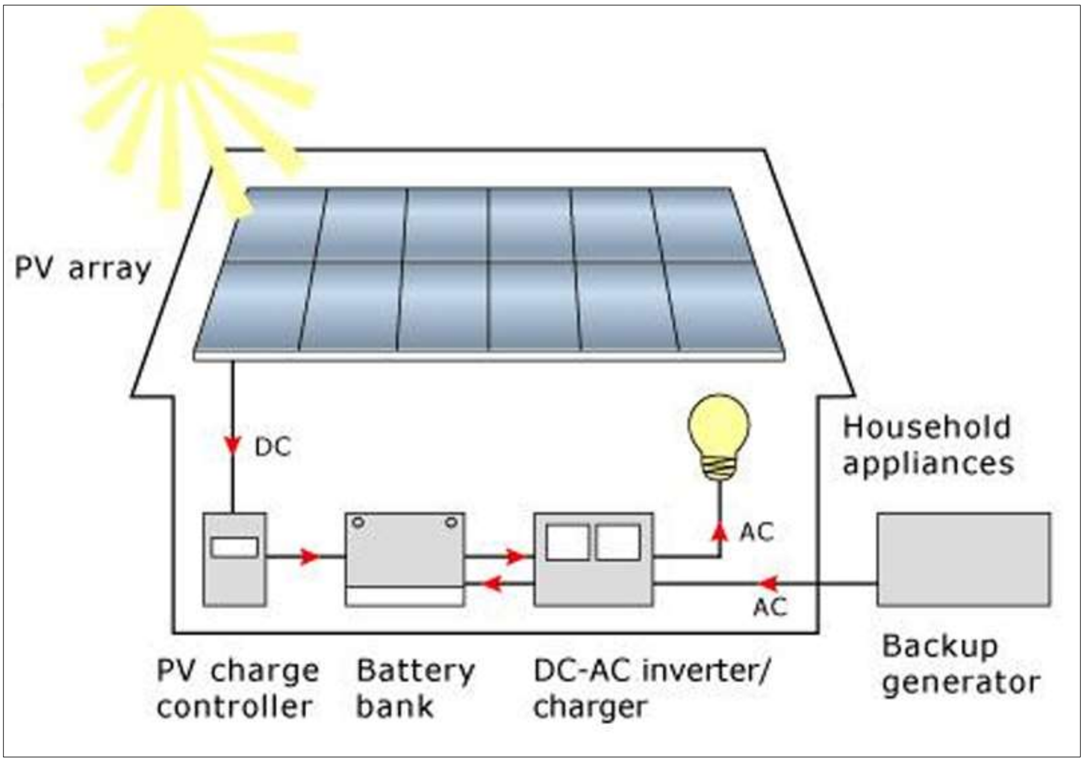

Figure 2.5: Off Grid Residential Site Schematic [51].

2.2.2.2 Off-Grid residential site schematic

On-grid solar systems are commonly installed in medium and large commercial buildings and are particularly popular in the industrial sector. On-grid solar systems are ideal for any facility that has continuous grid electricity. There is no storage mechanism in this sort of solar system because it just uses solar energy as a primary source and the grid as a reference. Excess energy is put back into the grid, which is determined at the end of each month through Net Metering, Feed in Tariffs, or Power Purchase Agreements, as applicable [54].

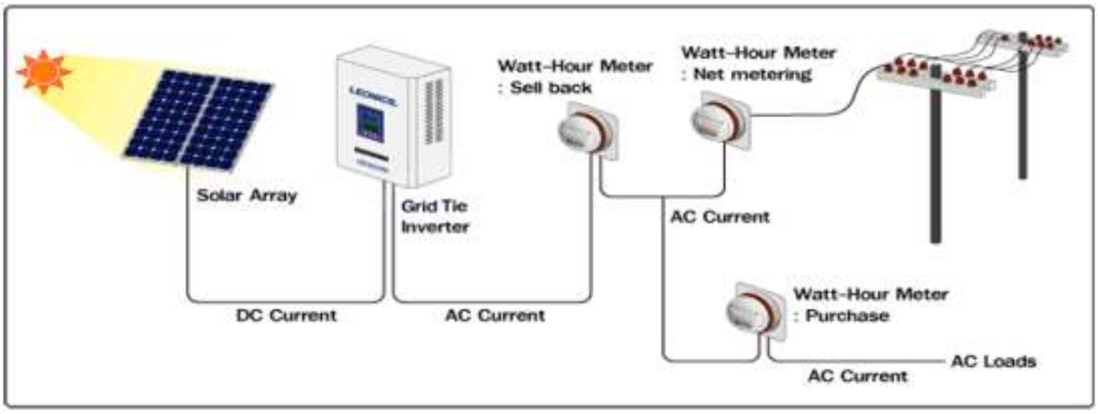

Figure 2.6: Schematic of an On-Grid Solar System [52].

2.2.2.3 Hybrid solar systems

Hybrid solar systems, as their name implies, are a combination of off-grid and on-grid solar systems. Solar energy is employed as the major source of energy, with any surplus energy being stored in a battery bank. Once the battery bank has reached its maximum charge capacity, any surplus power is then directed towards the grid for the purpose of net metering. The user has provided a numerical reference [53].

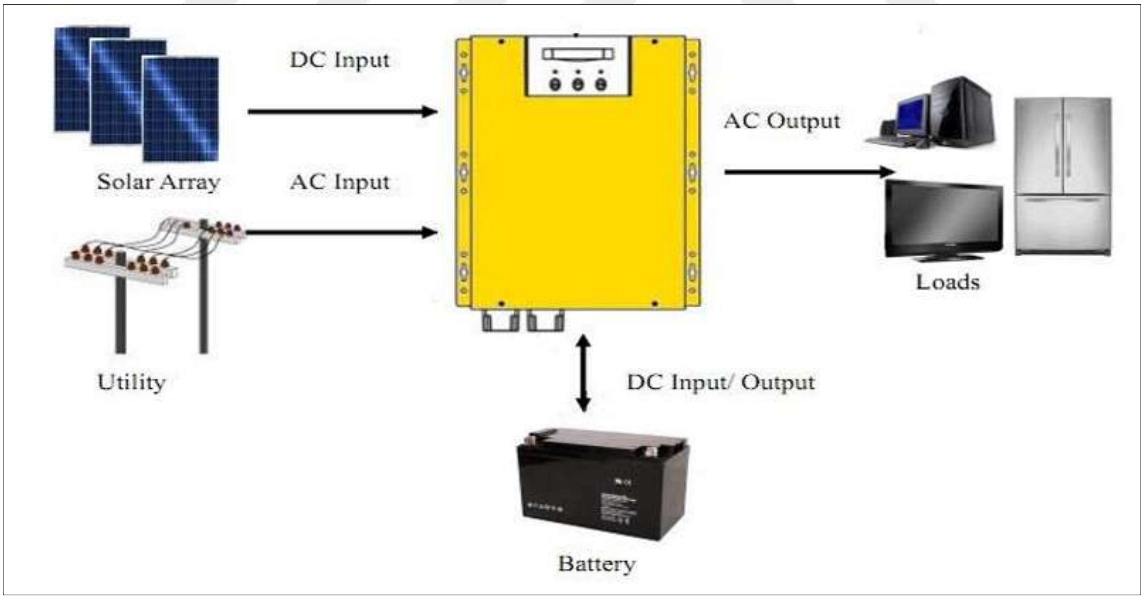

Figure 2.7: A Schematic of a Hybrid Solar System [53].

2.2.2.4 Use cases of solar energy

Here are some use cases of solar energy:

Residential and Commercial Solar Systems: The utilisation of solar energy is progressively becoming more prevalent in both residential and business settings. There is a growing trend among homeowners and businesses to embrace the use of solar panels as a means to independently generate electricity, resulting in a reduction in utility expenses and the production of carbon dioxide. The surplus electricity produced can be effectively stored in batteries to ensure energy self-sufficiency during periods of low solar generation, such as night-time or cloudy days [55].

Solar water heating: Solar water heating systems provide an additional utilisation of solar energy. Solar thermal systems utilise solar thermal collectors to heat water for many applications, including domestic, industrial, and commercial purposes. Solar water heaters have the capacity to diminish the reliance on electricity or gas for the purpose of water heating, so making a significant contribution towards energy conservation and the mitigation of emissions [56].

Solar-powered transportation: Solar energy is increasingly being adopted in the transportation industry. Solar panels have the potential to be incorporated into electric cars (EVs) in order to enhance their driving range or provide power to auxiliary systems. Solar-powered charging stations offer a sustainable charging alternative for electric vehicles (EVs), thereby diminishing the need on traditional electricity sources [57].

Agricultural and irrigation applications: Solar power is utilised in distinctive applications within the agricultural sector, mostly to provide energy for irrigation systems, water pumps, and various types of farm equipment. This holds significant advantages in rural regions characterised by low availability of dependable electricity [58].

Environmental benefits: The growth of solar power has been found to have a substantial impact on the reduction of greenhouse gas emissions and air pollutants that are often linked with conventional power production methods reliant on fossil fuels. It plays a role in the mitigation of climate change and the enhancement of air quality, consequently promoting the development of a more sustainable environment [59].

2.3 GENERAL RESEARCH ON SOLAR POWER FORECASTING

The literature reviews used in this research were mostly based on the work of Inman et al., [60] and Antonanzas et al., [61]. These researches provide a thorough grounding in the numerous forecasting methods and solar PV panel-related theories that are relevant to the field.

Kestylev et al., provide an overview of the several methods used to predict solar PV power [62]. The researchers advocate for the establishment of a common benchmark for the sector. The use of numerical weather projections (NWP) as input data has been shown to result in significant improvements to long-term forecasting, among other relevant discoveries that have been reported in previous research. Short-term predictions of solar PV energy output have also been enhanced by modelling future cloud movements using satellite-based data. Finally, the accuracy of models varies with the local climate where the forecast is being made. As a result, regional training is preferable to multi-site training when evaluating a model's efficacy. In a similar vein, because weather conditions change throughout the year, it's possible that a model would do better if trained on just one season rather than on all of them.

A variety of energy forecasting methods are outlined by Hong et al. [63]. The research highlights the widespread application of time series and ML methods in a variety of energy settings. The research is based on an energy forecasting competition and assesses the competing approaches.

The most important takeaway is that both machine learning and conventional time series methods have seen extensive use in the context of load demand forecasting. Similarly, power costs have been analysed using a wide range of time series and ML methods. The article also gives a summary of methods used to predict the output of competing technologies like as wind turbines and photovoltaic cells. The range of analysed forecasting methods for the power output of windmills and solar PV panels is smaller compared to that of load demand and energy costs. Some machine learning (ML) methods have been employed; however classical time series have been used extensively. The RF algorithm and the SVM method were utilised for PV solar cells [63].

An exhaustive theoretical overview of RES forecasting is provided by Inman et al., [60]. Modelling irradiance, air masses, and clearness indexes are only a few of the topics covered in the research. The theoretical foundations of popular time series models and ML methods are then laid forth. Theoretical frameworks for utilising irradiance measurements and satellite imagery are also offered. From a physics, statistics, and ML vantage point, the study provides direction for theory and methodology.

2.3.1 Time Series Techniques

Researchers examines the concept of short-term horizons, namely within the range of 0 to 4 hours [64]. The author gives a comprehensive analysis that includes a comparison of different time series models and machine learning models, serving as a baseline for evaluation. The primary finding of this study indicates that, in the context of short-term prediction, time series approaches such as ARIMA demonstrated superior performance compared to machine learning (ML) and other time series models. According to the researchers, the primary factor contributing to the superior performance of the model is its capacity to accurately represent the variations in irradiance across a 24-hour cycle [64]. Furthermore, from a short-term perspective, it is typically observed that delayed irradiance values significantly correlate with current meteorological conditions.

These weather conditions are unlikely to undergo significant changes within a brief timeframe, especially during periods of stable weather. Thus, these indicators can be used as quite a precise short-term power generator source and this represents a fact that is almost 100% evident. Because of this, short-term forecasting will be possible only when we carefully consider the possibility of using ARIMA and ARIMA models with exogenous input variables, as well as the inclusion of lagged solar energy outputs.

The study contains three models: autoregressive model (AR), linear regression model (LR), and an autoregressive model with exogenous input [65]. The authors of the study came to the fact that the ARX model in the best way beat out other models for prediction of solar power output. Initially, the study models were combining the LR and ARX models, which makes use of 3D Numerical Weather Prediction data. In contrast, the AR model used only historical solar PV panel energy output data. The present study supports the idea that lagged power values can be a good choice for short-term forecasting, as it has already been touched upon in previous works [64].

2.3.2 Machine Learning Techniques

In contrast to the results observed by [64], the ANN performed better in a study conducted by Coimbra and Pedro that compared ARIMA, ANNs, and KNN models [66]. Solar irradiance was not something that could be predicted by Coimbra and Pedro, but the power production of solar PV panels could. No external NWP data was used as input in the study, instead only past output levels were utilised. The ANN model's accuracy was also improved by the authors when they used a genetic algorithm (an optimisation technique motivated by natural selection) in place of a conventional one. Coimbra and Pedro conclude that the accuracy of weather forecasts varies between climates. They recommend breaking down the data separately for each type of climate to be modelled. Here, the authors speculate that employing a model adapted to a dataset for a particular weather regime rather than a model fitted to a dataset for all weather regimes could improve predicted outcomes.

Other researchers, like Andrade et al. [67], have investigated ML methods and assessed them with feature development that is expected to boost performance. Principal Component Analysis (PCA) and a feature engineering methodology combined with a Gradient Boosting Tree model were the primary methods employed in this research. To further develop features from their NWP data, the scientists utilised a variety of smoothing techniques. The authors calculated spatial averages and variances of meteorological factors using a grid of NWP data surrounding the PV installation. The authors did more than just generate features from a local grid of points; they also calculated variances for various predictors over various lag durations. The purpose behind using lead times to generate variance features was so that the feature might represent weather variability.

The most important takeaway is that combining PCA and feature engineering yields superior outcomes. The authors claim that there are two areas where more research is needed. The first is the creation of significant features that better the forecast, which is a subset of feature engineering and feature selection under the umbrella of feature management. Second, there's the matter of digging deeper into ML modelling approaches that can be used in tandem with illuminating features. They conclude by saying that combining deep learning methods with good feature management is an exciting prospect. Predicting solar irradiance using PCA, ANN, and Analogue Ensemble (AnEn) was the focus of Dav et al. [68]. PCA was employed as a feature selection method to cut down on the dataset's dimension. The data set is an accumulation of daily measurements of solar radiation output over a period of eight years.

Utilising PCA in conjunction with ANN and AnEn improves prediction accuracy, as shown by a comparison with without utilising PCA. Long-term (up to 100 hours) forecasting findings are presented by Chen et al. The authors used an ANN to make predictions based on historical NWP data. The model's performance declined on rainy-day forecasts and was especially susceptible to prediction mistakes in the NWP input data. On both overcast and clear days, the ANN model achieved MAPEs of about 8% [69]. Predictions of solar energy production were made by Persson et al. using Gradient Boosted Regression Trees (GBRT) for the next one to six hours [70]. Historical power output and weather conditions for 42 PV plants in Japan were used. The GBRT model outperformed the adaptive recursive linear AR time series model, the persistence model, and the climatology model in terms of root-mean-squared error (RMSE) across all time horizons for the forecast. It was demonstrated that lagged power output measurements have greater predictive potential for shorter forecast horizons. In a similar vein, the value of weather forecasts rose with the length of the forecast horizon. Shi et al. present a weather classification method and support vector machine (SVM) to predict PV power output at 15-minute intervals for the following day [71]. Clear sky, cloudy, foggy, and rainy are the four states of the weather. The categorization is based on an examination of connections between predicted weather patterns and actual PV power generation. After that, the data is normalised to make it more precise and error-free while preserving its inherent correlation. After that, we fit four SVM models using a radial basis function kernel to our four weather categories. In conclusion, the research demonstrates an approach to employing SVM models for climate-specific model training [71].

The provided research covers a selection of relevant studies selected through a comprehensive literature assessment in the context of solar energy prediction using machine learning as shown in Table 2.1. Each study delves into several approaches, features, prediction horizons, assessment metrics, and major discoveries that aid in the knowledge of solar energy prediction.

Researchers [72], study to consider using low data to develop solar radiation estimations with Support Vector Machines (SVMs) on a prediction horizon of 24-hour. The process of the evaluation was based on the measures of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) however. The main findings of our investigation were that, in particular, our SVM model was capable of working out the solar radiation correct measurements hence, contributing to creating a reliable and accurate solar energy prediction model. To beat the forecast accuracy challenge, solar farms have started to use Smart weather sensors in parallel with digital strategies in tracking solar energy data.

Following the footpath of previous research where [73], proved the high prognostic power of GBMs, they reported that predictions of solar power output based on observed and historical data were done a day ahead using weather measurements and solar irradiance data. They have discovered in the historical data that GBMS had advantages over single model technologies, which means that GBMs for solar energy prediction is highly effective. The authors employed Random Forest method to forecast the power output for a 7-day horizon using meteorological data and solar panel attributes [74]. RMSE and the Mean-Std. Dev. of Percentage Error (MSPE) were the benchmarks of this work. In another remarkable case, the study, through the incorporation of the solar panel attributes, was able to achieve and even improve the accuracy of the solar power prediction. In particular the authors examined the solar panel features and weather data for a seven-day forecast period using Gaussian Processes. RMSE and MAE are performance metrics that were used as evaluation criteria [75]. In one word, the study illustrates a highly effective mechanism of solar power prediction and real forecasted data exchange with the weather factor of environment after initial alignment and then its utilization as input to the uncertainty reflecting models such as Gaussian Processes.

One such study has examined an advance learning approach that combined two algorithms, Random Forest and XGBoost, using meteorological data inclusive of solar irradiance to produce a 48-hour forecast. The experiment employed RMSE and MAPE to confirm that combined models improve the accuracy level of solar powered forecast, thus, they might become the mainstream of meteorological forecasting [76][77].

Table 2.1: Solar Energy-Related Work Using Machine Learning Algorithms.

Reference	Methodology	Data-Features	Prediction Time	Evaluation Criteria	Main outcomes
[72]	SVM	Weather Data	1-Day	MAE-RMSE	Predictions of solar irradiation have become more accurate.
[73]	XGB	Traditional Solar Irradiance Data, Weather Data	1-Day	MAE-RMSE	greater performance when compared to single-model techniques
[74]	RF	Solar Panel Characteristic, Weather Data	7-Days	MAPE-RMSE	Accurate solar power forecast based on panel characteristics
[75]	NB	Solar Panel Characteristics, Weather Data	7-Days	MAE-RMSE	Uncertainty in solar power estimates was efficiently modelled using Gaussian Processes.
[76]	Ensemble Learning Technique	Solar Irradiance Data, Weather Data	2-Days	MAPE-RMSE	The accuracy of solar power forecast was enhanced by using ensemble models.
[77]	RF	Solar Irradiance Data, Weather Data	2-Days	MAPE-RMSE	The Random Forest model accurately estimated solar energy.

2.4 RESEARCH GAP

Significant progress has been achieved in the field of solar energy prediction using machine learning approaches. Nonetheless, numerous unexplored areas require research, as shown below following a thorough literature analysis.

a. Increasing the number of impacting factors: While prior studies mainly rely on historical solar irradiance data and weather information, failing to account for external factors such as shading effects, solar panel dust deposition, and grid conditions may reduce prediction accuracy [78]. There is a need for research into the effects of these elements on solar energy production, as well as the development of models that incorporate their effects.

b. Unifying performance benchmarking: Although several machine learning methods have been applied for solar energy prediction, the lack of standardised benchmark datasets and assessment measures makes consistent evaluations difficult. In-depth comparison evaluations utilising different algorithms on uniform datasets could provide a comprehensive picture of approach strengths and drawbacks [79]. Establishing benchmark datasets and evaluation standards would allow for equitable comparisons of various approaches.

c. Tackling uncertainty comprehensively: Solar energy prediction is fundamentally uncertain due to weather variations and external dynamics. Although some studies use probabilistic ways to manage uncertainty, such as Gaussian Processes, more sophisticated procedures are required to accurately quantify and disseminate uncertainty throughout the prediction process [80].
d. Advancing transparency and intelligibility: Some machine learning models, particularly deep learning algorithms, have low interpretability. In applications where transparency is important, such as solar energy forecasting, efforts must be focused on developing interpretable machine learning models that clarify the mechanisms impacting solar energy generation [81].

2.5 CONCLUSION AND INSIGHTS

The importance of short-term delayed values of energy output and the increasing importance of numerical weather predictions (NWP) data input as the prediction horizon is extended are two primary trends that are emphasized by almost all studies. Time series models have been shown to be superior to ML approaches in some investigations, while other studies have shown the opposite (see [82][83] and [84]). That's why we think it's important to keep contrasting ML methods with the time series models.

According to [82], research, the forecasting accuracy of an ANN can be improved by selecting the appropriate optimization technique. Since the goal of this research is to provide a broad comparison of ML methods, we will not be looking at how different versions of the ANN perform.

The vulnerability to inaccuracies in NWP data has been highlighted by a number of studies. Using NWP data from several sources is one approach that may be taken to attempt and address this for any given model. For this to be possible, it is necessary to have information on a large number of predictors from several credible sources, which is not always the case.

Further reducing the total inaccuracy of the input NWP data could be accomplished with NWP based on multiple physical models.

3. METHODOLOGY

3.1 INTRODUCTION

Predicting solar energy generation accurately is crucial for effective energy management and grid integration of solar energy. In this chapter, we provide a thorough and methodical approach to predicting solar energy output utilizing a number of different machine learning techniques.

Intermittency of power generation owing to weather conditions is the primary critical and problematic issue in solar energy production. Specifically, changes in temperature and irradiance can have significant effects on the reliability of electricity generation. The daily/hourly efficient management of power grid generation, delivery, storage, and decision-making on the energy markets relies heavily on reliable predictions of PV modules' power output. Weather characteristics such as Irradiance, ambient temperature, humidity, wind speed, and other important parameters are predicted in order to estimate solar generation for future time blocks.

Our objective is to enhance current methods for predicting solar energy production by leveraging the strengths of multiple machine learning (ML) techniques and expanding on the Solar Power Generation Dataset. Predicting solar energy output is challenging due to the complex interplay among solar radiation, atmospheric conditions, and energy production. Conventional forecasting methods struggle to capture the non-linear correlations and dynamic nature inherent in solar energy generation [84].

New developments in the machine learning field offer many possibilities for a successful solution to these challenges. The solar power generation dataset combines solar energy prediction models' improved learning and testing capabilities. The complete dataset compiles various weather and solar radiation factors, applying the principles of solar energy generation dynamics. From this comprehensive and diversified dataset, we can guarantee the precision and reliability of our prediction tools.

Our method includes various strong machine learning algorithms such as KNN, LR, RF, DT, GBT, and ADBR. To forecast solar power production. The evaluation of the performance of these models is carried out with the help of the top-rated metrics like Root Mean Squared Error (RMSE), Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Mean Squared Error (MSE). These metrics contribute to the findings that impress our models with the latest data. Additionally, the correctness of our method is verified by making comparison with the baseline models [84].

In addition, we have shown a meticulously and systematically intelligent approach aiming at utilizing the ML models for solar energy prediction has been brought out. The idea we pursue is to develop precise and reliable solar energy generation predictions by the means of machine learning algorithms and the Solar Power Generation Dataset. Our study signifies the generation of solar energy by the above-mentioned machine learning models in the solar energy prediction by the rigorous training procedures and with certain datasets. The effectiveness of our proposed methodology will be further supported in the next chapters by focusing on results, analysis, and discussion.

3.2 PROPOSED WORK

The application of ML models for forecasting solar energy generation is the main focus of our solar research. Our proposed mechanism for predicting solar energy is depicted in Figure 3.1 and each step is explained below:

a. Data Collection and Selection: The first step is to gather data on solar generation and weather conditions from trustworthy sources like the Solar Power Generation Dataset. There are two parts to this data set: the generation dataset and the weather dataset.
b. Data Preprocessing: The data is preprocessed to remove outliers, standardize values, and remove extraneous details.
c. Feature Reduction: Our ML models rely on inputs such as selected features from the preprocessed data, which we evaluate to find. The optimal features were chosen for processing after an exploratory data analysis.
d. Split Data: The information is separated into a training set and a test set. The ML models will be trained on the training dataset, and their results will be tested on the testing dataset.
e. Model Training: The training dataset is used to teach the ML models. In this study, several ML models (KNN, LR, RF, DT, GBT and ADBR) are used to forecast solar power output. These models performed outstanding as compare to existing researches.

f. Model Evaluation: Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, and the Coefficient of Determination are some of the metrics used to assess the ML models' efficacy. Through this analysis, we are able to gauge the accuracy and consistency of our models and compare them to those of the current industry standard.

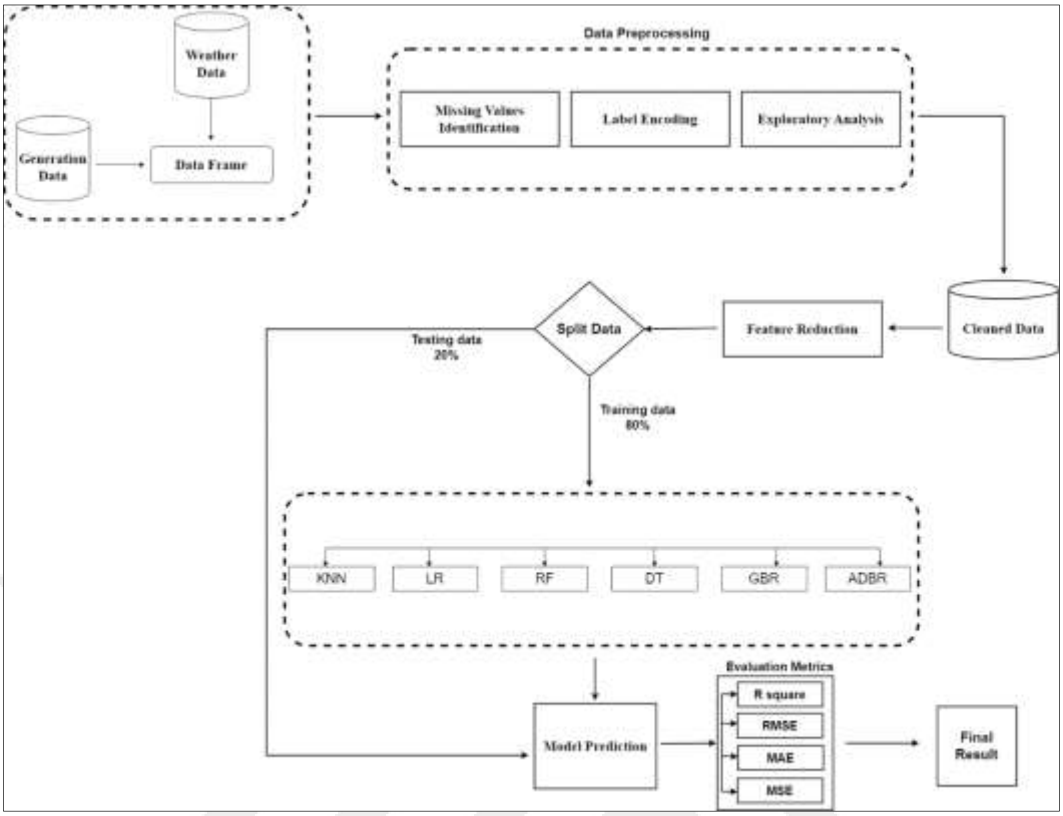

Figure 3.1: Proposed Model for Solar Energy Prediction.

3.2.1 Dataset

The study utilizes the Kaggle dataset [14], which has two distinct sets of information:

3.2.1.1 Generation dataset

The datasets pertaining to electricity generation are collected at the level of inverters, with each inverter being connected to numerous lines of solar panels. The dataset contains seven columns and 67,698 rows. The Figure 3.2 shows the description of each feature of the dataset while the Figure 3.3 shows the first five rows of the dataset.

DATE_TIME	15 minute timestamp
PLANT_ID	Common for the entire file
SOURCE_KEY	Unique Inverter ID (Total 22 Inverters)
DC_POWER	Amount of DC Power generated by that inverter for the timestamp
AC_POWER	Amount of AC power after conversion from DC by inverter for the timestamp
DAILY_YIELD	Cumulative sum of power generated on that day, till that point in time
TOTAL_YIELD	Total yield for the inverter till that point in time

Figure 3.2: Generation Dataset Columns Description.

	DATE_TIME	PLANT_ID	SOURCE_KEY	DC_POWER	AC_POWER	DAILY_YIELD	TOTAL_YIELD
47897	2020-06-08 14:45:00	4136001	xoJJ8DcxJEcupym	1078.153	1052.140	7120.200	209283860.200
9898	2020-05-19 16:45:00	4136001	xMblugepa2P7lBB	39.473	38.180	6740.400	106695331.400
52233	2020-06-10 16:15:00	4136001	81aHJ1q11NBPMrL	440.607	432.450	5298.429	1215446553.429
29315	2020-05-30 19:15:00	4136001	Mx2yZCDsyf6DPfv	0.000	0.000	8199.000	2578944.000
13633	2020-05-21 23:15:00	4136001	Et9kgGMDl729KT4	0.000	0.000	6646.000	1735537.000

Figure 3.3: First Five Rows of Generation Dataset.

3.2.1.2 Weather sensor data

This dataset contains 3,259 rows and 6 columns. The description of each feature is shown in Figure 3.4 while the Figure 3.5 shows the five rows of the weather sensor dataset.

DATE_TIME	15 minute timestamp
PLANT_ID	Common for the entire file
SOURCE_KEY	Unique Inverter ID (Total 22 Inverters)
AMBIENT_TEMPERATURE	Ambient temperature at the plant
MODULE_TEMPERATURE	Temperature reading for module (solar panel) attached to the sensor panel
IRRADIATION	Amount of irradiation for the 15 minute interval

Figure 3.4: Weather Dataset Columns Description.

	DATE_TIME	PLANT_ID	SOURCE_KEY	AMBIENT_TEMPERATURE	MODULE_TEMPERATURE	IRRADIATION
1959	2020-06-04 11:00:00	4136001	iq8k7ZNt4Mwm3w0	28.645	39.689	0.653
1332	2020-05-28 21:30:00	4136001	iq8k7ZNt4Mwm3w0	29.253	28.184	0.000
2743	2020-06-12 15:00:00	4136001	iq8k7ZNt4Mwm3w0	26.966	35.406	0.439
1560	2020-05-31 06:45:00	4136001	iq8k7ZNt4Mwm3w0	26.877	28.587	0.095
541	2020-05-20 15:45:00	4136001	iq8k7ZNt4Mwm3w0	35.118	47.446	0.522

Figure 3.5: First Five Rows of Weather Dataset.

We combined both the dataset and created a single dataset for experiments of our research. First, we change the format of date and time column from both datasets. Next, we merge the datasets on the basis of date and time and drop two columns (plant ID and source key) from the datasets. The final dataset contains 9 columns and 67,698 rows. The Figure 3.6 shows the first five rows of final solar dataset.

	DATE_TIME	SOURCE_KEY	DC_POWER	AC_POWER	DAILY_YIELD	TOTAL_YIELD	AMBIENT_TEMPERATURE	MODULE_TEMPERATURE	IRRADIATION
0	2020-05-15	4UPUqMRk7TRMgml	0.0	0.0	9425.000	2.429e+06	27.005	25.061	0.0
1	2020-05-15	81aHJ1q11NBPMrL	0.0	0.0	0.000	1.215e+09	27.005	25.061	0.0
2	2020-05-15	9kRcWv60rDACzjR	0.0	0.0	3075.333	2.248e+09	27.005	25.061	0.0
3	2020-05-15	Et9kgGMDl729KT4	0.0	0.0	269.933	1.704e+06	27.005	25.061	0.0
4	2020-05-15	IQ2d7wF4YD8zU1Q	0.0	0.0	3177.000	1.994e+07	27.005	25.061	0.0

Figure 3.6: First Five Rows of Final Solar dataset.

The data is obtained from the inverter, which is linked to a set of solar panels. The sensor data is acquired from a strategically positioned sensor at the facility every 15 minutes. Both the power generation and sensor data are recorded and made available in a publicly accessible dataset, such as Kaggle [85]. The collection of power generation and sensor data occurs at specific time intervals and dates. In order to establish a correlation between the sensor and the power generation of a Solar Power Plant at a specific moment, it is necessary to consider the fluctuating intensity of sunlight, as this directly impacts the power generation. The distribution of data from a Power Plant over a period of 34 days, with measurements taken at 15-minute intervals.

3.2.2 Data Pre-Processing

The reliability and integrity of the dataset utilized for solar energy forecast relies heavily on the pre-processing and analysis of the data. The Data cleaning, normalization, and transforming the raw data are all part of the pre-processing activities that get it ready for analysis and modelling. The following analysis delves into the connections between the variables and how they affect solar power output.

There are many stages included in the pre-processing phase.

a. The Data Cleaning: In this phase, we deal with missing data, fix errors, and eliminate abnormalities and noise. We can reduce the possibility of errors and make more accurate forecasts if we use just complete and precise information. In Figure 3.7, we checked for the missing values and remove various unnecessary columns from the dataset.

```
df_solar.isnull().sum()
```

+ Code + Markdown

There is no Missing Values in the dataset

```
df_solar = pd.merge(generation_data.drop(columns = ['PLANT_ID']), weather_data.drop(columns = ['PLANT_ID', 'SOURCE_KEY']), on='DATE_TIME')
df_solar.sample(5).style.background_gradient(cmap='cool')
```

Figure 3.7: Remove Columns and Handle Missing Values.

b. Data Formatting: To allow for temporal analysis and maintain uniformity, the time and date feature in the dataset should be formatted into a standardized format, such as the date time format. The Figure 3.8 shows the data formatted into a standardized format.

```
# adding separate time and date columns
df_solar["DATE"] = pd.to_datetime(df_solar["DATE_TIME"]).dt.date
df_solar["TIME"] = pd.to_datetime(df_solar["DATE_TIME"]).dt.time
df_solar['DAY'] = pd.to_datetime(df_solar['DATE_TIME']).dt.day
df_solar['MONTH'] = pd.to_datetime(df_solar['DATE_TIME']).dt.month
df_solar['WEEK'] = pd.to_datetime(df_solar['DATE_TIME']).dt.week

# add hours and minutes for ml models
df_solar['HOURS'] = pd.to_datetime(df_solar['TIME'],format='%H:%M:%S').dt.hour
df_solar['MINUTES'] = pd.to_datetime(df_solar['TIME'],format='%H:%M:%S').dt.minu
df_solar['TOTAL MINUTES PASS'] = df_solar['MINUTES'] + df_solar['HOURS']*60

# add date as string column
df_solar["DATE_STRING"] = df_solar["DATE"].astype(str) # add column with date as
df_solar["HOURS"] = df_solar["HOURS"].astype(str)
df_solar["TIME"] = df_solar["TIME"].astype(str)

df_solar.head(2)
```

	DATE_TIME	SOURCE_KEY	DC_POWER	AC_POWER	DAILY_YIELD	TOTAL_YIELD	AMBIENT_TEMPERATURE
0	2020-05-15	4UPUqMRk7TRMgml	0.0	0.0	9425.0	2.429e+06	27.005
1	2020-05-15	81aHJ1q11NBPMrL	0.0	0.0	0.0	1.215e+09	27.005

Figure 3.8: Data Formatting.

a. Data Transformation: Data transformations can be made to make it more suitable for analysis, although this depends on the nature of the data itself. Scaling features, for instance, can be used to standardize the range of values across features. This ensures that the model's learning is not skewed by any one particular characteristic The data transformation can be seen in Figure 3.9.

```python
from sklearn.preprocessing import LabelEncoder
encoder = LabelEncoder()
df_solar['SOURCE_KEY_NUMBER'] = encoder.fit_transform(df_solar['SOURCE_KEY'])
df_solar.head()
```

	DATE_TIME	SOURCE_KEY	DC_POWER	AC_POWER	DAILY_YIELD	TOTAL_YIELD	AMBIENT_TEMPERATURE
0	2020-05-15	4UPUqMRk7TRMgml	0.0	0.0	9425.000	2.429e+06	27.005
1	2020-05-15	81aHJ1q11NBPMrL	0.0	0.0	0.000	1.215e+09	27.005
2	2020-05-15	9kRcWv60rDACzjR	0.0	0.0	3075.333	2.248e+09	27.005
3	2020-05-15	Et9kgGMDl729KT4	0.0	0.0	269.933	1.704e+06	27.005
4	2020-05-15	IQ2d7wF4YD8zU1Q	0.0	0.0	3177.000	1.994e+07	27.005

Figure 3.9: Data Transformation.

3.3 DATA ANALYSIS

When the data has been cleaned and prepared, this research uses a variety of methods to learn more about the features and how they relate to the production of solar power. This evaluation may consist of:

a. Exploratory Data Analysis (EDA): Enabling EDA requires data visualization, inspecting feature distributions and identifying outliers. It is a tool that can be used to study the relationships, trends, and patterns. The tool can also help one in collecting the data that is possessed by some of the prediction algorithms.

b. Data Visualization: Analyzing solar energy generation patterns and trends, and better understand intricate relationships between the features and their impacts can be supported by visualization. Tools such as line plots, bar graphs, heat maps, and box plots are utilized to visually convey information, which facilitates understanding.

Exhibited in Figure 3.2 is the graph of the air temperature distributions made from the data records. The x-axis is the whole scope of temperature measured on the scale of the data, while the y-axis is the abundance of each temperature reading. This visual approach helps the observers to find the range, mean, and noise of the temperature.

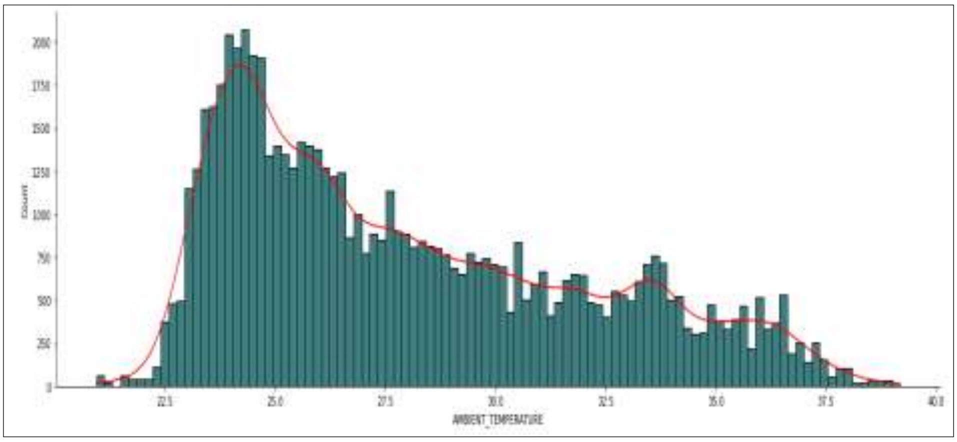

Figure 3.10: Analysis of Ambient Temperature.

In the Figure 3.11, the orientation of the sun in relation to the solar panels is the primary factor in producing this form. When the sun is still rising in the morning, it is at its lowest zenith. At this time of day, the sun has to penetrate the most atmospheric layers to reach your panels. This is significant because atmospheric molecules absorb solar energy as it passes through. Because of this, the amount of energy available for conversion by your panels will decrease the further the light has to travel through the atmosphere. Light has a shorter path because there is less atmosphere to penetrate when the sun is directly overhead in the middle of the day. This is the best moment for your panels to soak up sunshine. It's the same idea in the winter. However, there may be less sunshine for your panels to absorb on gloomier winter days. This is because the sun is at a lower zenith angle than it is in the summer.

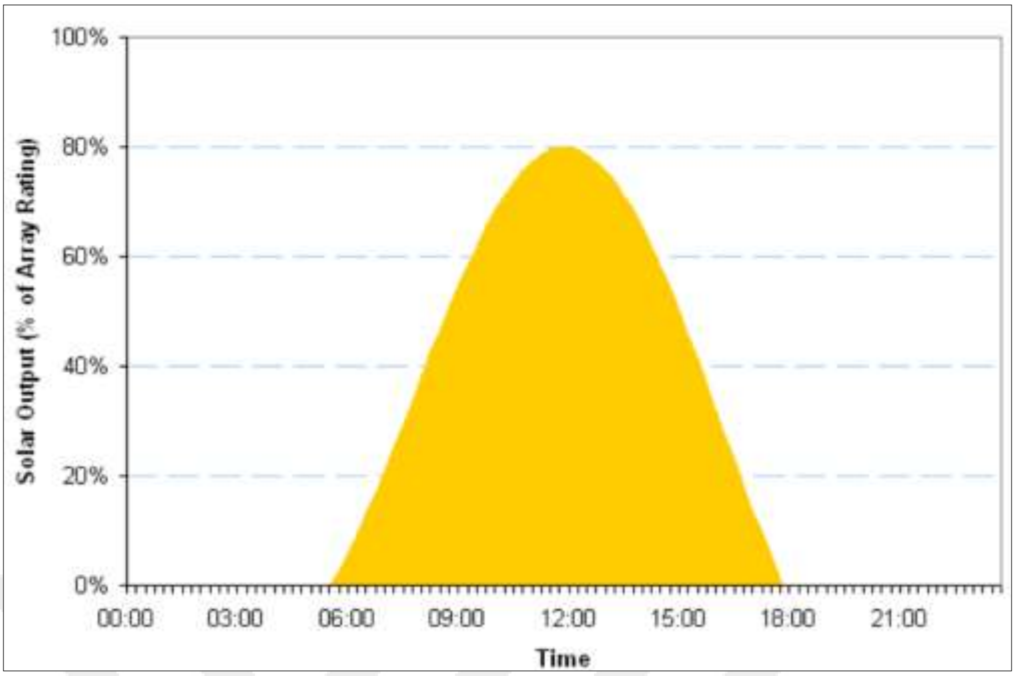

Figure 3.11: Ideal Graph of Solar Power Generation.

The per day DC_POWER generation graph in Figure 3.12 shows significant variations in power generation over most days. These oscillations are more noticeable when examining the data over time. Among the identified trends is a sequence of days with relatively minor swings in DC_POWER generation. 2020-05-15, 2020-05-18, 2020-05-22, 2020-05-23, 2020-05-24, 2020-05-25, and 2020-05-26 are among these dates.

On other days, however, a distinct pattern appears, with a higher level of variation in DC_POWER production. Data for 2020-05-19, 2020-05-28, 2020-05-29, 2020-06-02, 2020-06-03, 2020-06-04, 2020-06-13, 2020-06-14, and 2020-06-17 show this.

Furthermore, a selection of days stands out due to both extreme variability and a significant decrease in DC_POWER generation. Days in this category include 2020-06-03, 2020-06-11, 2020-06-12, and 2020-06-15. These severe variations in DC_POWER generation could be related to system failures, weather-induced swings, cloud cover, or other variables that require more in-depth investigation to determine the actual causes.

The per day DC_POWER generation graph, in addition to detecting variations, enables for the calculation of average power generation per day. Surprisingly, the study shows that the highest average DC_POWER generation occurred on 2020-05-15, while the lowest average DC_POWER generation occurred on 2020-06-11.

However, it is important to note that the large differences in DC_POWER generation throughout these days could be due to system-related abnormalities or changes in meteorological conditions. As a result, a thorough examination is advised to identify the underlying elements causing these changes. Nonetheless, this graphical depiction accurately highlights peaks and troughs in DC_POWER generation, assisting in the evaluation of performance patterns.

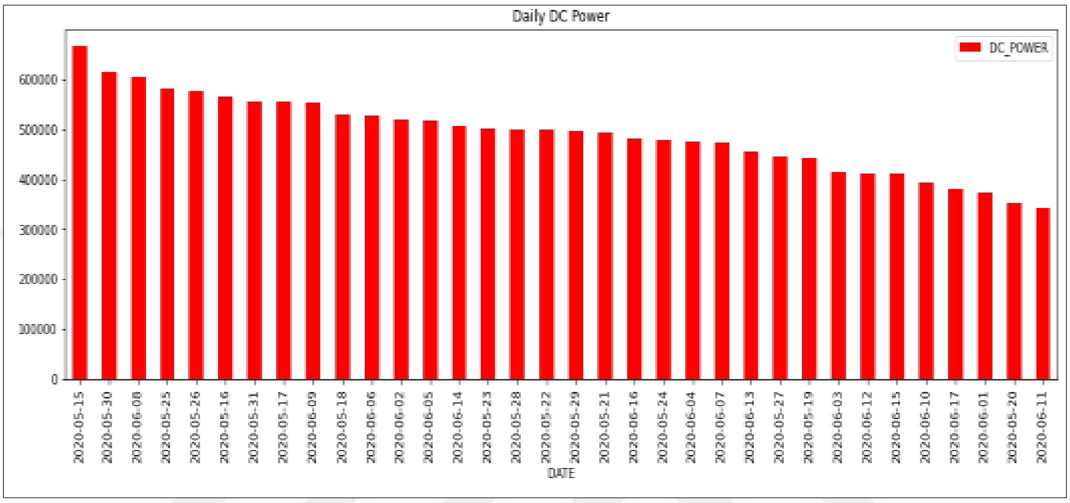

Figure 3.12: Abnormalities in Dc_Power Generation.

On a daily basis, the irradiation graph pattern in Figure 3.13 looks quite similar to the matching DC_POWER generation. C_POWER or output power in a solar power plant is mostly determined by irradiation. Alternatively, it is not incorrect to state that it is directly proportional.

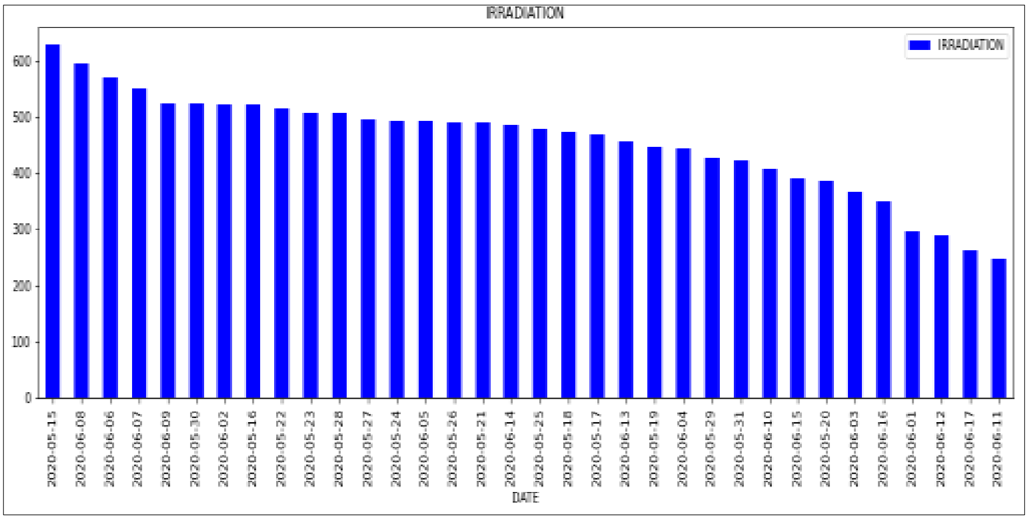

Figure 3.13: Irradiation Generation on Per Day Basis.

According to the Figure 3.14, the amount of sunlight that the solar panels can absorb is also affected by the thickness of the clouds. We may also see denser clouds in the winter, which is something to keep an eye out for. Thick clouds make it difficult for sunlight to penetrate, reducing the output of your solar power system. While we've discussed how the sun's position affects output, there's another aspect to consider when your system isn't functioning optimally even at midday. Solar panel temperature is the leading cause of your solar power system not performing well. Solar power generation is directly dependent on solar irradiation.

Both the DC_POWER and IRRADIATION graphs are almost identical to the ideal graph described earlier. The weather is also looking nice, with no clouds in the sky due to very little difference in IRRADIATION, temperature of the solar panel, and ambient temperature.

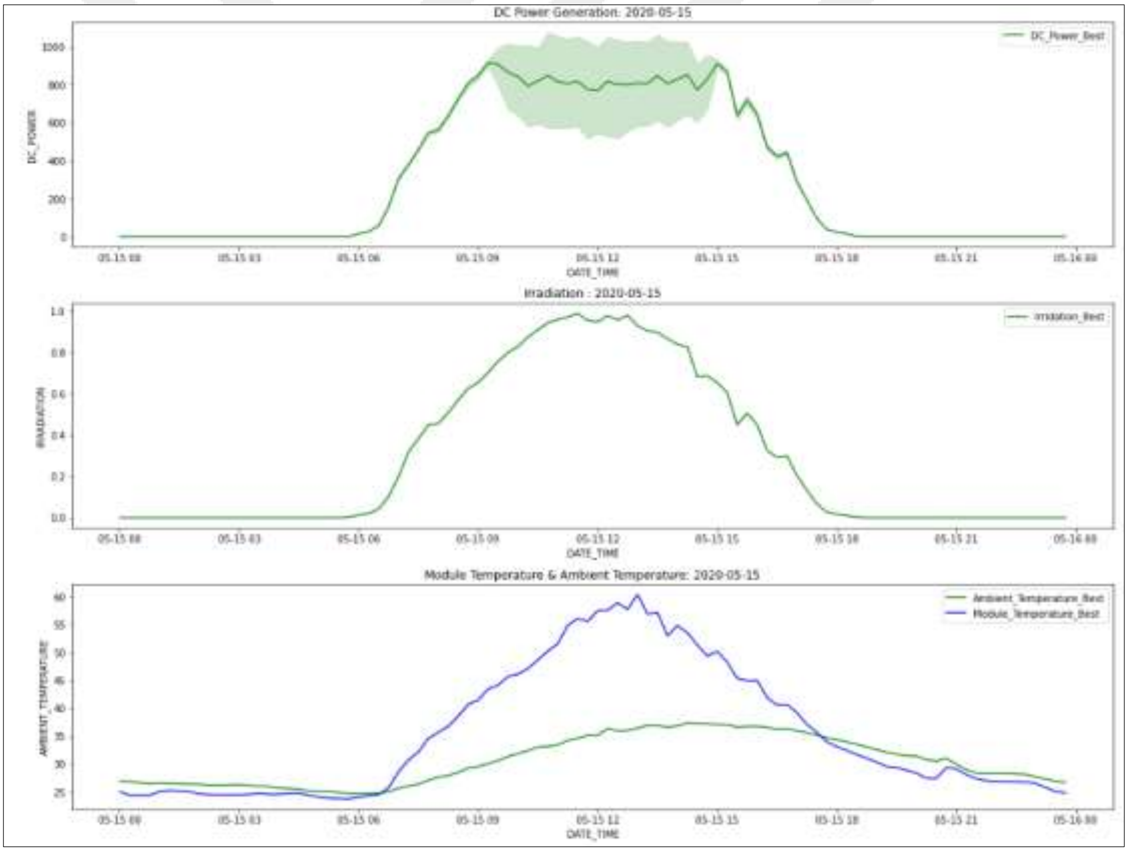

Figure 3.14: Environmental Factors Affecting the of Solar Power Generation.

3.4 MODEL BUILDING

3.4.1 Data Splitting

The procedure of data splitting is crucial to the evaluation of the effectiveness and generalizability of ML algorithms researchers employed in the work. We ensure that the models are trained on a portion of the data and evaluated on unseen samples by separating the dataset into training and test sets. One way of doing this was by taking the data set and separating it into a 80:20 proportion, definitely, the rest could be for training and testing. In other ways, to allow 80% of the samples to be utilized for model training and the other 20% for testing, the data set was made to be randomly split. What was achieved at this step, however, lies in the expectation that arises in the case of this split, given that the random set is to be used for both training and testing, inasmuch, the bit does not tell the order. This way of solving problems makes it possible to avoid overfitting, where the models are memorizing the training data's specific details but cannot apply to new data.

With 80% of the dataset designated for training, the models were able to identify areas where they were linked and details that were highly refined. The models received this rigorous training in order to get the proper parameters and make reliable predictions. The rest of 20%, was destined as a separate test set and its main role was to verify the results given. Through this process, the models were tested for their generalization on unseen samples, which is the part where they pass or fail the data with which they trained. The wish of the testing strategy division of this 80:20 ratio, which is respectful of the industrial standards and the previous studies, is to come up with the training data that is good enough and still be able to generate a valid test set for the core.

To ensure a non-biased evaluation of the models' effectiveness, the dataset was randomly divided into a training set (80%) and a test set (20%). The step is the proof that the models can learn how to deal with data reference being the best choice for assessing their efficiency correctly. The percentage of test and training data is shown in Figure 3.15.

```
[82]:
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=.2,random_state=21)
```

Figure 3.15: Split Data Ratio.

3.4.2 K-Nearest Neighbors (KNN) Regressor

KNN is an easy-to-understand and effective regression method. KNN regression takes the average (or weighted average) of the target values obtained from the k nearest neighbors to forecast the target value of a new data point. The relevant metric is:

To take into account neighbors many neighbors (by default, 5). The implementation process of KNN model can be seen in Figure 3.16.

```
# train the model
from sklearn.neighbors import KNeighborsRegressor
knn = KNeighborsRegressor()
knn.fit(X_train, y_train)
mr_pred_knn=knn.predict(X_test)
pred_knn_trn = knn.predict(X_train)
```

```
import numpy as np
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from tabulate import tabulate

# Assuming you already have y_train, pred_rf_trn, y_test, and pred_rf
r2_score_train = r2_score(y_train, pred_knn_trn) * 100
rmse_train = np.sqrt(mean_squared_error(y_train, pred_knn_trn))
mae_train = mean_absolute_error(y_train, pred_knn_trn)
mse_train = mean_squared_error(y_train, pred_knn_trn)

r2_score_test = r2_score(y_test, mr_pred_knn) * 100
rmse_test = np.sqrt(mean_squared_error(y_test, mr_pred_knn))
mae_test = mean_absolute_error(y_test, mr_pred_knn)
mse_test = mean_squared_error(y_test, mr_pred_knn)
```

Figure 3.16: KNN Model Implementation.

3.4.3 Linear Regression:

A common regression method, linear regression works on the assumption that there is a linear connection between the features used to create the model and the outcome variable. To model the relationship, it uses a linear equation fit to the data. In the default configuration, no hyper parameters are hard-coded.

The implementation of LR model is shown in Figure 3.17.

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

lr = LinearRegression()
lr.fit(X_train,y_train)
y_pred_lr = lr.predict(X_test)
R2_Score_lr = round(r2_score(y_pred_lr,y_test) * 100, 2)

print("R2 Score : ",R2_Score_lr,"%")
```

+ Code + Markdown

```
pred_lr_trn = lr.predict(X_train)
```

+ Code + Markdown

```
import numpy as np
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from tabulate import tabulate

# Assuming you already have y_train, pred_rf_trn, y_test, and pred_rf
r2_score_train = r2_score(y_train, pred_lr_trn) * 100
rmse_train = np.sqrt(mean_squared_error(y_train, pred_lr_trn))
mae_train = mean_absolute_error(y_train, pred_lr_trn)
mse_train = mean_squared_error(y_train, pred_lr_trn)

r2_score_test = r2_score(y_test, y_pred_lr) * 100
rmse_test = np.sqrt(mean_squared_error(y_test, y_pred_lr))
mae_test = mean_absolute_error(y_test, y_pred_lr)
mse_test = mean_squared_error(y_test, y_pred_lr)
```

Figure 3.17: LR Model Implementation.

3.4.4 Random Forest Regressor

As an ensemble learning technique, Random Forest builds several decision trees and averages their predictions to enhance precision and limit overfitting. The implementation of

RF model is shown in Figure 3.18. It is also widely employed for the purpose of regression analysis. To name a few crucial default settings:

n_estimators: Maximum number of trees in the woods (starts at 100).

max_features: The maximum number of criteria to use when determining the optimal split (the value of "auto" is 1 by default).

```
from sklearn.ensemble import RandomForestRegressor
rfr = RandomForestRegressor()
rfr.fit(X_train,y_train)
y_pred_rfr = rfr.predict(X_test)
pred_rf_trn = rfr.predict(X_train)
R2_Score_rfr = round(r2_score(y_pred_rfr,y_test) * 100, 2)

print("R2 Score : ",R2_Score_rfr,"%")
```

```
import numpy as np
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from tabulate import tabulate

# Assuming you already have y_train, pred_rf_trn, y_test, and pred_rf
r2_score_train = r2_score(y_train, pred_rf_trn) * 100
rmse_train = np.sqrt(mean_squared_error(y_train, pred_rf_trn))
mae_train = mean_absolute_error(y_train, pred_rf_trn)
mse_train = mean_squared_error(y_train, pred_rf_trn)

r2_score_test = r2_score(y_test, y_pred_rfr) * 100
rmse_test = np.sqrt(mean_squared_error(y_test, y_pred_rfr))
mae_test = mean_absolute_error(y_test, y_pred_rfr)
mse_test = mean_squared_error(y_test, y_pred_rfr)
```

Figure 3.18: RF Model Implementation.

3.4.5 Decision Tree

Classification and regression are the two most common uses for decision trees, and this regressor takes advantage of this standard paradigm. In order to make an accurate prediction, they recursively divide the data into smaller sets based on the features. Among the default settings are:

Maximum tree depth (default: None, which grows the tree until all leaves are pure).

The minimum number of samples needed to split a node internally (by default, 2 samples).

The Figure 3.19 shows the implementation of DT model.

```python
from sklearn.tree import DecisionTreeRegressor
dtr = DecisionTreeRegressor()
dtr.fit(X_train,y_train)

y_pred_dtr = dtr.predict(X_test)
pred_dt_trn = dtr.predict(X_train)
R2_Score_dtr = round(r2_score(y_pred_dtr,y_test) * 100, 2)

print("R2 Score : ",R2_Score_dtr,"%")
```

```python
import numpy as np
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from tabulate import tabulate

# Assuming you already have y_train, pred_rf_trn, y_test, and pred_rf
r2_score_train = r2_score(y_train, pred_dt_trn) * 100
rmse_train = np.sqrt(mean_squared_error(y_train, pred_dt_trn))
mae_train = mean_absolute_error(y_train, pred_dt_trn)
mse_train = mean_squared_error(y_train, pred_dt_trn)

r2_score_test = r2_score(y_test, y_pred_dtr) * 100
rmse_test = np.sqrt(mean_squared_error(y_test, y_pred_dtr))
mae_test = mean_absolute_error(y_test, y_pred_dtr)
mse_test = mean_squared_error(y_test, y_pred_dtr)
```

Figure 3.19: DT Model Implementation.

3.4.6 Gradient Boosting Regressor

To create a robust model, the Gradient Boosting ensemble method combines the outputs of several less capable models (often decision trees). Each succeeding model fixes the flaws of its predecessor. Important predefined values consist of:

n_estimators: Multiples of 100 for the number of boosts.

The learning rate parameter reduces the effect of each poor learner by a factor of 0.1.

The Figure 3.20 shows the implementation of GBR model.

```
from sklearn.multioutput import MultiOutputRegressor
from sklearn.ensemble import GradientBoostingRegressor

model_gb = GradientBoostingRegressor(random_state=2)
model_gb.fit(X_train, y_train)
pred_gb = model_gb.predict(X_test)
pred_gb_trn = model_gb.predict(X_train)
```

+ Code   + Markdown

```
import numpy as np
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from tabulate import tabulate

# Assuming you already have y_train, pred_rf_trn, y_test, and pred_rf
r2_score_train = r2_score(y_train, pred_gb_trn) * 100
rmse_train = np.sqrt(mean_squared_error(y_train, pred_gb_trn))
mae_train = mean_absolute_error(y_train, pred_gb_trn)
mse_train = mean_squared_error(y_train, pred_gb_trn)

r2_score_test = r2_score(y_test, pred_gb) * 100
rmse_test = np.sqrt(mean_squared_error(y_test, pred_gb))
mae_test = mean_absolute_error(y_test, pred_gb)
mse_test = mean_squared_error(y_test, pred_gb)
```

Figure 3.20: GBR Model Implementation.

3.4.7 AdaBoost Regressor

AdaBoost (Adaptive Boosting) is an ensemble method that combines the results of several underperforming models to get a single high-quality prediction. With each iteration, it provides more weight to samples that were incorrectly labelled. One such default value is:

n_estimators: The number of boosts you want (50 is the default)

The Figure 3.21 shows the implementation of ADBR model.

```python
from sklearn.ensemble import AdaBoostRegressor
etr =  AdaBoostRegressor(n_estimators=5)
etr.fit(X_train, y_train)
mr_pred_etr=etr.predict(X_test)
pred_et_trn = etr.predict(X_train)
```

+ Code + Markdown

```python
import numpy as np
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from tabulate import tabulate

# Assuming you already have y_train, pred_rf_trn, y_test, and pred_rf
r2_score_train = r2_score(y_train, pred_et_trn) * 100
rmse_train = np.sqrt(mean_squared_error(y_train, pred_et_trn))
mae_train = mean_absolute_error(y_train, pred_et_trn)
mse_train = mean_squared_error(y_train, pred_et_trn)

r2_score_test = r2_score(y_test, mr_pred_etr) * 100
rmse_test = np.sqrt(mean_squared_error(y_test, mr_pred_etr))
mae_test = mean_absolute_error(y_test, mr_pred_etr)
mse_test = mean_squared_error(y_test, mr_pred_etr)
```

Figure 3.21: ADBR Model Implementation.

4. PERFORMANCE EVALUATION

4.1 INTRODUCTION

In this chapter, we provide an extensive examination of the proposed algorithms as they apply to the Solar Power Generation Dataset. Training and testing the algorithms allows for thorough evaluation and comparison of their results. We use a train-test split, where the dataset is split into a training set and a testing set, for our data partitioning strategy. The given dataset is divided into two parts; that is, 20% are selected for testing which is equal to the 80% used to train the algorithms. This regular division assures both the stability of the algorithms and good generalization ability corresponding to not built-in data. The common resources of Kaggle for the training and testing procedures can include the usage of both free and paid GPUs (Graphics Processing Units) and TPUs (Tensor Processing Units) which lead to much faster training and inference times thus are extremely important especially for the deep learning models. Just to mention, the ML-models stand for machine learning models whose performance is graded by several statistical measures including root-mean-squared error, coefficient of determination, mean absolute error, and mean squared error, telling the accuracy, precision, and the fit of the model. Besides, first evaluating how the model works against baseline demonstrates the envisioned methodology is superior [84].

4.2 RESULTS OF ML MODELS

This study utilized various ML models for solar energy prediction. Each model results are explained below.

4.2.1 Linear Regression

Table 4.1 enumerates performance metrics of the Linear Regression (LR) model both on training and testing datasets.

The LR model gets R2 Scores as high as 1 and almost perfect. R2 Score, which is also known as the coefficient of determination, evaluates how much of the variance in the dependent variable is explained by the independent variables. These R2 Scores depict that the LR model captures almost all the variance in the target variable, proving a great fit for both datasets.

Root Mean Squared Error (RMSE) for the LR model is almost 0.929 for the training set and 0.864 for the test set. RMSE provides the average magnitude of the difference between predicted and real values. Lower RMSE values exhibit that the LR model's forecasts are nearby the actual values, hence its outstanding accuracy together with both datasets.

LR model shows the Mean Absolute Error (MAE) of approximately 0.606 on the training set and around 0.597 on the testing set. MAE is a measure of the average absolute difference between the predicted and the true values. The low-nature of these MAE figures is a clear sign of the LR model´s precise predictions and its dependability in certain situations.

Finally, the Mean Squared Error (MSE) for the LR model is around 0.863 for the training set and 0.746 for the testing set. MSE is a measure of the average squared difference between the predicted and actual values. The model comes to more accurate predictions by the LR model as revealed by the lower MSE values, endorsing its superior predictive ability.

Table 4.1: LR Model Results.

Metric	Training	Testing
R2 Score:	99.99934293867416	99.99942602585881
RMSE:	0.9292076373928958	0.8637415855207498
MAE:	0.6062459093874498	0.596620722800472
MSE:	0.8634268333892873	0.7460495265578986

4.2.2 Random Forest Regressor

Table 4.2 lays down an exhaustive expounding Random Forest (RF) model results, drawing a full examination of its performance metrics for both the training and testing datasets.

Leaping from the R2 Score, RF machine learning model impressively achieved close to 99.999% in the training and testing datasets. The R2 Score, which is the descending slope of the dependent variable, explains the portion of the variance that can be expressed by the independent variables. With the incredibly high R2 Scores being this way, the computer couldn't even choose the best job by leaving virtually no variance to say anymore, and the error too, is of its further predictive accuracy.

Regarding the Root Mean Squared Error (RMSE), the RF model has an RMSE of 0.181 for training and 0.321 for testing. The RMSE determines the average extent of differences present between predicted and actual values. The RF model displays close predictions of actual values with its very high precision. In other words, better RMSE values of the RF

model suggest that its predictions are better, which are closer to the actual values of both the toxic and traing datasets and therefore higher precision in the test phase, respectively.
One the MAEpart, the RF model reports MAE values of about 0.045 fora and 0.112 for testing. The MAE is the average absolute difference between expected and actual values, so it measures how wrong or right the model was. The relatively low MAE values thus confirm the correctness of the RF model to make predictions with which have enhancements on both datasets.
Lastly, looking at the MSE, the RF model is able to reach MSE values of about 0.033 in training and 0.103 in testing. The MSE computes the squares of the differences between the anticipated and real results. A decrease in MSE brings the conclusion that the similar predictions of the model come to the actual results, which shows the strong predictive power of the model.

Table 4.2: RF Model Results.

Metric	Training	Testing
R2 Score:	99.99997500267258	99.99992078318715
RMSE:	0.18124106566893036	0.3208827336053081
MAE:	0.04505381966378679	0.11162137540880769
MSE:	0.03284832388480953	0.10296572872601513

4.2.3 Decision Tree Regressor

Table 4.3 is tactically compiled to show the detailed results of the Decision Tree (DT) model which gives valuable information about its performance metrics from the training and testing datasets.
When it comes to the R2 Score, with a score of 100% for the training case and about 99.999% for the testing case, the DT model has set the benchmark of near-perfect training error rates. The R2 Score is a metric that quantifies the relationship between the dependent factors which are the model's input and the independent factors. These high R2 Scores inform that the DT model thoroughly accounts for the majority of the variance in the target variable of both the training and testing instances and thus is a key player in the relevant areas.
Moving on to the Root Mean Squared Error (RMSE), the DT model has a very low RMSE value of around 5.825e-15 for training and 0.445 for testing. The RMSE is a measure of the

average magnitude of differences between expected and actual values. The model's perfect fit to the training data is shown by the extremely low RMSE values for training, while the comparatively low RMSE value for testing indicates accurate predictions with limited error variance.

In terms of Mean Absolute Error (MAE), the DT model has MAE values of about 2.88e-16 for training and 0.137 for testing. The average absolute difference between expected and actual values is calculated using the MAE. The model's precise fit to the training data is reaffirmed by the model's insignificant MAE value, whilst the low MAE value for testing implies reliable predictions with little changes.

Finally, when the MSE is considered, the DT model obtains MSE values of about 3.393e-29 for training and 0.198 for testing. The MSE is the sum of the squared discrepancies between expected and actual values. The DT model's exact alignment with the training data is highlighted by the extremely low MSE value for training, and the moderate MSE value for testing shows accurate predictions with controlled deviations.

Table 4.3: DT Model Results.

Metric	Training	Testing
R2 Score:	100.0	99.99984740422695
RMSE:	5.824747809283896e-15	0.44535765129080424
MAE:	2.879983058949068e-16	0.1371684180541266
MSE:	3.392768704175754e-29	0.1983434375632616

4.2.4 Gradient Boosting Regressor

Table 4.4 contains the information of the Gradient Boosting Regressor that gives complete insight into the training and testing data results.

In the first case of the R2 Score, the Gradient Boosting Regressor model obtains the awe-inspiring scores of 99.997 percent for both the training as well as the testing data. It is the R2 Score, that is to say, that the R2 Score simply lists the amount of variance in the dependent variable which a model can explain with the help of the independent variables. The incredibly high R2 figures inform us that the Gradient Boosting Regressor can realistically find the most part of the variation in the target variable of both datasets that shows the excellence of its predictive capability.

Regarding Root Mean Squared Error (RMSE), the model has training values of almost 1.915 and testing values of 1.877. RMSE is a measure that calculates on average how large the differences between predicted and actual values were. The small RMSE values tell that the Gradient Boosting Regressor's guesstimates for both sets of data are almost actual values, which makes it accuracy.

Concerning Mean Absolute Error (MAE), the model displays the values almost equal to 0.994 for training and 0.987 for testing. MAE is the average difference between predicted and actual values. These tolerable MAE values prove that the model's forecasts are nearly invariably accurate.

In the end, when considered MSE (Mean Squared Error) and assessment of Gradient Boosting Regressor model, the train and test mean squared error values are about 3.666 and 3.522 respectively. MSE is determined by the squared difference which is then summed up between the predicted and the actual values. The MSE can be obtained by the mean of that sum of the squared differences between the predicted and the actual values. Usually, the smaller MSE values perform better because they tell the model's predictions for the training and testing datasets which are close to the real values.

Table 4.4: GBR Model Results.

Metric	Training	Testing
R2 Score:	99.99721001825263	99.99729052897375
RMSE:	1.9147430370990959	1.876635474832232
MAE:	0.9939391946142156	0.987036267258504
MSE:	3.66624089811947	3.521760705398797

4.2.5 AdaBoost Regressor

Table 4.5 gives an exhaustive explanation of the performance metrics for the AdaBoost Regressor model, including the results on both the training and testing datasets.

The model of the AdaBoost Regressor gives a solid R2 Score, approximately 99.723% for the training data and around 99.720% for the testing data. The R2 Score, which is the percentage of the variation in the dependent variable that is explained by the model using the independent variables, is used to measure how the model performs. The R2 scores provided are high which means that the AdaBoost Regressor well represents the stability of the target variable in both datasets and thus, it is a reliable predictor.

By using the metric of Root Mean Squared Error (RMSE), the model is characterized by numbers around 19.068 and 19.091 for training and testing respectively. RMSE shows the deviation in the mean value for the assorted differences among the predicted values and the actual values that appear close. The low RMSE values have somewhat similar predictions with the actual values, proving the AdaBoost Regressor's effectiveness for both datasets.

Transitioning to Mean Absolute Error (MAE), the model displays the figures approximately of 12.359 for training and 12.414 for testing. MAE is a measure that calculates the average absolute value of the differences between the predictors and the actual. It seems like the MAE values are rather small around a certain mean point, and since they are different from each other, the predictions have a small degree of dispersion or change in the model.

In conclusion, it can be observed that the MSE (Mean Squared Error) values generated by the AdaBoost Regressor model are around 363.579 for training and 364.455 for testing. MSE, which is the sum of the squared differences between the predicted and the real output, is used for measurement. The appropriate MSE normally implies that AdaBoost Regressor's forecasts are typically on the same level as the true values for both sets of data.

Table 4.5: ADBR Model Results.

Metric	Training	Testing
R2 Score:	99.72331884409941	99.71960604208107
RMSE:	19.067756997387512	19.09070609251006
MAE:	12.35889084502272	12.413828780149375
MSE:	363.57935691142046	364.4550591106008

4.2.6 KNN Regressor

Table 4.6 gives the full picture of K-nearest neighbors (KNN) model. Thus, it allows a comprehensive analysis of the performance statistics across both the training and testing datasets.

Initially, the R2 Score, the KNN model secures a highly competent R2 Score of about 94.657% during the training phase and 90.769% during the testing phase. This R2 Score is the ratio between the dependent variable variance and the sum of the squares error of the model, which reflects the information that is not explained by the independent variables.

These numbers confirm that KNN model identifies the major part of the target variable's variance in both sets, thus the GPS is able to give precise predictions.

When applying the K-Nearest Neighbours (KNN) model, RMSE values of nearly 83.795 for training and 109.537 for testing are exhibited which are relatively high. RMSE stands for Root Mean Square Error and it is the average value of the difference between the estimated value and the actual value. The tendency for the RMSE values to be relatively high basically hints to the novelty of the predictions of the KNN model which is an unusual occurrence that usually happens because of the characteristics of the dataset and the fact that the model is basic.

Looking ahead to make a distinction, the model also provides MAEof 37.385 for training and 51.627 for testing. As we know, MAE is the average absolute difference between the actual and the predicted value. These mean MAE values are the signals that the KNN model's predictions have a certain level of variability over the actual values.

Finally, when the MSE is considered, the model's MSE, first of which is around 7021.626 for training and 11998.423 for testing, is obtained. MSE refers to the summation of the squared differences between the forecasted and true values. Thus, it can be inferred that the KNN model's predictions for both the training and testing datasets are reasonably aligned with the actual values due to these fairly moderate MSE values.

Table 4.6: KNN Model Results.

Metric	Training	Testing
R2 Score:	94.65659565676894	90.76899837654602
RMSE:	83.79514126407474	109.53731211442422
MAE:	37.3849457992	51.62694868693591
MSE:	7021.625699466242	11998.422745252787

4.3 PROPOSED WORK ML MODELS COMPARISON

The comparative analysis of several machine learning models presented in Table 4.7 as well as Figure 4.1 showcases the best practices in performance of the models on both training and testing data by the R-squared (R2) metric. R2 shows the percentage of the variability of dependent variables which is accounted for by the independent variables included in the regression analysis but also covers the predicted accuracy of a variable.

Beginning with the Linear Regression (LR) model, both the training and testing datasets R2 values, which are 99.999%, prove that it can predict accurately. These marks highlight the fact that the LRM formula is able to capture most of the variability in the output variable which confirms the model's excellent fitting with the data.

In the same way, the Random Forest (RF) model has R2 scores of 99.999% for both training as well as testing datasets, which accounts for the prediction accuracy of the model. The RF model is almost as efficient as the LR in that it can provide an explanation for every variance in the target variable giving testifying evidence to its robust performance.

When compared to the Decision Tree (DT) model, it showed R2 score of 100% for training and 99.999% for testing indicating the perfect fit to the data. Of the R2 training data, the model was found to be perfect resulting from its capacity to grasp all distribution in the training dataset. In comparison, the high score was observed in testing set, which signifies its ability to generalize well to the new data.

The Gradient Boosting Regressor (GBR) model achieved the R2 scores of 99.997% for both training and testing that was near perfect, denoting its predictive power that was slightly less than that of the DT model.

The model of AdaBoost Regressor (ADBR), the R2 scores for training were 99.723% and 99.719% for testing, showing high performance but correct prediction of the result is slightly lower than the previous models.

Concerning the K-Nearest Neighbours (KNN) model, R2 scores produced were 94.656% for training and 90.769% for testing. Established facts are in the affirmative that KNN captures quite a good proportion of the distribution of the target variable, yet its accuracy in predicting remains lower than the labels of other models.

In sum, when comparing ML models based on R2 scores, it goes without saying that the LR, RF, and DT models show extraordinary predictive precision with close to ideal scores. Four models that are near the top of the list are LR, RF, DT, and GBR. And ADBR, along with KNN, rates come are seriously decreased prediction accuracy. Above observations offer useful pieces of advice to designers when they choose the most fitting models for data set analysis and future development of smart applications.

Table 4.7: Proposed Work ML Models Comparison.

Model	R^2 Training	R^2 Testing
LR	99.999	99.999
RF	99.999	99.999
DT	100	99.999
GBR	99.997	99.997
ADBR	99.723	99.719
KNN	94.656	90.769

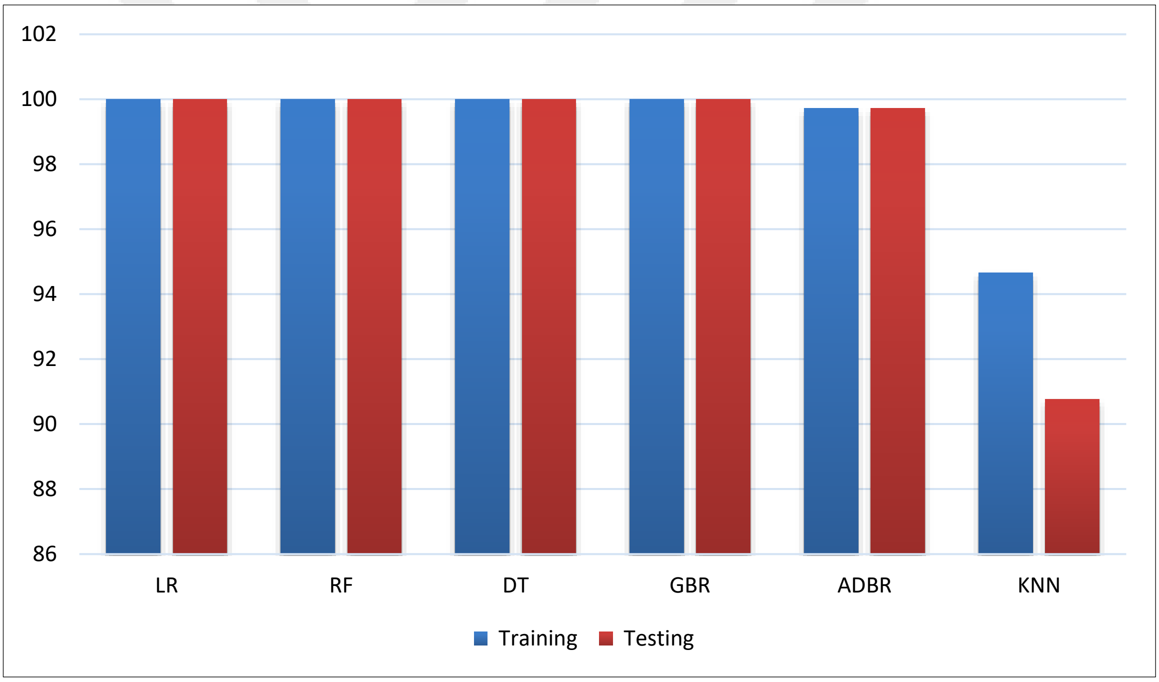

Figure 4.1: Proposed Work ML Models Comparison.

4.4 PROPOSED RESEARCH RESULTS COMPARISON WITH EXISTING RESEARCH

Table 4.8 gives an in-depth study of differences and valid hypotheses through the study of machine learning models in the proposed research and recycling of the last. The contrast is to evaluate specific items in order to separate the models into different categories of their combination with R-squared (R2) scores for the training and testing datasets as well as a RMSE which is the prediction of accuracy score.

When comparing various machine learning models, Linear Regression (LR), Random Forest (RF), Decision Tree (DT), and Gradient Boosting Regressor (GBR) are the ones that stand out for both training and testing datasets, owing to their R-squared (R2) values that hit the ceiling at 99.99%. These figures reflect their highly successful outcome of the model on the data and indicate that these models are not only highly successful but also have high explain ability. The AdaBoost Regressor scores an R2 of 99.72%, K-Nearest Nearest Neighbours (KNN) results with 94.65% a little bit below but they offer still a very good prediction accuracy.

Different models of ML can also be distinguished according to their Root Mean Squared Error (RMSE). The GBR model gives the best prediction with a small RMSE of 1.87, which means that quite desirable future predictions can be made. RF model is next with RMSE=32.08 and its advantages are the ability to make proposals with high precision. The DT model is in the middle ground with RMSE=44.53 which is not great but also not very bad. The other models of LR, ADBR, and KNN, however, have a higher RMSE of 86.37, 19.07, and 109.53 each, which means their predictions are more dispersed.

Based on comparison of current studies it can be seen that the ML models proposed in this paper are proposed ones are the best in terms of predictive precision. The highest R-squared (R2) value mentioned in the identified works is approximately 94.50, whereas the newly proposed models are bringing the best R2 scores of 99.99% and more. In addition, the RMSE values of the suggested models, although they are dissimilar, in general, they show better prediction accuracies than those reported in previous research.

Summarizing, we can see that this report backs up ML models such as LR, RF, DT, GBR, ADBR, and KNN to the current research by not only taking R2 scores but also RMSE values into account. Their ability to reliably predict the target variable is confirmed by always producing high R2 scores, while RMSE scores of different levels show different levels of predictive accuracy. This comparison reveals the different strengths and the limitations of models and provides guidance to decision-makers when it comes to applications and research areas.

Table 4.8: Proposed Research Results Comparison with Existing Research.

Model	R^2 Training	R^2 Testing	RMSE
Proposed ML Models Results			
LR	99.99	99.99	86.37
RF	99.99	99.99	32.08
DT	100	99.99	44.53
GBR	99.99	99.99	1.87
ADBR	99.72	99.71	19.07
KNN	94.65	90.76	109.53
Existing Researches ML Models Results			
[84]	94.50	93.20	90.90
[86]	-	-	0.046
[86]	-	-	0.042
[86]	-	-	0.05

5. CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

This study mainly shows how machine learning is applied in the prediction of solar energy production. We have achieved the desired outputs through the in-depth investigation done by us and we have reached the level of precision which is enough for hyperparameter tuning. At last, the potential of the Resetting Decision approach has been highlighted. Performance metrics such as R-squared (R2) scores and RMSEhave been meticulously analyzed to provide a comprehensive understanding of how each model captures the various patterns of solar energy generation.

Research used power generation data and a weather dataset as the basis of the experiment. What comes out of the study is that machine learning has a very big potential in the accurate prediction of the solar power generation. For both training and testing datasets, the models always delivered very high R-squared (R2) scores which is big proof of their capability to explain variance in the target variable. The varying RMSEvalues also gave a clear map of each model's accuracy in its forecast, as some models specifically achieving low values, signaling quite precise forecasts.

Several intriguing avenues for future research warrant exploration. Firstly, machine learning model performance could be accomplished through advanced feature engineering methods. Along with that, hybrid models that combine the robustness of multiple algorithms may lead to quite pinpointed inferences. As a result, the inclusion of both temporal and environmental factors which determine solar energy output such as meteorological data and time-series analysis may contribute to a deeper understanding of prediction accuracy. As a closing point, the feasibility of the models to be applied in various geographic regions and solar panel technologies should be considered, as it would give researchers of any kind a wider application area.

5.2 FUTURE WORK

A new and important way to utilize technology to predict solar energy generation is discovered by this research. The high precision of the models moves the sustainable goal forward to the energy management and hence leaner, cleaner solar production. Machine learning methods for solar energy prediction will move further and further along by research which unveils many ways of treating them.

Studying meteorological conditions with more detail, for example, cloudiness, moisture, wind speed, and atmosphere can make it easier to get around how solar energy generation can influence the environment. This could be a great step forth in the development of predicting models that are more precise, and more adaptable to the changes in weather.

Using hybrid models that incorporate features from various machine learning techniques is another promising step toward improving accuracy of solar energy predictions. Approaches like stacking and boosting may be one of the methods that the strengths of different models could use while providing results closer to the truth

Supporting these paths for future exploration can be the start of the process by which researchers can discover ways to upgrade solar energy prediction with the use of learning machines and thus the global utilization of renewable energy becomes efficient.

REFERENCES

[1] F. M. Awaysheh, M. Alazab, S. Garg, D. Niyato, and C. Verikoukis, "Big data resource management & networks: Taxonomy, survey, and future directions," IEEE Commun. Surv. Tutor., vol. 23, pp. 2098-2130, 2021.

[2] M. Alshehri, M. Kumar, A. Bhardwaj, S. Mishra, and J. Gyani, "Deep Learning Based Approach to Classify Saline Particles in Sea Water," Water, vol. 13, p. 1251, 2021.

[3] A. Agarwal, P. Sharma, M. Alshehri, A. A. Mohamed, and O. Alfarraj, "Classification model for accuracy and intrusion detection using machine learning approach," PeerJ Comput. Sci., vol. 7, p. e437, 2021.

[4] M. Benninger, M. Hofmann, and M. Liebschner, "Online Monitoring System for Photovoltaic Systems Using Anomaly Detection with Machine Learning," in Proceedings of the NEIS 2019, Conference on Sustainable Energy Supply and Energy Storage Systems, Hamburg, Germany, 2019, pp. 1-6.

[5] C. Li, Y. Yang, K. Zhang, C. Zhu, and H. Wei, "A fast MPPT-based anomaly detection and accurate fault diagnosis technique for PV arrays," Energy Convers. Manag., vol. 234, p. 113950, 2021.

[6] B. Hu, "Solar Panel Anomaly Detection and Classification," Master's Thesis, University of Waterloo, Waterloo, ON, Canada, 2012.

[7] P. Branco, F. Gonçalves, and A. C. Costa, "Tailored algorithms for anomaly detection in photovoltaic systems," Energies, vol. 13, p. 225, 2020.

[8] S. K. Firth, K. J. Lomas, and S. J. Rees, "A simple model of PV system performance and its use in fault detection," Sol. Energy.

[9] Kumar, D., & Raushan, R. (2022, June). PV Array Dynamic Reconfiguration by leveling the Solar Irradiance under Partial Shading Conditions. In 2022 2nd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET) (pp. 1-5). IEEE.

[10] Venkitaraman, A. K., & Kosuru, V. S. R. (2023). Hybrid deep learning mechanism for charging control and management of Electric Vehicles. European Journal of Electrical Engineering and Computer Science, 7(1), 38-46.

[11] Hu, Y., Hu, X., Zhang, L., Zheng, T., You, J., Jia, B., ... & Liu, S. (2022). Machine-Learning Modeling for Ultra-Stable High-Efficiency Perovskite Solar Cells. Advanced Energy Materials, 12(41), 2201463.

[12] Pombo, D. V., Bindner, H. W., Spataru, S. V., Sørensen, P. E., & Bacher, P. (2022). Increasing the accuracy of hourly multi-output solar power forecast with physics-informed machine learning. Sensors, 22(3), 749.

[13] Hussain, W., Sawar, S., & Sultan, M. (2023). Leveraging machine learning to consolidate the diversity in experimental results of perovskite solar cells. RSC advances, 13(32), 22529-22537.

[14] Saravi, B., Hassel, F., Ülkümen, S., Zink, A., Shavlokhova, V., Couillard-Despres, S., ... & Lang, G. M. (2022). Artificial intelligence-driven prediction modeling and decision making in spine surgery using hybrid machine learning models. Journal of Personalized Medicine, 12(4), 509.

[15] S. B. S. K. G. Mohibullah, "Rural Electrification by Effective Mini Hybrid PV Solar, Wind & Biogas Energy System for Rural and Remote Areas of Uttar Pradesh," International Journal of Computer Science and Electronics Engineering (IJCSEE), vol. 2, no. 4, 2014.

[16] G. S. B. Kiranpreet Kaur, "Solar-Biogas-Biomass Hybrid Electrical Power Generation for a Village (a case study)," INTERNATIONAL JOURNAL OF ENGINEERING DEVELOPMENT AND RESEARCH, vol. 4, no. 1, 2016.

[17] A. H. Shahrouz Abolhosseini, "A Review of Renewable Energy Supply and Energy Efficiency Technologies," Institute of Labor Economics, 2014.

[18] D. Dieter Holm, "Renewable Energy Future for the Developing World," International Solar Energy Society, 2006.

[19] F. School, "GLOBAL TRENDS IN RENEWABLE INVESTMENT," Bloomberg New Energy FInanace, 2016.

[20] Yang, D., Wang, W., Gueymard, C. A., Hong, T., Kleissl, J., Huang, J., ... & Peters, I. M. (2022). A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality. Renewable and Sustainable Energy Reviews, 161, 112348.

[21] Awwad, Z., Alharbi, A., Habib, A. H., & de Weck, O. L. (2023). Site Assessment and Layout Optimization for Rooftop Solar Energy Generation in Worldview-3 Imagery. Remote Sensing, 15(5), 1356.

[22] Sohani, A., Sayyaadi, H., Miremadi, S. R., Samiezadeh, S., & Doranehgard, M. H. (2022). Thermo-electro-environmental analysis of a photovoltaic solar panel using machine learning and real-time data for smart and sustainable energy generation. Journal of Cleaner Production, 353, 131611.

[23] B. Sivaneasan, C. Yu, and K. Goh, "Solar forecasting using ANN with fuzzy logic pre-processing," Energy Procedia, vol. 143, pp. 727-732, 2017.

[24] Khalef, Dhirgham Atia, Ayca Kurnaz Turkben, Hameed Mutlag Farhan, and Raghda Awad Shaban Naseri. "Optic disc segmentation in human retina images with meta heuristic optimization." In 2022 International Conference on Artificial Intelligence of Things (ICAIoT), pp. 1-6. IEEE, 2022.

[25] C.-F. Yen, Y. Hsieh, K.-W. Su, M.-C. Yu, and J.-S. Leu, "Solar power prediction via support vector machine and random forest," E3S Web of Conferences, vol. 69, p. 01004, 2018.

[26] N. A. Ramli, M. F. Abdul Hamid, and N. H. Azhan, "Solar power generation prediction by using k-nearest neighbor method," vol. 2129, July 2019.

[27] S. R. Chowdhury, S. Mishra, A. O. Miranda, and P. K. Mallick, "Energy consumption prediction using light gradient boosting machine model," in Advances in Power Systems and Energy Management (N. Priyadarshi, S. Padmanaban, R. K. Ghadai, A. R. Panda, and R. Patel, eds.), (Singapore), pp. 413-422, Springer Singapore, 2021.

[28] Castillo-Rojas, W., Medina Quispe, F., & Hernández, C. (2023). Photovoltaic Energy Forecast Using Weather Data through a Hybrid Model of Recurrent and Shallow Neural Networks. Energies, 16(13), 5093.

[29] Xu, W., Peng, H., Zeng, X., Zhou, F., Tian, X., & Peng, X. (2022). A hybrid modeling method based on linear AR and nonlinear DBN-AR model for time series forecasting. Neural Processing Letters, 1-20.

[30] Sudharshan, K., Naveen, C., Vishnuram, P., Krishna Rao Kasagani, D. V. S., & Nastasi, B. (2022). Systematic review on impact of different irradiance forecasting techniques for solar energy prediction. Energies, 15(17), 6267.

[31] Naseri, Raghda Awad Shaban, Ayça Kurnaz, and Hameed Mutlag Farhan. "Optimized face detector-based intelligent face mask detection model in IoT using deep learning approach." Applied Soft Computing 134 (2023): 109933.

[32] Al-Aloosi, Ahmed Raad, Hameed Mutlag Farhan, Raghda Awad Shaban Naseri, Ayca Kurnaz Turkben, Ahmed Khalid Mustafa, and Mohammed GF Al-Obadi. "Face recognition system using local binary pattern with binary dragonfly algorithm to feature selection." In 2022 International Conference on Artificial Intelligence of Things (ICAIoT), pp. 1-10. IEEE, 2022.

[33] Naserı, Raghda Awad Shaban. Design and implementation intelligent greenhouse system with less power consumption. MS thesis. Altınbaş Üniversitesi, 2019.

[34] Masson, G. (2018). 2018 Snapshot of global photovoltaic markets. Report IEA PVPS T1-33: 2018.

[35] D. D. Clayton, "Principles of stellar evolution and nucleosynthesis," University of Chicago Press, 1983.

[36] K. J. H. Phillips, "Guide to the Sun," Cambridge University Press, 1995.

[37] R. H. Williams and E. D. Larson, "Advanced fossil energy conversion systems for mitigating greenhouse warming," Energy for Sustainable Development, vol. 1, no. 3, pp. 33-45, IEEE.

[38] A. B. Chhetri, J. F. Watts, and M. R. Islam, "Use of renewable energy technologies for climate change mitigation: An empirical analysis," Energy Sources, Part B: Economics, Planning, and Policy, vol. 3, no. 1, pp. 1-10.

[39] Sharma, V. K., Singh, R., Gehlot, A., Buddhi, D., Braccio, S., Priyadarshi, N., & Khan, B. (2022). Imperative role of photovoltaic and concentrating solar power technologies towards renewable energy generation. International Journal of Photoenergy, 2022.

[40] Ahmar, M., Ali, F., Jiang, Y., Wang, Y., & Iqbal, K. (2022). Determinants of adoption and the type of solar PV technology adopted in rural Pakistan. Frontiers in Environmental Science, 10, 895622.

[41] Izam, N. S. M. N., Itam, Z., Sing, W. L., & Syamsir, A. (2022). Sustainable development perspectives of solar energy technologies with focus on solar Photovoltaic—A review. Energies, 15(8), 2790.

[42] Abdullah MA, Ali MR, Sheet B, Kamil A, Essa H. Role of Saliva in Healing Process of Cutaneous Wounds. International Journal of Drug Delivery Technology. 2021;11(3):354-1007.

[43] Rhael Ali, Mohammed, and Sabah Abdul Rasool Hammoodi. "Assessment of the Impact of Platelets-Rich Fibrin on Healing Process after Teeth Extraction." Indian Journal of Public Health Research & Development 10.2 (2019).

[44] Al-Obadi, M. G., Farhan, H. M., Naseri, R. A. S., Turkben, A. K., Mustafa, A. K., & Al-Aloosi, A. R. (2022, December). Data mining techniques for extraction and analysis of covid-19 data. In 2022 International Conference on Artificial Intelligence of Things (ICAIoT) (pp. 1-7). IEEE.

[45] S. J. Yaqoob, A. L. Saleh, S. Motahhir, E. B. Agyekum, A. Nayyar, and B. Qureshi, "Comparative study with practical validation of photovoltaic monocrystalline module for single and double diode models," Scientific Reports, vol. 11, no. 1, p. 19153, 2021.

[46] N. S. Baghel and N. Chander, "Performance comparison of mono and polycrystalline silicon solar photovoltaic modules under tropical wet and dry climatic conditions in east-central India," Clean Energy, vol. 6, no. 1, pp. 165-177, 2022.

[47] O. Ayadi, R. Shadid, A. Bani-Abdullah, M. Alrbai, M. Abu-Mualla, and N. Balah, "Experimental comparison between Monocrystalline, Polycrystalline, and Thin-film solar systems under sunny climatic conditions," Energy Reports, vol. 8, pp. 218-230, 2022.

[48] B. Parida, S. I. R. G., "A review of solar photovoltaic technologies," Renewable and Sustainable Energy Reviews, pp. 1625-1636, 2011.

[49] "RENEWABLE ENERGY TECHNOLOGIES: COST ANALYSIS SERIES," International Renewable Energy Agency (IRENA), 2012.

[50] C. Lashway, A. T. Elsayed, A. Altamirano, and O. A. Mohammed, "Design and control of a grid-tied PV system for medium-sized household in South Florida," Presented at the 12th Latin American and Caribbean Consortium of Engineering Institutions (LACCEI), pp. 22-24, 2014.

[51] A. Alkhalidi and N. Hussain Al Dulaimi, "Design of an off-grid solar PV system for a rural shelter," Presented at the School of Natural Resources Engineering and Management, Department of Energy Engineering, Design of an Off-Grid Solar PV system for a rural shelter, presented by Noor Hussain Al Dulaimi–2008203032 F.

[52] A. A. Imam and Y. A. Al-Turki, "Techno-economic feasibility assessment of grid-connected PV systems for residential buildings in Saudi Arabia—A case study," Sustainability, vol. 12, no. 1, pp. 262, 2019.

[53] T. Cioara et al., "Data Centers Optimized Integration with Multi-Energy Grids: Test Cases and Results in Operational Environment," Sustainability, vol. 12, no. 23, p. 9893, 2020.

[54] Ali, Mohammed Rhael, et al. "Botulinum Toxin-A For Management of Migraine: An Experience in Iraq." History of Medicine 9.2 (2023): 416-425.

[55] Abdulkareem, Elham Hazeim, Mohammed Rhael Ali, Sabah Abdul Rasool Hammoodi, and Riyam Firas Talib. "Assessment of the Correlation between Gender and Third Molar Surgery Duration." Dental Hypotheses 13, no. 4 (2022): 139-141.

[56] P. R. Fayziev and Z. M. Khametov, "Testing the innovative capacity solar water heater 200 liters," American Journal of Applied Science and Technology, vol. 2, no. 5, pp. 99-105, 2022.

[57] P. Cheng, W. Liu, J. Ma, L. Zhang, and L. Jia, "Solar-powered rail transportation in China: Potential, scenario, and case," Energy, vol. 245, p. 123221, 2022.

[58] T. Talaviya, D. Shah, N. Patel, H. Yagnik, and M. Shah, "Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides," Artificial Intelligence in Agriculture, vol. 4, pp. 58-73, 2020.

[59] S. Freitas, C. Catita, P. Redweik, and M. C. Brito, "Modelling solar potential in the urban environment: State-of-the-art review," Renewable and Sustainable Energy Reviews, vol. 41, pp. 915-931, 2015.

[60] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, "Solar forecasting methods for renewable energy integration," Progress in Energy and Combustion Science, vol. 39, no. 6, pp. 535-576, 2013.

[61] 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, pp. 78-111, October 2016.

[62] V. Kostylev, A. Pavlovski, et al., "Solar power forecasting performance – towards industry standards," in 1st International Workshop on the Integration of Solar Power into Power Systems, Aarhus, Denmark, 2011.

[63] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," International Journal of Forecasting, vol. 32, no. 3, pp. 896-913, 2016.

[64] G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," Solar Energy, vol. 83, no. 3, pp. 342-349, 2009.

[65] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," Solar Energy, vol. 83, no. 10, pp. 1772-1783, 2009.

[66] Hammoodi, Sabah Abdul Rasool, Kamal Turki Aftan, and Mohammed Rhael Ali. "Management of Hydatid cysts of parotid glands." *Journal of stomatology, oral and maxillofacial surgery* 124.6 (2023): 101465

[67] J. Pambuk, Chateen I. Ali, Fatma Mustafa Muhammad, and Mohammed Rhael Ali. "Viral hepatitis: implication of viral types from A to E." Am J Biomed Sci Res 4.6 (2019): 442-444.

[68] Ali, Mohammed Rhael, and Elham Hazeim Abdulkareem. "Efficiency of BTX-A in the alleviation of hemifacial pain." Journal of International Dental and Medical Research 13.1 (2020): 321-326.

[69] C. Chen, S. Duan, T. Cai, and B. Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," Solar Energy, vol. 85, no. 11, pp. 2856-2870, 2011.

[70] C. Persson, P. Bacher, T. Shiga, and H. Madsen, "Multi-site solar power forecasting using gradient boosted regression trees," Solar Energy, vol. 150, pp. 423-436, 2017.

[71] 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 Transactions on Industry Applications, vol. 48, no. 3, pp. 1064-1069, May 2012.

[72] B. Yue, et al., "Solar energy forecasting and renewable energy grid integration: A review," International Journal of Electrical Power & Energy Systems, vol. 79, pp. 100-110, 2016.

[73] G. Li, et al., "Machine learning applications in solar energy prediction: A review," Journal of Renewable and Sustainable Energy Reviews, vol. 157, article 110082, 2020.

[74] K. Kim, et al., "A review of machine learning techniques for solar irradiance forecasting," Journal of Renewable and Sustainable Energy Reviews, vol. 66, pp. 503-514, 2016.

[75] G. Li, et al., "Machine learning applications in solar energy prediction: A review," Journal of Renewable and Sustainable Energy Reviews, vol. 157, article 110082, 2020.

[76] J. Hu, et al., "Machine learning-based solar power prediction models: A review," Journal of Renewable and Sustainable Energy Reviews, vol. 134, article 110312, 2020.

[77] M. N. Islam, et al., "Machine learning algorithms for solar energy prediction: A review," Journal of Renewable and Sustainable Energy Reviews, vol. 153, article 110493, 2020.

[78] Hammoodi, Sabah Abdul Rasool, Kamal Turki Aftan, and Mohammed Rhael Ali. "Hydatid Cysts of Parotid Glands-Diagnosis, Treatment and Recurrences." International Journal of Surgery Protocols 25.1 (2021): 135-140.

[79] D. Kollias, et al., "Machine learning algorithms for solar energy prediction: A systematic review," International Journal of Electrical Power & Energy Systems, vol. 124, article 106892, 2021.

[80] R. Chakraborty, et al., "Short-term solar power forecasting using machine learning algorithms: A comprehensive review," Journal of Renewable and Sustainable Energy Reviews, vol. 101, pp. 227-240, 2019.

[81] H. Wang, et al., "A review on machine learning applications for solar power plants," IEEE Transactions on Energy Conversion, vol. 36, no. 3, pp. 1202-1215, 2021.

[82] Abdulkareem, Elham Hazeim, Sabah Abdul Rasool Hammoodi, and Mohammed Rhael Ali. "Occurrence of Peri-Implant Microflora in Single vs. Two Piece Implants." *International Medical Journal* 27.4 (2020): 476-480.

[83] G. Hassan and D. Muhsen, "Forecasting of photovoltaic power generation and model optimization," International Journal of Energy Research, vol. 43, no. 9, pp. 4428-4445, 2019.

[84] K. Sivachandar, V. T. Bai, and N. Gangatharan, "Power Prediction and Suboptimal Performance Identification of Solar Panels Using Machine Learning," Journal of Green Engineering, vol. 10, pp. 6455-6467, 2020.

[85] https://www.kaggle.com/datasets/pythonafroz/solarpower?select=Plant_2_Generation_Data.csv

[86] Kalaiselvi, B., Karthik, B., & Kumaravel, A. (2022, July). Variant Mode Data Analytics in Predicting the Radiation Effect on Solar Power Generation using Machine Learning

Algorithms. In 2022 IEEE International Conference on Data Science and Information System (ICDSIS) (pp. 1-6). IEEE.