REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

# ENHANCING MAXIMUM POWER POINT TRACKING THROUGH ENSEMBLE LEARNING TECHNIQUES

**Hayder Husam Mahmood AL-MAYYAH**

Master's Thesis

Supervisor

Asst. Prof. Dr. Zaid HAMODAT

İstanbul, 2024

# ENHANCING MAXIMUM POWER POINT TRACKING THROUGH ENSEMBLE LEARNING TECHNIQUES

**Hayder Husam Mahmood AL-MAYYAH**

Electrical and Computer Engineering

Master’s Thesis

ALTINBAŞ UNIVERSITY

2024

The thesis titled ENHANCING MAXIMUM POWER POINT TRACKING THROUGH ENSEMBLE LEARNING TECHNIQUES prepared by HAYDER HUSAM MAHMOOD AL-MAYYAH and submitted on 26/01/2024 has been **accepted unanimously** for the degree of Master of Science in Electrical and Computer Engineering.

Asst. Prof. Dr. Zaid HAMODAT

Supervisor

Thesis Defense Committee Members:

| | | |
|---|---|---|
| Asst. Prof. Dr. Zaid HAMODAT | Department of Electrical and Computer Engineering, Altınbaş University | |
| Asst. Prof. Dr. Mesut ÇEVİK | Department of Electrical and Electronics Engineering, Altınbaş University | |
| Asst. Prof. Dr. Abbas UĞURENVER | Department of Electrical and Electronics Engineering, İstanbul Aydın University | |

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

I hereby declare that all information/data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Hayder Husam Mahmood AL-MAYYAH

Signature

# DEDICATION

I want to express my deepest gratitude to the Almighty for blessing me with the mental acuity, health, and endurance required to successfully navigate and complete this course of study.

A profound thank you is due to my parents, the bedrock of my life, whose unwavering support has been a guiding light from my earliest days to the present. Their boundless dedication and sacrifices are immeasurable, and words cannot adequately convey my appreciation.

I extend sincere thanks to all those who played a role in shaping my academic path, with special recognition for my supervisor, Asst. Prof. Dr. Zaid HAMODAT. His invaluable guidance and collaboration were integral to the fruition of this thesis.

In heartfelt acknowledgment of their consistent encouragement, I dedicate this thesis to my parents. Their enduring support has been the driving force behind my academic journey, seamlessly weaving together the tapestry of age-old traditions and modern opportunities.

# PREFACE

I want to thank Allah for helping me complete my master's degree. Without his blessings, this achievement wouldn't have been possible.

I dedicate my efforts to my wonderful family and friends. Their love, encouragement, care, and prayers, day and night, made it possible for me to succeed.

I'm grateful to my family for their sacrifices and unwavering belief in my abilities. Their support has been the driving force behind every step in my academic journey.

To my friends, thanks for being there and making the journey enjoyable. Your support has turned challenges into opportunities and made success a shared celebration.

Acknowledging the love and prayers that surrounded me, I realize the impact of a supportive community. This gratitude is a way of recognizing the importance of faith, family, and friends in reaching educational milestones.

# ABSTRACT

## ENHANCING MAXIMUM POWER POINT TRACKING THROUGH ENSEMBLE LEARNING TECHNIQUES

AL-MAYYAH, Hayder Husam Mahmood

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

Supervisor: Asst. Prof. Dr. Zaid HAMODAT

Date: January / 2024

Page: 88

Maximum Power Point Tracking (MPPT) is an essential method in photovoltaic (PV) solar systems for optimizing the extraction of available power. This technique enhances energy conversion efficiency and aligns with ongoing efforts to improve the effectiveness of renewable energy sources. This thesis presents a systematic investigation into the predictive modeling of solar energy phenomena, specifically focusing on solar power generation and solar radiation prediction. Through the comprehensive evaluation of various individual machine learning models, including Linear Regression (LR), Support Vector Regression (SVR), and XGBoost Regressor, as well as an Ensemble Learning (EL) approach, the study elucidates the complexities and nuances of modeling solar energy systems. The analysis was conducted on two distinct datasets: Solar Power Generation and Solar Radiation Prediction. The individual models were rigorously assessed using multiple statistical metrics, revealing varying degrees of accuracy, fit, and performance. A remarkable discovery was the efficacy of the Ensemble Learning model, employing techniques such as Bagging Regressor, which consistently outperformed the individual models across both datasets. By adeptly aggregating the predictions of multiple underlying models, the EL approach achieved superior predictive accuracy, explaining an impressive proportion of the variance in both solar power generation and solar radiation. The findings of this research contribute significantly to the understanding of solar energy modeling, endorsing ensemble learning as a potent and versatile tool for enhancing prediction accuracy. Moreover, the comparative analysis sheds light on the trade-offs between different modeling techniques, offering guidance for future research and practical applications within the renewable energy sector. This thesis not only sets a new benchmark in the field of solar energy forecasting but also aligns with the broader imperatives of sustainable energy management and climate stewardship.

**Keywords:** Maximum Power Point Tracking, Renewable Energy, Ensemble Learning, Regression, PV.

# TABLE OF CONTENTS

**Pages**

# LIST OF TABLES

**Pages**

# LIST OF FIGURES

**Pages**

# ABBREVIATIONS

| | | |
|---|---|---|
| AI | : | Artificial Intelligence |
| DL | : | Deep Learning |
| ML | : | Machine Learning |
| SVR | : | Support Vector Regression |
| RF | : | Random Forest |
| LR | : | Linear Regression |
| MPPT | : | Maximum Power Point Tracking |
| PV | : | Photovoltaic |
| XGBoost | : | Extreme Gradient Boosting |
| EL | : | Ensemble Learning |
| PCA | : | principal Component Analysis |
| ANN | : | Artificial Neural Networks |
| DT | : | Decision Tree |
| EDA | : | Exploratory Data Analysis |
| $R^2$ | : | R-squared |
| RMSE | : | Root Mean Squared Error |
| MAE | : | Mean Absolute Error |
| MAPE | : | Mean Absolute Percentage Error |
| MSE | : | Mean Squared Error |

# 1. INTRODUCTION

## 1.1 BACKGROUND AND MOTIVATION

The challenge of energy production holds immense significance for the future. Changes in the energy market have been substantial, particularly concerning the utilization of both renewable energy sources and fossil fuels. These changes are a response to a growing consciousness about environmental harm, the finite availability of fossil fuel, various political and economic pressures from different countries, and an increase in investments in green energy [1]. Currently, a large portion of global energy production is derived from fossil fuels, leading to greenhouse gas emissions that contribute to pollution. Concurrently, excessive use of these resources diminishes their reserves. With rising fuel costs and mounting environmental issues, renewable energies are gaining importance as a vital source for providing electricity to residential and industrial areas[2].

Renewable energy sources, including wind, hydro, geothermal, and solar energies, have a significant impact on electric power generation. These energy forms have been around since the inception of Earth and were utilized by the earliest human generations in various manners. The primary benefit of harnessing these renewable energies is the reduction in pollution and emission of greenhouse gases like carbon dioxide and nitrogen oxides, which contribute to global warming [3]. In recent times, the utilization of these energy sources has seen substantial growth, benefiting from rapid technological advancements. Solar energy is regarded as one of the fastest-growing technologies and has witnessed a significant decrease in the costs of equipment [4].

Photovoltaic (PV) technology, which converts solar energy into electricity using semiconducting materials, has emerged as a vital renewable energy source due to its eco-friendly nature. In real-world PV applications, Partial Shading Conditions (PSC) caused by passing clouds, nearby trees, or buildings, are quite frequent. This can result in power loss, hotspots, and reliability issues within the PV system [5]. Consequently, there is a need to both prevent hotspot damage and optimize the output power for partially shaded PV arrays.

The problem of maximizing the transfer of power from a photovoltaic generator to the load arises from the nonlinear electrical characteristics (V–I) of PV cells. These characteristics are dependent on varying environmental conditions such as temperature and solar irradiance, which in turn affect the cell's solar irradiance and temperature. Efficiency is often reduced in PV due to the energy conversion from solar to electrical, posing major hurdles to maximizing solar energy utilization. Therefore, achieving the highest efficiency in PV systems becomes paramount. To enhance the power output of a photovoltaic system, it is necessary to operate the photovoltaic panel at its maximum power point (MPP). As such, maximum power point tracking (MPPT) techniques have been developed to extract the utmost power from PV panels, optimizing the use of the panel's power [6], [7]. Figure 1.1 illustrates the P-V characteristic curves of a PV system under both Uniform Irradiation Conditions (UIC) and PSC. While only a single peak is present in the P-V characteristics under UIC, there are multiple peaks under PSC. Among these, the one with the highest power is referred to as the Global Maximum Power Point (GMPP), while the others are known as the Local Maximum Power Points (LMPP). Tracking the GMPP is considered one of the most cost-effective methods for enhancing the output power of a PV array.

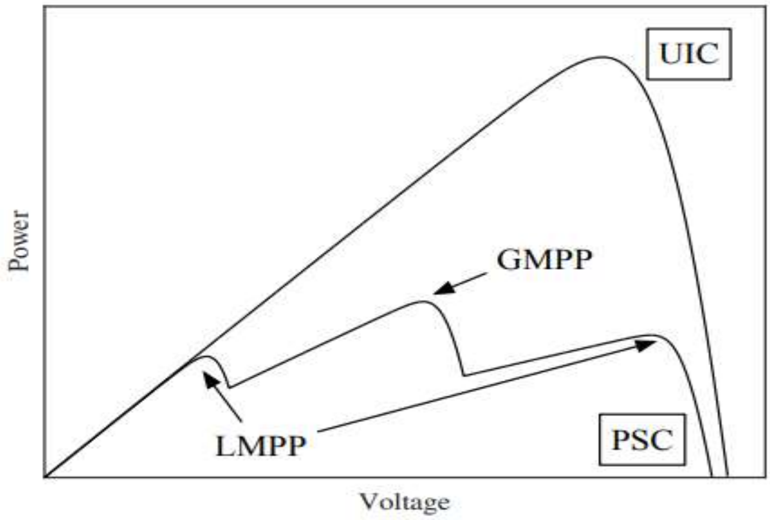



**Figure 1.1:** Comparison of P-V Characteristic Curves UIC and PSC in a PV System.

The initial costs for implementing photovoltaic (PV) systems are notably high, and the efficiency of converting energy within these systems is comparatively low [8]. Furthermore, the energy output of PV systems is acutely affected by fluctuations in weather conditions [9]. Factors such as changes in cloud cover, the presence of fog or heat, and airborne dust and particles on the panel can all substantially reduce the efficiency of the power conversion process [10]. As a result, a considerable amount of research is being directed toward optimizing PV panels to reach their maximum output power under all weather conditions, which is key to making them profitable [8].

A closer examination of the nonlinear characteristics of PV systems reveals that for any specific set of solar insolation and ambient temperature values, there is a singular operating point (output voltage) where the PV output power reaches its peak. Alterations in the output voltage, whether due to load changes or other factors, result in the panel producing power below this maximum [11]. The output power is further contingent on factors such as solar radiation and temperature [12]. Together, these complexities make the constant operation of the system at the maximum output power point an exceptionally challenging task. The goal is to achieve a balance in operating points where the PV panel's maximum power point coincides with the load requirements [6].

Historically, various attempts have been made to regulate PV systems to operate at maximum power. Techniques like Perturb and Observe (PaO) [8] and Incremental Conductance (IC) have been among the early solutions to this challenge. Although these methods have been widely adopted due to their ease of implementation, they are not without drawbacks. These include a level of uncertainty tied to their sensitivity to sudden weather changes [[8],[13], and oscillations around the operating point [14], [15] both of which have created room for continued exploration and improvement in the field.

Several diverse strategies have been explored in the literature to address the specific challenge of enhancing the results achieved by the PaO. These include the application of fuzzy logic-based techniques [16]–[19] as well as computational optimization methods like swarm optimization [20] and ant colony optimization [21], all aimed at boosting the speed and stability in response to changes in irradiance.

Artificial neural networks (ANNs), a well-known category of machine learning (ML) algorithms, have also been employed to tackle this problem[22]–[25] . Owing to their capacity to model nonlinear functions, ANNs have been instrumental in accurately estimating the PV panel reference voltage that corresponds to the maximum output power. Beyond merely pinpointing the maximum power point, approaches that center on ANN have demonstrated a quick response during testing phases.

Support vector regression (SVR), another ML algorithm, has been utilized in the context of renewable energy [9], [26] . Specifically, in one study [11], SVR was applied successfully to predict the PV reference voltage that correlates with maximum output power. Unlike ANNs, which tend to converge at local minima, SVR has the advantage of converging at global minima. This distinct characteristic positions SVR as an appealing choice for modeling functions and prediction tasks within the renewable energy sector [11].

ML in renewable energy not only focuses on individual models like SVR and ANNs but also explores the promising field of ensemble learning. Individual models have their unique strengths and weaknesses, and their performance can be highly dependent on specific characteristics of the data. Ensemble learning, on the other hand, combines predictions from multiple ML models to produce a more robust and accurate prediction. For example, an ensemble of SVR, ANNs, and other algorithms might be used to counter the weaknesses of one model with the strengths of another. By Levera [27]ging the diverse characteristics of individual models, ensemble learning can lead to enhanced stability, reduced bias, and improved generalization [28]. This approach has been gaining traction in the renewable energy sector, especially in the optimization of photovoltaic systems, as it provides a sophisticated tool to navigate the complexities and variabilities inherent in renewable energy sources. The integration of both individual models and ensemble learning techniques represents an innovative frontier in maximizing the efficiency and reliability of renewable energy systems [29], [30].

## 1.2 REREARCH MOTIVATION

The growing urgency to transition from conventional fossil fuels to more sustainable energy sources propels the necessity for extensive research into renewable energy technologies. Among them, solar energy, captured through PV systems, stands as a beacon of untapped potential and ecological responsibility. However, the complex nature of maximizing the power extraction from PV systems, coupled with the influence of environmental variables, presents a multifaceted challenge. While traditional techniques and individual ML models have made strides in this direction, they are not without limitations. Ensemble learning techniques emerge as a promising frontier, with the potential to enhance the accuracy, reliability, and efficiency of PV systems. This thesis is motivated by the imperative to explore and harness these advanced methods, unifying the strengths of various models to overcome their individual weaknesses. In an era marked by escalating energy demands, environmental degradation, and economic pressures, the pursuit of optimized renewable energy solutions is not merely an academic endeavor; it is a societal imperative. The innovations and insights gleaned from this research have the potential to transform the renewable energy landscape, contributing to a future where clean, efficient, and sustainable energy is not just an aspiration but a tangible reality. By delving into the intricacies of ensemble learning techniques, this thesis seeks to break new ground in the field, driving technological advancements that resonate with global sustainability goals and align with the urgent call for responsible energy consumption and environmental stewardship. It's a pursuit that extends beyond technical mastery, reaching into the realms of global impact, ecological balance, and the legacy we leave for future generations.

### 1.3 PROBLEM STATEMENT

The efficient utilization of solar energy through PV systems is hampered by a complex set of challenges related to maximizing power output. The nonlinear characteristics of PV cells, coupled with the fluctuating environmental conditions such as temperature and solar irradiance, make the task of continuous operation at the MPP highly intricate. Traditional techniques like PaO and Incremental Conductance (IC) offer some solutions but are prone to uncertainties, oscillations, and sensitivity to sudden weather changes. While individual ML models like ANNs and SVR have made advances in this field, they too have inherent limitations and may not always provide the optimal solution. Ensemble learning has been identified as a promising approach to overcoming these challenges but remains underexplored in the context of renewable energy. The problem this thesis aims to address, therefore, is the need for an innovative, robust, and efficient method to track the MPP in PV systems, leveraging the full potential of ensemble learning. The research seeks to navigate the intricate landscape of solar energy conversion, creating a model that synergizes individual strengths of diverse algorithms, thereby leading to enhanced stability, reduced bias, and improved generalization in the context of PV systems. In addressing this problem, the thesis contributes to the broader goal of optimizing renewable energy sources, a critical need in the global shift towards sustainable energy solutions.

## 1.4 RESEARCH HYPOTHESES

The research hypotheses of this study encapsulate a comprehensive and innovative approach to renewable energy modeling.

a. Hypothesis 1: posits that the fusion of individual ML models with ensemble learning techniques will transcend the confines of singular models, orchestrating a synergistic collaboration that can address the unique challenges of solar energy forecasting.

b. Hypothesis 2: extends this line of thinking to real-world applications, foreseeing how this blend of models can drive significant advancements in the renewable energy sector.

c. Hypothesis 3: looks beyond the immediate scope, envisioning the potential of the methodologies to lay new foundations in the field of solar energy, making both theoretical and practical breakthroughs. Together, these hypotheses establish a multifaceted framework that guides the research toward uncharted territories in renewable energy exploration and utilization.

## 1.5 RESEARCH OBJECTIVES

The principal objective of this research is to devise and evaluate a robust methodology that leverages various individual ML models, such as Support Vector Regression (SVR), Linear Regression, and XGBoost Regression, in conjunction with ensemble learning techniques for Solar Power Generation and predicting solar radiation. Recognizing that each individual model possesses its unique strengths and weaknesses, the research aims to explore the synergistic integration of these models through ensemble learning. This innovative approach is designed to capitalize on the complementary attributes of the individual models, potentially overcoming their standalone limitations and enhancing both accuracy and efficiency. The research focuses on comprehensive experimentation with these models, examining their performance under diverse environmental and operational conditions. In particular, the study seeks to assess how different ensemble strategies can be tailored to specific scenarios, thus optimizing the prediction of solar radiation and the power generated from PV systems. By pursuing this multifaceted objective, the thesis is committed to advancing the state-of-the-art in renewable energy modeling, contributing valuable insights that could revolutionize the way solar energy is harnessed, predicted, and utilized. The aspiration is not just to foster theoretical understanding but to translate this knowledge into practical applications that could have a profound impact on the renewable energy sector.

## 1.6 REPORT STRUCTURE

The chapters of our research is as follows:

Chapter 1: Introduction

This chapter introduces the background of the study, highlighting the importance of renewable energy and the role of ML and ensemble learning techniques in optimizing solar power generation. It outlines the research objectives, hypotheses, and the rationale behind the study.

Chapter 2: Literature Review

A comprehensive review of existing literature related to solar energy prediction, individual ML models, and ensemble learning techniques. This chapter identifies gaps in current knowledge and sets the stage for the research.

Chapter 3: Methodology

Detailed presentation of the research design, including the selection and description of individual models and ensemble learning techniques, data collection, and preprocessing. The chapter also discusses the rationale behind the chosen methods.

Chapter 4: Results and Analysis

Presentation and analysis of the experimental results, with a focus on the evaluation of the models under different environmental and operational conditions. This chapter includes statistical analysis and interpretation of the findings in relation to the research hypotheses.

Chapter 5: Conclusion and Future Work

Summary of the key findings, conclusions drawn from the research, implications, limitations, and recommendations for future research in the field of renewable energy modeling.

# 2. LITERATURE REVIEW

## 2.1 INTRODUCTION

The literature review chapter delves into the existing research and studies relevant to our investigation of renewable energy and MPPT algorithms. Renewable energy sources are crucial in addressing environmental concerns and meeting energy demands sustainably. In this context, we explore various methods and algorithms for maximizing the efficiency of renewable energy systems, with a particular focus on EL and ML techniques such as SVR, LR, XGBoost Regression, DT, and RF. Understanding these approaches will provide valuable insights for our research on optimizing renewable energy systems and advancing MPPT methods.

## 2.2 LITERATURE REVIEW

In [2] Renewable technologies are clean, sustainable, and minimize environmental impacts. The sun is the primary source of energy, with heat and light being transformed into biomass and wind energy. These technologies offer opportunities to mitigate greenhouse gas emissions and reduce global warming by replacing conventional energy sources. This article reviews CO2 mitigation through solar cookers, water heaters, dryers, biofuels, improved cookstoves, and hydrogen.

In [4] Solar energy conversion is a growing trend in industrial applications, with a study by the International Energy Agency revealing that solar array installations will supply 45% of global energy demand by 2050. Solar thermal is particularly popular for generating electricity, processing chemicals, and space heating. Solar electricity is widely used in telecommunications, agriculture, water desalination, and building industries for various purposes. The study aims to explore the integration of solar energy systems in industrial applications, aiming to improve energy stability, sustainability, and conversion reduction.

This paper [6] evaluates maximum power point tracking (MPPT) techniques for photovoltaic panels, comparing energy extraction and tracking factor. It uses MatLab/Simulink and dSPACE platforms, a digitally controlled boost dc-dc converter, and an Agilent Solar Array E4350B simulator. Results show conventional MPPT algorithms and improved algorithms like IC based on proportional-integral (PI) and perturb and observe based on PI. A user-friendly interface evaluates dynamic response and TF. A typical daily insulation is used to verify experimental results.

This study [23] presents a direct neural control (DNC) scheme for maximum power point tracking (MPPT) in photovoltaic systems. The intelligent controller uses a hybrid learning mechanism, including an on-line learning rule based on gradient decent method and an off-line learning rule based on Big Bang-Big Crunch algorithm. The system's effectiveness is tested under partial shading conditions and compared to conventional perturbation and observation methods.

The researcher in this paper [31] uses machine learning techniques to estimate energy yield in a photovoltaic system under partial shading. Three gradient boosted regression tree models are trained and evaluated, showing improved performance in both average and worst-case scenarios. The methods are fast to train and deploy. Moreover, we show that even a small number of training examples is sufficient to achieve improved global MPP estimation. allowing for further improvements with more computational resources.

In [32] the integration of renewable energy sources like solar and wind can decrease system reliability and stability. Solar power forecasting can improve system stability by providing future power generation estimates and facilitating optimal hydro power plant dispatch. Machine Learning (ML) algorithms have shown great performance in time series forecasting, making them suitable for forecasting power using weather parameters. This paper compares the performance of ML algorithms with the Smart Persistence (SP) method, revealing that ML models outperform SP models.

This study [34] Machine learning has become a popular approach for hourly solar forecasting, but there is a myth about forecast accuracy. The article evaluates the hourly forecasting performance of 68 machine learning algorithms for three sky conditions, seven locations, and five climate zones in the continental United States. Tree-based methods perform well in two-year overall results but rarely stand out during daily evaluation. No universal model can be found, but some preferred ones for each sky and climate condition are advised.

In [35] the researchers introduced a groundbreaking technique for MPPT in PV systems with the goal of nearly perfect efficiency. Several existing MPPT algorithms were recognized, including Fractional Open Circuit Voltage (FOCV), P&O, Fractional Short-Circuit Current (FSCC), INC, FLC, and NN. However, the study's emphasis was on refining the FSCC approach. Traditional use of FSCC, despite its simplicity and speed, has the drawback of requiring a short-circuit of the solar panel to determine the short-circuit current, resulting in significant power losses and fluctuations. The authors tackled these challenges by devising a new strategy that directly detects the short-circuit current by frequently measuring the solar panel's output.

In [36] this paper presents a novel GMPPT method using a machine-learning algorithm, which can detect the global maximum power point (GMPP) operating point in the shortest possible search time. This method doesn't require knowledge of PV module or array structure characteristics and can detect the GMPP in fewer search steps, making it suitable for PV applications with quick changing shading patterns. The algorithm reduces detection time by 80.5%-98.3%.

The paper [37] explores the use of machine learning based MPPT techniques to maximize power on a photovoltaic (PV) system under power supply conditions (PSC). Nine ML-based MPPT techniques, including Decision Tree, Multivariate Linear Regression, Gaussian Process Regression, Weighted K-Nearest Neighbors, Linear Discriminant Analysis, Bagged Tree, Naïve Bayes classifier, Support Vector Machine, and Recurrent Neural Network, were tested using MATLAB SIMULINK simulation software. The experimental results demonstrated that WK-NN performs significantly better when compared with other proposed ML based algorithms.

In [27] Microgrids are crucial for smart grids, combining distributed renewable energy sources (RESs), energy storage devices, and load management methodologies. However, the intermittent nature of RESs presents challenges like reliability, power quality, and balance between supply and demand. Forecasting power generation from RESs is essential for efficient operations and optimal utilization. Machine learning and deep learning (DL)-based models are promising solutions for predicting consumer demands and energy generation from RESs. This study provides a comprehensive survey of DL-based approaches for power forecasting of wind turbines, solar panels, and electric power load forecasting, focusing on both residential and commercial sectors. The efficiency of these forecasting methods depends on historical data, necessitating large data storage devices and high processing power devices.

In [38] The Maximum Power Point Tracking (MPPT) algorithm is used to maximize solar photovoltaic system efficiency, but traditional methods like P&O, FPA, and PSO struggle due to partial shading and atmospheric conditions. A Regression controller based MPPT system is developed to achieve maximum peak voltage during partial shading effects. This controller predicts the duty cycle for boost converters based on stored data from the PV system, resulting in increased efficiency by about 20%, 16.96%, and 15% compared to traditional algorithms.

This study [39] proposes a new ensemble model for hourly global horizontal irradiance forecast, combining two advanced base models, extreme gradient boosting forest and deep neural networks (XGBF-DNN). The model is integrated using ridge regression to avoid over-fitting and ensures diversity among base models. To enhance accuracy, a subset of input features is selected, including temperature, clear-sky index, relative humidity, and hour of the day. The model is validated with data from three different climatic Indian locations and analyzed for seasonal variations. The model shows the best combination of stability and prediction accuracy, with a forecast skill score ranging from 33% to 40% in prediction error. This makes XGBF-DNN an ideal and reliable model for hourly global horizontal irradiance prediction, making it suitable for various energy sectors.

This paper [40] proposes a new approach to Maximum Power Point Tracking (MPPT) for photovoltaic systems, aiming to increase efficiency close to 100%. MPPT algorithms like Fractional Open Circuit Voltage (FOCV), Perturb and Observe (P&O), Fractional Short-Circuit Current (FSCC), Incremental Conductance (INC), Fuzzy Logic Controller (FLC), and Neural Network (NN) can be used to achieve this. However, the FSCC algorithm, which requires a single current sensor, has limitations due to strong oscillations and power losses due to the Joule effect. The new approach directly detects short-circuit current by reading the output current of the solar panel, which is then used to calculate short-circuit current by increasing or decreasing global solar irradiance. Experimental results show reduced energy losses, temporal response attenuation, low oscillations, and better accuracy compared to the classic FSCC algorithm.

Researchers in [41] proposed an off-grid Solar Photovoltaic Water Pumping System (SPVWPS) that integrates a DC-DC boost converter, a three-phase DC-AC inverter, and a three-phase induction motor (IM) with a centrifugal pump. The novelty lies in a control strategy that marries an improved fractional open-circuit voltage (FOCV) method for MPPT with closed-loop scalar control, obviating the need for specific sensors in the IM. Tested on a 1KVA rated prototype, the system showed excellent response time, minimal oscillations around the MPP, 99% efficiency, and a high flow rate, validating the method's potency in maximizing power extraction.

In [42] researchers proposed a new approach to improve PV system performance by optimizing MPPT algorithms using AI techniques. They employed ANN and SVM to overcome the limitations of traditional methods. The researchers simulated the PO based MPPT algorithm to generate the MPP under various weather conditions, using the data to train the ANN and SVM models. The ANN-based MPPT algorithm achieved an MSE of 0.5, while the cubic SVM algorithm performed best with an MSE of 0.15926 in predicting the MPP. The results suggest that the SVM-based MPPT approach is more effective in tracking the MPP, particularly in challenging weather conditions. The study emphasizes the potential of AI-based techniques like ANN and SVM in enhancing PV system performance.

Researchers in [43] proposed a cutting-edge hybrid MPPT algorithm, PSO_ML-FSSO, which combines Particle-Swarm-Optimization-trained ML with Flying Squirrel Search Optimization to enhance solar PV system efficiency. This algorithm, modeled in MATLAB/Simulink, was benchmarked against prominent methods such as P&O, INC, PSO, and others under various conditions, like temperature shifts and partial shading. The results indicated that the new approach boosted efficiency by 0.72% and cut the settling time by 76.4%.

Researchers in [44] proposed a DL framework for high-precision photovoltaic power generation (PVPG) forecasting. Recognizing the limitations of current ML algorithms, this study introduces a physics-constrained LSTM (PC-LSTM) approach for the hourly dayahead PVPG prediction. The PC-LSTM leverages domain knowledge, ensuring more accurate and reasonable forecasts compared to models that rely solely on extensive data. Through real-world PV dataset evaluations and sensitivity analysis on input feature selection, the PC-LSTM outperformed the standard LSTM and other conventional methods, proving especially effective in scenarios with sparse data. Overall, the PC-LSTM offers a more reliable and accurate approach to PVPG forecasting.

This paper [45] presents a novel approach to operating solar PV panels at maximum power point (MPP) using linear and nonlinear regression-type machine learning algorithms. The models predict the maximum available power and voltage for specific quantities of irradiation and temperature, determining the boost converter's duty cycle. Simulation results show that regression algorithms force the PV panel to work at the predicted MPP, with efficiency higher than 95.21% with linear regression models and more than 95.32% with nonlinear regression models.

This paper [46] proposes a hybrid model for predicting future solar energy generation, integrating machine learning and statistical approaches. The model, which uses an ensemble of machine learning models, reduces placement cost and outperforms conventional models. The study also shows that a hybrid model that uses both machine learning and statistics outperforms a machine learning-only model.

This paper [47] explores the use of Machine Learning (ML) in solar plant systems, focusing on their advantages and shortcomings compared to classical methods. It discusses intelligent, self-adaptive models for sizing, forecasting, maintenance, and control, sets performance benchmarks, proposes an effective integration scheme, and estimates the impact of ML technologies on the solar plant value chain. The research also highlights the need for a simple and effective integration scheme for a smart solar plant.

This paper [48] compares various maximum power point tracking techniques for PV systems, including P&O, INC, FLC, NN, and ANFIS. The hybrid method, which combines neural network and P&O, is used for PV systems. These methods differ in their characteristics such as convergence speed, ease of implementation, sensors used, cost, and range of efficiencies. The hybrid method offers a spontaneous recovery during dynamic environmental changes and improves performance compared to other techniques. These methods differ in their characteristics such as convergence speed, ease of implementation, sensors used, cost, and range of efficiencies. The study also examines drawbacks and wastage of panel output energy.

This paper [49] discusses the importance of solar energy conversion for power generation and pollution reduction. It highlights the potential of PV control and MPPT in increasing module efficiency and lifespan. It also discusses the integration of artificial intelligence methods with PV systems.

The study [50] proposes an automated method for calibrating a maximum power point tracking (MPPT) algorithm in photovoltaic (PV) systems using supervised machine learning. The method involves designing and simulating a schematic diagram of a PV system, creating a dataset in MATLAB/Simulink, and analyzing the output characteristics of a solar cell. The improved MPPT algorithm, based on a neural network (NN) method, is then implemented in a hardware setup, and verified with empirical data. The proposed model shows an even higher efficiency (99.8%) without additional computational cost.

In [51] The monitoring of Maximum Power Point (MPP) is a crucial method for improving Photo Voltaic (PV) power extraction under various weather conditions. The MPPT capability directly impacts system performance. Artificial Intelligence (AI) algorithms have been proposed as a potential method for large-scale MPP location. This study provides a comprehensive review of AI-based MPPT strategies, categorizing them into subcategories based on application strategy. The study analyzes the merits and demerits of different AI-incorporated MPPT techniques, creating a set of techniques for users to choose the best approach for their needs and system constraints.

In [11] Photovoltaic panels are a promising renewable energy source, but their energy output depends on factors like irradiation and ambient temperature. This paper introduces a method using Support Vector Machines to improve panel efficiency. The dataset, obtained from a real photovoltaic setup in Spain, includes temperature, radiation, output current, voltage, and power for a year. The results show the system accurately drives the photovoltaic panel to produce optimal output power for specific temperatures and irradiation levels.

Researchers in [52] introduces a machine-learning algorithm for maximum power point tracking (MPPT) in isolated photovoltaic (PV) systems. The energy generation in PV systems is non-linear due to weather conditions. The proposed strategy uses a decision-tree regression ML algorithm to determine the MPP of a PV system. The algorithm predicts the maximum power available and module voltage for a defined amount of irradiance and temperature. The boost converter duty cycle is determined using predicted values. Simulations show that the proposed method improves PV panel efficiency by over 93.93% in steady state despite erratic irradiance and temperatures, outperforming existing topologies like β-MPPT, cuckoo search, and artificial neural network results.

In [53] the increasing population and energy demand necessitate the use of artificial intelligence models and algorithms in energy systems. This study explores the applications of Machine Learning (ML) and Deep Learning (DL) in energy systems, focusing on the development of DL algorithms that are more accurate and less error prone. The research uses knowledge discovery in research databases to understand the status and future of ML and DL applications in energy systems. It also investigates the most common and efficient applications in this field, such as optimization, forecasting, and fault detection. The study covers most algorithms and their evaluation metrics, including both important and newer ones. The research aims to identify critical areas and research gaps in the field.

Researchers in [54] proposed the use of an XGBoost regression algorithm for predicting solar power with minimal error. They trained the model using 80% of one-year and five years of historical hourly solobioraar irradiance data from Johannesburg city. The XGBoost algorithm's performance was compared with that of the SVM model to determine the model with the least forecast errors. The results showed that the XGBoost model achieved a lower nRMSE value of 6.63% compared to the SVM model's 6.81%. Implementing the XGBoost algorithm for solar irradiance forecasts could significantly enhance the stability of solar electric power generation and facilitate optimal connection to the power grid.

In [55] proposed a novel variable boundary reinforcement learning (VBRL) algorithm for maximum power point tracking. This innovative method autonomously segments the appropriate state range during its interaction with the environment. As a model-free approach, VBRL learns and controls without requiring prior knowledge. It excels in rapidly tracking the MPP and maintains high accuracy under varying conditions. Notably, VBRL can also find the extremum of analogous unimodal curve models. When compared to other techniques in fluctuating environments, the VBRL method showcases superior performance, particularly when irradiance changes swiftly.

Researchers in [56] proposed an investigation into the feasibility of using ML based MPPT techniques to optimize power harnessing in a PV system under Partial Shading Conditions (PSC). They introduced nine ML-based MPPT techniques, including DT, Multivariate Linear Regression (MLR), Gaussian Process Regression (GPR), Weighted K-Nearest Neighbors (WK-NN), Linear Discriminant Analysis (LDA), Bagged Tree (BT), Naïve Bayes classifier (NBC), SVM, and RNN. Through three experiments under varying weather conditions, and validations using MATLAB SIMULINK simulation software, the study found that the WK-NN method performed significantly better compared to the other proposed ML-based algorithms.

Researchers in [57] proposed ML-based methods to forecast solar power generation in Badabenakudi, Orissa, India, considering the unpredictability of renewable resources. The study compared ML models both with and without an MPPT controller. Results showed that the Coarse Tree model was the best for forecasting with an MPPT controller, achieving an RMSE of 1.675, while the Rational Quadratic Gaussian Process Regression (RQGPR) model was optimal without the MPPT controller, with an RMSE of 1.628. The findings reinforced the superiority of ML techniques over statistical ones in solar power generation forecasting.

In [58] researchers introduced an innovative MPPT solution utilizing ML and DL with a LSTM model. Unlike traditional MPPT approaches relying on current and voltage sensors, this study estimated the PV module's output resistance using environmental parameters like temperature and radiation, coupled with artificial intelligence algorithms. Comparative assessments involving MSE, RMSE, and MAE) revealed the LSTM model's superior performance in tracking the MPP compared to ANN and regression methods. Specifically, LSTM exhibited a substantial advantage with 37% lower MSE, 21% lower RMSE, and 31% lower MAE compared to the ANN method.

In [59] researchers propose novel ML techniques to enhance Photovoltaic integrated Shunt Active Power Filter performance. They introduced a hybrid Support Vector Machine

Regression Perturb and Observe (SVM regression-P&O) algorithm for efficient maximum power point tracking (MPPT). SVM accelerates tracking by predicting an initial duty cycle, while a fixed-step P&O algorithm ensures high MPPT accuracy. Additionally, they improve harmonics detection using Adaline's intelligent learning combined with ML, creating a novel SVM regression-Adaline PQ strategy. Simulations demonstrate significant reductions in PV energy losses (up to 99%) and harmonics extraction response time (up to 98.8%), along with enhanced power quality under varying conditions and nonlinear loads compared to traditional methods.

Researchers in this paper [60] proposed a DL model using a back propagation neural network (BPNN) to achieve maximum power point. Designed to optimize power output from solar grids when connected to a boost converter, the BPNN-DL model predicts reference voltage under various weather conditions, ensuring consistent power output and voltage stability. Testing demonstrated the model's resilience to disturbances and revealed superior power output compared to other methods based on regression analysis metrics for training, testing, and validation.

In [61] researchers proposed an advanced MPPT Controller leveraging deep data representation from diverse sources to precisely compute Control Points in microgrids. This DL architecture aims to analyze extensive data from photovoltaic-based DGs, employing techniques like the Moore-Penrose method for comprehensive PV panel insights. The resulting PV-DG data informs the MPPT Tracker, estimating the peak power point and voltage. The innovative Extreme Learning Machine (ELM) utilizes the Ridge Regression algorithm to curtail initial randomization. This MPPT system enhances DC-DC converter control points accuracy and improves stability margins by expediting pulse width modulation duty calculations. The deep representation method's versatility is showcased in hardware-in-loop and simulation applications.

Researchers in [62] proposed a ML based methodology for predicting power output from building-integrated photovoltaic systems in relation to various building orientations. The methodology includes data quality assessment, a ML algorithm, weather clustering, and accuracy evaluation. By applying linear regression coefficients to the neural network's forecast outputs, the prediction accuracy of the PV power generation improved. The resulting model displayed a root mean square error of 4.42% in NN, 16.86% in QSVM, and 8.76% in TREE across different building facets such as flat roofs and various facades.

In [63]. researchers develop a Reinforcement Learning-based torque controller for a variable-speed wind turbine, addressing the Optimal Torque - Maximum Power Point Tracking problem. The reward optimization function, primarily based on rotor power variation, guides the optimal action (electromagnetic torque adjustment) to regulate turbine speed. The controller successfully manages the operation of a simulated 1.5 MW wind turbine under varying wind conditions after Proximal Policy Optimization training. Results demonstrate the controller's ability to maintain optimal power generation without relying on empirical parameters typically used in wind turbine control systems.

In [64] researchers tackle a multi-objective challenge with innovative optimization and classification techniques. They optimize power output from solar PV panels using a maximum peak point tracking method. The first phase employs endowed multi-verse optimization (EMVO) to enhance power prediction by estimating the best fitness value for faster optimization convergence. The second phase concentrates on fault detection to ensure the PV system operates normally, leveraging an augmented radial basis functional network (ARBFN). EMVO and ARBFN are applied for power forecasting and fault prediction. Experimental results show reduced absolute error (0.2%), relative error (3%), improved prediction accuracy (98%), and decreased fault mis-classification rate (2.5%). The proposed system outperforms existing approaches.

## 2.3 RENEWABLE ENERGY

Renewable energy refers to the power derived from nature's abundant and regenerative resources, including sunlight, wind, rain, tides, and geothermal heat. Unlike conventional fossil fuels, which are limited and depleted, these renewable resources can be naturally replenished, making them virtually inexhaustible for human utilization. The growing global energy crisis has galvanized a renewed emphasis on the advancement and adoption of clean and renewable energy sources. Clean Development Mechanisms (CDMs) [65] are now being implemented by organizations worldwide, fostering a global transition towards more sustainable energy solutions. In contrast to the increasingly scarce fossil fuels, renewable energy sources also offer the advantage of environmental sustainability. Traditional fossil fuels, besides being finite, are associated with pollution and environmental degradation due to their combustion. Renewable energy, on the other hand, offers a cleaner alternative, generating power without the detrimental pollution associated with conventional energy sources. This contrast underscores the growing importance and potential of renewable energy as a viable and responsible choice for future energy needs.

### 2.3.1 Sources of Renewable Energy

Renewable energy, celebrated for its sustainability and environmental friendliness, is sourced from various natural phenomena. Below are more detailed descriptions of some key sources of renewable energy:

Wind Power**:** Modern wind turbines, ranging in rated power from 600 kW to 5 MW [66], are designed to harness the kinetic energy found in airflows. The energy extracted from wind increases dramatically with the wind's velocity, being a cubic function of speed. Recent innovations in aerodynamic design have given rise to aerofoil wind turbines, further optimizing the efficiency of energy capture [67].

Solar Power: With roots tracing back to British astronomer John Herschel [68], solar energy has evolved into a versatile source of renewable power. It can be harnessed in two primary ways: as solar thermal energy for applications such as space heating or converted into electrical energy. The latter can be achieved with solar photovoltaic cells or concentrating solar power plants, making solar power an adaptable and abundant resource [69].

Small Hydropower: Classified as renewable energy sources, small hydropower installations, up to 10MW, convert the potential energy of stored water into electrical energy [70]. This can include run-of-the-river hydroelectricity, where the kinetic energy of flowing water is utilized without the need for dams or reservoirs, thus reducing environmental impact.

Biomass: Is another renewable energy form where plants, capturing sunlight through photosynthesis, function as natural batteries to store solar energy [71], [72]. When these plants are combusted, they release the trapped energy, offering a replenishable energy source that can be used as required.

Geothermal: Is the heat stored within the Earth's layers [73]. This continuous conduction of heat from the core to the surface can be harnessed to heat water, produce superheated steam, and power steam turbines to generate electricity. Historically, the use of geothermal energy was confined to regions near tectonic plate boundaries; however, recent technological advancements have expanded its accessibility, albeit with some geographical limitations.

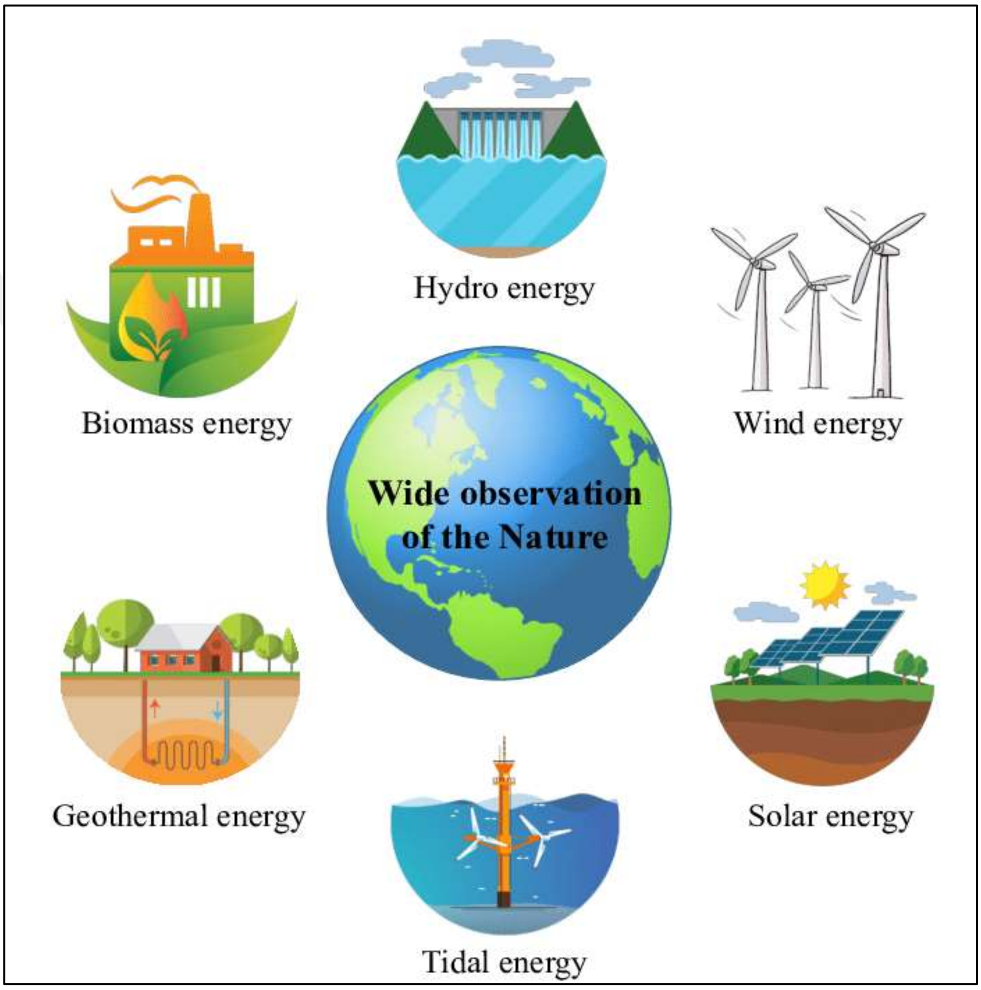



**Figure 2.1**: Renewable Energy Sources.

Each of these sources represents a facet of the broad spectrum of renewable energy options. They collectively embody the movement towards greener energy solutions, responding to the urgent needs of environmental conservation, and resource sustainability. Their diverse nature allows for adaptability and innovation, fostering continued growth in the renewable energy sector.

## 2.4 MAXIMUM POWER POINT TRACKING ALGORITHMS

Maximum power point tracking (MPPT) represents a critical aspect within a photovoltaic (PV) system, facilitating the extraction of the optimal power output at the maximum power point (MPP) [74]. This MPP is heavily influenced by external factors such as irradiation and temperature, both of which are subject to continual fluctuation throughout the course of a day. As these conditions shift, the MPP correspondingly moves, and any deviation from this optimum point may result in significant power losses. Consequently, the utilization of MPPT algorithms becomes imperative in PV applications to ensure the harnessing of maximum power from a solar array. These algorithms acknowledge that for every given set of environmental conditions, there exists a unique optimum terminal voltage at which the PV array should operate to achieve the greatest possible power output, thereby enhancing the system's efficiency. The determination of this specific point, although contingent primarily on irradiance and temperature, is not straightforward, as its precise location remains elusive. However, the MPP can be pinpointed through the deployment of search algorithms or intricate calculation models, each serving to dynamically adapt to the changing conditions and guide the PV system to operate as close to the MPP as possible. This ongoing tracking of the maximum power point ensures that the PV system continually operates at its fullest potential, a feat made possible only through the sophisticated and responsive application of MPPT algorithms.

## 2.5 MACHINE LEARNING METHODS

Machine learning is a field of artificial intelligence that leverages statistical techniques to give computer systems the ability to learn from data and improve performance on a specific task without being explicitly programmed [75]. Within the broad domain of ML, various types and methods can be distinguished, each with unique characteristics and applications.

Supervised Learning: This approach involves training a model on a labelled dataset, meaning that the input data is paired with corresponding output labels [76], [77]. The goal is to learn a mapping from inputs to outputs. Common algorithms within SL include LR, DT, SVM, and NN. This method is widely used in applications like spam detection, image recognition, and sales forecasting.

Unsupervised Learning: eals with datasets without labelled responses. The focus is on discovering patterns and relationships in the data. Clustering (e.g., K-means) and Association (e.g., Apriori) are prevalent techniques in UL [78], [79]. This category is often employed in market segmentation, anomaly detection, and recommendation systems.

Semi-supervised Learning: This method combines aspects of both SL and UL. It utilizes both labelled and unlabelled data for training, often leading to improved model performance [80], [81]. SSL is beneficial when acquiring a fully labelled dataset is expensive or time-consuming.

### 2.5.1 Support Vector Regression

The Support Vector Regression (SVR) algorithm, a specialized form of the Support Vector Machine (SVM), operates on the principle of finding a hyperplane that optimizes the separation margin between classes. Unlike SVM, which focuses on classification, SVR is used for regression and prediction tasks. The SVR works by defining a decision boundary that not only fits the data but also maintains a maximal margin from the data points, known as support vectors [82], [83].

The key components of SVR include the selection of a kernel type (such as linear, polynomial, Radial Basis Function 'rbf', or 'sigmoid'), the tolerance level for the stopping criteria, the degrees for the polynomial kernel function, the kernel coefficient, and the regularization parameter. These components can be configured in various ways to fit specific data scenarios.

Mathematically, SVR solves the following optimization problem:

$$\text{Minimize : } \frac{1}{2}\,||w||^2 + C\sum_{i=1}^{n}(\xi i + \xi i\,*) \tag{2.1}$$

Here, w is the weight vector, b is the bias, ϵ is the tube sensitivity, C is the regularization parameter that controls the trade-off between the complexity of the model and the degree to which deviations larger than ϵ are tolerated, xi are the input features, yi are the actual output values, and ξi and ξi∗ are slack variables.

In essence, SVR strives to find the optimal hyperplane that fits the data, considering the selected kernel function, while also accommodating flexibility through the fine-tuning of various parameters. This makes SVR an adaptable and robust method for regression, capable of handling complex and non-linear relationships in the data.

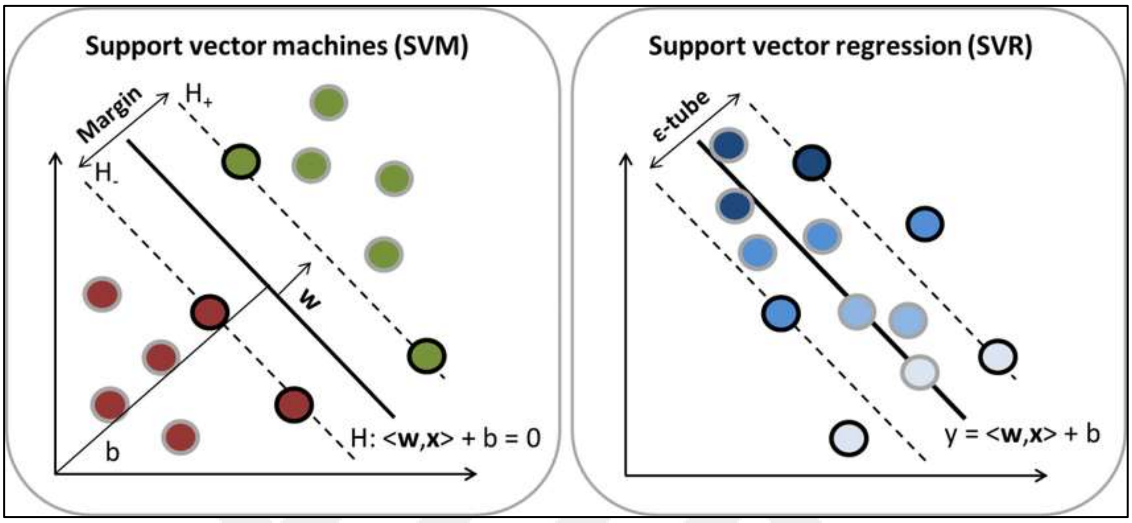



**Figure 2.2:** Example of SVM and SVR.

### 2.5.2 Linear Regression

LR is a statistical method that models the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. The architecture of LR is relatively straightforward, emphasizing simplicity and interpretability [84].

In the simplest form of LR, the relationship between the dependent variable y and a single independent variable x is modelled as a straight line. The mathematical equation representing this relationship is:

$$y = \beta_0 + \beta_1 x + \epsilon \quad (2.2)$$

Here, y is the dependent variable we're trying to predict, x is the independent variable, β0 is the y-intercept of the line, β1 is the slope of the line, and ϵ represents the error term, capturing the deviations of the observed values from the values predicted by the model.

In Multiple LR, where there are more than one independent variables, the equation becomes:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon \quad (2.3)$$

Here, p is the number of independent variables, and β1, β2…, βp are the coefficients that represent the effect that each independent variable x1, x2…, xp has on the dependent variable.

The coefficients are typically estimated using a method called Ordinary Least Squares (OLS), which minimizes the sum of the squared differences between the observed values and the values predicted by the model [85]. The LR model assumes several conditions, such as linearity, independence, homoscedasticity, and normality of errors. If these assumptions hold, linear regression provides a powerful yet simple method for modeling relationships, interpreting the effect of variables, and predicting outcomes.

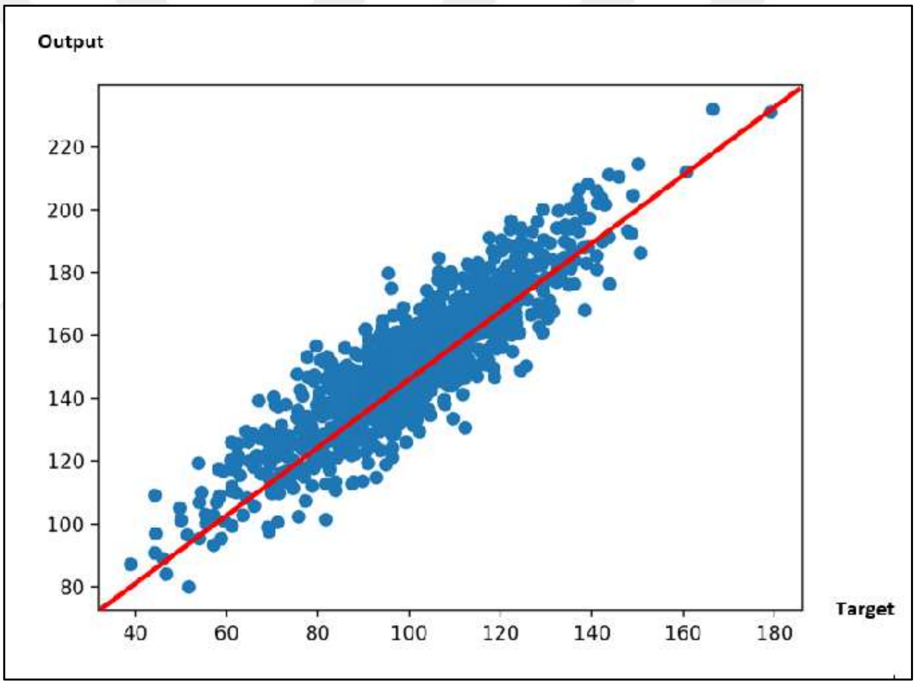



**Figure 2.3:** Linear Regression.

### 2.5.3 XGBoost Regression

XGBoost, standing for "Extreme Gradient Boosting" is an advanced gradient boosting technique that integrates the principles of weak prediction modeling, primarily utilizing decision trees. Its design aims to maximize speed, accuracy, and reliability, exceeding the capabilities of traditional decision tree models.

The foundation of XGBoost lies in the application of gradient boosting algorithms, which function to minimize the differential loss function of weak learners through gradient descent optimization techniques. The XGBoost algorithm can be considered greedy in nature, as it tends to overfit the training dataset rapidly. However, this tendency is counterbalanced by employing methods such as shrinkage, tree constraint, regularization, and stochastic gradient boosting [86].

XGBoost employs gradient boosting algorithms. These algorithms endeavor to diminish the discrepancy of the loss function in weak learners via gradient descent optimization techniques. Mathematically, if L represents the loss function, and $h_t$ represents the prediction of the weak learner at iteration t, the update rule in the gradient boosting framework is:

$$h_{t+1} = h_t - \eta \nabla L(h_t) \tag{2.4}$$

where η is the learning rate.

XGBoost's architecture is distinct in its hybrid ensemble methodology, amalgamating various techniques to enhance the model's accuracy, robustness, and computational efficiency. This amalgamation includes innovative strategies such as clever penalization of trees, proportional shrinking of leaf nodes, the implementation of the Newton boosting algorithm, extra randomization parameters, the application on single distributed systems, out-of-core computation, and an automated feature selection process. XGBoost begins with a weak prediction model and iteratively refines it by concentrating on the mistakes of the initial model. This iterative enhancement continues until predefined parameter limits are attained, creating a learning approach that is adaptive or incremental, learning progressively from its errors.

One of the outstanding characteristics of XGBoost is its performance, often delivering higher accuracy levels than standard decision tree models. However, this increased accuracy comes at the expense of interpretability. While decision trees are praised for their transparent nature and ease of interpretation through a simple traversal of tree nodes, XGBoost's complexity grows with the expansion of trees and nodes. This heightened complexity affects the ability to codify results and articulate an explanation for the achieved accuracy [87].

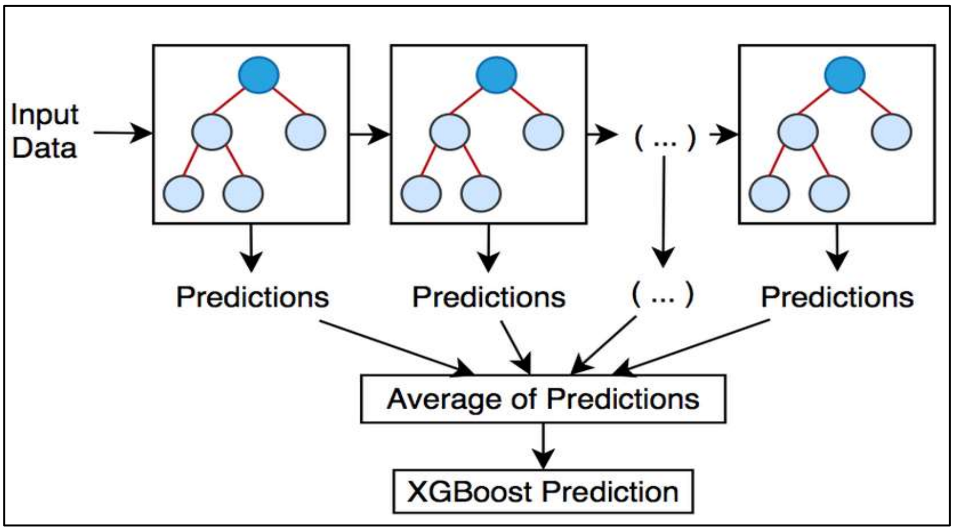



**Figure 2.4:** XGBoost Architecture.

### 2.5.4 Random Forest (RF)

RF regression, a concept pioneered by Breiman [88], stands as a complex and flexible nonparametric ML technique. In contrast to parametric models that require specific functional assumptions, RF regression is free from such restrictions, allowing it to mold itself to a wide array of functional forms and thus adapt to a multitude of regression scenarios.

Constructed as a collection of decision trees, the architecture of RF is both novel and effective. Each tree within the ensemble takes part in binary splitting to discern the output value from the given inputs. Uniquely, RF builds each tree by randomly selecting samples from the dataset with replacement, and at every node, a subset of input variables smaller than the total 'p' is considered for splitting. This subset size, referred to as 'mtry', often corresponds to p/3, but this ratio can vary according to the specific problem [89]. 'Mtry' functions as a vital tuning parameter, shaping the individual decision trees within the ensemble.

Mathematically, if T represents the number of trees in the RF model, and p is the total number of input variables, the aggregate prediction $\hat{y}$ for an instance x is given by:

$$\hat{y}(x) = \frac{1}{T} \sum_{i1}^{T} \hat{y}_i \, (x) \tag{2.5}$$

where $\hat{y}$i (x) is the prediction of the $i^{\text{th}}$ tree. The aggregate prediction from the RF model is derived by averaging the numerical results yielded by its multiple decision trees. With hundreds of trees each limited to a subset of variables, the RF model manifests robust characteristics, minimizing the risk of overfitting. This robust nature also equips the model to efficiently handle data that might deviate from a normal distribution.

A significant facet of the success of RF is the deliberate design that inhibits the trees from invariably relying on the most potent predictor variable for the initial split. By compelling the usage of varied predictors, the ensemble's trees are rendered less correlated, fostering a more stable and trustworthy aggregated prediction. Consequently, this lessens the variations and boosts the reliability of the model, distinguishing it from other ensembles that might lean heavily on a singular, primary predictor variable [89]. In essence, Random Forest regression encapsulates a blend of innovation, adaptability, and reliability, offering a robust solution to complex regression tasks.

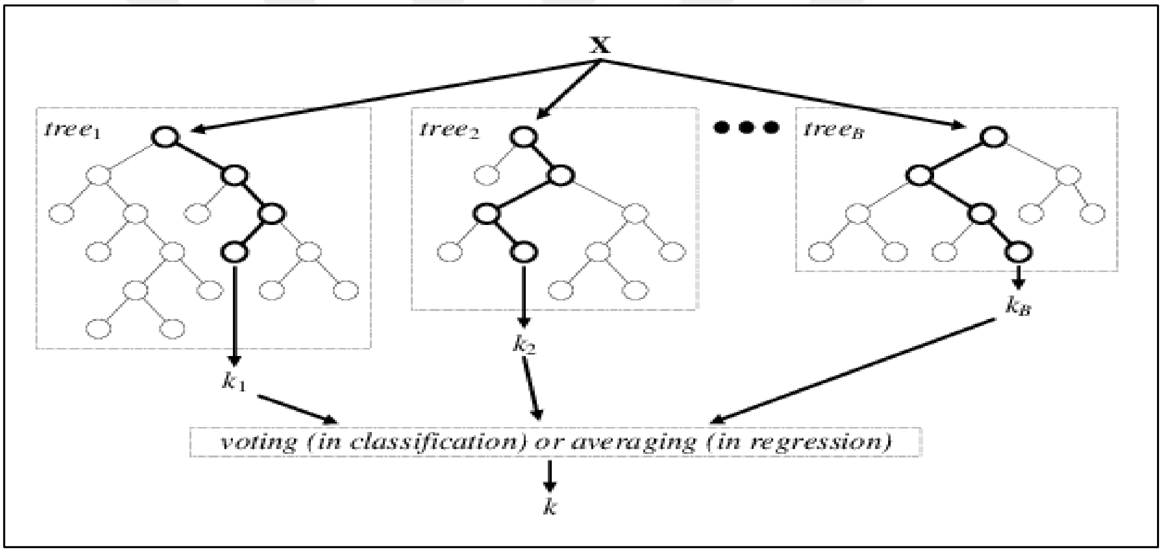



**Figure 2.5:** Random Forest Architecture.

### 2.6 ENSEMBLE LEARNING METHODS

Ensemble models represent an attractive alternative to conventional ML algorithms, leveraging the collective power of multiple diverse predictors. These models function by integrating the output values from various predictors, culminating in a singular prediction. The primary advantage of this approach lies in its ability to enhance the accuracy of both classification and regression tasks, frequently achieving this improvement with a reduced computational time. Unlike many state-of-the-art ML algorithms, which often demand intricate tuning and extensive domain expertise, ensemble methods tend to be more user-friendly, requiring less specialized tuning. Whether utilized in classifiers or regressors, ensemble techniques have proven to be highly effective across an array of applications. From image recognition to medical diagnostics, network security, and beyond, ensemble learning has demonstrated a robust performance, underscoring its versatility and potential as a tool in modern ML practices [90].

Ensemble learning models offer a sophisticated method for improving the performance of ML tasks, as explored by various scholars such as Bauer and Kohavi [91] and Rokach [92]. These models primarily incorporate two vital techniques: bagging and boosting. Bagging, short for bootstrap aggregating, is an approach first introduced by Breiman [88]. It's a relatively straightforward method that involves creating independent predictors from random samples of the training set and then averaging their outputs. On the other hand, boosting techniques, like AdaBoost [93], utilize consecutive training of predictors with continuous adjustments of the ensemble weights, thus making them dependent models, as classified by Rokach [92]. Hastie [94] succinctly captures ensemble learning as a two-fold process: The development of a population of base learners from the training data and the combination of these base learners into a composite predictor.

A fundamental element in the successful construction of ensembles is the concept of diversity. It is imperative that the weak predictors within the ensemble are diverse or at least uncorrelated [92], [95], and this diversity can be achieved through various means. Manipulating the training samples, the features, or the weak predictors' parameters are common strategies, and the hybrid utilization of different learning algorithms gives rise to what's known as heterogeneous ensembles. Such diversity has been shown to offer a superior trade-off between bias and variance, as evidenced by Oza and Tumer [90]. Ensemble learning's architecture, as illustrated in Figure 2.6.

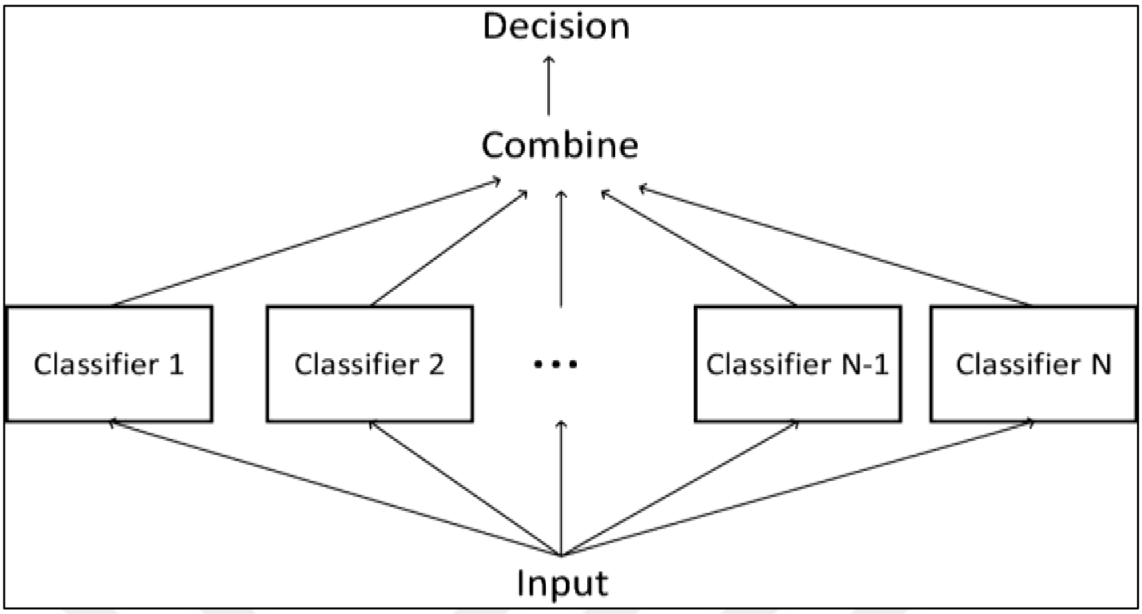



**Figure 2.6:** Ensemble Learning Architecture.

### 2.6.1 Bagging

Bootstrap aggregating, commonly known as bagging, is a renowned method introduced by Breiman [88]. This technique hinges on the construction of independent predictors, each developed from random samples of the training set. The final prediction is derived by either averaging (for regression problems) or voting (for classification problems) the outputs of these individual predictors. Specifically, in the context of regression, the prediction is formulated as:

$$f(x) = \frac{1}{T} \sum_{i=1}^{T} f_i(x_0) \tag{2.6}$$

where each model $f_i$ is trained on the i-th bootstrap sample.

Breiman's work illuminates the fact that, for both classification and regression tasks, bagging ensembles of DTs perform markedly better than singular trees. He also provides insightful reasoning behind the success of bagging, emphasizing that due to the bootstrap method, every predictor garners a unique and confined perspective of the training data. The term "bootstrap" refers to a statistical resampling technique where random samples are repeatedly selected with replacement from a single dataset, providing an accurate representation of the underlying distribution [94].

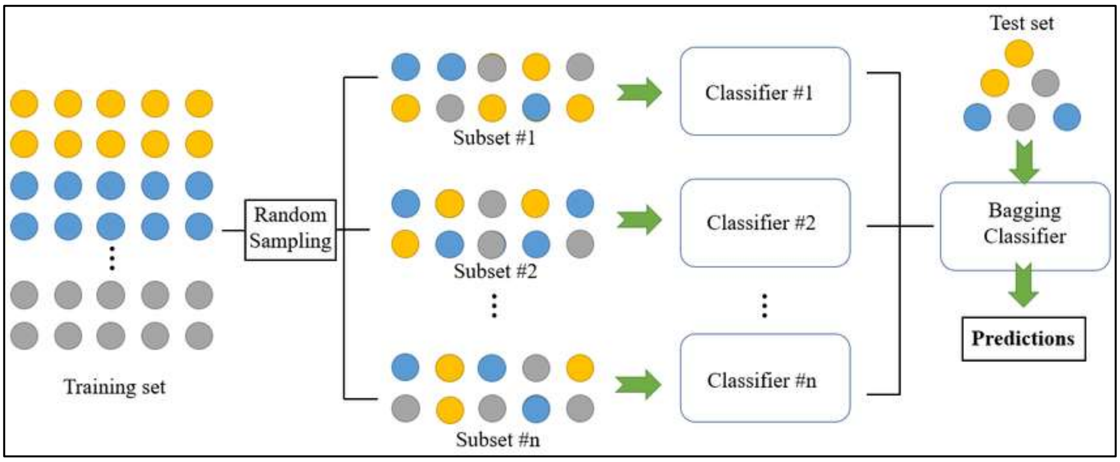



**Figure 2.7:** Bagging Architecture.

### 2.6.2 Boosting

Boosting techniques, such as AdaBoost [93], operate by integrating T individual predictors, yet the process diverges significantly from methods like bagging. Instead of training the predictors independently, boosting engages in an iterative procedure where each training instance is assigned a weight, which is then modified in each iteration. With every iteration, a new predictor is added to the ensemble, and the quality of the ensemble's predictions is subsequently evaluated against the training set instances.

Incorrectly classified instances are given increased weights compared to the correctly classified ones, enhancing the overall classification with additional iterations. This approach has often led AdaBoost and similar boosting techniques to surpass the predictive quality of bagging algorithms. However, this improvement comes with certain caveats, such as a tendency for boosting algorithms to suffer from overfitting problems [91], [92]. Moreover, AdaBoost is not universally superior to bagging, as indicated in some studies [96].

One significant disadvantage, particularly relevant for certain applications, is the inherently sequential nature of the adaptive boosting algorithm. Unlike other methods that can train weak predictors in parallel, boosting requires a step-by-step process, which may hinder efficiency and scalability in particular scenarios. This limitation might influence the choice of algorithm, depending on the specific requirements and constraints of the task at hand.

### 2.6.3 Stacking

Stacked generalization, or stacking, is a specialized method of blending the outputs of ensemble members [90], [94], [97]. Unlike simpler techniques that might rely on merely averaging or weighting the output values, Stacking introduces an additional layer of complexity by training another model on the individual predictions, as depicted in Figure 2-9. In Stacking, the top layer model calculates the final output using linear regression or any suitable ML algorithm, with the aim of accounting for the behaviour of the individual predictors. By training on the errors of the single models, stacking endeavours to evaluate and correct these errors, leading to potentially more accurate predictions.

However, despite these potential advantages, Stacking was not employed in this work due to its inherent complexities and uncertainties. It's a more intricate method compared to other ensemble techniques, and its success in enhancing predictions is not guaranteed. Additionally, the computational burden of Stacking is considerable. The optimization required for both the ensemble members and the regression in the final layer can lead to high computational costs, making it a less appealing option in scenarios where efficiency and simplicity are paramount.

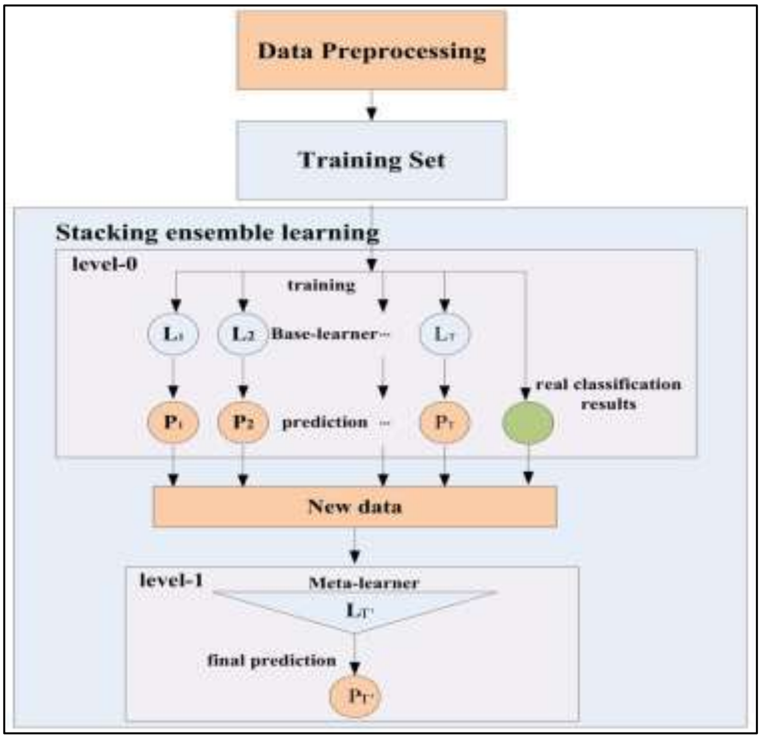



**Figure 2.8:** Stacking Architecture.

## 2.7 CONCLUSION

chapter has provided a comprehensive overview of the existing research and studies related to renewable energy and MPPT algorithms. Renewable energy sources play a vital role in addressing environmental challenges and ensuring sustainable energy generation. The exploration of various ML and EL techniques, has shed light on their potential in maximizing the efficiency of renewable energy systems.

# 3. METHODOLOGY

## 3.1 INTRODUCTION

The advancement of renewable energy, particularly solar energy, represents a cornerstone in the global shift towards sustainable and eco-friendly power solutions. The ability to accurately predict solar energy production is not just a scientific endeavor; it's a necessity that could revolutionize energy management and distribution. This recognition serves as the impetus for the research undertaken in this thesis.

This chapter delineates the comprehensive approach adopted to investigate and model solar energy production. It provides an in-depth look into the research design, detailing the rationale behind the chosen methods, the data description, EDA, the application of individual ML models, and the incorporation of ensemble learning strategies.

## 3.2 THESIS APPROACH

The proposed model for the thesis is a comprehensive approach designed to predict solar energy production, employing an intricate blend of ML algorithms and ensemble learning techniques. The process begins with the utilization of two distinct datasets related to solar energy. These datasets are subject to Exploratory Data Analysis (EDA), a crucial step that helps in understanding the underlying structure of the data, identifying patterns, and uncovering anomalies or inconsistencies that may need correction.

Once the EDA is completed, the data is systematically divided into two sets: the training set, used to fit the models, and the test set, employed to evaluate their predictive performance. At this stage, individual ML models are applied to the training data, including established algorithms like LR, XGBoost Regressor, and SVR. Each of these models offers unique advantages and can provide insights into different aspects of the solar energy prediction task.

The process doesn't stop with individual model training; instead, it advances to a more complex stage, employing EL techniques. Specifically, the model leverages a Bagging Regressor, an ensemble meta-estimator that fits base regressors on random subsets of the original dataset and then aggregates their individual predictions to form a final prediction.

This approach brings together the strengths of multiple underlying models, potentially improving predictive accuracy and robustness.

The culmination of this well-structured process is the evaluation of the ensemble model, where its performance is rigorously assessed using appropriate metrics and validation techniques. The proposed model aims to deliver a more nuanced and robust understanding of solar energy production, showcasing the potential and flexibility of modern EL methodologies in the renewable energy sector.

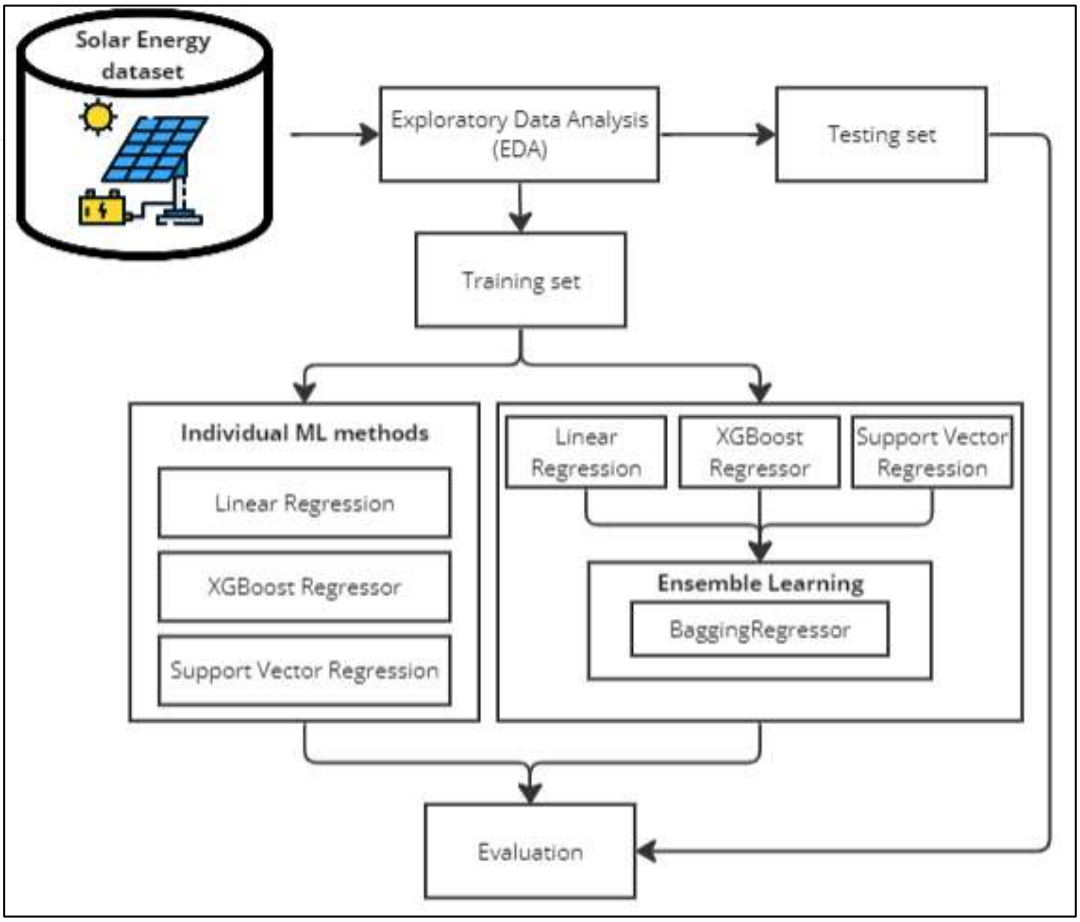



**Figure 3.1:** The Proposed Model.

### 3.2.1 Data Description

This section focuses on the Data Description, offering a detailed insight into two distinct datasets related to solar power. These datasets are vital in understanding and analyzing various aspects of solar energy, including power generation and solar radiation prediction. The information derived from these datasets serves as the foundation for further exploration and analysis, opening doors to forecast energy production, evaluate the efficiency of solar panels, and understand weather-related dependencies that influence solar radiation. The detailed features and attributes within these datasets provide a comprehensive view of the elements influencing solar energy, setting the stage for an in-depth discussion.

#### 3.2.1.1 Solar power generation data

The "Solar Power Generation Data" dataset is an extensive collection of information comprising 8760 rows and 8 columns, as visualized in Figure 3.2. The dataset is available for access and download from Kaggle. The power generation datasets are obtained at the inverter level, where each inverter is responsible for multiple lines of solar panels. On the other hand, the sensor data is collected at the plant level, utilizing a single array of sensors strategically positioned throughout the plant. With this dataset, several areas of concern at solar power plants can be addressed. One significant aspect is the ability to predict power generation for the upcoming days. Accurate predictions in this regard can greatly assist in efficient grid management, allowing for better distribution and utilization of the generated electricity. Additionally, the dataset provides an opportunity to identify the need for panel cleaning or maintenance. As solar panels are exposed to various environmental factors, such as dust, debris, and weather conditions, they can gradually accumulate dirt and other substances, potentially reducing their efficiency. By analyzing the sensor readings and power generation data, it becomes possible to detect any patterns or correlations that indicate a decline in performance due to panel contamination. This information can prompt the scheduling of cleaning or maintenance activities, ensuring optimal functioning of the solar panels. Moreover, the dataset allows for the identification of faulty or sub optimally performing equipment within the solar power plants. By analyzing the power generation data and comparing it with the expected values, any deviations or anomalies can be detected, indicating potential faults or inefficiencies in the equipment, such as inverters or solar panels. This information is crucial for proactive maintenance and timely repairs, minimizing downtime and maximizing the overall output of the solar power plants.

| | Date-Hour(NMT) | WindSpeed | Sunshine | AirPressure | Radiation | AirTemperature | RelativeAirHumidity | SystemProduction |
|---|---|---|---|---|---|---|---|---|
| 0 | 01.01.2017-00:00 | 0.6 | 0 | 1003.8 | -7.4 | 0.1 | 97 | 0.0 |
| 1 | 01.01.2017-01:00 | 1.7 | 0 | 1003.5 | -7.4 | -0.2 | 98 | 0.0 |
| 2 | 01.01.2017-02:00 | 0.6 | 0 | 1003.4 | -6.7 | -1.2 | 99 | 0.0 |
| 3 | 01.01.2017-03:00 | 2.4 | 0 | 1003.3 | -7.2 | -1.3 | 99 | 0.0 |
| 4 | 01.01.2017-04:00 | 4.0 | 0 | 1003.1 | -6.3 | 3.6 | 67 | 0.0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 8755 | 31.12.2017-19:00 | 4.1 | 0 | 988.2 | -4.8 | -0.7 | 94 | 0.0 |
| 8756 | 31.12.2017-20:00 | 2.1 | 0 | 987.3 | -5.0 | -0.3 | 95 | 0.0 |
| 8757 | 31.12.2017-21:00 | 1.8 | 0 | 986.7 | -5.3 | 0.2 | 93 | 0.0 |
| 8758 | 31.12.2017-22:00 | 2.2 | 0 | 986.0 | -5.4 | 0.3 | 92 | 0.0 |
| 8759 | 31.12.2017-23:00 | 2.4 | 0 | 985.6 | -5.9 | 0.4 | 96 | 0.0 |

8760 rows × 8 columns

**Figure 3.2:** Solar Power Generation Data.

In the given dataset, the features are as follows:

a. Date-Hour (NMT): This feature represents the date and hour of the data point, indicating the specific time at which the measurements were taken.

b. WindSpeed: This feature indicates the speed of the wind, which can have an impact on the efficiency and performance of the solar panels.

c. Sunshine: This feature represents the amount of sunshine or sunlight received during the specified time, which is an essential factor in determining solar power generation.

d. AirPressure: This feature denotes the atmospheric pressure, which can indirectly influence solar panel performance.

e. Radiation: This feature represents the solar radiation or the amount of solar energy reaching the solar panels, which is a crucial factor in power generation.

f. AirTemperature: This feature indicates the ambient air temperature at the solar power plant location, which can affect the efficiency of the solar panels.

g. RelativeAirHumidity: This feature represents the relative humidity of the air, which can also impact the performance of the solar panels.

The label or the target variable in this dataset is the "system prediction" or the predicted power generation. It is the variable that we aim to predict based on the given features.

### 3.2.1.2 Solar radiation prediction

The dataset provided is from the Space Apps Moscow event, specifically focused on measurements related to weather conditions. It includes various columns such as "wind direction," "wind speed," "humidity," and "temperature." The objective of the dataset is to predict the level of solar radiation, with the assumption that solar energy batteries are being used and the goal is to determine if it would be reasonable to utilize them in the future. The dataset encompasses measurements taken over the course of the past four months, providing a temporal aspect to the data. By analyzing the relationships between different factors, one can explore dependencies and their impact on solar radiation. Two specific dependencies that have been identified are the relationship between humidity and solar radiation, as well as temperature and solar radiation.

The "Solar Radiation Prediction" dataset, encompassing an extensive collection of 32,686 rows and 11 columns, provides detailed insights into various weather conditions and their effects on solar radiation.

| | UNIXTime | Data | Time | Radiation | Temperature | Pressure | Humidity | WindDirection(Degrees) | Speed | TimeSunRise | TimeSunSet |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1475229326 | 9/29/2016 12:00:00 AM | 23:55:26 | 1.21 | 48 | 30.46 | 59 | 177.39 | 5.62 | 06:13:00 | 18:13:00 |
| **1** | 1475229023 | 9/29/2016 12:00:00 AM | 23:50:23 | 1.21 | 48 | 30.46 | 58 | 176.78 | 3.37 | 06:13:00 | 18:13:00 |
| **2** | 1475228726 | 9/29/2016 12:00:00 AM | 23:45:26 | 1.23 | 48 | 30.46 | 57 | 158.75 | 3.37 | 06:13:00 | 18:13:00 |
| **3** | 1475228421 | 9/29/2016 12:00:00 AM | 23:40:21 | 1.21 | 48 | 30.46 | 60 | 137.71 | 3.37 | 06:13:00 | 18:13:00 |
| **4** | 1475228124 | 9/29/2016 12:00:00 AM | 23:35:24 | 1.17 | 48 | 30.46 | 62 | 104.95 | 5.62 | 06:13:00 | 18:13:00 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **32681** | 1480587604 | 12/1/2016 12:00:00 AM | 00:20:04 | 1.22 | 44 | 30.43 | 102 | 145.42 | 6.75 | 06:41:00 | 17:42:00 |
| **32682** | 1480587301 | 12/1/2016 12:00:00 AM | 00:15:01 | 1.17 | 44 | 30.42 | 102 | 117.78 | 6.75 | 06:41:00 | 17:42:00 |
| **32683** | 1480587001 | 12/1/2016 12:00:00 AM | 00:10:01 | 1.20 | 44 | 30.42 | 102 | 145.19 | 9.00 | 06:41:00 | 17:42:00 |
| **32684** | 1480586702 | 12/1/2016 12:00:00 AM | 00:05:02 | 1.23 | 44 | 30.42 | 101 | 164.19 | 7.87 | 06:41:00 | 17:42:00 |
| **32685** | 1480586402 | 12/1/2016 12:00:00 AM | 00:00:02 | 1.20 | 44 | 30.43 | 101 | 83.59 | 3.37 | 06:41:00 | 17:42:00 |

32686 rows × 11 columns

**Figure 3.3:** Solar Radiation Prediction.

Features:

a. Temperature: The temperature recorded during the measurements.

b. Pressure: The atmospheric pressure recorded during the measurements.

c. Humidity: The humidity level recorded during the measurements.

d. WindDirection (Degrees): The wind direction in degrees during the measurements.

e. Speed: The wind speed recorded during the measurements.
f. Month: The month in which the measurements were taken.
g. Day: The day of the month when the measurements were taken.
h. Hour: The hour of the day when the measurements were taken.
i. Minute: The minute of the hour when the measurements were taken.
j. Second: The second of the minute when the measurements were taken.
k. SunPerDayHours: The duration of sunlight per day in hours.

**Label:**

Radiation: The level of solar radiation recorded during the measurements.

The features provide information about the weather conditions and time at which the measurements were taken. These features are used to predict the corresponding level of solar radiation, which serves as the label or target variable in this dataset.

The "Solar Radiation Prediction" dataset, essential for the analysis of weather-related factors influencing solar energy, is publicly available on Kaggle, a renowned platform for data science competitions and datasets.

### 3.2.2 Exploratory Data Analysis (EDA)

#### 3.2.2.1 EDA of solar power generation data

An exploratory data analysis (EDA) technique is used to visualize the total power production of a solar power plant during the year 2017. The line plot is created using the seaborn library, with time (year-month) on the x-axis and the system's power production in megawatts (MW) on the y-axis. The line plot helps in understanding the trend and pattern of power production over time. By visualizing the data in this way, any noticeable variations, seasonality, or trends in power production can be identified. This information can be valuable for analyzing the performance of the solar power plant and understanding the factors that influence power generation. The x-axis, representing time, allows for the identification of any patterns or cycles in power production throughout the year. The y-axis, representing power production, provides an insight into the overall level of power generated by the system at different points in time.

Figure 3.4 provides a visual representation of the power production across the year 2017. It illustrates the total amount of energy generated, often broken down by months or specific intervals, allowing for a clear understanding of the production trends, peaks, and troughs within that year. Whether segmented by different energy sources, geographical regions, or other categorizations, this figure serves as an essential reference for analyzing the efficiency, sustainability, and potential growth or challenges in the power production sector for that particular year. It can be an instrumental tool for policymakers, energy companies, and researchers to glean insights into the energy landscape of 2017.

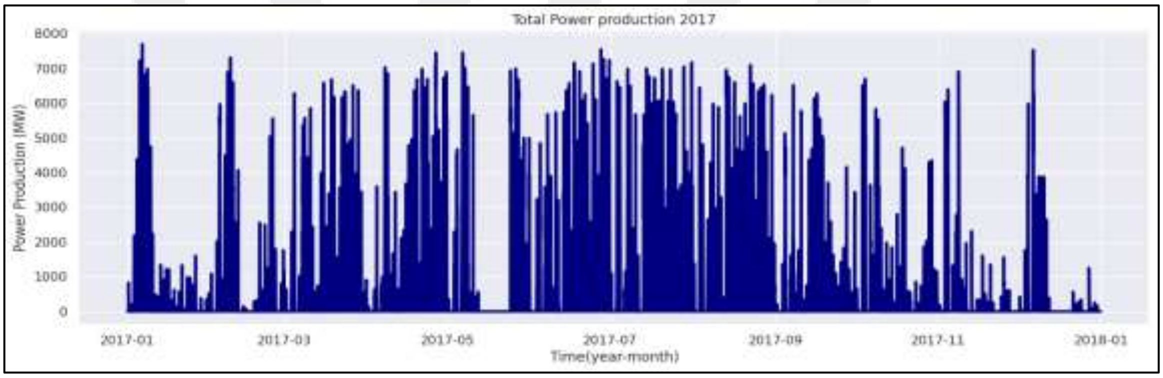



**Figure 3.4:** Total Power Production in 2017.

#### 3.2.2.2 EDA of solar radiation prediction

During the EDA, a feature correlation analysis was conducted to understand the relationships between different variables in the dataset. The variables used for the correlation analysis were "Temperature," "Pressure," "Humidity," "WindDirection (Degrees)," and "Speed." These variables were selected to investigate their potential influence on solar radiation levels. To visualize the distributions of these variables, histograms with kernel density estimation (KDE) plots were created. Each variable was plotted in a separate subplot within a larger figure. The overall figure size was set to 20 by 10 to ensure clear and informative visualizations. The histograms with KDE plots provided insights into the distribution patterns and the central tendencies of the selected variables. This analysis helped in understanding the spread and concentration of data points for each variable. By examining the shape of the distributions, it became possible to identify any significant skewness or outliers present in the data. Through this EDA process, the aim was to gain a better understanding of the individual variables and their potential relationship with solar radiation. By visualizing the distributions, it became possible to identify any noteworthy trends, patterns, or anomalies that might exist within the data. These insights could then be used to inform further analysis and model building, aiding in the prediction of solar radiation levels based on the given features.

The figure 3.5, presents the correlation matrix of the dataset provides a concise visual representation of how different features such as Temperature, Pressure, Humidity, WindDirection (Degrees), and Speed are interrelated. By displaying the correlation coefficients between these variables, the matrix offers insights into their linear relationships, highlighting the degree to which they either positively or negatively influence each other. For instance, strong positive correlations between variables could indicate that as one feature increases, the other tends to as well, while negative correlations signify an inverse relationship. This visualization aids in understanding the underlying patterns and potential multicollinearity within the dataset, which can be vital in the subsequent modeling and predictive analysis of solar radiation.

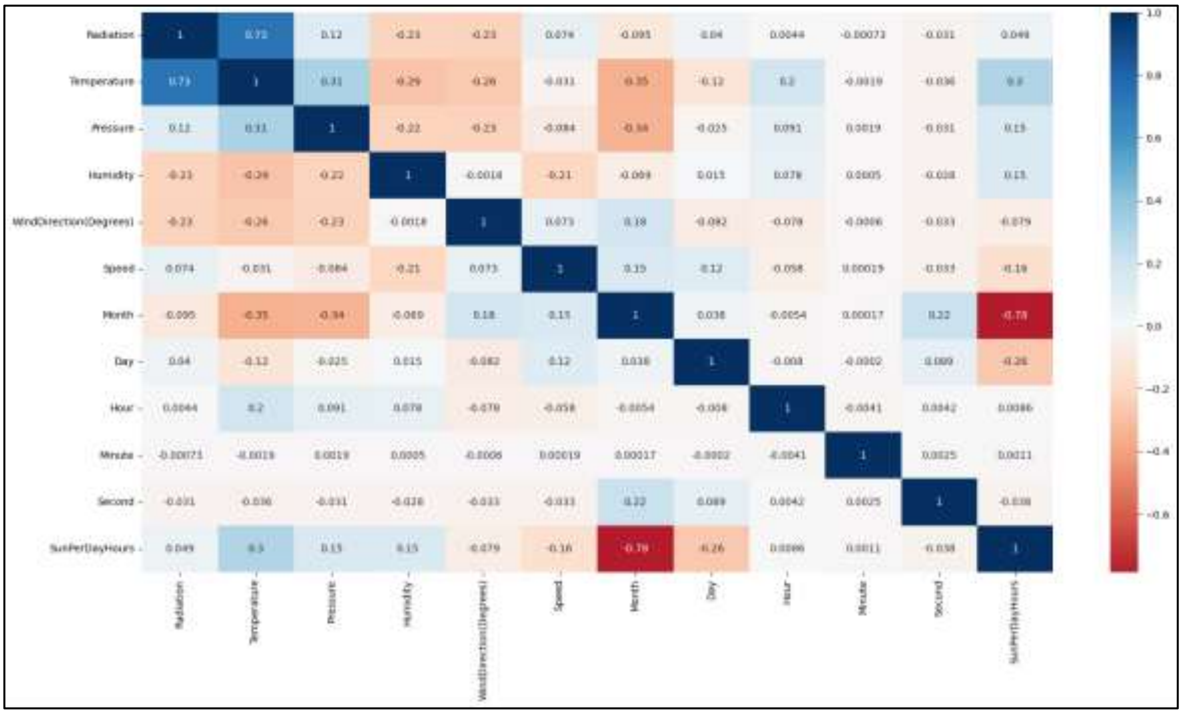



**Figure 3.5:** Feature Correlation.

The figure 3.6 that illustrates the plot of Solar Radiation Prediction, incorporating features such as Temperature, Pressure, Humidity, WindDirection (Degrees), and Speed, offers a multifaceted view of the factors affecting solar radiation. This visualization effectively showcases the interplay and correlations between these essential meteorological variables and how they collectively influence solar radiation levels. The visual representation allows for an in-depth understanding of the dependencies between these features, revealing how fluctuations in these elements may alter solar energy potential.

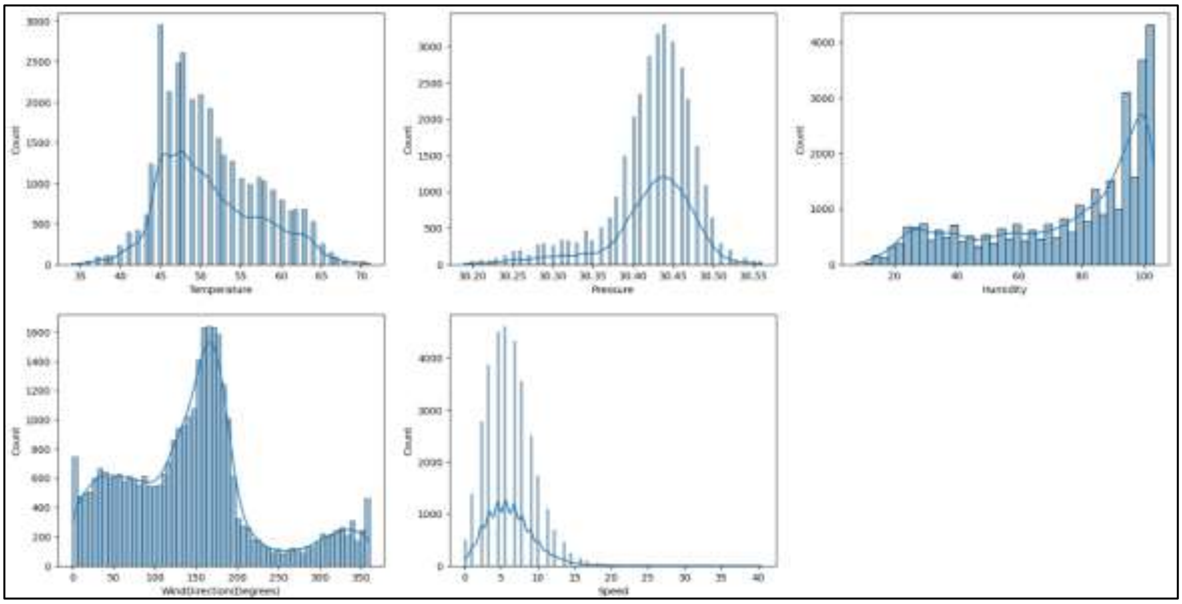



**Figure 3.6:** Plot of Solar Radiation Prediction.

### 3.2.3 Data Splitting

In the process of preparing the data for modeling, the dataset was divided into two distinct subsets: 80% was allocated for training, and the remaining 20% was set aside for testing. This split ensures that the model has a substantial amount of data to learn from, while still reserving a portion for evaluation, allowing for an unbiased assessment of the model's predictive performance. By segmenting the data in this manner, it facilitates a more rigorous validation of the model's effectiveness, as it will be tested on unseen data, reflecting a more realistic scenario of how the model might perform in practice.

### 3.2.4 Individual Model

In the proposed model, several individual ML methods were employed to comprehend and predict the underlying patterns in the data. Among these methods are LR, XGBoost Regressor, and SVR.

#### 3.2.4.1 Linear regression method

In the solar energy-focused model crafted for the thesis, Linear Regression is a key component that enables the comprehension of the intricate relationship between various meteorological attributes and solar power generation. The architecture is founded on the LinearRegression() function, marking the initiation of this linear analytical process. During the training phase, the fit (X_train, y_train) method is employed to align the model with the training data, establishing the optimal linear equation that symbolizes the connection between the independent variables in X_train and the corresponding dependent solar energy outputs in y_train. This part of the process is essential, as it sets the stage for accurately mapping the correlations within the data, reflecting how changes in factors like temperature, wind speed, or pressure linearly translate into alterations in solar power production. Following this training, the model's predictive capabilities are exercised using the predict(X_test) method, producing a series of predictions represented by y_pred. These predictions are more than mere estimations; they are insights into how the solar energy system may respond to specific environmental conditions, all derived from the linear relationship uncovered during the model's training. The elegance of Linear Regression, encapsulated in its simplicity, efficiency, and interpretability, resonates well with the application at hand. It not only elucidates the underlying patterns governing solar energy production but also offers a methodologically robust and scientifically grounded approach. It's a chosen tool within the thesis, facilitating both the understanding of the current dynamics of solar energy and the potential to enhance future applications in the renewable energy field.

#### 3.2.4.2 XGBoost regressor method

In the context of solar energy modeling within the thesis, the XGBoost regression architecture is employed as a potent ML tool, providing both precision and efficiency. Utilizing the xgb.XGBRegressor() function, the architecture builds upon the principles of gradient boosting frameworks by combining the predictions of several base estimators (typically decision trees) in order to improve accuracy and robustness.

XGBoost operates by sequentially adding predictors to an ensemble, correcting the predecessors' errors. Each tree in the ensemble is constructed to correct the errors made by the previous trees, hence reducing both bias and variance. Through a process of iterative refinement, the XGBoost model tunes itself to the underlying complexities and patterns in the data, such as those found in the meteorological variables affecting solar power generation.

One of the remarkable features of XGBoost is its ability to handle different types of predictor variables, allowing for integration and analysis of diverse data such as temperature, wind speed, humidity, and more, all of which are critical factors in solar energy prediction. The architecture also allows for fine-grained control over hyperparameters, enabling customization and optimization for specific datasets or prediction requirements. Additionally, XGBoost provides advantages in computational efficiency by utilizing parallel processing and offering an option to handle missing data, which can be a common occurrence in real-world solar energy datasets. This ensures that the model is not only theoretically sound but also practically viable, capable of handling large and complex datasets relevant to solar energy applications.

In terms of the solar energy domain, the application of XGBoost can lead to nuanced insights and accurate predictions regarding solar radiation and energy production. By modeling the intricate relationships between weather conditions and solar output, it contributes to a deeper understanding of how various environmental factors come together to influence the energy harvested from the sun. This can further aid in optimizing the operation of solar power plants, scheduling maintenance, enhancing grid management, and ultimately contributing to the broader goals of renewable energy efficiency and sustainability.

#### 3.2.4.3 Support vector regression method

In the exploration of solar energy within the thesis, Support Vector Regression (SVR) with a Radial Basis Function (RBF) kernel offers a sophisticated methodology to predict and model the complex relationships within the data. The SVR architecture with an RBF kernel, instantiated by SVR (kernel='rbf'), is particularly attuned to handling nonlinear relationships, a common occurrence in solar energy data.

Support Vector Regression operates by finding a hyperplane that best fits the data in a higher-dimensional space. The RBF kernel serves as a transformation function that maps the input features into this higher-dimensional space, enabling the model to capture nonlinear patterns without explicitly transforming the data. This is crucial for modeling solar energy, where the relationship between various features like temperature, humidity, pressure, wind speed, and solar radiation can be highly intricate and nonlinear. By employing SVR with an RBF kernel, the model focuses on defining a boundary that incorporates a specified margin of tolerance, known as the epsilon-insensitive tube. Data points falling within this tube are considered error-free, allowing the model to become more robust to minor fluctuations and noise. This makes SVR an appealing choice for solar energy data, which may contain subtle variations and inconsistencies arising from environmental and instrumental factors. One of the standout features of SVR is its ability to manage both underfitting and overfitting through careful tuning of hyperparameters such as the cost parameter (C) and the kernel parameter (gamma). This fine-grained control ensures that the model can be tailored to the specific characteristics of the solar energy data, achieving an optimal balance between bias and variance.

In practical terms, the application of SVR with an RBF kernel in solar energy modeling provides valuable insights into the factors influencing solar power generation. For instance, understanding the nonlinear interactions between atmospheric conditions and solar panel output can aid in predictive maintenance, efficiency optimization, and more accurate energy forecasting. The ability to model complex relationships also supports decision-making in areas such as solar farm placement, equipment selection, and energy grid management, translating into real-world impact and value.

### 3.2.5 Ensemble Learning Method

In the context of solar energy modeling, Ensemble learning represents an innovative approach that leverages the strengths of multiple individual regression models to create a more robust and accurate predictive system. The specific Ensemble learning architecture employed in the thesis combines Linear Regression, Support Vector Regression (SVR) with an RBF kernel, and XGBoost Regressor. These models are integrated using the BaggingRegressor, providing a comprehensive framework to analyze and predict solar energy-related variables. The ensemble model is instantiated by the BaggingRegressor method with 20 estimators and a specific random state for reproducibility. This bagging technique operates by training each base estimator on a random subset of the training data and then aggregating their predictions. The base estimators consist of Linear Regression, SVR with an RBF kernel, and XGBoost Regressor, each offering unique contributions to the ensemble.

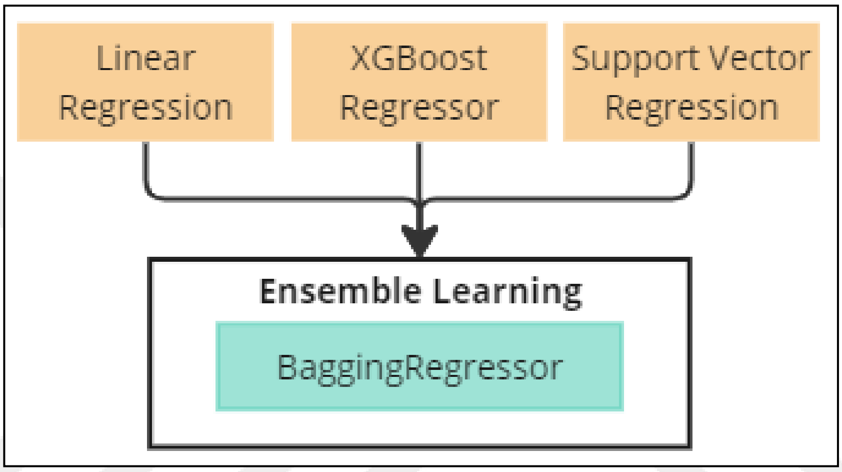



**Figure 3.7:** Ensemble Learning Method.

Linear Regression: This model contributes simplicity and interpretability, providing a linear perspective on the relationships within the data. It can capture straightforward patterns and offer an easy-to-understand baseline for the ensemble.

Support Vector Regression: The SVR model, with its ability to handle nonlinear relationships, brings flexibility and adaptability to the ensemble. It can capture complex patterns that may not be evident to a linear model, enhancing the ensemble's ability to generalize from the data.

XGBoost Regressor: XGBoost adds a sequential learning dimension to the ensemble. It builds trees iteratively, focusing on areas where previous trees have made mistakes, thereby boosting the model's performance on challenging aspects of the data.

By integrating these diverse models into an ensemble using bagging, the architecture seeks to create a consensus that leverages the best aspects of each model. Bagging helps in reducing the variance of the individual models, resulting in a more stable and resilient predictive system. The random subsampling of data for each base estimator ensures that the ensemble can capture a broader perspective on the underlying relationships in the solar energy data. This ensemble approach is particularly powerful for solar energy modeling, where various factors such as temperature, humidity, radiation, wind speed, and pressure interact in complex ways. The multifaceted nature of the ensemble allows it to embrace these complexities, providing a nuanced understanding that transcends what any single model might achieve.

## 3.3 CONCLUSION

The methodology chapter of this thesis has meticulously detailed the multifaceted approach undertaken to predict solar energy production, a critical aspect in the broader context of renewable energy research. The chapter began with data description, followed by exploratory data analysis to glean insights into the underlying structure of the data. The use of established ML models, coupled with innovative ensemble learning techniques, signifies a sophisticated and well-rounded approach.

# 4. RESULTS AND DISCUSSION

## 4.1 INTRODUCTION

In the following chapter, the results obtained from the comprehensive analysis of various ML models applied to the Solar Radiation Prediction dataset are presented and critically discussed. Utilizing a diverse set of statistical metrics, the performance of individual models such as LR, SVR, XGBoost Regressor, and EL has been rigorously evaluated. The discussions delve into interpreting the significance of these results, and implications of each model in the context of solar radiation prediction. This analysis provides pivotal insights that contribute to the overarching goal of enhancing the accuracy and efficiency of solar energy forecasting, thus aligning with the broader objectives of sustainable energy management and climate stewardship.

## 4.2 EVALUATION METRICS

The assessment of the model's performance is a critical step in understanding its efficacy and precision in making predictions. This evaluation is achieved through the utilization of standard metrics that are grounded in the observed and predicted values produced by the model. By quantifying the performance, these metrics facilitate a more tangible and objective comparison with other models. This numerical representation simplifies the task of measuring how well the model has learned from the training data and how effectively it generalizes to unseen data, providing essential insights into its reliability and accuracy.

R2 Score (R2): also known as the coefficient of determination, is a crucial metric in evaluating the accuracy of a forecasting model. It quantifies the proportion of the variance in the dependent variable that is predictable from the independent variables. The R2 Score ranges from 0 to 1, with a value of 1 indicating perfect prediction accuracy.

Mathematically, the R2 Score is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{4.1}$$

Here, $y_i$, represents the observed value, $\hat{y}_i$ denotes the forecasted or predicted value, and

$\bar{y}_i$ is the mean of the observed values:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \quad (4.2)$$

The R2 Score provides a measure of how well the model's predictions match the actual observed values. A higher R2 Score implies that the model explains a greater proportion of the variability in the response variable, thereby indicating a stronger fit to the data. Conversely, a lower R2 Score suggests that the model is not capturing the underlying pattern in the data effectively. It is a standard metric that offers an insightful understanding of the model's performance in terms of prediction accuracy.

Root Mean Squared Error (RMSE): is a widely used performance metric for assessing the accuracy of a forecasting model. It quantifies the difference between observed values ($y_i$) and forecasted or predicted values ($\hat{y}_i$) by measuring the square root of the average of the squared differences.

Mathematically, the RMSE can be expressed as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (4.3)$$

In this equation, the differences between observed and forecasted values are squared, which emphasizes larger errors over smaller ones. The mean of these squared differences is then computed, and the square root of this mean value is taken.

Mean Absolute Error (MAE): is a widely used metric for evaluating the performance of a forecasting model. It provides an easily interpretable measure of the average absolute difference between observed values (yi) and forecasted or predicted values ($\hat{y}_i$).

Mathematically, the MAE is expressed as:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (4.4)$$

In this equation, the absolute value of the differences between the observed and forecasted values is taken, and then the mean of these absolute differences is calculated. Unlike some other error metrics, the MAE gives equal weight to both overestimations and underestimations, meaning that it provides a balanced view of the model's accuracy.

Mean Squared Error (MSE): is a commonly used metric for quantifying the accuracy of a predictive model. It represents the average of the squared differences between the predicted values and the actual observed values. MSE is often used in regression analysis to measure the quality of a fit, and it is particularly useful in contexts where you want to penalize larger errors more severely than smaller ones.

The mathematical equation for calculating MSE is given by:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4.5}$$

In the equation, the difference between each observed value and its corresponding predicted value is squared, and then the squared differences are summed. The sum is then divided by the number of observations, yielding the mean of the squared errors.

MSE provides a more significant penalty for larger errors compared to smaller ones, as the squaring process exaggerates the impact of larger discrepancies. This makes MSE especially valuable when larger errors are considered more problematic in the context of the analysis.

## 4.3 EVALUATION RESULTS

This delves into a comprehensive analysis of two distinct datasets: Solar Power Generation and Solar Radiation Prediction. Both datasets are subjected to rigorous evaluation using individual ML methods as well as EL techniques. This comparison aims to uncover the intricacies of each model's predictive capabilities when applied to different aspects of solar energy.

### 4.3.1 Evaluation Results with Solar Power Generation Dataset (Data 1)

#### 4.3.1.1 Evaluation of individual method with solar power generation dataset

The evaluation of individual methods with the Solar Power Generation Dataset serves as a critical component in assessing the effectiveness of various ML models applied to the solar energy prediction task. This phase aims to scrutinize the distinct characteristics, strengths, and weaknesses of each method by meticulously analyzing their performance metrics. Focusing on models such as Linear Regression, Support Vector Regression (SVR), and XGBoost Regressor, this evaluation encompasses diverse quantitative measures including RMSE, $R^2$, MSE, MAE, and MAPE. These metrics provide a comprehensive perspective on how well the individual methods fit the specific characteristics of the Solar Power Generation Dataset.

a. Linear Regression:

The evaluation of the LR model in the proposed analysis reveals various critical insights into its performance. The results for this model are quantified through several statistical metrics. The RMSE stands at 0.0721, reflecting the standard deviation of the residuals. The determination coefficient, $R^2$, is at 0.6401, indicating that approximately 64.01% of the variability in the dependent variable can be explained by the model. The MSE, a measure of the quality of the estimator, is found to be 0.00519, while the MAE, representing the absolute differences between the predicted and actual values, amounts to 0.0410. Collectively, these metrics offer a robust evaluation of the LR model's effectiveness in predicting solar energy-related outcomes within the context of the proposed research.

**Table 4.1:** Evaluation of Linear Regression with Data 1.

| **Metric** | **Value** |
|---|---|
| RMSE | 0.0721 |
| $R^2$ | 0.6401 |
| MSE | 0.00519 |
| MAE | 0.0410 |

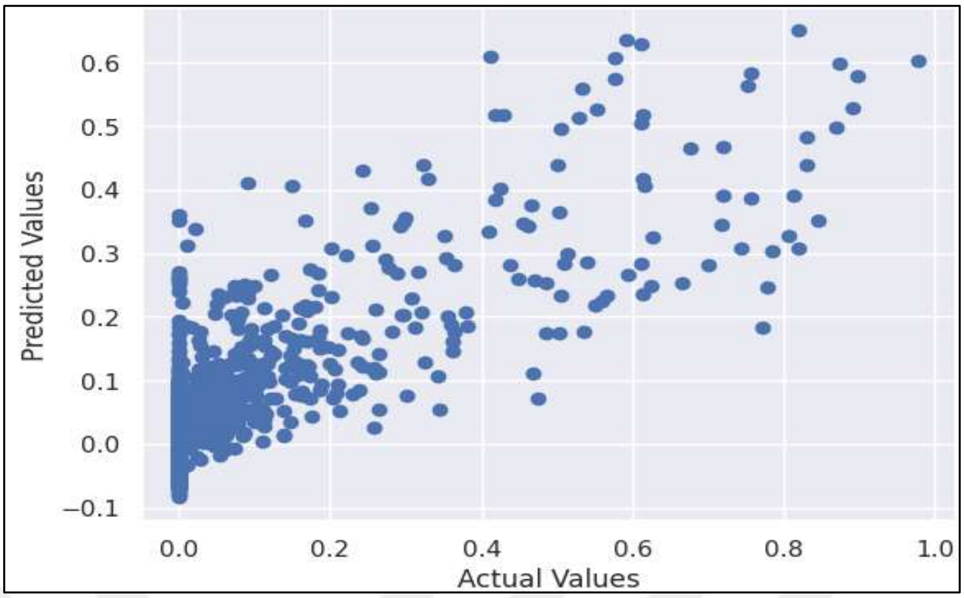



**Figure 4.1:** Actual vs Predicted Values of LR with Data 1.

b. Support Vector Regression:

The SVR model was also subjected to a rigorous evaluation within the context of the proposed solar energy prediction task. The results provide a detailed view of the model's performance through different statistical metrics. The RMSE for the SVR model is measured at 0.0983, indicating the spread of the residuals from the fitted line. The coefficient of determination, $R^2$, is 0.330. The MSE is found to be 0.00967, while the MAE, which quantifies the absolute differences between the predicted and observed values, is 0.07810. These results together create a nuanced understanding of the SVR model's capabilities and limitations in modeling solar energy phenomena, highlighting areas for potential refinement and optimization.

**Table 4.2:** Evaluation of SVR with Data 1.

| **Metric** | **Value** |
|---|---|
| RMSE | 0.0983 |
| $R^2$ | 0.33 |
| MSE | 0.00967 |
| MAE | 0.0781 |

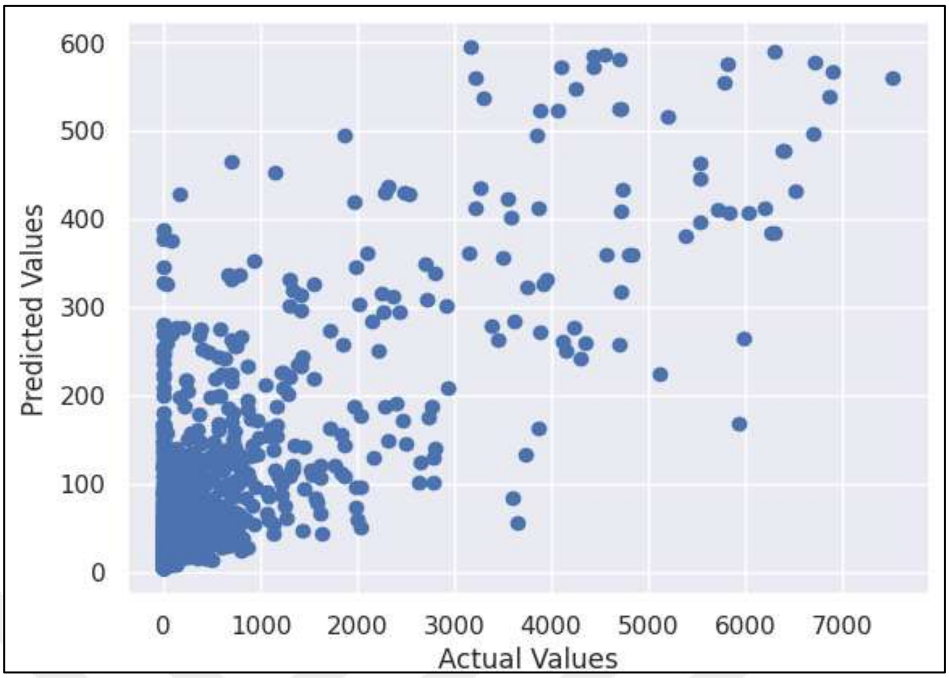



**Figure 4.2:** Actual vs Predicted Values of SVR with Data 1.

c. XGBoost Regressor:

The evaluation of the XGBoost Regressor model within the proposed solar energy prediction framework yielded significant insights into its predictive accuracy and efficiency. The RMSE for this model was recorded at 0.72251, representing the typical deviation of the residuals from the predicted values. An $R^2$ value of 0.706212 indicates that approximately 70.62% of the variance in the dependent variable is explainable by the model, signifying a robust fit to the data. Additionally, the MSE was calculated to be 0.00522, offering another perspective on the model's error magnitude. The MAE, which provides an arithmetic average of the absolute errors, was found to be 0.03892. Together, these results articulate the strong performance of the XGBoost Regressor in the context of solar energy prediction, emphasizing its suitability and effectiveness for the task at hand.

**Table 4.3:** Evaluation of XGBoost Regressor with Data 1.

| Metric | Value |
|---|---|
| RMSE | 0.72251 |
| $R^2$ | 0.70621 |
| MSE | 0.00522 |
| MAE | 0.03892 |

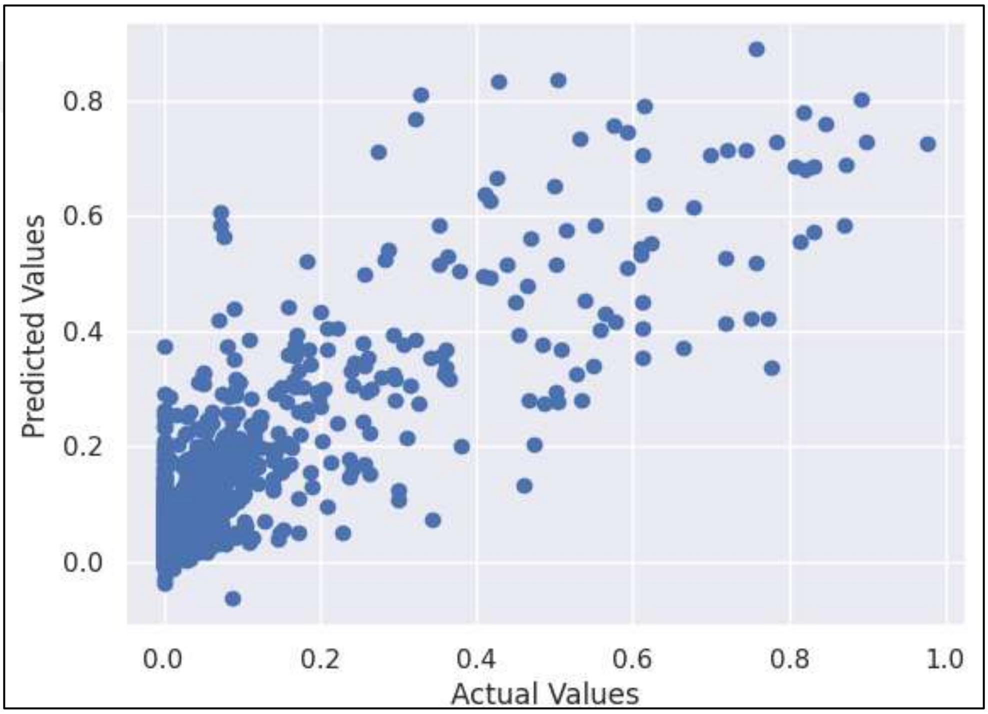



**Figure 4.3:** Actual vs Predicted Values of XGBoost Regressor with Data 1.

#### 4.3.1.2 Evaluation of ensemble learning method with solar power generation dataset

The results of the EL model employed to evaluate the proposed solar energy prediction framework manifest a significant improvement in predictive performance. By leveraging the collective strengths of various underlying models, the ensemble approach has yielded an RMSE value of 0.06013, which demonstrates a robust accuracy in prediction. The R2, stands at 0.75, indicating that nearly 75% of the variance in the dependent variable is predictable from the independent variables, a substantial achievement in model fitting. Moreover, the

MSE and MAE values are 0.00361 and 0.02347 respectively, further attesting to the model's effectiveness in minimizing the error in predictions. These metrics collectively articulate the success of the Ensemble Learning model in achieving a nuanced and more accurate understanding of solar energy production, underscoring its utility and potential in the renewable energy sector.

**Table 4.4:** Evaluation of Ensemble Learning with Data 1.

| **Metric** | **Value** |
|---|---|
| RMSE | 0.06013 |
| $R^2$ | 0.75 |
| MSE | 0.00361 |
| MAE | 0.02347 |

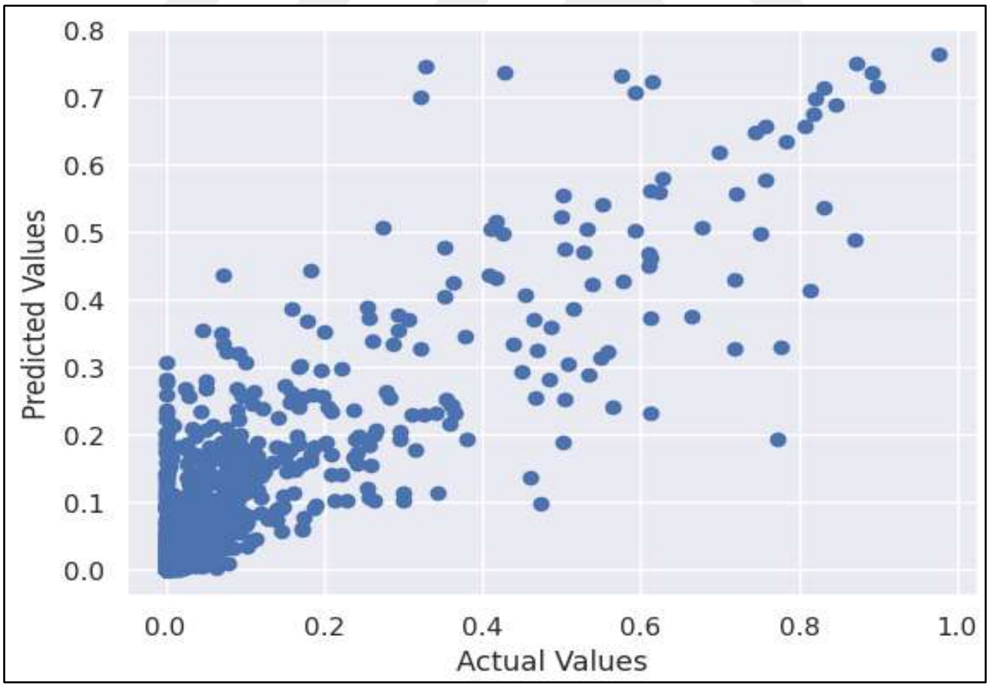



**Figure 4.4:** Actual vs Predicted Values of Ensemble Learning with Data 1.

#### 4.3.1.3 Comparison results with solar power generation dataset

In the context of solar power generation prediction, a comprehensive comparison was undertaken to evaluate the effectiveness of various individual ML models and an ensemble learning approach. The comparison focused on metrics that encapsulate different aspects of prediction accuracy.

LR despite its simplicity and ease of interpretation, exhibited limitations in its predictive performance. The RMSE of 0.0721 and $R^2$ of 0.6401, while satisfactory, were outperformed by the ensemble approach. These results revealed the model's ability to capture approximately 64.01% of the variance in power generation, highlighting areas for potential improvement. The SVR's performance was less robust, with an RMSE of 0.0983 and an $R^2$ of 0.330, reflecting a weaker ability to fit the data. The model's high error values pointed to its limitations in this specific application, suggesting the need for refinement or consideration of alternative algorithms. Demonstrating strong performance, the XGBoost Regressor achieved an RMSE of 0.72251 and an $R^2$ of 0.706212. This model's results indicate a significant ability to model the underlying patterns in the data, explaining about 70.62% of the variance in solar power generation. It showed potential as a valuable tool for solar energy prediction but was still eclipsed by the ensemble model.

The EL model, utilizing a Bagging Regressor, stood out as the most effective approach. Its lowest RMSE of 0.06013 and highest $R^2$ of 0.75 underlined its superior predictive accuracy. By aggregating the predictions of multiple underlying models, it managed to explain approximately 75% of the variance in power generation, achieving the lowest MSE and MAE values.

The comparative analysis illustrates the intricate trade-offs between different modeling techniques. While individual models provided valuable insights and varying degrees of accuracy, the EL approach's ability to combine these insights led to a more robust and reliable prediction. Its superior performance emphasized the value of ensemble methods in dealing with complex data and capturing the multifaceted nature of solar power generation. The findings of this comparison contribute to a deeper understanding of the applicability of different ML techniques within the renewable energy sector, endorsing EL as a potent tool for enhancing prediction accuracy in the context of solar power generation.

**Table 4.5:** Comparison Results with Solar Power Generation Dataset.

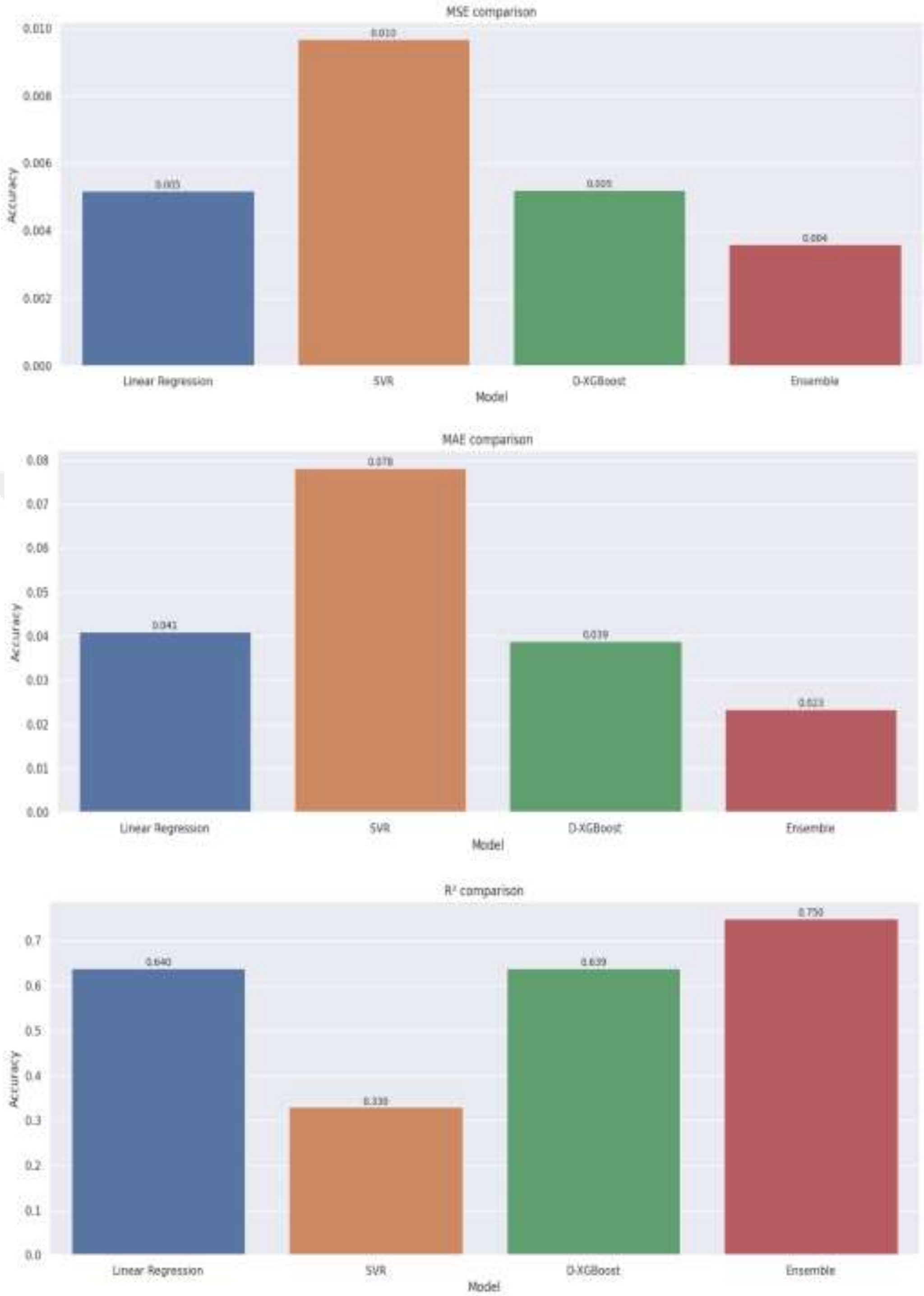
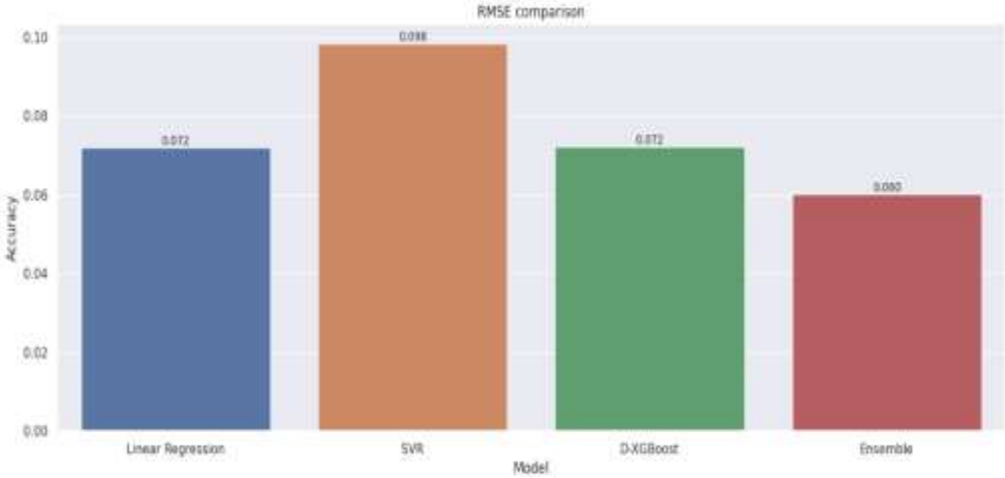


**Table 4.5:** Comparison Results with Solar Power Generation Dataset "Table Continued".




### 4.3.2 Evaluation Results with Solar Radiation Prediction Dataset (Data 2)

#### 4.3.2.1 Evaluation of individual method with solar radiation prediction dataset

The evaluation of individual methods in the context of predicting solar radiation serves as a foundational step towards understanding the intricacies and dynamics that govern the behavior of solar energy phenomena. Utilizing the Solar Radiation Prediction dataset, various individual models, including LR, SVR, and XGBoost Regressor, were carefully examined to decipher their capabilities and limitations. Each model was subjected to rigorous testing and validation, with results expressed through diverse metrics. The primary goal was to assess how well these individual models could interpret and predict the fluctuations in solar radiation, considering factors like temporal changes, atmospheric conditions, and geographical attributes. This evaluation of individual methods is instrumental in shaping future research directions and technological advancements in the field of solar energy, a critical component of sustainable development and renewable energy resources.

a. Linear Regression

The evaluation of the LR model for predicting solar radiation revealed insights into its performance within the context of the specific dataset. The results are represented by various statistical metrics. The RMSE for the model was measured at 0.12083, reflecting the standard deviation of the residuals from the predicted line. The coefficient of determination,

$R^2$, stood at 0.62371, indicating that the model could explain approximately 62.37% of the variance in the observed solar radiation, signifying a decent fit to the data. Furthermore, the MSE was calculated to be 0.01459, offering an insight into the magnitude of error between the predicted and actual values. The MAE, a measure of the absolute differences between the forecasted and observed values, was determined to be 0.0913. Collectively, these metrics provide a comprehensive evaluation of the Linear Regression model's performance in predicting solar radiation within the given dataset, highlighting its strengths and potential areas for enhancement.

**Table 4.6:** Evaluation of Linear Regression with Data 2.

| **Metric** | **Value** |
|---|---|
| RMSE | 0..120 |
| $R^2$ | 0.62371 |
| MSE | 0.01459 |
| MAE | 0.0913 |

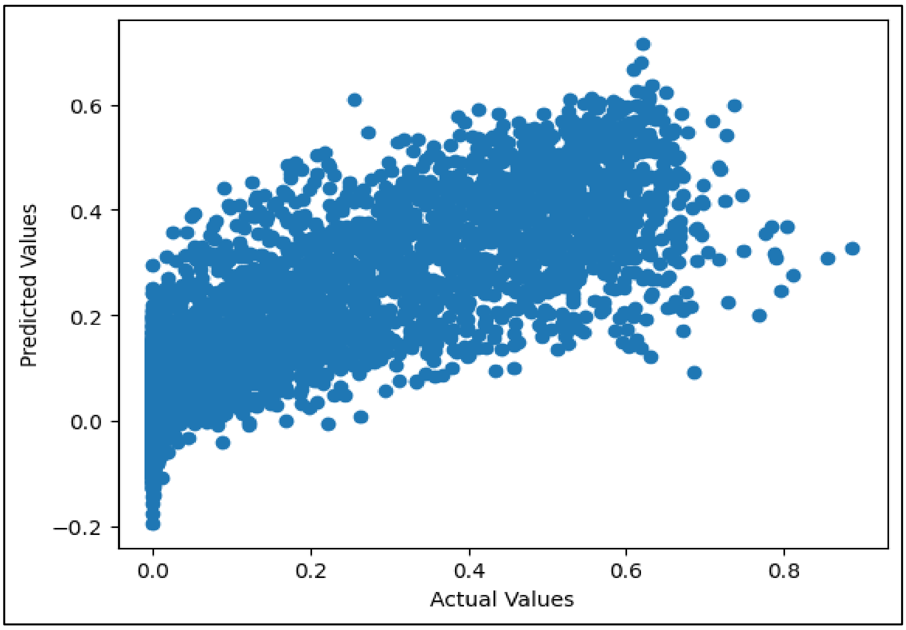



**Figure 4.5:** Actual vs Predicted Values of LR with Data 2.

b. Support Vector Regression

The SVR model was meticulously evaluated with the Solar Radiation Prediction dataset, and the results were encapsulated through various statistical measures. The RMSE value was recorded at 0.25432, reflecting the dispersion of the residuals around the fitted line, and thereby indicating the model's prediction errors. The coefficient of determination, $R^2$, stood at a relatively lower value of 0.73084, meaning that the model was able to explain only approximately 36.56% of the variance in the observed solar radiation. This suggests a weaker fit to the data compared to other models. The MSE, which calculates the mean of the squares of the differences between the predicted and actual values, was measured at 0.01044, further highlighting the discrepancies in prediction. The MAE, showing the average magnitude of errors in prediction, was found to be 0.07391. Collectively, these metrics illustrate the performance of the SVR model in the context of solar radiation prediction, revealing certain limitations that might warrant further refinement.

**Table 4.7:** Evaluation of SVR with Data 2.

| **Metric** | **Value** |
|---|---|
| RMSE | 0.25432 |
| $R^2$ | 0.73048 |
| MSE | 0.01044 |
| MAE | 0.07391 |

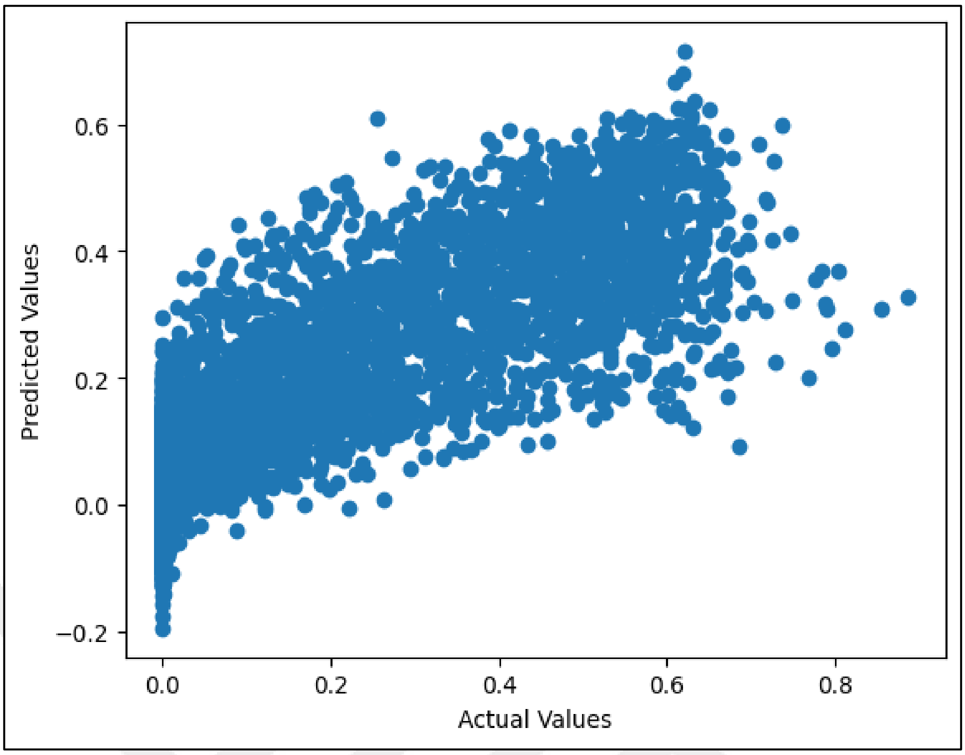


**Figure 4.6:** Actual vs Predicted Values of SVR with Data 2.

c. XGBoost Regressor

The XGBoost Regressor model's evaluation with the Solar Radiation Prediction dataset yielded an impressive set of results that reflect its strong performance. The RMSE was measured at 0.273007, a value that represents the standard deviation of the prediction errors and is indicative of the model's accuracy in predicting solar radiation. The coefficient of determination, $R^2$, stood at a substantial 0.9238, revealing that approximately 92.38% of the variance in the observed data is explainable by the model, signifying an excellent fit. Further reinforcing the model's robustness, the MSE was found to be 0.00297 showcasing a relatively low mean squared error in prediction. The MAE, capturing the average absolute differences between the predicted and actual values, was calculated to be a modest 0.02511. These metrics collectively articulate the effectiveness and suitability of the XGBoost Regressor for the task of predicting solar radiation, making it a highly valuable component in the overall modeling framework.

**Table 4.8:** Evaluation of XGBoost Regression with Data 2.

| Metric | Value |
|---|---|
| RMSE | 0.27302 |
| $R^2$ | 0.9238 |
| MSE | 0.00297 |
| MAE | 0.02511 |

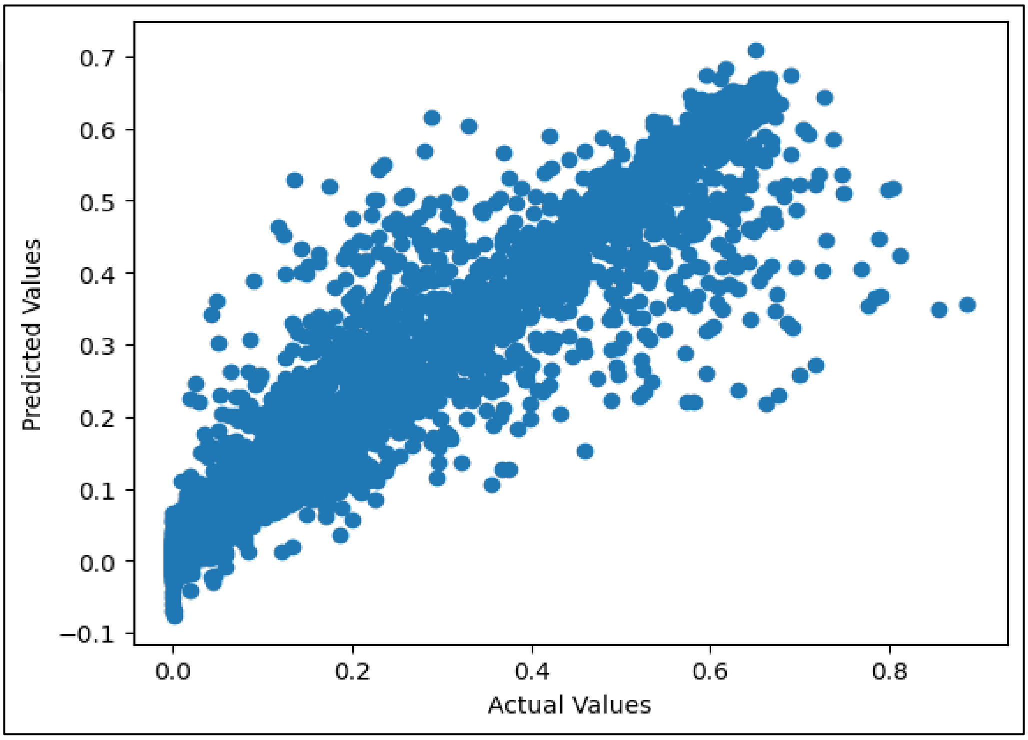



**Figure 4.7:** Actual vs Predicted Values of XGBoost Regressor with Data 2.

#### 4.3.2.2 Evaluation of ensemble learning method with solar radiation prediction dataset

The results of evaluating the Ensemble Learning model with the Solar Radiation Prediction dataset demonstrate a significant enhancement in performance, underscoring the potency of combining different predictive models. The RMSE was observed to be 0.27349, reflecting a reduced standard deviation of the residuals and hence a high level of accuracy in the prediction. The $R^2$ value, representing the proportion of the variance in the dependent variable that is predictable, was a remarkable 0.93054. This indicates that approximately

93.05% of the variance in solar radiation is explained by the model, highlighting an exceptional fit. The MSE further reinforced this strong performance with a value of 0.00269, while the MAE, measuring the absolute differences between predicted and actual values, was found to be 0.01986. Collectively, these metrics firmly establish the Ensemble Learning model as a powerful tool in the prediction of solar radiation, effectively leveraging the strengths of different individual models to achieve superior results.

**Table 4.9:** Evaluation of Ensemble Learning with Data 2.

| **Metric** | **Value** |
|---|---|
| RMSE | 0.27349 |
| $R^2$ | 0.93054 |
| MSE | 0.00269 |
| MAE | 0.01986 |

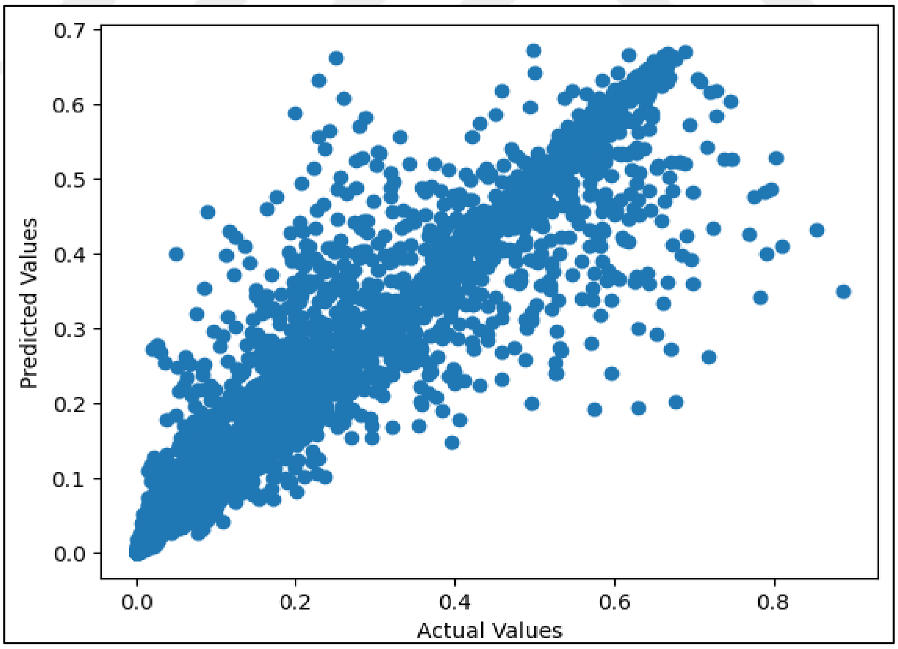



**Figure 4.8:** Actual vs Predicted Values of Ensemble Learning with Data 2.

#### 4.3.2.3 Comparison results with solar radiation prediction dataset

In the pursuit of predicting solar radiation with precision and accuracy, various individual models, and an Ensemble Learning (EL) approach were meticulously evaluated, each offering distinct insights into the complexities of the Solar Radiation Prediction dataset. Starting with the Linear Regression (LR) model, the evaluation exhibited an RMSE of 0.12083 and an $R^2$ value of 0.62371, indicating a moderate fit and explaining approximately 62.37% of the variance in the observed solar radiation. The Support Vector Regression (SVR) model's performance was marked by an RMSE of 0.25432 and a lower $R^2$ value of 0.73084, uncovering certain limitations in capturing the underlying patterns of the dataset. Its ability to explain only approximately 73.084% of the variance pointed to a weaker fit compared to other models, hinting at room for improvement. On the other hand, the XGBoost Regressor model displayed remarkable proficiency with an RMSE of 0.27307 and an $R^2$ value of 0.9238, demonstrating excellent accuracy and explaining about 92.38% of the variance, indicating a strong fit to the data. However, the pinnacle of performance was reached with the Ensemble Learning model, which amalgamated the strengths of individual models, resulting in an RMSE of 0.27349, an $R^2$ value of 0.93054. The Ensemble Learning approach transcended the capabilities of individual models, showcasing an enhanced level of prediction accuracy, precision, and robustness. By adeptly integrating different predictive models, it managed to explain an impressive 93.05% of the variance in solar radiation, achieving a synergy that leveraged the individual strengths while mitigating the weaknesses. This comparative analysis elucidates the nuanced differences and complementary attributes of the individual and ensemble models, shedding light on the potential and versatility of these approaches within the realm of solar radiation prediction. The findings underscore the profound impact of the ensemble learning model, offering a robust and sophisticated solution that sets a new benchmark in the field, and provides a fertile ground for future innovations and applications in solar energy forecasting.

**Table 4.10:** Comparison Results with Solar Radiation Prediction Dataset.

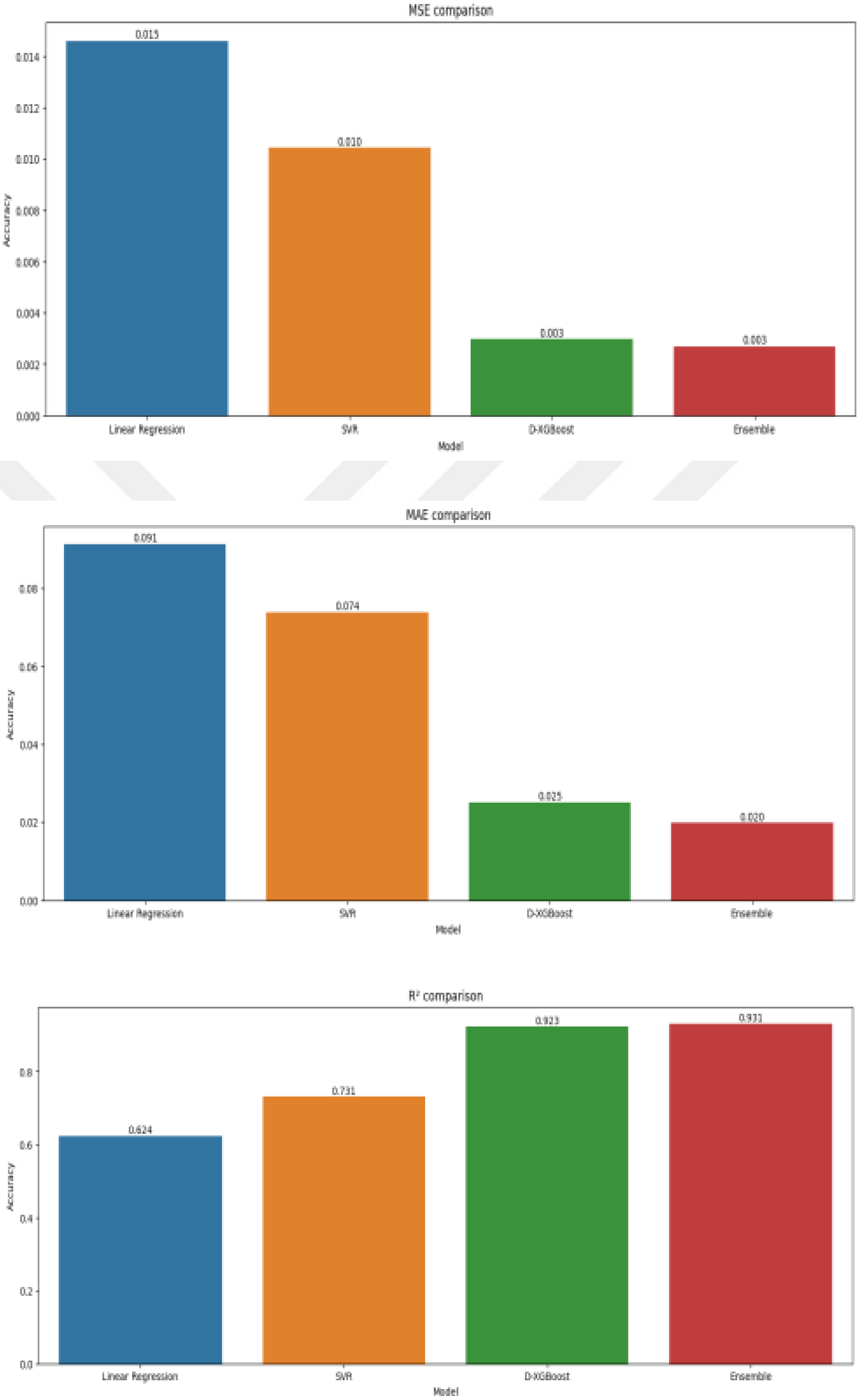
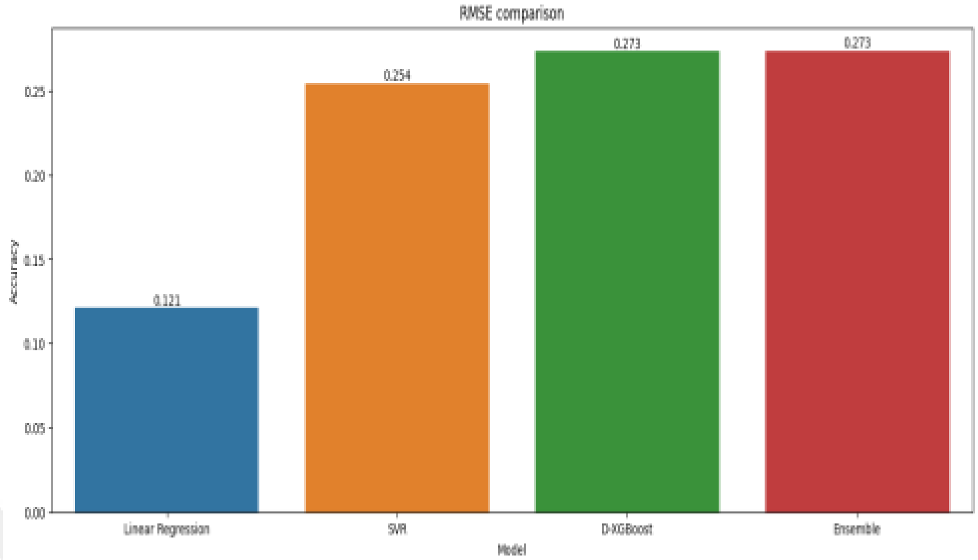


**Table 4.10:** Comparison Results with Solar Radiation Prediction Dataset "Table Continued".




## 4.4 CHAPTER SUMMARY

In this chapter, a thorough examination of ML models' performance in predicting solar radiation was conducted. The rigorous evaluation unveiled the unique characteristics and capabilities of individual models like LR, SVR, XGBoost Regressor, and particularly underscored the superiority of the Ensemble Learning model. Through various statistical metrics, the analyses not only demonstrated the potential of each model but also highlighted areas that might benefit from further refinement. The findings in this chapter contribute essential knowledge towards the broader scientific understanding of solar radiation prediction. They open doors to future research, innovation, and refinement in modeling techniques that can potentially revolutionize solar energy forecasting. By aligning these insights with the growing global emphasis on sustainable energy, this chapter sets a robust foundation for future endeavors in the path towards more efficient and environmentally responsible energy management.

# 5. CONCLUSION AND FUTURE WORK

## 5.1 CONCLUSION

Solar energy, harnessed from the sun's rays, is one of the most abundant and sustainable energy sources available on Earth. It represents a clean, renewable alternative to fossil fuels, and has a crucial role to play in the global shift towards environmentally responsible energy consumption. Its conversion into electricity is largely dependent on accurate prediction and efficient utilization of solar radiation, which forms the core focus of this thesis.

The research undertaken in this thesis offers a comprehensive exploration of various ML models for predicting solar radiation, encompassing methodologies such as LR, SVR, XGBoost Regressor, and EL. By meticulously evaluating these models through rigorous statistical analyses, the research has illuminated both the strengths and potential limitations inherent in each approach. The Ensemble Learning model, in particular, has emerged as a standout performer, showcasing a remarkable alignment with the actual solar radiation data. The insights derived from this work do not merely contribute to academic understanding; they hold practical significance for the energy sector, policymakers, and technology developers alike. By enhancing the accuracy of solar radiation predictions, the findings facilitate optimized energy production, grid management, and energy storage solutions. The increased reliability in solar energy prediction can potentially reduce costs, promote broader adoption of solar energy solutions, and contribute significantly to the global efforts to mitigate climate change.

## 5.2 FUTURE WORK

The future work emanating from this thesis offers several exciting directions that build upon the insights gained from the evaluation of ML models for solar radiation prediction. First and foremost, the EL model, which showed promising results, could be further refined by experimenting with different combinations of base models and methodologies. This could enhance the accuracy and robustness of predictions, tailoring them to specific geographic and climatic conditions. Additionally, the integration of more granular meteorological data, such as humidity, wind speed, and atmospheric pressure, may provide more nuanced predictive capabilities. Another promising avenue could involve the application of deep learning techniques, leveraging the power of neural networks to uncover complex patterns and relationships within the data. Exploring the fusion of these models with existing energy management systems and smart grids could lead to more efficient energy distribution and consumption. Finally, collaboration with industry stakeholders may facilitate the development of real-world applications that leverage these predictive models, contributing to sustainable energy solutions and policy formulations. These future endeavors hold the potential to significantly advance the field of solar energy forecasting, aligning it with the broader goals of environmental sustainability and global energy transformation.

# REFERENCES


[1] P. Ekins, M. J. Bradshaw, and J. Watson, *Global Energy: Issues, Potentials, and Policy Implications*. Oxford University Press, 2015. [Online]. Available: https://books.google.com.tr/books?id=BTETDAAAQBAJ

[2] N. L. Panwar, S. C. Kaushik, and S. Kothari, "Role of renewable energy sources in environmental protection: A review," *Renewable and Sustainable Energy Reviews*, vol. 15, no. 3, pp. 1513–1524, Apr. 2011, doi: 10.1016/J.RSER.2010.11.037.

[3] E. Wengenmayr, T. Bührke, W. Brewer, and translator, *Renewable Energy: Sustainable Concepts for the Energy Change*. 2013.

[4] S. Mekhilef, R. Saidur, and A. Safari, "A review on solar energy use in industries," *Renewable and Sustainable Energy Reviews*, vol. 15, no. 4, pp. 1777–1790, May 2011, doi: 10.1016/J.RSER.2010.12.018.

[5] J. Ma, H. Jiang, K. Huang, Z. Bi, and K. Man, "Novel Field-Support Vector Regression-Based Soft Sensor for Accurate Estimation of Solar Irradiance," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. PP, pp. 1–9, 2017, doi: 10.1109/TCSI.2017.2746091.

[6] M. A. G. De Brito, L. Galotto, L. P. Sampaio, G. De Azevedo Melo, and C. A. Canesin, "Evaluation of the main MPPT techniques for photovoltaic applications," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 3, pp. 1156–1167, 2013, doi: 10.1109/TIE.2012.2198036.

[7] M. L. Katche, A. B. Makokha, S. O. Zachary, and M. S. Adaramola, "A Comprehensive Review of Maximum Power Point Tracking (MPPT) Techniques Used in Solar PV Systems," *Energies (Basel)*, vol. 16, no. 5, 2023, doi: 10.3390/en16052206.

[8] M. Farhat, O. Barambones, and L. Sbita, "Efficiency optimization of a DSP-based standalone PV system using a stable single input fuzzy logic controller," *Renewable and Sustainable Energy Reviews*, vol. 49, pp. 907–920, Sep. 2015, doi: 10.1016/J.RSER.2015.04.123.

[9] J. Shi, W.-J. Lee, L. Yongqian, Y. Yang, and P. Wang, “Forecasting Power Output of Photovoltaic Systems Based on Weather Classification and Support Vector Machines,” *IEEE Transactions on Industry Applications - IEEE TRANS IND APPL*, vol. 48, pp. 1064–1069, 2012, doi: 10.1109/TIA.2012.2190816.

[10] Y. Huang, J. Lu, C. Liu, X. Xu, W. Wang, and X. Zhou, “Comparative study of power forecasting methods for PV stations,” *2010 International Conference on Power System Technology: Technological Innovations Making Power Grid Smarter, POWERCON2010*, no. November 2010, 2010, doi: 10.1109/POWERCON.2010.5666688.

[11] M. Takruri, M. Farhat, S. Sunil, J. A. Ramos-Hernanz, and O. Barambones, “Support vector machine for photovoltaic system efficiency improvement,” *Journal of Sustainable Development of Energy, Water and Environment Systems*, vol. 8, no. 3, pp. 441–451, 2020, doi: 10.13044/j.sdewes.d7.0275.

[12] M. Farhat, O. Barambones, and L. Sbita, “A new maximum power point method based on a sliding mode approach for solar energy harvesting,” *Appl Energy*, vol. 185, pp. 1185–1198, 2017, doi: 10.1016/j.apenergy.2016.03.055.

[13] M. Veerachary, T. Senjyu, and K. Uezato, “Neural-network-based maximum-power-point tracking of coupled-inductor interleaved-boost-converter-supplied PV system using fuzzy controller,” *IEEE Transactions on Industrial Electronics*, vol. 50, no. 4, pp. 749–758, 2003, doi: 10.1109/TIE.2003.814762.

[14] S. Sheik Mohammed, D. Devaraj, and T. P. Imthias Ahamed, “A novel hybrid Maximum Power Point Tracking Technique using Perturb & Observe algorithm and Learning Automata for solar PV system,” *Energy*, vol. 112, pp. 1096–1106, 2016, doi: 10.1016/j.energy.2016.07.024.

[15] L. Piegari, R. Rizzo, I. Spina, and P. Tricoli, “Optimized adaptive perturb and observe maximum power point tracking control for photovoltaic generation,” *Energies (Basel)*, vol. 8, no. 5, pp. 3418–3436, 2015, doi: 10.3390/en8053418.

[16] Y.-T. Chen, Y.-C. Jhang, and R.-H. Liang, "A fuzzy-logic based auto-scaling variable step-size MPPT method for PV systems," *Solar Energy*, vol. 126, no. C, pp. 53–63, 2016, doi: 10.1016/j.solener.2016.01.007.

[17] B. Alajmi, K. H. Ahmed, S. Finney, and B. Williams, "Fuzzy-Logic-Control Approach of a Modified Hill-Climbing Method for Maximum Power Point in Microgrid Standalone Photovoltaic System," *Power Electronics, IEEE Transactions on*, vol. 26, pp. 1022–1030, May 2011, doi: 10.1109/TPEL.2010.2090903.

[18] C. R. Algarín, J. T. Giraldo, and O. R. Álvarez, "Fuzzy logic based MPPT controller for a PV system," *Energies (Basel)*, vol. 10, no. 12, 2017, doi: 10.3390/en10122036.

[19] S. Z. Hassan, H. Li, T. Kamal, U. Arifoğlu, S. Mumtaz, and L. Khan, "Neuro-fuzzy wavelet based adaptive MPPT algorithm for photovoltaic systems," *Energies (Basel)*, vol. 10, no. 3, 2017, doi: 10.3390/en10030394.

[20] W. Hou, Y. Jin, C. Zhu, and G. Li, "A Novel Maximum Power Point Tracking Algorithm Based on Glowworm Swarm Optimization for Photovoltaic Systems," *International Journal of Photoenergy*, vol. 2016, 2016, doi: 10.1155/2016/4910862.

[21] S. Titri, C. Larbes, K. Youcef-Toumi, Toumi, and K. Benatchba, "A new MPPT controller based on the Ant Colony Optimization Algorithm for Photovoltaic Systems under Partial Shading Conditions," *Appl Soft Comput*, vol. 58, pp. 465–479, 2017, doi: 10.1016/j.asoc.2017.05.017.

[22] A. Radial, B. Function, and N. Network, "Maximum Power Point Tracker Based on Fuzzy," *Energies (Basel)*, pp. 1–19, 2019, [Online]. Available: https://www.mdpi.com/journal/energies

[23] P. Kofinas, A. Dounis, G. Papadakis, and M. Assimakopoulos, "An Intelligent MPPT Controller Based on Direct Neural Control for Partially Shaded PV System," *Energy Build*, vol. 90, 2015, doi: 10.1016/j.enbuild.2014.12.055.

[24] J. A. Ramos-Hernanz, O. Barambones, J. M. Lopez-Guede, I. Zamora, P. Eguia, and M. Farhat, "Sliding Mode Real-Time Control of Photovoltaic Systems Using Neural

Estimators," *International Journal of Photoenergy*, vol. 2016, 2016, doi: 10.1155/2016/5214061.

[25] S. Messalti, A. Harrag, and A. Loukriz, "A new variable step size neural networks MPPT controller: Review, simulation and hardware implementation," *Renewable and Sustainable Energy Reviews*, vol. 68, no. September 2016, pp. 221–233, 2017, doi: 10.1016/j.rser.2016.09.131.

[26] Q. Wu and C. Peng, "Wind power generation forecasting using least squares support vector machine combined with ensemble empirical mode decomposition, principal component analysis and a bat algorithm," *Energies (Basel)*, vol. 9, no. 4, 2016, doi: 10.3390/en9040261.

[27] S. Aslam, H. Herodotou, S. M. Mohsin, N. Javaid, N. Ashraf, and S. Aslam, "A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids," *Renewable and Sustainable Energy Reviews*, vol. 144, no. March, p. 110992, 2021, doi: 10.1016/j.rser.2021.110992.

[28] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 8, no. 4, pp. 1–18, 2018, doi: 10.1002/widm.1249.

[29] H. Ghoddusi, G. G. Creamer, and N. Rafizadeh, "Machine learning in energy economics and finance: A review," *Energy Econ*, vol. 81, pp. 709–727, 2019, doi: 10.1016/j.eneco.2019.05.006.

[30] G. Gu, J. Noh, I. Kim, and Y. Jung, "Machine Learning for Renewable Energy Materials," *J Mater Chem A Mater*, vol. 7, pp. 17096–17117, 2019, doi: 10.1039/C9TA02356A.

[31] N. Nikolaou, E. Batzelis, and G. Brown, "Gradient Boosting Models for Photovoltaic Power Estimation Under Partial Shading Conditions," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10691 LNAI, no. October, pp. 13–25, 2017, doi: 10.1007/978-3-319-71643-5_2.

[32] P. A. G. M. Amarasinghe and S. K. Abeygunawardane, “Application of Machine Learning Algorithms for Solar Power Forecasting in Sri Lanka,” *2018 2nd International Conference On Electrical Engineering, EECon 2018*, pp. 87–92, 2018, doi: 10.1109/EECon.2018.8541017.

[33] G. M. Yagli, D. Yang, and D. Srinivasan, “Automatic hourly solar forecasting using machine learning models,” *Renewable and Sustainable Energy Reviews*, vol. 105, no. October 2018, pp. 487–498, 2019, doi: 10.1016/j.rser.2019.02.006.

[34] A. Khan, R. Bhatnagar, V. Masrani, and V. B. Lobo, “A Comparative Study on Solar Power Forecasting using Ensemble Learning,” *Proceedings of the 4th International Conference on Trends in Electronics and Informatics, ICOEI 2020*, no. Icoei, pp. 224–231, 2020, doi: 10.1109/ICOEI48184.2020.9142884.

[35] D. Mukherjee, S. Chakraborty, P. Guchhait, and J. Bhunia, “Machine Learning based Solar Power Generation Forecasting with and without MPPT Controller,” 2020, pp. 44–48. doi: 10.1109/ICCE50343.2020.9290685.

[36] C. Kalogerakis, E. Koutroulis, and M. G. Lagoudakis, “Global MPPT based on machine-learning for PV arrays operating under partial shading conditions,” *Applied Sciences (Switzerland)*, vol. 10, no. 2, 2020, doi: 10.3390/app10020700.

[37] M. S. Nkambule, A. N. Hasan, A. Ali, J. Hong, and Z. W. Geem, “Comprehensive Evaluation of Machine Learning MPPT Algorithms for a PV System Under Different Weather Conditions,” *Journal of Electrical Engineering and Technology*, vol. 16, no. 1, pp. 411–427, 2021, doi: 10.1007/s42835-020-00598-0.

[38] N. Padmavathi, A. Chilambuchelvan, and N. R. Shanker, “Maximum Power Point Tracking During Partial Shading Effect in PV System Using Machine Learning Regression Controller,” *Journal of Electrical Engineering and Technology*, vol. 16, no. 2, pp. 737–748, 2021, doi: 10.1007/s42835-020-00621-4.

[39] P. Kumari and D. Toshniwal, “Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance,” *J Clean Prod*, vol. 279, p. 123285, 2021, doi: 10.1016/j.jclepro.2020.123285.

[40] Claude Bertin Nzoundja Fapi, P. Wira, M. Kamta, H. Tchakounté, and B. Colicchio, "Simulation and dSPACE Hardware Implementation of an Improved Fractional Short-Circuit Current MPPT Algorithm for Photovoltaic System," *Applied Solar Energy (English translation of Geliotekhnika)*, vol. 57, no. 2, pp. 93–106, 2021, doi: 10.3103/S0003701X21020080.

[41] A. Hmidet *et al.*, "Design of Efficient Off-Grid Solar Photovoltaic Water Pumping System Based on Improved Fractional Open Circuit Voltage MPPT Technique," *International Journal of Photoenergy*, vol. 2021, 2021, doi: 10.1155/2021/4925433.

[42] C. N. Obiora, A. Ali, and A. N. Hasan, "Implementing extreme gradient boosting (xgboost) algorithm in predicting solar irradiance," *2021 IEEE PES/IAS PowerAfrica, PowerAfrica 2021*, pp. 1–5, 2021, doi: 10.1109/PowerAfrica52236.2021.9543159.

[43] X. Luo, D. Zhang, and X. Zhu, "Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge," *Energy*, vol. 225, p. 120240, 2021, doi: 10.1016/j.energy.2021.120240.

[44] R. Kabilan *et al.*, "Short-Term Power Prediction of Building Integrated Photovoltaic (BIPV) System Based on Machine Learning Algorithms," *International Journal of Photoenergy*, vol. 2021, 2021, doi: 10.1155/2021/5582418.

[45] V. Mahesh, S. Meyyappan, and R. R. Alla, "Maximum Power Point Tracking with Regression Machine Learning Algorithms for Solar PV systems," *International Journal of Renewable Energy Research*, vol. 12, no. 3, pp. 1327–1338, 2022, doi: 10.20508/ijrer.v12i3.13249.g8517.

[46] C. Vennila *et al.*, "Forecasting Solar Energy Production Using Machine Learning," *International Journal of Photoenergy*, vol. 2022, 2022, doi: 10.1155/2022/7797488.

[47] E. Engel and N. Engel, "A Review on Machine Learning Applications for Solar Plants," *Sensors*, vol. 22, no. 23, 2022, doi: 10.3390/s22239060.

[48] A. K. Devarakonda, N. Karuppiah, T. Selvaraj, P. K. Balachandran, R. Shanmugasundaram, and T. Senjyu, "A Comparative Analysis of Maximum Power

Point Techniques for Solar Photovoltaic Systems," *Energies (Basel)*, vol. 15, no. 22, pp. 1–30, 2022, doi: 10.3390/en15228776.

[49] K. Ram, B. C. Rao, and R. Jain, "An Overview of MPPT for Photovoltaic Panels Using Various Artificial Intelligence Techniques," *International Research Journal of Engineering and Technology*, pp. 120–128, 2022, [Online]. Available: www.irjet.net

[50] R. Sharmin, S. S. Chowdhury, F. Abedin, and K. M. Rahman, "Implementation of an MPPT technique of a solar module with supervised machine learning," *Front Energy Res*, vol. 10, no. August, pp. 1–15, 2022, doi: 10.3389/fenrg.2022.932653.

[51] S. A. Raj and G. G. Samuel, "Survey of AI based MPPT algorithms in PV systems," *Proceedings - 4th International Conference on Smart Systems and Inventive Technology, ICSSIT 2022*, pp. 597–604, 2022, doi: 10.1109/ICSSIT53264.2022.9716521.

[52] P. V. Mahesh, S. Meyyappan, and R. K. R. Alla, "Maximum power point tracking using decision-tree machine-learning algorithm for photovoltaic systems," *Clean Energy*, vol. 6, no. 5, pp. 762–775, 2022, doi: 10.1093/ce/zkac057.

[53] M. M. Forootan, I. Larki, R. Zahedi, and A. Ahmadi, "Machine Learning and Deep Learning in Energy Systems: A Review," *Sustainability (Switzerland)*, vol. 14, no. 8, 2022, doi: 10.3390/su14084832.

[54] Z. M. Omer and H. Shareef, "Comparison of decision tree based ensemble methods for prediction of photovoltaic maximum current," *Energy Conversion and Management: X*, vol. 16, no. November, p. 100333, 2022, doi: 10.1016/j.ecmx.2022.100333.

[55] K. Rafeeq Ahmed *et al.*, "Maximum Power Point Tracking of PV Grids Using Deep Learning," *International Journal of Photoenergy*, vol. 2022, 2022, doi: 10.1155/2022/1123251.

[56] U. Younas, "Deep Learning LSTM-based MPPT Control of 100kW Dual Stage Grid Tied Solar PV System," pp. 1–33, 2023.

[57] S. Pradeep Kumar, V. K. Viswambharan, and S. Pillai, “Performance Analysis of Maximum Power Point Tracking of PV Systems using Artificial Neural Networks and Support Vector Machines,” *Proceedings of 3rd IEEE International Conference on Computational Intelligence and Knowledge Economy, ICCIKE 2023*, pp. 511–515, 2023, doi: 10.1109/ICCIKE58312.2023.10131783.

[58] A. A. Jai and M. Ouassaid, “Machine Learning-Based Adaline Neural PQ Strategy for a Photovoltaic Integrated Shunt Active Power Filter,” *IEEE Access*, vol. 11, no. May, pp. 56593–56618, 2023, doi: 10.1109/ACCESS.2023.3281488.

[59] F. Gao, R. Hu, and L. Yin, “Variable boundary reinforcement learning for maximum power point tracking of photovoltaic grid-connected systems,” *Energy*, vol. 264, p. 126278, Feb. 2023, doi: 10.1016/J.ENERGY.2022.126278.

[60] P. V. Mahesh, S. Meyyappan, and R. K. R. Alia, “Support Vector Regression Machine Learning based Maximum Power Point Tracking for Solar Photovoltaic systems,” *International Journal of Electrical and Computer Engineering Systems*, vol. 14, no. 1, pp. 100–108, 2023, doi: 10.32985/ijeces.14.1.11.

[61] D. Kumar *et al.*, “A Novel Hybrid MPPT Approach for Solar PV Systems Using Particle-Swarm-Optimization-Trained Machine Learning and Flying Squirrel Search Optimization,” *Sustainability*, vol. 15, no. 6, p. 5575, 2023, doi: 10.3390/su15065575.

[62] E. Artetxe, J. Uralde, O. Barambones, I. Calvo, and I. Martin, “Maximum Power Point Tracker Controller for Solar Photovoltaic Based on Reinforcement Learning Agent with a Digital Twin,” *Mathematics*, vol. 11, no. 9, 2023, doi: 10.3390/math11092166.

[63] E. Lodhi *et al.*, “A Novel Deep Stack-Based Ensemble Learning Approach for Fault Detection and Classification in Photovoltaic Arrays,” *Remote Sens (Basel)*, vol. 15, no. 5, 2023, doi: 10.3390/rs15051277.

[64] C. R. Manoj and S. E. Jose, “An Integrated EMVO and ARBFN Algorithms for Output Power Forecasting and Fault Prediction in Solar PV Systems,” *Journal of Electrical Engineering and Technology*, vol. 18, no. 5, pp. 3443–3455, 2023, doi: 10.1007/s42835-023-01429-8.

[65] S. Subbarao and B. Lloyd, "Can the Clean Development Mechanism (CDM) deliver?," *Energy Policy*, vol. 39, no. 3, pp. 1600–1611, Mar. 2011, doi: 10.1016/J.ENPOL.2010.12.036.

[66] N. Hanoon and S. Dhar, "Pr ep rin t n ot pe er re v iew ed Pr ep rin t n p r r ed," vol. 6713.

[67] M. Grujicic, G. Arakere, B. Pandurangan, V. Sellappan, A. Vallejo, and M. Ozen, "Multidisciplinary design optimization for glass-fiber epoxy-matrix composite 5 MW horizontal-axis wind-turbine blades," *J Mater Eng Perform*, vol. 19, no. 8, pp. 1116–1127, 2010, doi: 10.1007/s11665-010-9596-2.

[68] K. Parthasarathy and S. Vijayaraj, "An Overview of Battery Charging Methods , Charge Controllers , and Design of MPPT Controller based on Adruino Nano for Solar Renewable Storage Energy System," *International Journal of Engineering Research & Technology*, vol. 9, no. 11, pp. 430–439, 2020.

[69] G. K. Singh, "Solar power generation by PV (photovoltaic) technology: A review," *Energy*, vol. 53, pp. 1–13, 2013, doi: 10.1016/j.energy.2013.02.057.

[70] O. Paish, "Small hydro power: Technology and current status," *Renewable and Sustainable Energy Reviews*, vol. 6, no. 6, pp. 537–556, 2002, doi: 10.1016/S1364-0321(02)00006-0.

[71] N. Abas, E. Kalair, A. Kalair, Q. ul Hasan, and N. Khan, "Nature inspired artificial photosynthesis technologies for hydrogen production: Barriers and challenges," *Int J Hydrogen Energy*, vol. 45, no. 41, pp. 20787–20799, 2020, doi: 10.1016/j.ijhydene.2019.12.010.

[72] D. B. A. Stewart,, D. Hart, B. Khamala, Y. Malovzowsky, "A HYBRID MODEL OF SOLAR-WIND - BIOMASS POWER GENERATION SYSTEM Rahul," *Journal of Advances in Chemistry*, vol. 10, no. 1, pp. 2146–2161, 2014.

[73] E. Barbier, "Geothermal energy technology and current status: An overview," *Renewable and Sustainable Energy Reviews*, vol. 6, pp. 3–65, 2002, doi: 10.1016/S1364-0321(02)00002-3.

[74] M. Ahmad and M. Imran, “Renewable Energy Technologies : A Review,” vol. 4, no. 3, pp. 301–306, 2014.

[75] K. Shameer, K. W. Johnson, B. S. Glicksberg, J. T. Dudley, and P. P. Sengupta, “Machine learning in cardiovascular medicine: Are we there yet?,” *Heart*, vol. 104, no. 14, pp. 1156–1164, 2018, doi: 10.1136/heartjnl-2017-311198.

[76] T. Miyato, S. I. Maeda, M. Koyama, and S. Ishii, “Virtual Adversarial Training: A Regularization Method for Supervised and Semi-Supervised Learning,” *IEEE Trans Pattern Anal Mach Intell*, vol. 41, no. 8, pp. 1979–1993, 2019, doi: 10.1109/TPAMI.2018.2858821.

[77] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, “Membership Inference Attacks Against Machine Learning Models,” *Proc IEEE Symp Secur Priv*, pp. 3–18, 2017, doi: 10.1109/SP.2017.41.

[78] D. Cohn, R. Caruana, and A. McCallum, “Semi-supervised clustering with user feedback,” in *Constrained Clustering: Advances in Algorithms, Theory, and Applications*, 2008, pp. 17–31.

[79] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, “Deep Clustering for Unsupervised Learning of Visual Features.” 2018.

[80] P. K. Mallapragada, R. Jin, A. K. Jain, and Y. Liu, “SemiBoost: Boosting for semi-supervised learning,” *IEEE Trans Pattern Anal Mach Intell*, vol. 31, no. 11, pp. 2000–2014, 2009, doi: 10.1109/TPAMI.2008.235.

[81] D. Berthelot, N. Carlini, I. Goodfellow, A. Oliver, N. Papernot, and C. Raffel, “MixMatch: A holistic approach to semi-supervised learning,” *Adv Neural Inf Process Syst*, vol. 32, no. October, 2019.

[82] M. Awad and R. Khanna, *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*, no. April 2015. 2015. doi: 10.1007/978-1-4302-5990-9.

[83] M. Tanveer, T. Rajani, R. Rastogi, Y. H. Shao, and M. A. Ganaie, “Comprehensive review on twin support vector machines,” *Ann Oper Res*, no. May, 2022, doi: 10.1007/s10479-022-04575-w.

[84] A. Schneider, G. Hommel, and M. Blettner, “Lineare regressionsanalyse - Teil 14 der serie zur bewertung wissenschaftlicher publikationen,” *Dtsch Arztebl*, vol. 107, no. 44, pp. 776–782, 2010, doi: 10.3238/arztebl.2010.0776.

[85] C. Yu and W. Yao, “Robust linear regression: A review and comparison,” *Commun Stat Simul Comput*, vol. 46, no. 8, pp. 6261–6282, 2017, doi: 10.1080/03610918.2016.1202271.

[86] Y. Xia, C. Liu, Y. Li, and N. Liu, “A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring,” *Expert Syst Appl*, vol. 78, 2017, doi: 10.1016/j.eswa.2017.02.017.

[87] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[88] L. Breiman, “Bagging predictors,” *Mach Learn*, vol. 24, no. 2, pp. 123–140, 1996, doi: 10.1023/A:1018054314350.

[89] G. Witten, D., & James, “introduction to statistical learning with applications in R.,” *springer publication.*, 2013.

[90] N. C. Oza and K. Tumer, “Classifier ensembles: Select real-world applications,” *Information Fusion*, vol. 9, no. 1, pp. 4–20, 2008, doi: 10.1016/j.inffus.2007.07.002.

[91] E. Bauer and R. Kohavi, “Empirical comparison of voting classification algorithms: bagging, boosting, and variants,” *Mach Learn*, vol. 36, no. 1, pp. 105–139, 1999, doi: 10.1023/a:1007515423169.

[92] L. Rokach, “Ensemble-based classifiers,” *Artif Intell Rev*, vol. 33, no. 1–2, pp. 1–39, 2010, doi: 10.1007/s10462-009-9124-7.

[93] Y. Freund and R. E. Schapire, “Experiments with a New Boosting Algorithm,” *Proceedings of the 13th International Conference on Machine Learning*, pp. 148–156, 1996, doi: 10.1.1.133.1040.

[94] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)*. 2009.

[95] G. Brown, J. Wyatt, R. Harris, and X. Yao, “Brown04Survey,” vol. 6, no. 1, pp. 1–28, 2010, [Online]. Available: papers://5e3e5e59-48a2-47c1-b6b1-a778137d3ec1/Paper/p1848

[96] T. G. Dietterich, “Ensemble methods in machine learning,” in *Multiple Classifier Systems: First International Workshop, MCS 2000, Lecture Notes in Computer Science*, 2000, pp. 1–15.

[97] H. Liu, J. Xu, S. Chen, and T. Guo, “Compiler Optimization Parameter Selection Method Based on Ensemble Learning,” *Electronics (Switzerland)*, vol. 11, no. 15, pp. 1–26, 2022, doi: 10.3390/electronics11152452.