### Wind Speed Prediction using ML

##### A PROJECT REPORT

***Submitted by***

**PRITHIVI RAJ P (210104203)**

**SAKTHIVEL A (211420104238)**

**DHINAKARAN R (211420104323)**

***in partial fulfillment for the award of the degree***

***of***

#### BACHELOR OF ENGINEERING

***IN***

**COMPUTER SCIENCE AND ENGINEERING**



#### PANIMALAR ENGINEERING COLLEGE

**(An Autonomous Institution, Affiliated to Anna University, Chennai)**

**MARCH 2024**

### PANIMALAR ENGINEERING COLLEGE

**(An Autonomous Institution, Affiliated to Anna University, Chennai)**

#### BONAFIDE CERTIFICATE

Certified that this project report Wind Speed Prediction using ML is the bonafide work of **PRITHIV RAJ P (211420104203), SAKTHIVEL A (211420104238),**

**DHINAKARAN R (211420104308)**, who carried out the project work under my supervision.

##### Signature of the HOD with date Signature of the Supervisor with date Dr. L. JABASHEELA M.E., Ph.D., Dr. M. KRISHNAMOORTHY M.E.,

**MBA., Ph.D.,**

##### PROFESSOR AND HEAD,

Department of Computer Science and Engineering,

Panimalar Engineering College, Chennai - 600123

##### ASSOCIATE PROFESSOR,

Department of Computer Science and Engineering,

Panimalar Engineering College, Chennai - 600123

Certified that the above candidates were examined in the End Semester Project Viva- Voce Examination held on...........................

**INTERNAL EXAMINER EXTERNAL EXAMINER**

#### DECLARATION BY THE STUDENT

##### We PRITHIVI RAJ P (211420104203), SAKTHIVEL A (211420104238),

**DHINAKARAN R (211420104323),** hereby declares that this project report

Titled **WIND SPEED PREDICTION USING ML**, under the guidance of

**Dr. M. KRISHNAMOORTHY M.E., MBA., Ph.D.,** is the original work done by us and we have not plagiarized or submitted to any other degree in any university by us.

##### PRITHIVI RAJ P SAKTHIVEL A

**DHINAKARAN R**

#### ACKNOWLEDGEMENT

Our profound gratitude is directed towards our esteemed Secretary and Correspondent, **Dr. P. CHINNADURAI, M.A., Ph.D**., for his fervent encouragement. His inspirational support proved instrumental in galvanizing our efforts, ultimately contributing significantly to the successful completion of this project

We want to express our deep gratitude to our Directors,**Tmt. C. VIJAYARAJESWARI, Dr. C. SAKTHI KUMAR, M.E., Ph.D., and Dr. SARANYASREE SAKTHI KUMAR, B.E., M.B.A., Ph.D.,** for graciously

affording us the essential resources and facilities for undertaking of this project.

Our gratitude is also extended to our Principal, **Dr. K. MANI, M.E., Ph.D.,** whose facilitation proved pivotal in the successful completion of this project.

We express our heartfelt thanks to **Dr. L. JABASHEELA, M.E., Ph.D.,** Head of the Department of Computer Science and Engineering, for granting the necessary facilities that contributed to the timely and successful completion of the project.

We would like to express our sincere thanks to **Dr. V. SUBEDHA M.Tech., Ph.D.,** and **Dr. M. KRISHNAMOORTHY M.E., MBA., Ph.D.,** and all the faculty members of the Department of CSE for their unwavering support for the successful completion of the project.

##### PRITHIVI RAJ P (211420104203) SAKTHIVEL A (211420104238) DHINAKARAN R (211420104323)

ABSTRACT

This study explores the application of machine learning techniques in predicting renewable energy generation. Leveraging historical energy production data and relevant weather parameters, such as solar irradiance and wind speed, a predictive model is developed. Various machine learning algorithms, including regression and time series analysis, are employed to capture complex relationships between weather conditions and energy output. The proposed model demonstrates the ability to forecast renewable energy generation accurately and efficiently. Through extensive experimentation and validation, the study showcases the potential of machine learning in optimizing energy resource management, aiding in grid stability, and facilitating the integration of renewable sources into the power system.

# TABLE OF CONTENTS

| **CHAPTER** | **TITLE** | **PAGE** |
| --- | --- | --- |
| **NO.** |  | **NO.** |
|  | **ABSTRACT** | v |
|  |  |  |
|  | **LIST OF FIGURES** | vii |
| **1.** | **INTRODUCTION** | 01 |
|  | 1.1 Overview | 01 |
|  | 1.2 Problem Definition | 02 |
| **2.**  **3.** | **LITERATURE SURVEY**  **SYSTEM ANALYSIS** | 03  07 |
|  | 3.1 Existing System | 07 |
|  | 3.2 Proposed System  3.3 Development environment | 08  10 |
|  |  |  |
| **4.** | **SYSTEM DESIGN** | 11 |
|  | 4.1 UML Diagrams | 11 |
|  | 4.2 Preparing the Dataset | 16 |
|  | 4.3 ER Diagram | 17 |
|  | 4.4 Predictive model | 18 |
| **5.** | **SYSTEM ARCHITECTURE** | 20 |
|  | 5.1 Architecture Overview | 20 |
|  | 5.2 Module Description | 21 |
| **6.** | **SYSTEM IMPLEMENTATION** | 22 |
|  | 6.1 Front-end code | 22 |

6.2 Back-end code 30

| **CHAPTER NO.**  **7.** | **TITLE**  **ALGORITHMS** | **PAGE NO.**  39 |
| --- | --- | --- |
|  |  |  |
| **8.** | **CONCLUSION** | 49 |
|  | 8.1Conclusion | 49 |
|  | 8.2 Future enhancement | 49 |
|  | **APPENDICES** | 50 |
|  | Sample Screenshots | 50 |
| **9.** | **REFERENCES** | 52 |

# LIST OF FIGURES

| **FIG NO** | **FIGURE DESCRIPTION** | **PAGE NO** |
| --- | --- | --- |
| 4.1.1 | Use Case Diagram | 11 |
| 4.1.2 | Class Diagram | 12 |
| 4.1.3 | Sequence Diagram | 13 |
| 4.1.4 | Work flow diagram | 14 |
| 4.1.5 | Activity Diagram | 15 |
| 4.3 | ER Diagram | 17 |
| 4.4 | Module diagram | 18 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## CHAPTER 1

### INTRODUCTION

* 1. **OVERVIEW**

Due to two main factors-the industry's orientation toward privatizations (reforms) and the shift in electricity generation toward clean, pollution-free renewable energy sources-the electricity sector, particularly in the supply industry, has undergone numerous structural and systematic changes over the past several years throughout the world. Electricity forecasting becomes one of the most crucial tasks in managing the power systems in this dynamic climate. In operation planning, scheduling, and real-time power system balance, forecasting is crucial. In today's power systems, electricity load, pricing, and renewable energy sources are the three main forecasting concerns. The wind power industry has experienced significant growth and has assumed a leadership position among the recently developed renewable sources of energy (solar energy).

Additionally, the penetration of electricity produced by renewable energy sources into the power system is growing quickly, which is seen as an alternative source of energy. Wind energy has had enormous growth among new renewable energy sources in recent years; in many nations, it represents real replacement for fossil fuels. Additionally, the capability for producing wind energy is stochastic, intermittent, and connected to the production of other ramp occurrences. Despite this, it is a free and pollution-free source of energy, making it a popular source of energy.

### PROBLEM DEFINITION

Despite growing awareness of the environmental and economic benefits of renewable energy sources, there persists a significant reliance on non-renewable energy sources such as fossil fuels and nuclear power. This reliance poses a multitude of challenges, including but not limited to environmental degradation, energy insecurity, and exacerbation of climate change. Despite the availability of renewable energy technologies, many individuals, businesses, and governments continue to prioritize non-renewable energy sources. Understanding the factors contributing to this reluctance and identifying effective strategies to promote the adoption of renewable energy is essential for mitigating the adverse impacts of continued dependence on non-renewable energy sources.

The continued predominance of non-renewable energy sources presents a pressing challenge to global sustainability efforts. Despite advancements in renewable energy technologies and increasing recognition of the environmental consequences of fossil fuel consumption, significant barriers persist in the widespread adoption of renewable energy.

These barriers encompass technological limitations, insufficient policy support, economic considerations, and entrenched societal norms. Addressing these obstacles is paramount to fostering a transition towards a more sustainable energy future, characterized by reduced greenhouse gas emissions

## CHAPTER 2

### LITERATURE SURVEY

**TITLE:** Wind Power Forecasting

**AUTHOR:** Q. Chen, K. A. Folly

**YEAR:** 2018

Accurate short-term wind power forecasts are very important for reliable and efficient operation of power systems with high wind power penetration. Many conventional and artificial intelligence methods have been developed to achieve accurate wind power forecasting. Time-series-based algorithms are known to be simple, and robust, and have been used in the past for forecasting with some level of success. Recently some researchers have advocated for artificial-intelligence-based methods such as Artificial Neural Networks (ANNs), Fuzzy Logic, etc., for forecasting because of their flexibility. This paper presents a comparison of conventional and two artificial intelligence methods for wind power forecasting. The conventional method discussed in this paper is the Autoregressive Moving Average (ARMA) which is one of the most robust and simple time-series methods. The artificial intelligence methods are Artificial Neural Networks (ANNs) and Adaptive Neuro-fuzzy Inference Systems (ANFIS). Simulation results for very short-term and short-term forecasting show that ANNs and ANFIS are suitable for the very short-term (10 minutes ahead) wind speed and power forecasting, and the ARMA is suitable for the short-term (1 hour ahead) wind speed and power forecasting.

**TITLE:** Wind Power Forecasting with Deep Learning Networks: Time-Series Forecasting

**AUTHOR:** Wen-Hui Lin, Ping Wang, Kuo-Ming Chao, Hsiao-Chung Lin, Zong-Yu Yang and Yu-Huang Lai

**YEAR:** 2021

Studies have demonstrated that changes in the climate affect wind power forecasting under different weather conditions. Theoretically, accurate prediction of both wind power output and weather changes using statistics-based prediction models is difficult. In practice, traditional machine learning models can perform long-term wind power forecasting with a mean absolute percentage error (MAPE) of 10% to 17%, which does not meet the engineering requirements for our renewable energy project. Deep learning networks (DLNs) have been employed to obtain the correlations between meteorological features and power generation using a multilayer neural convolutional architecture with gradient descent algorithms to minimize estimation errors. This has wide applicability to the field of wind power forecasting. Therefore, this study aimed at the long-term (24–72-h ahead) prediction of wind power with a MAPE of less than 10% by using the Temporal Convolutional Network (TCN) algorithm of DLNs. In our experiment, we performed TCN model pretraining using historical weather data and the power generation outputs of a wind turbine from a Scada wind power plant in Turkey. The experimental results indicated a MAPE of 5.13% for 72-h wind power prediction, which is adequate within the constraints of our project. Finally, we compared the performance of four DLN-based prediction models for power forecasting, namely, the TCN, long short-term memory (LSTM), recurrent neural network (RNN), and gated recurrence unit (GRU) models. We validated that the TCN outperforms the other three models for wind power prediction in terms of data input volume, stability of error reduction, and forecast accuracy.

**TITLE:** Deep Learning-Based Prediction of Wind Power for Multi-turbines in a Wind Farm

**AUTHOR:** Xiaojiao Chen, Xiuqing Zhang, Mi Dong, Liansheng Huang, Yan Guo and Shiying He

**YEAR:** 2021

The prediction of wind power plays an indispensable role in maintaining the stability of the entire power grid. In this paper, a deep learning approach is proposed for the power prediction of multiple wind turbines. Starting from the time series of wind power, it is presents a two-stage modeling strategy, in which a deep neural network combines spatiotemporal correlation to simultaneously predict the power of multiple wind turbines. Specifically, the network is a joint model composed of a Long Short-Term Memory Network (LSTM) and a Convolutional Neural Network (CNN). Herein, the LSTM captures the temporal dependence of the historical power sequence, while the CNN extracts the spatial features among the data, thereby achieving the power prediction for multiple wind turbines. The proposed approach is validated by using the wind power data from an offshore wind farm in China, and the results in comparison with other approaches show the high prediction preciseness achieved by the proposed approach.

**TITLE:** Wind Power Prediction with Machine Learning Ensembles

**AUTHOR:** Justin Philipp Heinermann

**YEAR:** 2016

For a sustainable integration of wind power into the electricity grid, precise and robust predictions are required. With increasing installed capacity and changing energy markets, there is a growing demand for short-term predictions. Machine learning methods can be used as a purely data-driven, spatio-temporal prediction model that yields better results than traditional physical models based on weather simulations. However, there are two big challenges when applying machine learning techniques to the domain of wind power predictions. First, when applying state-of-the-art algorithms to big training data sets, the required computation times may increase to an unacceptable level. Second, the prediction performance and reliability have to be improved to cope with the requirements of the energy markets.

**TITLE:** Short-term wind power prediction based on extreme learning machine with error correction

**AUTHOR:** Zhi Li, Lin Ye, Yongning Zhao, Xuri Song, Jingzhu Teng and Jingxin Jin

**YEAR:** 2016

Large-scale integration of wind generation brings great challenges to the secure operation of the power systems due to the intermittency nature of wind. The fluctuation of the wind generation has a great impact on the unit commitment. Thus, accurate wind power forecasting plays a key role in dealing with the challenges of power system operation under uncertainties economically and technically.In this paper, a combined approach based on an Extreme Learning Machine (ELM) and an error correction model is proposed to predict wind power in the short-term time scale. Firstly, an ELM is utilized to forecast short-term wind power. Then the ultra-short-term wind power forecasting is acquired based on processing the short-term forecasting error by persistence method.

## CHAPTER 3

### SYSTEM ANALYSIS

* 1. **EXISTING SYSTEM**

The article proposes a solution to address the challenge of using renewable energy for datacenters, which can be environmentally friendly but suffer from energy instability. The instability might lead to interruptions and failures in job execution. Existing efforts to predict energy generation have limitations due to energy instability. The proposed system introduces a renewable energy allocation approach for cloud datacenters. The goal is to prevent Service Level Objective (SLO) violations caused by insufficient renewable energy supply while minimizing energy costs and carbon emissions. The system employs deep learning to predict renewable energy tail distributions for future time slots, forecasts energy demand, and predicts CPU utilization for Physical Machines (PMs).

Using these predictions, the system allocates renewable energy sources to PM areas, utilizing a Reinforcement Learning (RL)-based method. The proposed system aims to enhance the utilization of renewable energy in datacenters while ensuring consistent operation and sustainability.

Issues in Existing System

* They analyze the renewable supply demand.
* They use less effective machine learning approach.
* They did not implement deploy.
* Limited scalability.

### PROPOSED SYSTEM

The proposed system focuses on predicting renewable energy generation using machine learning techniques. It collects historical data on variables such as weather conditions, solar radiation, wind speed, and energy production. By employing regression algorithms like linear regression or advanced methods like neural networks, the system creates models to forecast renewable energy output. These models are trained on past data and continuously refined with new information to improve accuracy. Real-time data integration allows the system to provide up-to-date predictions, aiding energy grid management. Regular maintenance ensures the models remain effective in capturing changing patterns and trends. Ultimately, this system facilitates optimized energy distribution and utilization by enabling proactive decision-making based on reliable renewable energy predictions.

The goals of minimizing the total energy monetary cost and total carbon emission. We assume that the cost of stored renewable energy is higher than that of directly using the renewable energy since there is an additional cost for energy storage. Thus, we define SLO as the successful running probability by only using supplied renewable energy and aim to minimize the number of SLO violations.

**MERITS**

* We use an advanced regression method for this problem.
* We build a user-friendly web application.
* We analyze the renewable power generation & radiation.
* High scalability.

Supervised Machine Learning **is the** majority of practical machine learning uses supervised learning. Supervised learning is where have input variables (X) and an output variable (y) and use an algorithm to learn the mapping function from the input to the output**is y = f(X).** The goal is to approximate the mapping function so well that when you have new input data (X) that you can predict the output variables (y) for that data. Techniques of Supervised Machine Learning algorithms include **logistic regression**, **multi-class classification**, **Decision Trees** and **support vector machines etc**. Supervised learning requires that the data used to train the algorithm is already labelled with correct answers. Supervised learning problems can be further grouped into **Classification** problems

This problem has as goal the construction of a succinct model that can predict the value of the dependent attribute from the attribute variables. The difference between the two tasks is the fact that the dependent attribute is numerical for categorical for classification. A classification model attempts to draw some conclusion from observed values. Given one or more inputs a classification model.

### DEVELOPMENT ENVIRONMENT

* + 1. **SOFTWARE REQUIREMENT**
       1. Windows/Mac OS
       2. HTML
       3. CSS
       4. Anaconda
       5. Jupyter notebook
       6. VS code
       7. Browser

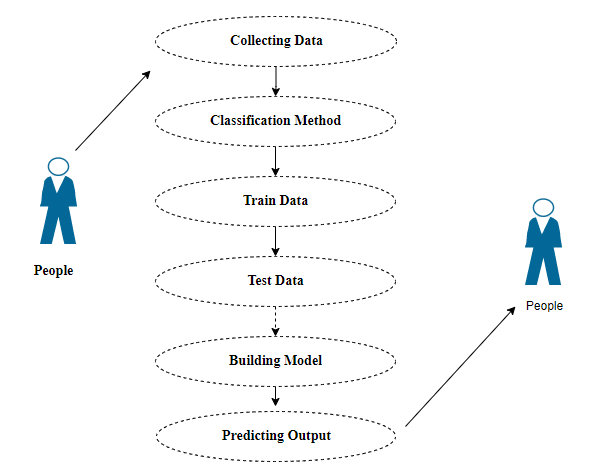
### HARDWARE REQUIREMENT

* + - 1. Processor: Pentium III/IV
      2. Memory (RAM): 4GB
      3. Hard Drive: Minimum 80 GB

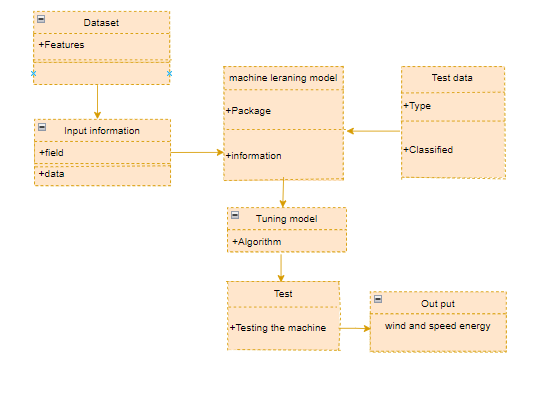
**CHAPTER 4**

**SYSTEM DESIGN**

### UML DIAGRAMS

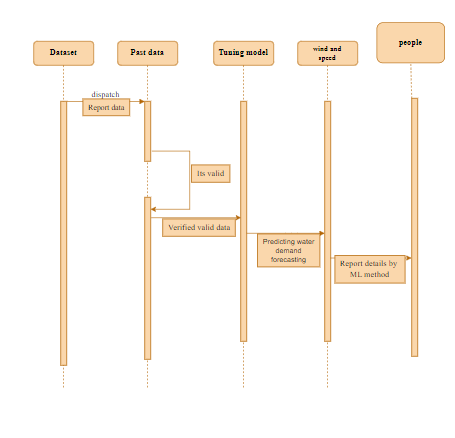


##### Fig 4.1.1 Use case diagram



##### Fig 4.1.2 Class diagram

The class diagram refers to relationships between different classes.



##### Fig 4.1.3 Sequence diagram

The sequence diagram is used to transfer the dataset to the training model.



##### 

##### 

##### 





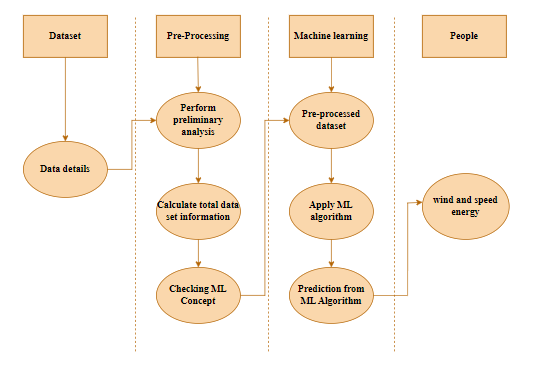






##### Fig 4.1.4 Workflow diagram

The Workflow diagram shows how work done in the model



##### Fig 4.1.5 Activity diagram

The activity diagram sharing the datasets with the training models

### 4.2 PREPARING THE DATASET

This Dataset contains various weather features which can be analyzed and used as predictors. It Contains

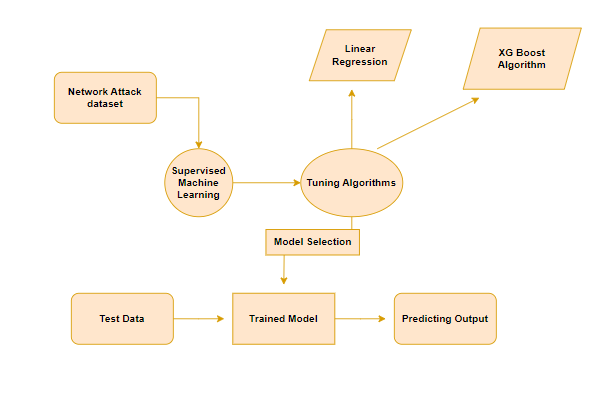
* System Power Generated
* Wind Speed
* Wind Direction

Pressure

It is a secondary source and discusses published information in a particular subject area and sometimes information in a particular subject area within a certain period. Its ultimate goal is to bring the reader up to date with current literature on a topic and form the basis for another goal, such as future research that may be needed in the area and precedes a research proposal and maybe just a simple summary of sources. Usually, it has an organizational pattern and combines both summary and synthesis.

### A summary is a brief recap of crucial information found in a source, whereas a synthesis involves rearranging or reorganizing information. A synthesis could offer a fresh perspective on previously known material by combining new and old interpretations or by tracing the intellectual advancements and major debates in the field. Depending on the context, the literature review may assess the sources and suggest the most relevant ones to the reader.

### 4.3 ER DIAGRAM



### An Entity Relationship Diagram (ERD), also referred to as an Entity-Relationship Model, is a visual representation of an information system that showcases the relationships between individuals, objects, places, concepts, or events within the system. ERD is a technique of data modeling that can be used to define business processes and serve as the basis for a relational database.

### 4.4 PREDICTIVE MODEL



#### 

#### 

#### 

#### 

#### 

#### 

#### 

#### 

#### 

#### 

#### 

#### 

#### Fig 4.4 Module diagram

### CONSTRUCTION

## 

## Machine learning needs data gathering have lot of past data’s. Data gathering have sufficient historical data and raw data. Before data pre-processing, raw data can’t be used directly. It’s used to pre-process then, what kind of algorithm with model. Training and testing this model working and predicting correctly with minimum errors. Tuned model involved by tuned time to time with improving the accuracy.

## The data which was collected might contain missing values that may lead to inconsistency. To gain better results data need to be preprocessed so as to improve the efficiency of the algorithm. The outliers have to be removed and also variable conversion need to be done.

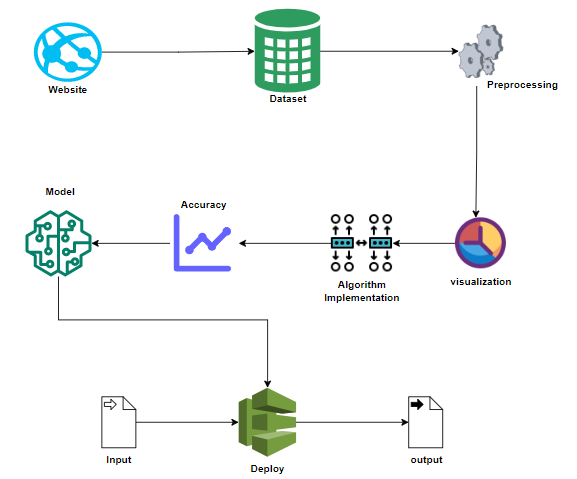
The software requirements specification is a technical specification of requirements for the software product. It is the first step in the requirements analysis process. It lists requirements of a particular software system. The following details to follow the special libraries like sk-learn, pandas, numpy, matplotlib and seaborn.

Navigator is an easy, point-and-click way to work with packages and environments without needing to type conda commands in a terminal window. You can use it to find the packages you want, install them in an environment, run the packages, and update them – all inside Navigator

## CHAPTER 5

**SYSTEM ARCHITECTURE**

* 1. **ARCHITECTURE OVERVIEW**

**s**

#### Fig 5.1 Architecture diagram

### MODULE DESCRIPTION

Validation techniques in machine learning are used to get the error rate of the Machine Learning (ML) model, which can be considered as close to the true error rate of the dataset. If the data volume is large enough to be representative of the population, you may not need the validation techniques. However, in real-world scenarios, to work with samples of data that may not be a true representative of the population of given dataset. To finding the missing value, duplicate value and description of data type whether it is float variable or integer. The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters.

The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration. The validation set is used to evaluate a given model, but this is for frequent evaluation. It as machine learning engineers use this data to fine-tune the model hyper parameters. Data collection, data analysis, and the process of addressing data content, quality, and structure can add up to a time-consuming to-do list. During the process of data identification, it helps to understand your data and its properties; this knowledge will help you choose which algorithm to use to build your model.

Several different data cleaning tasks use Python’s Pandas library and specifically, it focuses on probably the biggest data cleaning task, missing values and it canmore [quickly clean data](https://www.dataoptimal.com/data-cleaning-with-python-2018/)**.** It wants to spend less time cleaning data**,** and more time exploring and modeling. Some of these sources are just simple random mistakes. Other times, there can be a deeper reason why data is missing. It’s important to understand these different types of missing data from a statistics point of view.

## CHAPTER 6

**SYSTEM IMPLEMENTATION**

**FRONT END CODE:**

**GENERATIONAL.HTML**

{% extends 'base.html' %}  
  
{% block content %}  
  
    <div class="bg-gradient-to-r from-gray-500 to-red-500 mx-auto max-w-6xl bg-white py-20 px-12 lg:px-24 shadow-xl mb-24">  
      <form method="POST" action="{% url 'generation\_model' %}">  
        {% csrf\_token %}  
        <div class="bg-gradient-to-r from-gray-500 to-red-500 shadow-md rounded px-8 pt-6 pb-8 mb-4 flex flex-col">  
          <center><h1 class="text-6xl py-6 text-blue-300">Generation Model</h1></center>  
          <div class="-mx-3 md:flex mb-6">  
             
            <div class="md:w-1/2 px-3 mb-6 md:mb-0">  
              <label class="uppercase tracking-wide text-black text-xs font-bold mb-2" for="company">T\_amb\_Degree</label>  
              <input class="w-full bg-gray-200 text-black border border-gray-200 rounded py-3 px-4 mb-3" id="company" step="0.01" name="T\_amb\_Degree" type="number" required placeholder="Please enter your amb degree value">  
  
              <label class="uppercase tracking-wide text-black text-xs font-bold mb-2" for="company">DewPoint\_Degree</label>  
              <input class="w-full bg-gray-200 text-black border border-gray-200 rounded py-3 px-4 mb-3" name="DewPoint\_Degree" step="0.01" id="company" type="number" required placeholder="Please enter   
             

**BASE.HTML**

{% load static %}

<!DOCTYPE html>

<html lang="en">

<head>

    <meta charset="UTF-8">

    <meta name="viewport" content="width=device-width, initial-scale=1.0">

    <title>Document</title>

    <script src="[https://cdn.tailwindcss.com](https://cdn.tailwindcss.com/)"></script>

</head>

<body class="bg-gradient-to-r from-gray-500 to-red-500">

    <header class="text-gray-700 body-font border-b border-gray-200">

        <div class="container mx-auto flex flex-wrap p-5 flex-col md:flex-row items-center">

          <a class="flex title-font font-medium items-center text-gray-900 mb-4 md:mb-0" href="#" >

            <img src="{% static 'images/doctor-q.svg' %}" class="w-10 h-10 text-white p-2 bg-indigo-500 rounded-full" alt="">

            <span class="ml-3 text-xl">Renewable Energy</span>

          </a>

          <nav class="md:ml-auto flex flex-wrap items-center text-base justify-center">

            <a href="{% url 'home' %}" class="mr-5 hover:text-gray-900">Home</a>

            <a href="{% url 'about' %}" class="mr-5 hover:text-gray-900">Problem statement</a>

            <a href="{% url 'generation\_model' %}" class="mr-5 hover:text-gray-900">Generation Model</a>

            <a href="{% url 'speed\_model' %}" class="mr-5 hover:text-gray-900">Speed Model</a>

          </nav>

          {% if user.is\_authenticated %}

          <button class="inline-flex items-center bg-gray-200 border-0 py-1 px-3 focus:outline-none hover:bg-gray-300 rounded text-base mt-4 md:mt-0"><a href="{% url 'logout' %}">Logout</a>

            <svg fill="none" stroke="currentColor" stroke-linecap="round" stroke-linejoin="round" stroke-width="2" class="w-4 h-4 ml-1" viewBox="0 0 24 24">

              <path d="M5 12h14M12 5l7 7-7 7"></path>

            </svg>

          </button>

          {% endif %}

        </div>

      </header>

      {% block content %} {% endblock %}

      <section class="text-gray-700 body-font relative">

        <div class="absolute inset-0 bg-gray-300">

          <iframe width="100%" height="100%" frameborder="0" marginheight="0" marginwidth="0" title="map" scrolling="no" src="[https://maps.google.com/maps?width=100%](https://maps.google.com/maps?width=100%25)&amp;height=600&amp;hl=en&amp;q=%C4%B0zmir+(My%20Business%20Name)&amp;ie=UTF8&amp;t=&amp;z=14&amp;iwloc=B&amp;output=embed" style="filter: grayscale(1) contrast(1.2) opacity(0.4);"></iframe>

        </div>

        <form method="post">

<div class="container px-5 py-24 mx-auto flex">

            <div class="lg:w-1/3 md:w-1/2 bg-white rounded-lg p-8 flex flex-col md:ml-auto w-full mt-10 md:mt-0 relative z-10">

              <h2 class="text-gray-900 text-lg mb-1 font-medium title-font">Feedback</h2>

              <p class="leading-relaxed mb-5 text-gray-600">Post-ironic portland shabby chic echo park, banjo fashion axe</p>

              <input class="bg-white rounded border border-gray-400 focus:outline-none focus:border-indigo-500 text-base px-4 py-2 mb-4" placeholder="Email" type="email">

              <textarea class="bg-white rounded border border-gray-400 focus:outline-none h-32 focus:border-indigo-500 text-base px-4 py-2 mb-4 resize-none" placeholder="Message"></textarea>

              <button class="text-white bg-indigo-500 border-0 py-2 px-6 focus:outline-none hover:bg-indigo-600 rounded text-lg">Button</button>

              <p class="text-xs text-gray-500 mt-3">Chicharrones blog helvetica normcore iceland tousled brook viral artisan.</p>

            </div>

          </form>

        </div>

      </section>

        <!-- <div class="bg-gray-200">

          <div class="container mx-auto py-4 px-5 flex flex-wrap flex-col sm:flex-row">

            <p class="text-gray-500 text-sm text-center sm:text-left">2023 (chadjayasuriya)

              <a href="<https://twitter.com/knyttneve>" class="text-gray-600 ml-1" target="\_blank" rel="noopener noreferrer">Subscribe mychannel</a>

            </p>

            <span class="sm:ml-auto sm:mt-0 mt-2 sm:w-auto w-full sm:text-left text-center text-gray-500 text-sm">Founder & CEO Jayasurya</span>

          </div>

        </div> -->

      </footer>

</body>

</html>

### CONTACT.HTML

{% extends 'base.html' %}

{% block content %}

  <!-- Credit: Componentity.com -->

  <section class="text-gray-700 body-font border-t border-gray-200">

    <div class="container px-5 py-24 mx-auto">

      <div class="flex flex-wrap w-full mb-20 flex-col items-center text-center">

        <h1 class="sm:text-3xl text-2xl font-medium title-font mb-2 text-gray-900">Neural Doctors club</h1>

        <p class="lg:w-1/2 w-full leading-relaxed text-base">Neurologists say that our brains are programmed much more for stories than for abstract ideas. Tales with a little drama are remembered far longer than any slide crammed with analytics.</p>

      </div>

      <div class="flex flex-wrap -m-4">

        <div class="xl:w-1/3 md:w-1/2 p-4">

          <div class="border border-gray-300 p-6 rounded-lg">

            <div class="w-10 h-10 inline-flex items-center justify-center rounded-full bg-indigo-100 text-indigo-500 mb-4">

              <svg fill="none" stroke="currentColor" stroke-linecap="round" stroke-linejoin="round" stroke-width="2" class="w-6 h-6" viewBox="0 0 24 24">

                <path d="M22 12h-4l-3 9L9 3l-3 9H2"></path>

              </svg>

            </div>

            <h2 class="text-lg text-gray-900 font-medium title-font mb-2">Doctors</h2>

            <p class="leading-relaxed text-base">As we get involved in the emotional healing of others, we get healed

                emotionally ourselves.</p>

          </div>

        </div>

        <div class="xl:w-1/3 md:w-1/2 p-4">

          <div class="border border-gray-300 p-6 rounded-lg">

            <div class="w-10 h-10 inline-flex items-center justify-center rounded-full bg-indigo-100 text-indigo-500 mb-4">

              <svg fill="none" stroke="currentColor" stroke-linecap="round" stroke-linejoin="round" stroke-width="2" class="w-6 h-6" viewBox="0 0 24 24">

                <circle cx="6" cy="6" r="3"></circle>

                <circle cx="6" cy="18" r="3"></circle>

                <path d="M20 4L8.12 15.88M14.47 14.48L20 20M8.12 8.12L12 12"></path>

              </svg>

            </div>

            <h2 class="text-lg text-gray-900 font-medium title-font mb-2">Surgeon</h2>

            <p class="leading-relaxed text-base">A surgeon should give as little pain as possible while he is treating the patient, and no pain at all when he charges his fee.</p>

          </div>

        </div>

        <div class="xl:w-1/3 md:w-1/2 p-4">

          <div class="border border-gray-300 p-6 rounded-lg">

            <div class="w-10 h-10 inline-flex items-center justify-center rounded-full bg-indigo-100 text-indigo-500 mb-4">

              <svg fill="none" stroke="currentColor" stroke-linecap="round" stroke-linejoin="round" stroke-width="2" class="w-6 h-6" viewBox="0 0 24 24">

                <path d="M20 21v-2a4 4 0 00-4-4H8a4 4 0 00-4 4v2"></path>

                <circle cx="12" cy="7" r="4"></circle>

              </svg>

            </div>

            <h2 class="text-lg text-gray-900 font-medium title-font mb-2">Patients</h2>

            <p class="leading-relaxed text-base">A patient is a person who is receiving medical treatment from a doctor or hospital. A patient is also someone who is taken care of by a particular doctor.</p>

          </div>

        </div>

  </section>

  <section class="text-gray-700 body-font border-t border-gray-200">

    <div class="container px-5 py-24 mx-auto">

      <div class="flex flex-col text-center w-full mb-20">

        <h1 class="sm:text-3xl text-2xl font-medium title-font mb-4 text-gray-900">Our Team</h1>

        <p class="lg:w-2/3 mx-auto leading-relaxed text-base">Teamwork begins by building trust. And the only way to do that is to overcome our need for invulnerability.</p>

      </div>

      <div class="flex flex-wrap -m-2">

        <div class="p-2 lg:w-1/3 md:w-1/2 w-full">

          <div class="h-full flex items-center border-gray-200 border p-4 rounded-lg">

            <img alt="team" class="w-16 h-16 bg-gray-100 object-cover object-center flex-shrink-0 rounded-full mr-4" src="<https://dummyimage.com/80x80/edf2f7/a5afbd>">

            <div class="flex-grow">

              <h2 class="text-gray-900 title-font font-medium">Team member 1</h2>

              <p class="text-gray-500">Your contribution</p>

            </div>

          </div>

        </div>

        <div class="p-2 lg:w-1/3 md:w-1/2 w-full">

          <div class="h-full flex items-center border-gray-200 border p-4 rounded-lg">

            <img alt="team" class="w-16 h-16 bg-gray-100 object-cover object-center flex-shrink-0 rounded-full mr-4" src="<https://dummyimage.com/84x84/edf2f7/a5afbd>">

            <div class="flex-grow">

              <h2 class="text-gray-900 title-font font-medium">Team member 2</h2>

              <p class="text-gray-500">Your contribution</p>

            </div>

          </div>

        </div>

        <div class="p-2 lg:w-1/3 md:w-1/2 w-full">

          <div class="h-full flex items-center border-gray-200 border p-4 rounded-lg">

            <img alt="team" class="w-16 h-16 bg-gray-100 object-cover object-center flex-shrink-0 rounded-full mr-4" src="<https://dummyimage.com/88x88/edf2f7/a5afbd>">

            <div class="flex-grow">

              <h2 class="text-gray-900 title-font font-medium">Team member 3</h2>

              <p class="text-gray-500">Your contribution</p>

            </div>

          </div>

        </div>

  </section>

### SPEED MODEL.HTML

{% extends 'base.html' %}

{% block content %}

    <div class="bg-gradient-to-r from-gray-500 to-red-600 mx-auto max-w-6xl bg-white py-20 px-12 lg:px-24 shadow-xl mb-24">

      <form method="post">

        {% csrf\_token %}

        <div class="bg-gradient-to-r from-gray-500 to-red-500 shadow-md rounded px-8 pt-6 pb-8 mb-4 flex flex-col">

          <center><h1 class="text-6xl py-6 text-blue-300">Speed Model</h1></center>

          <div class="-mx-2 md:flex mb-6">

            <div class="md:w-1/2 px-3 mb-6 md:mb-0">

              <label class="uppercase tracking-wide text-black text-xs font-bold mb-2" for="company">Wind Speed (m/s)</label>

              <input class="w-full bg-gray-200 text-black border border-gray-200 rounded py-3 px-4 mb-3" id="company" name="Wind Speed (m/s)" type="number" step="0.000001" placeholder="Please enter your wind speed value">

              <label class="uppercase tracking-wide text-black text-xs font-bold mb-2 px-4" for="company">Theoretical\_Power\_Curve (KWh)</label>

              <input class="w-full bg-gray-200 text-black border border-gray-200 rounded py-3 px-4 mb-3" name="Theoretical\_Power\_Curve (KWh)" id="company" step="0.000001" type="number" placeholder="Please enter your power curve kwh ">

              <label class="uppercase tracking-wide text-black text-xs font-bold mb-2" for="company">Wind Direction (°)</label>

              <input class="w-full bg-gray-200 text-black border border-gray-200 rounded py-3 px-4 mb-3" name="Wind Direction (°)" id="company" step="0.000001" type="number" placeholder="Please enter your wind direction">

            </div>

          </div>

          <div class="-mx-3 md:flex mt-2">

            <div class="md:w-full px-3">

              <button class="md:w-full bg-gray-900 text-white font-bold py-2 px-4 border-b-4 hover:border-b-2 border-gray-500 hover:border-gray-100 rounded-full">

                Submit

              </button>

            </div>

          </div>

          <h1>{{ prediction\_text }}</h1>

        </div>

      </form>

    </div>

  </body>

{% endblock content %}

**TEST.HTML**

<!DOCTYPE html>

<html lang="en">

<head>

    <meta charset="UTF-8">

    <meta name="viewport" content="width=device-width, initial-scale=1.0">

    <title>Document</title>

</head>

<style>

    \* {

        margin: 0;

        padding: 0;

        box-sizing: border-box;

      }

      body {

        display: flex;

        justify-content: center;

        align-items: center;

        min-height: 100vh;

        background-color: #23242a;

      }

      .box {

        position: relative;

        width: 380px;

        height: 420px;

        background: #1c1c1c;

        border-radius: 8px;

        overflow: hidden;

      }

      .box::before, .box::after {

        content: "";

        position: absolute;

        top: -50%;

        left: -50%;

        width: 380px;

        height: 420px;

        z-index: 1;

        transform-origin: bottom right;

        animation: animate 6s linear infinite;

        background: linear-gradient(0deg, transparent, blue, blue, blue);

      }

      .box::after {

        animation-delay: -3s;

      }

      .box\_\_borderline {

        position: absolute;

        top: 0;

        inset: 0;

      }

      .box\_\_borderline::before, .box\_\_borderline::after {

        content: "";

        position: absolute;

        top: -50%;

        left: -50%;

        width: 380px;

        height: 420px;

        transform-origin: bottom right;

        animation: animate 6s linear infinite;

        z-index: 1;

        background: linear-gradient(0deg, transparent, red, red, red);

      }

      .box\_\_borderline::before {

        animation-delay: -1.5s;

### BACK-END CODING

*# Data Pre-Processing*

*# import the necessary libraries.*

**import** pandas **as** pd

**import** numpy **as** np

*# Avoid unnessary warnings in this module(Ex: software updation, version-mismatch, cache errors)*

**import** warnings

warnings**.**filterwarnings('ignore')

*# Read the datasets*

df**=**pd**.**read\_csv('GENERATION.csv')

*# Check the head values.*

df**.**head()

*# Check our data how many columns they are.*

df**.**columns

*# Drop the unwanted features in this project.*

df**=**df**.**drop(["Unnamed: 8", "Unnamed: 9","Timeseries"],axis**=**1)

*# Drop the null values of in this datasets.*

df1**=**df**.**dropna()

*# Check the bottom line records of this data.*

df**.**tail()

*# Check the dimension of our datasets (ex: how many rows and columns)*

df**.**shape

*# Check the overall datasets value*

df**.**size

*# Check our data how many columns they are.*

df**.**columns

*# Check our datasets information (ex: data\_type, NUlL\_count, Columns, and so on)*

df**.**info()

*# Check if any null value in our datasets*

df**.**isnull()

*# Remove the null values*

df **=** df**.**dropna()

*# Check the unique value for specific columns*

df['Power Generated\n(kw)']**.**unique()

*# Describe our data in stastical point of view.*

df**.**describe()

*# Check the relationship between each individual columns.*

df**.**corr()

*# Check our datasets information (ex: data\_type, NUlL\_count, Columns, and so on)*

df**.**info()

*# Checkout the data is duplicated.*

df**.**duplicated()

*# Total sum of duplicated values.*

sum(df**.**duplicated())

**DATA VISUALIATION**

*# Data Visualization*

*# Import necessary libraries.*

**import** pandas **as** pd

**import** numpy **as** np

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

*# Avoid unnessary warnings in this module(Ex: software updation, version-mismatch, cache errors)*

**import** warnings

warnings**.**filterwarnings('ignore')

*# Read the datasets.*

df**=**pd**.**read\_csv('GENERATION.csv')

*# Check the head values.*

df**.**head()

*# Check our datasets information (ex: data\_type, NUlL\_count, Columns, and so on)*

df**.**info()

*# Drop unnessary features of this datasets*

df**=**df**.**drop(["Unnamed: 8", "Unnamed: 9","Timeseries"],axis**=**1)

*# Remove the null data*

df**=**df**.**dropna()

*# Check the bottom value of this datasets*

df**.**tail()

*# Plot a Histogram.*

plt**.**figure(figsize**=**(15,5))

plt**.**subplot(1,2,1)

plt**.**hist(df['Power Generated\n(kw)'],color**=**'DarkOrange')

plt**.**subplot(1,2,2)

plt**.**hist(df['Relative Humidity\n(%)'],color**=**'HotPink')

*# Plot the Histogram*

df**.**hist(figsize**=**(15,55),layout**=**(15,4), color**=**'MistyRose')

plt**.**show()

*# Plot the Histogram*

df['Pressure mBar']**.**hist(figsize**=**(10,5),color**=**'Magenta')

*# Scatter Plot*

plt**.**scatter(df['Power Generated\n(kw)'],df['DewPoint\n(Degree)'], color**=**'darkslategray') *# scatter, plot, triplot, stackplot*

*# Check the outlayers of this datasets*

plt**.**boxplot(df['WindDir (Degree)'], boxprops**=**{'color': 'Violet'})

plt**.**show()

*# Plot a density graph*

df['T amb (Degree)']**.**plot(kind**=**'density',color**=**"Plum")

*# Violin plot*

sns**.**violinplot(df['Wind Speed at 76.8 m\n(m/sec)'], color**=**'Plum')

*# barplot, boxenplot, boxplot, countplot, displot, distplot, ecdfplot, histplot, kdeplot, pointplot, violinplot, stripplot*

*# Check the relation between each columns using heatmap.*

fig, ax **=** plt**.**subplots(figsize**=**(20,15))

sns**.**heatmap(df**.**corr(),annot **=** **True**, fmt**=**'0.2%',cmap **=** 'autumn',ax**=**ax)

*# Visualise the distribution of data.*

sns**.**histplot(df['Power Generated\n(kw)'], kde**=True**)

plt**.**xlabel('Power Generated\n(kw)')

plt**.**ylabel('Frequency')

plt**.**title('Distribution Plot of Power Generated')

plt**.**show()

**IMPLEMENTING XG BOOST ALGORITHM**

*# Import the necessary libraries*

**import** pandas **as** pd

**import** numpy **as** np

*# Avoid unnessary warnings in this module(Ex: software updation, version-mismatch, cache errors)*

**import** warnings

warnings**.**filterwarnings('ignore')

*# Read the datasets*

df**=**pd**.**read\_csv('GENERATION.csv')

*# Check the head values.*

df**.**head()

*# Check our data how many columns they are.*

df**.**columns

*# Drop the unwanted features in this project.*

df**=**df**.**drop(["Unnamed: 8", "Unnamed: 9","Timeseries"],axis**=**1)

*# Drop the null values of in this datasets.*

df1**=**df**.**dropna()

*# Check the bottom line records of this data.*

df1**.**tail()

df **=** df1**.**rename({'T amb (Degree)':'T\_amb\_Degree','DewPoint\n(Degree)':'DewPoint\_Degree','Relative Humidity\n(%)':'Relative\_Humidity', 'Pressure mBar':'Pressure\_Bar','WindDir (Degree)':'WindDirection', 'Wind Speed at 76.8 m\n(m/sec)':'Wind\_Speed\_at\_76.8', 'Power Generated\n(kw)':'Power\_Generated\_kw'},axis**=**1)

*# Split the datasets features into two varibles (dependent, independent)*

*# Independent variables*

x **=** df**.**drop(labels**=**'Power\_Generated\_kw', axis**=**1)

*# Dependent variable*

y **=** df**.**loc[:,'Power\_Generated\_kw']

x

*# Split the data sets into two methods Training, Testing*

**from** sklearn.model\_selection **import** train\_test\_split

x\_train, x\_test, y\_train, y\_test **=** train\_test\_split(x, y, test\_size**=**0.20, random\_state**=**42)

print("NUMBER OF TRAIN DATASET : ", len(x\_train))

print("NUMBER OF TEST DATASET : ", len(x\_test))

print("TOTAL NUMBER OF DATASET : ", len(x\_train)**+**len(x\_test))

*# # Build a Learning pattern algorithm for regression problem.*

*# import xgboost as xgb*

**from** sklearn.ensemble **import** RandomForestRegressor

XGB **=** RandomForestRegressor()

*# Fit is the training funtion execute to learn*

XGB**.**fit(x\_train, y\_train)

*# Test the input features.*

y\_pred **=** XGB**.**predict(x\_test)

*# Check the Rand score*

**from** sklearn.metrics **import** rand\_score

RAND **=** rand\_score(y\_test, y\_pred)

print("MEAN RAND SCOORE OF XG BOOST REGRESSOR",RAND**\***100)

*# Checkout the mean squared error rate*

**from** sklearn.metrics **import** mean\_squared\_error

MSE **=** mean\_squared\_error(y\_test, y\_pred)

print("MEAN SQUARE ERROR SCOORE OF XG BOOST REGRESSOR",MSE**\***100)

*# Checkout the mean absolute error*

**from** sklearn.metrics **import** mean\_absolute\_error

MAE **=** mean\_absolute\_error(y\_test, y\_pred)

print("MEAN ABSOLUTE ERROR OF XG BOOST REGRESSOR",MAE)

*# Checkout the R2score*

**from** sklearn.metrics **import** r2\_score

R2 **=** r2\_score(y\_test, y\_pred)

print("R-SQUARE VALUE OF XG BOOST REGRESSOR",R2)

*# Save the model for deployment purpose.*

**import** joblib

joblib**.**dump(XGB, 'Generation2.pkl')

**IMPLEMENTING LINEAR REGRESSION**

*# Linear Regression*

*# Import necessary packages*

**import** pandas **as** pd

**import** numpy **as** np

pd**.**set\_option('display.max\_rows', **None**)

*# Avoid unnessary warnings in this module(Ex: software updation, version-mismatch, cache errors)*

**import** warnings

warnings**.**filterwarnings('ignore')

*# Read the datasets*

df**=**pd**.**read\_csv('SPEED.csv')

*# Check the head values.*

df**.**head()

*# Check our data how many columns they are.*

df**.**columns

*# Drop the null values of in this datasets.*

df**=**df**.**dropna()

*# Drop the unwanted features in this project*

df1**=**df**.**drop("Date/Time", axis**=**1)

*# Check the bottom line records of this data.*

df1**.**tail()

df **=** df1**.**rename({'LV ActivePower (kW)':'LV\_ActivePower\_kW','Wind Speed (m/s)':'Wind\_Speed\_ms','Theoretical\_Power\_Curve (KWh)':'Theoretical\_Power\_Curve\_KWh', 'Wind Direction (°)':'Wind\_direction'},axis**=**1)

df**.**head()

*# Split the datasets features into two varibles (dependent, independent)*

*# Independent variables*

x **=** df**.**drop(labels**=**'LV\_ActivePower\_kW', axis**=**1)

*# Dependent variable*

y **=** df**.**loc[:,'LV\_ActivePower\_kW']

*# Split the data sets into two methods Training, Testing*

**from** sklearn.model\_selection **import** train\_test\_split

x\_train, x\_test, y\_train, y\_test **=** train\_test\_split(x, y, test\_size**=**0.20, random\_state**=**42)

print("NUMBER OF TRAIN DATASET : ", len(x\_train))

print("NUMBER OF TEST DATASET : ", len(x\_test))

print("TOTAL NUMBER OF DATASET : ", len(x\_train)**+**len(x\_test))

*# Build a Learning pattern algorithm for regression problem.*

**from** sklearn.linear\_model **import** LinearRegression

LR **=** LinearRegression()

*# Fit is the training funtion execute to learn*

LR**.**fit(x\_train, y\_train)

*# Test the input features.*

y\_pred **=** LR**.**predict(x\_test)

*# Check the Rand score*

**from** sklearn.metrics **import** rand\_score

RAND **=** rand\_score(y\_test, y\_pred)

print("MEAN RAND SCOORE OF LINEAR REGRESSOR",RAND**\***100)

*# Checkout the mean squared error rate*

**from** sklearn.metrics **import** mean\_squared\_error

MSE **=** mean\_squared\_error(y\_test, y\_pred)

print("MEAN SQUARE ERROR SCOORE OF LINEAR REGRESSOR",MSE**\***100)

*# Checkout the mean absolute error*

**from** sklearn.metrics **import** mean\_absolute\_error

MAE **=** mean\_absolute\_error(y\_test, y\_pred)

print("MEAN ABSOLUTE ERROR OF LINEAR REGRESSOR",MAE)

*# Checkout the R2score*

**from** sklearn.metrics **import** r2\_score

R2 **=** r2\_score(y\_test, y\_pred)

print("R-SQUARE VALUE OF LINEAR REGRESSOR",R2)

*# Save the model for deployment purpose.*

**import** joblib

joblib**.**dump(LR, 'windspeed2.pkl')

context = { 'patients':

patients,

'medical\_records': medical\_records

}

return render(request, 'home.html', context)

## CHAPTER 7

## ALGORITHMS

The below 2 different algorithms are compared:

* Linear Regression
* XG Boost Algorithm

The K-fold cross validation procedure is used to evaluate each algorithm, importantly configured with the same random seed to ensure that the same splits to the training data are performed and that each algorithm is evaluated in precisely the same way. Before that comparing algorithm, Building a Machine Learning Model using install Scikit-Learn libraries. In this library package have to done preprocessing, linear model with logistic regression method, cross validating by KFold method, ensemble with random forest method and tree with decision tree classifier. Additionally, splitting the train set and test set. To predicting the result by comparing accuracy.

**REGRESSION:**

Regression is a statistical method used in finance, investing, and other disciplines that attempts to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables).

Regression helps investment and financial managers to value assets and understand the relationships between variables, such as [commodity prices](https://www.investopedia.com/terms/c/commodity.asp) and the stocks of businesses dealing in those commodities.

The two basic types of regression are simple linear regression and multiple linear regression, although there are non-linear regression methods for more complicated data and analysis. Simple linear regression uses one independent variable to explain or predict the outcome of the dependent variable Y, while multiple linear regression uses two or more independent variables to predict the outcome.

Regression can help finance and investment professionals as well as professionals in other businesses. Regression can also help predict sales for a company based on weather, previous sales, GDP growth, or other types of conditions. The [capital asset pricing model](https://www.investopedia.com/terms/c/capm.asp) (CAPM) is an often-used regression model in finance for pricing assets and discovering costs of capital.

The general form of each type of regression is:

* **Simple linear regression:** Y = a + bX + u
* **Multiple linear regression:** Y = a + b1X1+ b2X2 + b3X3 + ... + btXt + u

Where:

* Y = the variable that you are trying to predict (dependent variable).
* X = the variable that you are using to predict Y (independent variable).
* a = the intercept.
* b = the slope.
* u = the regression residual.

Regression is often used to determine how many specific factors such as the price of a commodity, interest rates, particular industries, or sectors influence the price movement of an asset. The aforementioned CAPM is based on regression, and it is utilized to project the expected returns for stocks and to generate costs of capital. A stock's returns are regressed against the returns of a broader index, such as the S&P 500, to generate a beta for the particular stock.

Beta is the stock's risk in relation to the market or index and is reflected as the slope in the CAPM model. The return for the stock in question would be the dependent variable Y, while the independent variable X would be the market risk premium.

Additional variables such as the market capitalization of a stock, valuation ratios, and recent returns can be added to the CAPM model to get better estimates for returns. These additional factors are known as the Fama-French factors, named after the professors who developed the multiple linear regression model to better explain asset returns.

Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression is to model the [linear relationship](https://www.investopedia.com/terms/l/linearrelationship.asp) between the explanatory (independent) variables and response (dependent) variables. In essence, multiple regression is the extension of ordinary least-squares (OLS) [regression](https://www.investopedia.com/terms/r/regression.asp) because it involves more than one explanatory variable.

Simple linear regression is a function that allows an analyst or statistician to make predictions about one variable based on the information that is known about another variable. Linear regression can only be used when one has two continuous variables—an independent variable and a dependent variable. The independent variable is the parameter that is used to calculate the dependent variable or outcome. A multiple regression model extends to several explanatory variables.

**The multiple regression model is based on the following assumptions:**

* There is a [linear relationship](https://www.investopedia.com/terms/l/linearrelationship.asp) between the dependent variables and the independent variables
* The independent variables are not too highly [correlated](https://www.investopedia.com/terms/c/correlation.asp) with each other
* yi observations are selected independently and randomly from the population
* Residuals should be [normally distributed](https://www.investopedia.com/terms/n/normaldistribution.asp) with a mean of 0 and [variance](https://www.investopedia.com/terms/v/variance.asp) σ

The [coefficient of determination](https://www.investopedia.com/terms/c/coefficient-of-determination.asp) (R-squared) is a statistical metric that is used to measure how much of the variation in outcome can be explained by the variation in the independent variables. R2 always increases as more predictors are added to the MLR model, even though the predictors may not be related to the outcome variable.

R2 by itself can't thus be used to identify which predictors should be included in a model and which should be excluded. R2 can only be between 0 and 1, where 0 indicates that the outcome cannot be predicted by any of the independent variables and 1 indicates that the outcome can be predicted without error from the independent variables.

When interpreting the results of multiple regression, beta coefficients are valid while holding all other variables constant ("all else equal"). The output from a multiple regression can be displayed horizontally as an equation, or vertically in table form.

The [least-squares](https://www.investopedia.com/terms/l/least-squares.asp) method is a form of mathematical regression analysis used to determine the [line of best fit](https://www.investopedia.com/terms/l/line-of-best-fit.asp) for a set of data, providing a visual demonstration of the relationship between the data points. Each point of data represents the relationship between a known independent variable and an unknown dependent variable.

## Least Squares Method:

This method of [regression](https://www.investopedia.com/terms/r/regression.asp) analysis begins with a set of data points to be plotted on an x- and y-axis graph. An analyst using the least-squares method will generate a line of best fit that explains the potential relationship between independent and dependent variables.

The least-squares method provides the overall rationale for the placement of the line of best fit among the data points being studied. The most common application of this method, which is sometimes referred to as "linear" or "ordinary," aims to create a straight line that minimizes the sum of the squares of the errors that are generated by the results of the associated equations, such as the squared residuals resulting from differences in the observed value, and the value anticipated, based on that model.

### The Line of Best Fit Equation:

The line of best fit determined from the least squares method has an equation that tells the story of the relationship between the data points. Line of best-fit equations may be determined by computer software models, which include a summary of outputs for analysis, where the coefficients and summary outputs explain the dependence of the variables being tested.

### Least Squares Regression Line:

If the data shows a leaner relationship between two variables, the line that best fits this linear relationship is known as a least-squares regression line, which minimizes the vertical distance from the data points to the regression line. The term “least squares” is used because it is the smallest sum of squares of errors, which is also called the "variance."

In regression analysis, dependent variables are illustrated on the vertical y-axis, while independent variables are illustrated on the horizontal x-axis. These designations will form the equation for the line of best fit, which is determined from the least-squares method.

In contrast to a linear problem, a non-linear least-squares problem has no closed solution and is generally solved by iteration. The discovery of the least-squares method is attributed to Carl Friedrich Gauss, who discovered the method in 1795.

An example of the least-squares method is an analyst who wishes to test the relationship between a company’s [stock returns](https://www.investopedia.com/articles/investing/011416/how-evaluate-stock-performance.asp), and the returns of the index for which the stock is a component. In this example, the analyst seeks to test the dependence of the stock returns on the index returns.

To achieve this, all of the returns are plotted on a chart. The index returns are then designated as the independent variable, and the stock returns are the dependent variable. The line of best fit provides the analyst with coefficients explaining the level of dependence.

The least-squares method is a mathematical technique that allows the analyst to determine the best way of fitting a curve on top of a chart of data points. It is widely used to make scatter plots easier to interpret and is associated with [regression analysis](https://www.investopedia.com/terms/r/regression.asp). These days, the least-squares method can be used as part of most statistical software programs.

The least-squares method is used in a wide variety of fields, including finance and investing. For financial analysts, the method can help to quantify the relationship between two or more variables such as a stock’s share price and its [earnings per share](https://www.investopedia.com/terms/e/eps.asp) (EPS). By performing this type of analysis, investors may attempt to forecast the future behavior of stock prices or other factors.

To illustrate, consider the case of an investment considering whether to invest in a gold mining company. The investor might wish to know how sensitive the company’s stock price is to changes in the market price of gold. To study this, the investor could use the least-squares method to trace the relationship between those two variables over time onto a scatter plot. This analysis could help the investor predict the degree to which the stock’s price would likely rise or fall for any given increase or decrease in the price of gold.

## The Difference Between Linear and Multiple Regression:

[Ordinary linear squares](https://www.investopedia.com/terms/l/least-squares-method.asp) (OLS) regression compares the response of a dependent variable given a change in some explanatory variables. However, a dependent variable is rarely explained by only one variable. In this case, an analyst uses multiple regression, which attempts to explain a dependent variable using more than one independent variable. Multiple regressions can be linear and nonlinear.

Multiple regressions are based on the assumption that there is a linear relationship between both the dependent and independent variables. It also assumes no major correlation between the independent variables.

A multiple regression considers the effect of more than one explanatory variable on some outcome of interest. It evaluates the relative effect of these explanatory, or independent, variables on the dependent variable when holding all the other variables in the model constant.

A dependent variable is rarely explained by only one variable. In such cases, an analyst uses multiple regression, which attempts to explain a dependent variable using more than one independent variable. The model, however, assumes that there are no major correlations between the independent variables.

In multiple linear regression, the model calculates the [line of best fit](https://www.investopedia.com/terms/l/line-of-best-fit.asp) that minimizes the variances of each of the variables included as it relates to the dependent variable. Because it fits a line, it is a linear model. There are also non-linear regression models involving multiple variables, such as logistic regression, quadratic regression, and probit models.

Any econometric model that looks at more than one variable may be a multiple. [Factor models](https://www.investopedia.com/terms/m/multifactor-model.asp) compare two or more factors to analyze relationships between variables and the resulting performance. The [Fama and French Three-Factor Mod](https://www.investopedia.com/terms/f/famaandfrenchthreefactormodel.asp) is such a model that expands on the [capital asset pricing model](https://www.investopedia.com/terms/c/capm.asp) (CAPM) by adding size risk and value risk factors to the market risk factor in CAPM (which is itself a regression model). By including these two additional factors, the model adjusts for this outperforming tendency, which is thought to make it a better tool for evaluating manager performance.

**Nonlinear Regression:**

Nonlinear regression is a form of regression analysis in which data is fit to a model and then expressed as a mathematical function. Simple linear [regression](https://www.investopedia.com/terms/r/regression.asp) relates two variables (X and Y) with a straight line (y = mx + b), while nonlinear regression relates the two variables in a nonlinear (curved) relationship.

**XG BOOST ALGORITHM:**

XG Boost (extreme Gradient Boosting) is a versatile machine learning algorithm used for regression tasks, where the goal is to predict a continuous numeric value rather than a categorical label. The XG Boost Regressor is a specific implementation of the XG Boost algorithm tailored for regression problems. Here's an explanation of how the XG Boost Regressor works:

Boosting Algorithm: Like other variants of the XG Boost algorithm, the XG Boost Regressor is a boosting algorithm. Boosting is an ensemble learning technique that combines multiple weak learners (typically decision trees) to create a strong learner. In the context of regression, XG Boost builds a series of decision trees sequentially, with each tree aiming to correct the errors made by the previous ones.

Gradient Boosting: XG Boost Regressor employs gradient boosting, a technique that minimizes a loss function (often mean squared error, but it can be customized) by iteratively adding decision trees. It optimizes the model's predictions by adjusting the weights of data points based on the errors made by the previous trees. This iterative process continues until a specified number of trees (boosting rounds) is reached or until the model reaches a predetermined level of accuracy.

Extreme Gradient Boosting: XG Boost Regressor extends traditional gradient boosting by adding regularization terms and using a more advanced optimization technique. This helps control overfitting, making the model more robust and capable of handling complex data.

Key Features of XG Boost Regressor:

Regularization: XG Boost Regressor includes L1 (Lasso) and L2 (Ridge) regularization terms in its objective function to prevent overfitting, just like in XG Boost for classification.

Gradient Descent Optimization: It employs gradient descent optimization to find the best parameters for its decision trees. This optimization technique enhances training efficiency and model accuracy.

**Cross-validation**: XG Boost Regressor supports cross-validation, allowing you to assess its performance effectively and fine-tune hyperparameters for better results.

**Handling Missing Data**: Similar to its classification counterpart, XG Boost Regressor can automatically handle missing values in the dataset, simplifying preprocessing.

**Parallel Processing**: It can leverage parallel processing, making it faster to train on multi-core machines.

**Tree Pruning**: XG Boost Regressor uses tree pruning techniques to remove unnecessary splits, resulting in more efficient and less complex trees.

Advantages of XG Boost Regressor:

**Excellent Predictive Performance**: XG Boost Regressor often outperforms other regression algorithms in terms of accuracy and generalization.

**Robustness**: It can handle noisy data and outliers effectively.

**Feature Importance**: Like the classification variant, XGBoost Regressor provides feature importance scores, helping you identify the most influential features in making predictions.

**Flexibility**: It is adaptable to various regression tasks and can handle both simple and complex relationships in the data.

**Community and Support**: XG Boost Regressor benefits from a strong and active community of users and contributors, ensuring it is well-maintained and well-documented.

**Use Cases**: XG Boost Regressor can be applied to a wide range of regression problems, including:

Predicting house prices based on features like square footage, number of bedrooms, and location.

Forecasting stock prices, commodity prices, or other financial metrics.

Estimating the demand for a product based on historical sales data and market factors.

Predicting the duration of a service call or delivery time in logistics.

Modeling and forecasting time series data.

In summary, the XG Boost Regressor is a powerful and versatile algorithm for regression tasks, known for its robustness, accuracy, and ability to handle complex data. It is widely used in data science and machine learning for various real-world applications where accurate numerical predictions are required.

## CHAPTER 8

##### CONCLUSION

**8.1 CONCLUSION**

The energy balance of wind energy is exceptionally certain. The energy consumed in the entire chain of wind plants is recuperated in a few normal functional months. The correlation of wind energy with ordinary innovations features the ecological benefits of wind energy. The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The Best accuracy on public test set is higher accuracy score is will be find out. This application can help out to find the Wind Power Prediction.

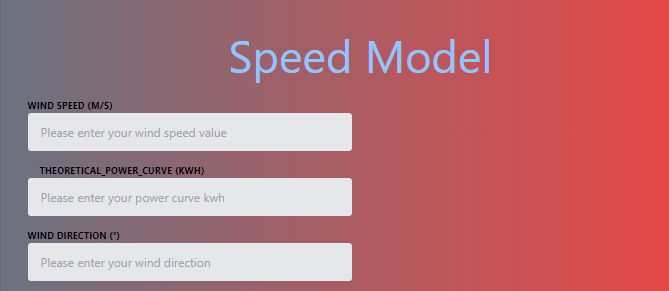
* 1. **FUTURE ENHANCEMENTS**
     1. Provide storage by using blockchain technology.
     2. Wind Power Generation Prediction to connect the AI Model
     3. To Optimize the Work to implement in AI Environment
     4. Enable access across all platforms .

The above mentioned are the future enhancements that can be done to make this project much more dynamic.

## CHAPTER 8 APPENDICES

##### 

##### GENERATIONAL MODEL



**SPEED MODEL**

## CHAPTER 9

##### REFERENCES

[1] A. Malekpour and A. Pahwa, “Stochastic networked microgrid energy management with correlated wind generators,” IEEE Trans. Power Syst., vol. 32, no. 5, pp. 3681–3693, Sep. 2017.

[2] C. Gu, C. Liu, J. Zhang, H. Huang, and X. Jia, “Green scheduling for cloud data centers using renewable resources,” in Proc. IEEE Conf. Comput. Commun. Workshops, 2015, pp. 354–359.

[3] Z. Liu et al., “Renewable and cooling aware workload management for sustainable data centers,” in Proc. ACM SIGMETRICS Joint Int. Conf. Meas. Model. Comput. Syst., 2012, pp. 175–186.

[4] L. Gu, J. Cai, D. Zeng, Y. Zhang, H. Jin, and W. Dai, “Energy efficient task allocation and energy scheduling in green energy powered edge computing,” Future Gener. Comput. Syst., vol. 95, pp. 89–99, 2019.

[5] C. Xu, K. Wang, P. Li, R. Xia, S. Guo, and M. Guo, “Renewable energy aware big data analytics in geo-distributed data centers with reinforcement learning,” IEEE Trans. Netw. Sci. Eng., vol. 7, no. 1, pp. 205–215, First Quarter 2020.

[6]I.DeCourchelle,T.Guérout,G.DaCosta,T.Monteil,andY.Labit,“Green energyefficientschedulingmanagement,”Simul.ModellingPract.Theory, vol. 93, pp. 208–232, 2019.

[7] H. Wang, J. Gong, Y. Zhuang, H. Shen, and J. Lach, “Healthedge: Task scheduling for edge computing with health emergency and human behavior consideration in smart homes,” in Proc. IEEE Int. Conf. Bigdata, 2017 , pp. 1213–1222.

[8] J. Gao, H. Wang, and H. Shen, “Smartly handling renewable energy instability in supporting a cloud datacenter,” in Proc. IEEE Int. Parallel Distrib. Process. Symp., 2020, pp. 769–778.

[9] J. Gao, H. Wang, and H. Shen, “Task failure prediction in cloud data centers using deep learning,” IEEE Trans. Serv. Comput., vol. 15, no. 3 , pp. 1411–1422, May/Jun. 2022.

[10] Sourcecode.Accessed:Jan.2023.[Online].Available:https://github.com/ pcl-projects/Renewable-Energy-Allocation-System

[11] G. Chen et al., “Energy-aware server provisioning and load dispatching for connection-intensive internet services,” in Proc. USENIX Symp. Netw. Syst. Des. Implementation, 2008, pp. 337–350.

[12] B. Heller et al., “ElasticTree: Saving energy in data center networks,” in Proc.USENIXSymp.Netw.Syst.Des.Implementation,2010,pp.249–264.

[13] Y. Lin and H. Shen, “EAFR: An energy-efficient adaptive file replication system in data-intensive clusters,” IEEE Trans. Parallel Distrib. Syst., vol. 28, no. 4, pp. 1017–1030, Apr. 2017.

[14] M. Dabbagh, B. Hamdaoui, M. Guizani, and A. Rayes, “Energy-efficient cloud resource management,” in Proc. IEEE Conf. Comput. Commun. Workshops, 2014, pp. 386–391.

[15] M. Xu and R. Buyya, “Managing renewable energy and carbon footprint in multi-cloud computing environments,” J. Parallel Distrib. Comput., vol. 135, pp. 191–202, 2020.

[16] Z. Liu, H. Yu, R. Liu, M. Wang, and C. Li, “Configuration optimization model for data-center-park-integrated energy systems under economic, reliability, and environmental considerations,” Energies, vol. 13, 2020 , Art. no. 448.

[17] S. Nayak, S. Panda, S. Das, and S. Pande, “An efficient renewable energybased scheduling algorithm for cloud computing,” in Proc. Int. Conf. Distrib. Comput. Internet Technol., 2021, pp. 81–97.

[18] Z. Wen, Y. Wang, and F. Liu, “StepConf: Slo-aware dynamic resource configuration for serverless function workflows,” in Proc. IEEE Conf. Comput. Commun., 2022, pp. 1868–1877.

[19] G. Safaryan, A. Jindal, M. Chadha, and M. Gerndt, “SLAM: SLO-aware memory optimization for serverless applications,” in Proc. IEEE 15th Int. Conf. Cloud Comput., 2022, pp. 30–39.

[20] S.K.Shukla,D.Ghosal,andM.K.Farrens,“Understandingandleveraging cluster heterogeneity for efficient execution of cloud services,” in Proc. IEEE 10th Int. Conf. Cloud Netw., 2021, pp. 56–64.

[21] C. Zhang, M. Yu, F. Yan, and W. Wang, “Enabling cost-effective, SLOaware machine learning inference serving on public cloud,” IEEE Trans. Cloud Comput., vol. 10, no. 3, pp. 1765–1779, Third Quarter 2022.

[22] S. K. Shukla and M. K. Farrens, “Leveraging network delay variability to improve QoE of latency critical services,” in Proc. IEEE Int. Conf. Netw., Archit. Storage, 2021, pp. 1 – 8.

[23] D. Alsadie, Z. Tari, E. J. Alzahrani, and A. Y. Zomaya, “Life: A predictive approach for VM placement in cloud environments,” in Proc. Int. Symp. Netw. Comput. Appl., 2017, pp. 1 – 8.

[24] R. K. Ramesh, H. Wang, H. Shen, and Z. Fan, “Machine learning for load balancing in cloud datacenters,” in Proc. IEEE/ACM 21st Int. Symp. Cluster, Cloud Internet Comput., 2021, pp. 186–195.

[25] E. Cortez, A. Bonde, A. Muzio, M. Russinovich, M. Fontoura, and R. Bianchini, “Resource central: Understanding and predicting workloads for improved resource management in large cloud platforms,” in Proc. Symp. Operating Syst. Princ., 2017, pp. 153–167.

[26] Y. Hua, Z. Zhao, R. Li, X. Chen, Z. Liu, and H. Zhang, “Deep learning with long short-term memory for time series prediction,” IEEE Commun. Mag., vol. 57, no. 6, pp. 114–119, Jun. 2019.

[27] Z. Chen, J. Hu, G. Min, A. Y. Zomaya, and T. El-Ghazawi, “Towards accurate prediction for high-dimensional and highly-variable cloud workloads with deep learning,” IEEE Trans. Parallel Distrib. Syst., vol. 31, no. 4 , pp. 923–934, Apr. 2020.

[28] J. Kumar, R. Goomer, and A. Singh, “Long short term memory recurrent neuralnetwork(LSTM-RNN)basedworkloadforecastingmodelforcloud datacenters,” Procedia Comput. Sci., vol. 125, pp. 676–682, 2018.

[29] W. Ding, F. Luo, C. Gu, H. Lu, and Q. Zhou, “Performance-to-power ratio aware resource consolidation framework based on reinforcement learning in cloud data centers,” IEEE Access, vol. 8, pp. 15472–15483, 2020.

[30] X. Zhou, K. Wang, W. Jia, and M. Guo, “Reinforcement learning-based adaptiveresourcemanagementofdifferentiatedservicesingeo-distributed data centers,” in Proc. Int. Symp. Qual. Serv., 2017, pp. 1 – 6.

[31] S. Telenyk, E. Zharikov, and O. Rolik, “Modeling of the data center resource management using reinforcement learning,” in Proc. Int. Sci.- Practical Conf. Problems Infocommunications Sci. Technol., 2018 , pp. 289–296.

[32] H. Shen and L. Chen, “Distributed autonomous virtual resource management in datacenters using finite-Markov decision process,” IEEE/ACM Trans. Netw., vol. 25, no. 6, pp. 3836–3849, Dec. 2017.

[33] R. Komp, Practical Photovoltaics: Electricity From Solar Cell, 3rd ed. Ann Arbor, MI, USA: Aatec publications, 2003.

[34] P. Gipe, Wind Power, Revised Edition: Renewable Energy for Home, Farm, and Business. Chelsea, VT, USA: Chelsea Green, 2004.

[35] H. Shen and L. Chen, “CompVM: A complementary VM allocation mechanism for cloud systems,” IEEE/ACM Trans. Netw., vol. 26, no. 3 , pp. 1348–1361, Jun. 2018.