Solar Radiation Analysis & Prediction Using Machine Learning Algorithms

Mohammad Hasan[1], Sabrina Sharmin[2] Mohammad Mizanur Rahman[3]
[1]Psotgraduate Researcher
[2]Assistant Professor
[3]Associate Professor
[1,2]Department of Computer Science and Engineering, Jahangirnagar University, Dhaka-1342, Bangladesh
[3]Department of Urban and Regional Planning, Jahangirnagar University, Dhaka-1342, Bangladesh

Corresponding author: Sabrina Sharmin (E-mail: sabrina.329@juniv.edu)

ABSTRACT: The solar radiation analysis and prediction model aims to develop and implement machine learning algorithms for analyzing and predicting solar radiation, with the goal of enhancing the efficiency and integration of solar power systems. Accurate solar radiation prediction is crucial for agriculture, weather forecasting, health awareness, effective energy management, grid stability and optimizing the utilization of solar energy resources. This project begins with an extensive collection of solar radiation data from Kaggle.com, including weather stations, satellite imagery, and ground-based sensors. The dataset is preprocessed and feature engineering techniques are applied to extract relevant meteorological and environmental parameters. Several machine learning techniques including Linear Regression Models, Decision Trees, Support Vector Regressors, KNN, XGBoost Regressor, Random Forests, and Gradient Boosting Regressor, are trained and evaluated using the prepared dataset. Then performance of these algorithms is assimilated and compared. This project will showcase the effectiveness of machine learning algorithms by predicting accurate solar radiation.

KEYWORDS: Decision Tree, Gradient Boosting Algorithm, Insolation, KNN, Linear Regression, Machine Learning Algorithms, Prediction, Random Forest, Solar Radiation, Training & Testing, XGBoost.

1. INTRODUCTION

Solar radiation, known by various terms such as solar insolation, solar resource, solar energy, or sunlight, encompasses the electromagnetic radiation emitted by the sun. [1] This radiant energy serves as the fundamental source of energy for Earth and plays a crucial role in numerous natural processes. It contributes significantly to the surface radiation balance, photosynthesis in vegetation, hydrological cycles, ecological equilibrium, and the occurrence of climate and weather extremes. Multiple technologies have been developed to harness solar radiation and convert it into effective means of energy, such as electricity and heat. Nevertheless, the practical viability and economic feasibility of these technologies at specific geographical locations depend on the availability of solar resources. The sheer number of solar insolation that is captured the Earth's exterior is influenced by various factors, including the atmosphere, biosphere, and hydrosphere.

Solar radiation has a substantial impact on global temperature, as even minor fluctuations in the sun's energy output can have significant repercussions on Earth's climate. Changes in solar energy can alter global mean sea level, global mean temperatures, and the occurrence of extreme climate events. As a result, [1] precise measurements and analyses of the spatial and temporal variability of solar radiation are essential for studies into the use of solar energy, the manufacture of construction materials, and the understanding of extreme weather and climatic events. The availability of solar radiation varies throughout the year and can be found to some level in every region of the world. The understanding and utilization of solar radiation are essential for a wide range of applications, making it a critical area of study and exploration. The following variables influence the extent to which solar insolation reaches any given location on Earth's outermost layer:

- Active Period
- Sense of Direction
- Duration of Sunlight
- Local Weather
- Angle of Incidence of Sun Rays
- Transparency of the Atmosphere

The interaction between the sun and our Earth's exterior results in varying angles of sunlight, influenced by the spherical shape of the Earth. As the sun's position changes throughout the day, the angle of sunlight can range from 0° (near the horizon) to 90° (directly overhead). Maximum energy absorption by the Earth's surface occurs when sunlight is perpendicular to it. However, as sunlight travels through the atmosphere for longer durations, it becomes more dispersed and obscured. The unique spherical nature of the Earth also means that polar regions never experience a high position of the sun. Additionally, due to the Earth's tilted axis of rotation, these polar regions go through extended periods without any sunlight during certain times of the year.

The Earth orbits around the sun in an elliptical pattern, drawing nearer to it at a certain period of the year. The proportion of solar insolation which the exterior of the Earth receives increases somewhat as the sun gets closer to the earth. when the northern hemisphere is experiencing winter and summertime in the southern hemisphere, the Earth is less distant from the sun. However, the presence of vast oceans tempers the increasingly hot summers and chilly winters that one could anticipate in the southern hemisphere.

The planet's rotary axis of rotation's 23.5° tilt provides an increasingly important part in influencing exactly how much sun rays hit the planet exterior at any given geographic location. The tilting of the Earth's axis has a significant impact on the length of days in different hemispheres throughout the year. In the northern hemisphere, tilting results in longer days from the vernal equinox in the spring to the autumnal equinox, while in the southern hemisphere, longer days occur during the remaining six months. During the equinoxes, which typically fall between March 23rd and September 22nd each year, the duration of day and night is precisely 12 hours. At the beginning of the day and late afternoon, the sun appears lower in the sky, while during midday, it reaches its highest point, allowing its radiation to travel a greater distance through the atmosphere. Consequently, on a clear day, a solar collector receives the highest amount of solar energy around solar noon. As sunlight traverses the atmosphere, it undergoes absorption, dispersion, and reflection, primarily influenced by various factors such as:

- Air molecules
- Clouds
- Volcanoes
- Dusts
- Pollutants
- Water Vapor
- Forest fires

These elements contribute to the modification of solar radiation. Diffuse sun radiation refers to the scattered sunlight, while direct beam solar insolation denotes the sunlight that directly captured by the Earth's exterior. Global solar power encompasses both diffuse and direct sun radiation. Atmospheric conditions can lead to a reduction of up to 10% in direct beam radiation on clear, dry days, and up to 100% on heavily clouded days. Machine learning models utilized in solar radiation analysis heavily rely on specific climatic variables and meteorological factors, including sunshine duration, visibility, and land surface temperature. These factors play a crucial role in accurately predicting solar radiation levels. Additionally, studies have demonstrated the correlation between solar radiation and extreme climate events by analyzing trends between extreme land surface heat and solar power levels.

1.1 OBJECTIVES

The objectives of Solar radiation analysis and prediction using machine learning algorithms are as follows:

a. **Weather forecasting:** Solar radiation prediction is integral to accurate weather forecasting. By utilizing meteorological data including cloud cover, humidity, and atmospheric conditions, solar radiation model can estimate the quantity of sunlight captured by the Earth's exterior. Precise solar radiation predictions enhance the accuracy of weather forecasts, benefiting diverse sectors such as agriculture, aviation, and disaster management. These predictions enable better agricultural planning, optimize flight operations, and support effective decision-making in mitigating the impact of severe weather events. Solar radiation prediction is closely tied to weather forecasting because the quantity of sunlight captured by the Earth's exterior is a key factor in determining the weather. Solar radiation models use meteorological data, such as cloud cover, humidity, and atmospheric conditions, to estimate the sheer numbers of sunlight reaching the Earth's exterior. By using the machine learning model, we will able to study the Earth's atmosphere. Solar radiation model can be used to study the effects of different atmospheric conditions on the quantity of sunlight reaching the Earth's exterior.

b. **Environmental research:** Solar radiation analysis and prediction model will play a significant role in environmental research and the study of climate change. By providing valuable insights into the Earth's energy balance, temperature patterns, and climate dynamics, it enables a deeper understanding of long-term climate trends. This model will help researchers assessing the impact of climate change on ecosystems and develop effective strategies for mitigating its effects. By analyzing solar radiation data, we can aid in improving understanding of the Earth's climate system and foster sustainable environmental practices. For example, solar radiation data can be used to track the amount of energy that is absorbed by the planet's atmosphere and oceans. The Earth's energy budget, which is an illustration of the balance between incoming and exiting energy, can then be calculated using this data. The planet's energy budget is a key factor in determining the Earth's climate, so tracking solar radiation levels is essential for understanding how climate change is affecting the planet. Solar radiation data can also be used to track changes in temperature patterns. For example, scientists can use solar radiation data to track the rise in global temperatures that has been observed over the past century. The consequences of climate change on ecosystems can then be evaluated using the mentioned information, such as the melting of glaciers and the spread of diseases. Finally, solar radiation model can be used to track climate dynamics. Climate dynamics is the study of the physical processes that drive the Earth's climate system.

c. **Energy production and planning:** Solar radiation model can serve as the fundamental energy source for solar power generation. Precise prediction and planning of solar radiation enable us to optimize the placement and efficiency of solar panels, thereby maximizing energy production and ensuring a reliable supply of clean energy. By leveraging accurate solar radiation predictions, we can enhance energy planning strategies, improve the integration of solar power into the grid, and promote sustainable energy practices. Solar radiation is the fundamental source of energy for solar power generation. It is a renewable and abundant source of energy, but its availability varies depending on the time of day, season, and weather conditions. Accurately predicting and planning solar radiation can help us to optimize the placement and efficiency of solar panels, maximizing energy production and ensuring a stable supply of clean energy. Overall, accurate solar radiation prediction is a valuable tool that can be used to improve energy production and planning. By using this tool, we can maximize the total amount of energy that is generated from solar power, ensure a stable supply of clean energy, and reduce the costs of solar power production.

d. **Grid management:** Solar power, being an intermittent energy source reliant on sunlight availability, benefits from accurate solar radiation predictions. These predictions aid grid operators in effectively managing the integration of solar power into the electrical grid. By anticipating fluctuations in solar generation, operators can balance it with other energy sources, ensuring a dependable and stable power supply. Predicting solar radiation supports efficient grid management, enabling optimal utilization of solar energy while maintaining grid reliability and stability. Solar power is an intermittent energy source, meaning that its availability depends on the weather. This can pose challenges for grid operators, who need to ensure a reliable and stable power supply at all times. Predicting solar radiation can help grid operators manage the integration of solar power into the grid more effectively. By knowing how much solar radiation is expected, grid operators can anticipate fluctuations in solar generation and balance it with other sources of energy. This can help to prevent blackouts and brownouts, and ensure that the grid remains stable. For example, if grid operators know that there is going to be a decrease in solar radiation, they can increase the generation from other sources, such as fossil fuels or hydroelectric power. This will help to ensure that there is enough electricity to meet demand. Predicting solar radiation can also help grid operators to optimize the use of solar power. By knowing when and where solar radiation is most likely to be high, grid operators can dispatch solar power plants to maximize their output. This can lead to decrease the cost of electricity and improve the efficiency of the grid. So, predicting solar radiation is a valuable tool for grid operators, businesses, and individuals. By knowing how much solar radiation is expected, these stakeholders can make better decisions about how to use solar power and ensure a reliable and stable power supply.

2. LITERATURE REVIEW

Solar energy or solar power is a renewable resource that can be used to balance climate and weather extremes, vegetation photosynthesis, hydrological cycles, electricity generation, ecological balance, and surface radiation. It can also be used to heat water and power other equipment. The efficient and economical use of this resource depends on the ability to estimate solar insolation accurately. The analysis and forecasting of solar insolation can be done using a variety of techniques, such as empirical models, time series models. But these physical models and statistical algorithms are frequently used in traditional methods for estimating solar power. These approaches, however, fall short in representing the intricate and nonlinear interactions present in solar radiation. Simple mathematical correlations between solar radiation and other meteorological factors like temperature, humidity, and cloud cover serve as the foundation for empirical models. To forecast future solar radiation levels, time series models are used. In recent years, machine learning approaches have emerged as a possible replacement for precise solar radiation monitoring and prediction. [3] Artificial neural networks are one type of machine learning approach that can learn from historical data and find trends that can be altered to predict future solar radiation levels.

2.1 RELATED WORK

Ali EtemGürel demonstrated that [1] machine learning methods are useful for predicting solar radiation. For the purpose of forecasting solar radiation in 2023, he published an article in https://www.sciencedirect.com/science/article and assessed the effectiveness of various machine learning techniques. With an average inaccuracy of 5%, his team

discovered that artificial neural networks outperformed all other systems.
Mr. Gabriel de Freitas Viscondi, demonstrates a comparison of artificial intelligence [2] algorithms for estimating sun radiation https://www.mdpi.com/1996-1073/14/18/5657. In this study, the efficacy of six machine learning algorithms: support vector regression, artificial neural networks, decision trees, random forests, gradient boosting machines, and extreme learning machines—for estimating solar radiation was compared. With an average error of 5.2%, the results demonstrated that artificial neural networks performed the best.
An analysis of machine learning methods [3] for predicting solar radiation Solar radiation prediction using machine learning techniques: a review, available at www.researchgate.net/publication/337043958_Solar_Radiation_Prediction_Using_Machine_Learning_Techniques_A_Review. An overview of the many machine learning methods that have been applied to forecast solar radiation is given in this review study. The paper addresses the benefits and drawbacks of various strategies and offers suggestions for further study.
LieXing Huang used various machine learning techniques, solar radiation prediction and consequences for extreme climatic events www.frontiersin.org/articles/10.3389/feart.2021.596860/full [4] The effectiveness of several machine learning methods for predicting solar radiation during extreme climate events was examined in this study. The outcomes demonstrated that support vector machines and artificial neural networks performed best, with average errors of 5.8% and 6.2%, respectively.
The area of forecasting solar radiation still faces certain difficulties, nevertheless. The restricted availability of historical data is one difficulty. The necessity to create machine learning algorithms that can be utilized to forecast solar radiation levels in various geographic regions and during various climatic circumstances is another difficulty. Despite these difficulties, research in the subject of solar radiation prediction is expanding quickly. The accuracy of solar radiation forecast is anticipated to increase as more data becomes accessible and machine learning algorithms advance.

3. METHODOLOGY

3.1 The theoretical material from the entire research is mostly covered in this chapter. Any individual who reads this thesis will be changed by it. To make things simpler to understand, we have compiled some fundamental details. In order to predict compensation, machine learning is essential. We have summarized regression methods since We have dealt with continuous values to predict solar radiation. The machine learning regression techniques that We utilized for this investigation are also explained in detail. We have gone through the data visualization techniques used in this study.We have broken down the methodology for implementing the solar radiation prediction model into several steps. The following thorough technique is used to implement the machine learning model:

3.1.1 Requirement Analysis: First, we have analyzed the requirements and gathered the expectations of stakeholders to ensure that the model meets the desired objectives and provides a comprehensive and user-friendly experience. This includes understanding the specific needs of the stakeholders, such as the desired accuracy of the predictions, the frequency of predictions, the format of the outputs, specific features that the model should be able to predict, as well as the accuracy and latency requirements.

3.1.2 DatasetCollection: We have collected my dataset for solar radiation prediction model from Kaggle.com. This dataset is collected using sensor traces and contains weather data recorded at regular intervals, typically every 5 minutes. The weather attributes include date, time, sunlight duration, [8] atmospheric pressure, wind direction and speed, solar radiation, humidity and temperature. Download and access of the dataset is openly available.

3.1.3 Dataset Reading: We have read and loaded the collected dataset for preparing it for training and testing machine learning models. The dataset is typically stored in csv file format, and the reading process involves extracting the necessary information and organizing it for further analysis.

3.1.4 Dataset Preparation: At this stage, we have performed essential data preprocessing steps to clean and transform the dataset for optimal utilization by machine learning models. This involves handling missing values, scaling independent variables, encoding continuous variables, handling outliers, and selecting relevant features. These preprocessing steps mentioned above, ensure that the dataset is suitable for training accurate prediction models and testing.

3.1.5 Dataset Segmentation: Dividing the preprocessed dataset into training and testing sets is mandatory. The training set will be manipulated to train the machine learning algorithms, while the testing set will be manipulated to evaluate their behavior. It can vary based on the dataset's size and characteristics. Here, we have divided the dataset into [9] training and testing dataset where the splitting ratio is 70% and 30%.

3.1.6 Data Scaling& Reshaping: Applying data scaling and reshaping techniques to improve the performance and training process of [10]models is a fundamental step in the process of building a machine learning model. Scaling ensures that features have comparable scales, while reshaping ensures that the data is in a convenient format for choosing machine learning algorithms. We have applied both data scaling and reshaping for my project.

3.1.7 Model Training: We have trained the machine learning model using the training dataset. Model training involves feeding the data to the models, allowing them to learn patterns and relationships within the dataset. Techniques such as cross-validation, regularization, and handling imbalanced data can be applied during training to enhance performance and mitigate overfitting.

3.1.8 Model Evaluation: Then, we have evaluated the trained models' performance and quality using various evaluation metrics and techniques. Assess how well the models generalize to unseen data and whether they meet the desired objectives. This evaluation process helps me in selecting the best-performing machine learning model.

3.2 METHODICAL APPROACHES

Machine Learning (ML) is a fascinating discipline within computer science that empowers computer systems to comprehend and interpret data much like humans do. At its core, machine learning involves employing algorithms or methods to unveil patterns and insights from raw data, thereby enabling systems to make informed decisions and predictions. By harnessing the power of artificial intelligence, machine learning has revolutionized the way computers perceive and process information, leading to advancements in various fields and unlocking new opportunities for solving complex problems.

3.3 WORKFLOW STEPS

Workflow is the process of preprocessing data and organizing it systematically so that one can quickly grasp all the techniques and have the option of analyzing the data to identify any constraints. There are two main sections to this study project. One half involves analyzing the training dataset's situation, while the other involves predicting solar radiation for the testing dataset using five independent variables: temperature, pressure, humidity, wind direction (degrees), and speed. We've broken out the method for data analysis and visualization for my job step by step in the figure below:

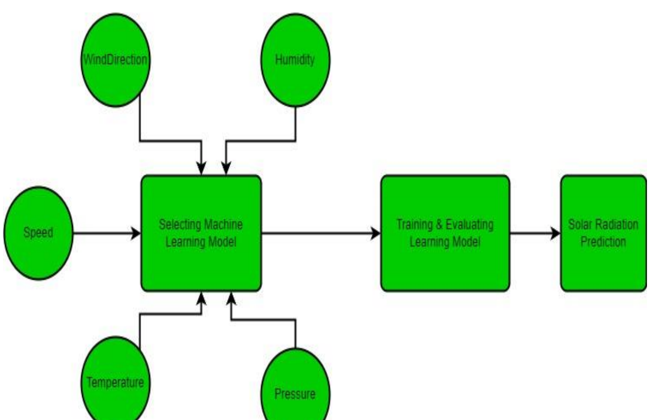

FIGURE 1: The Independent (Left) & Dependent variables (Right) of the Dataset

Here, the following diagram depicts the independent variables (temperature, pressure, humidity, wind direction, and speed) on the left portion of the diagram and the dependent variable (solar radiation) on the right portion.
Wehave trained & tested the machine learning model by manipulating the dataset. The following flowchart shows the sequential machine learning model building technique:

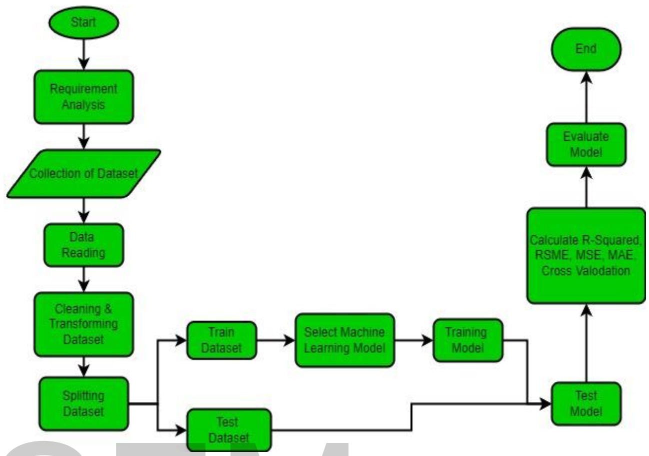

FIGURE 2: The Model Building Technique

At the end, I have visualized the necessary data. The following figure demonstrates the data visualization strategies that is used in this model in this model:

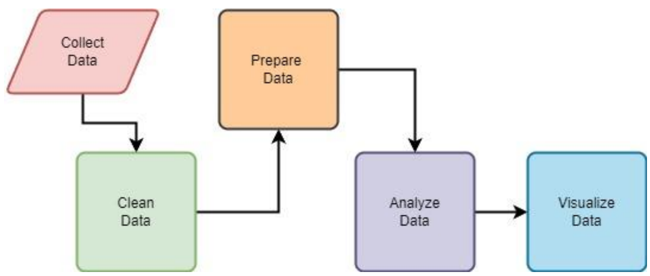

FIGURE 3: Data Visualization

4. RESULT ANALYSIS & PERFORMANCE EVALUATION

In this chapter, the findings and discussions are concisely described. [11] This section shows the results of the data analysis for the model that predicts solar radiation, as well as the main conclusions of the developed machine learning models, variable correlation depiction, test results, anticipated information visualizations etc.

4.1 Data Analysis Result: We have found the NASA solar forecasting dataset from Kaggle.com which is available for free to download. The dataset has several columns (UNIXTime, Date, Time, Radiation, Temperature, Pressure, Humidity, WindDirection, Speed, TimeSunRise, TimeSunSet) and some specific features which is depicted below:

	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

FIGURE4: The Dataset

The dataset has Eleven columns. The columns and their respective data types are shown in the figure 4.2 and 4.3 using columns function:
`

```
sun.columns

Index(['UNIXTime', 'Data', 'Time', 'Radiation', 'Temperature', 'Pressure',
       'Humidity', 'WindDirection(Degrees)', 'Speed', 'TimeSunRise',
       'TimeSunSet'],
      dtype='object')
```


FIGURE 5: The Columns of the Dataset

```
sun.dtypes

UNIXTime                    int64
Data                       object
Time                       object
Radiation                 float64
Temperature                 int64
Pressure                  float64
Humidity                    int64
WindDirection(Degrees)    float64
Speed                     float64
TimeSunRise                object
TimeSunSet                 object
dtype: object

[ ] sun.shape

(32686, 11)
```

FIGURE 6: The Datatypes of the Columns

The dataset possesses no null values which is depicted in the following hitmap:

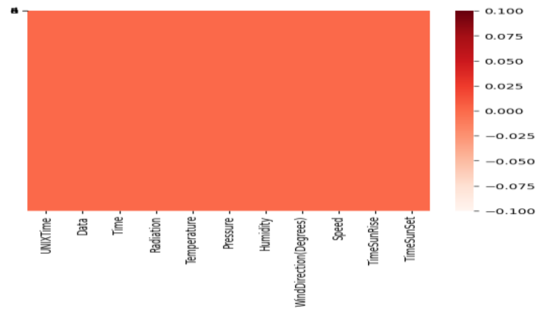

FIGURE 7: The Hitmap

As the dataset encompasses eleven columns, we have considered only five of them for further calculation. Rest of the columns are dropped as they aren't compatible with the research. The dropped columns (UNIXTime, Date, Time, TimeSunRise, TimeSunSet) are shown in the figure below:

```
[ ] x = sun.drop(['Data', 'Time', 'Radiation', 'TimeSunRise', 'TimeSunSet'], axis = 1)
    y = sun['Radiation']

[ ] x.shape

    (32686, 6)

[ ] y.shape

    (32686,)
```

FIGURE 8: Dropping Irrelevant Columns from the Dataset

The remaining five columns are considered as five independent variables (Temperature, Pressure, Humidity, WindDirection, Speed) for predicting Solar Radiation. Here, we considered Radiation column as a dependent variable. Then, we showed the density distributions of five independent variable.

The figure mentioned below shows 'Temperature' distribution in the dataset:

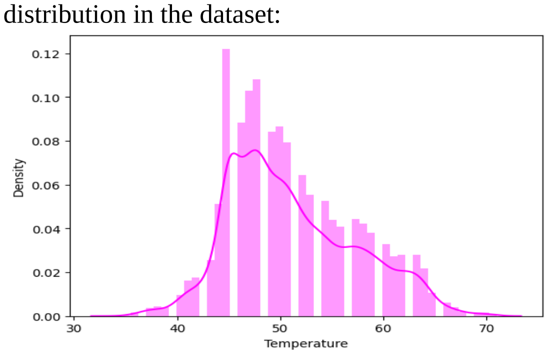

FIGURE 9: The Temperature Distribution

The next figure shows 'Humidity' distribution in the dataset:

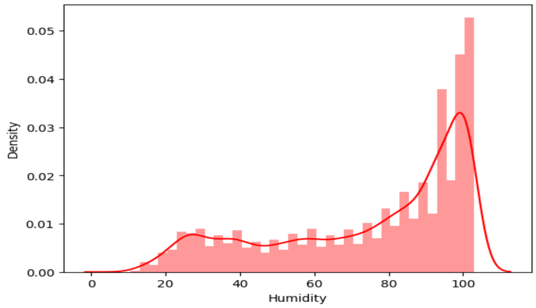

FIGURE 10: The Humidity Distribution

The figure mentioned below shows ‘Pressure’ distribution in the dataset:

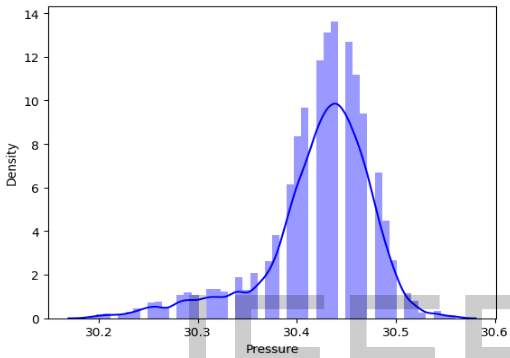

FIGURE 11: The Pressure Distribution

The figure mentioned below shows ‘Speed’ distribution in the dataset:

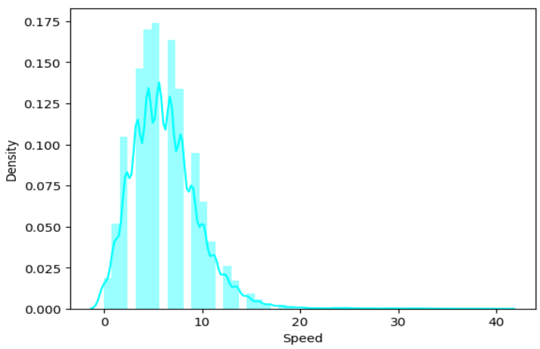

FIGURE 12: The Speed Distribution

The figure mentioned below shows ‘WindDirection’ distribution in the dataset:

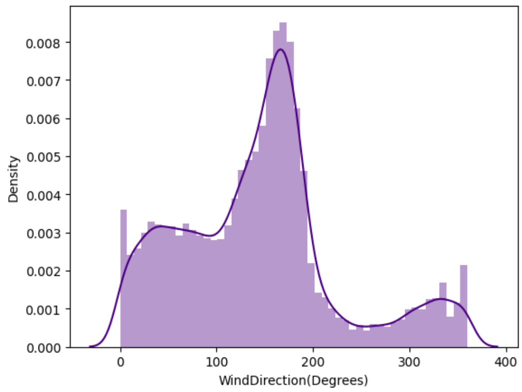

FIGURE 13: The Wind Direction Distribution

Then, we performed data scaling operation in order to enhance balance among variables and to improve performance of the model. The following figure is demonstrating data scaling operation:

```
[19] features = ['Temperature', 'Pressure', 'Humidity', 'WindDirection(Degrees)', 'Speed']
     label = ['Radiation']
     x = sun[features]
     y = sun[label]

[20] from sklearn.preprocessing import MinMaxScaler
     scaler = MinMaxScaler()
     x = scaler.fit_transform(x)
```

FIGURE 14: Data Scaling

Then we calculated each independent variable’s feature importance value in predicting solar radiation. The figure mentioned below shows feature importance graph:

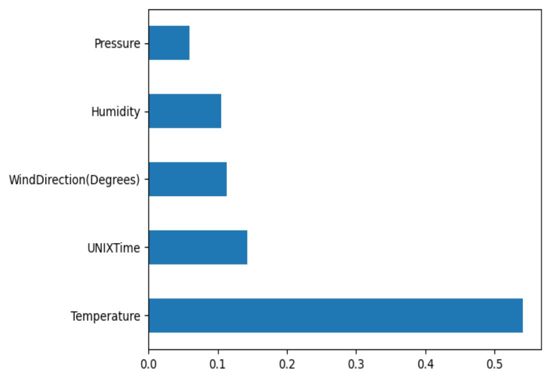

FIGURE 15: Feature Importance Graph

Then we prepared pairwise independent variable comparison graph which is shown below:

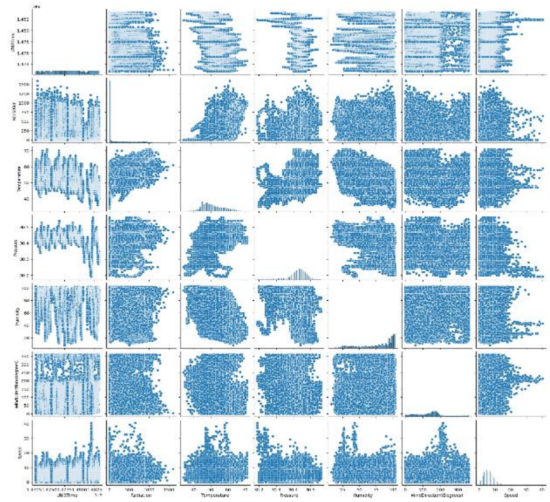

FIGURE 16: Pairwise Independent Variable Comparison Graph

We have prepared the correlation matrix to observe interdependencies and interrelation among variables which is demonstrated below:

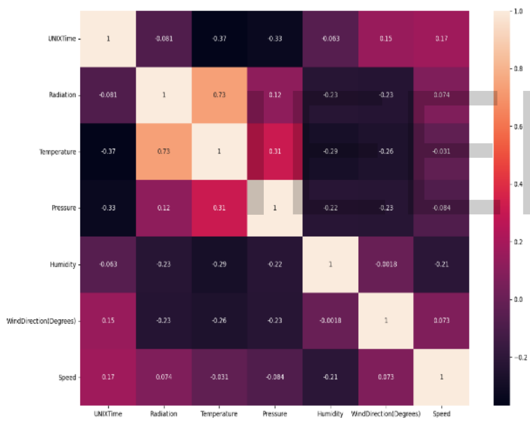

FIGURE 17: Correlation Matrix

In Finally, we performed result comparison among machine learning models which is given below:

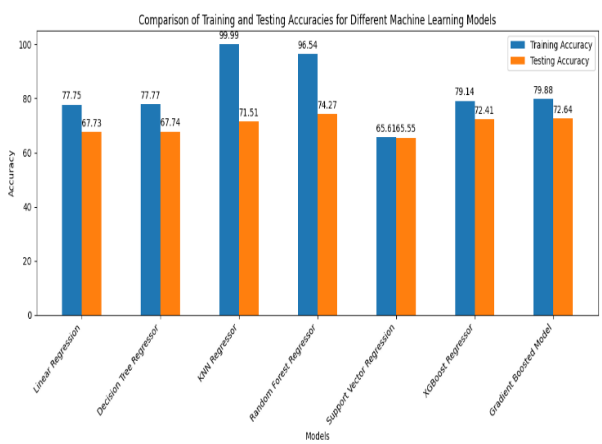

FIGURE 18: Machine Learning ModelsComparison Graph

First, we used the Linear Regression model to predict solar radiation by using five independent variables (Humidity), (Speed), (WindDirection), (Temperature), (Pressure) and by using LinearRegression() function. 70 % data from the dataset was selected for training purpose and 30% data for testing. The regression score found in training and testing is 77.75% and 67.73% respectively.

Secondly, we used the Decision Tree Regression model to predict solar radiation by using five independent variables and DecisionTreeRegressor() function. 70 % and 30% data are used for training and testing purposes respectively. The same train-test ratio is maintained for the rest of the models. The regression score found in training and testing is 77.77% and 67.74% respectively.

Then, we used the K-Nearest Neighbor Regression model to predict solar radiation by using KNeighborsRegressor() function. The regression score found in training and testing is 99.99% and 71.51% respectively.

After that, we used the [12] Random Forest Regression model to predict solar radiation by using RandomForestRegressor() function. The regression score found in training and testing is 96.54% and 74.27% respectively.

Then, we used the [13] Support Vector Machine Regression (SVM) model by using SVR() function. The regression score found in training and testing is 65.61% and 65.55% respectively.
After that, we used the XGBoost Regression model by using XGBRegressor() function. The regression score found in training and testing is 79.14% and 72.41% respectively.

Finally, we used the Gradient Boosting Regression model to predict solar radiation by using GradientBoostingRegressor() function. The regression score found in training and testing is 79.88% and 72.64% respectively.

4.2 RESULT SUMMERY:

In the table mentioned above, we can see that Support Vector Machine Regression (SVM) has a training accuracy of 65.61% which is the lowest among all machine learning models. On the other hand, K-Nearest Neighbor Regressor holds 99.99% accuracy, being the highest among the training set. If we consider testing accuracy, we can clearly see SVM Regressor has the lowest testing accuracy of 65.55%. Random Forest Regressor holds the highest testing accuracy of 74.27%.

5. CONCLUSION

Solar radiation is a sustainable and promising energy resource for the next generation. Machine learning algorithms offer a robust approach to predict solar radiation in specific timeframes and geographical regions. This report presents a methodology [5] that aims to achieve effective and precise implementation [2] of solar radiation prediction. By leveraging machine learning models, [6] the proposed approach holds [14] the potential to significantly enhance the accuracy and reliability of solar radiation forecasts. The proposed model underwent rigorous evaluation using a comprehensive dataset encompassing historical solar radiation data. The evaluation results unequivocally demonstrate that this model attains a considerable improvement in the accuracy of solar radiation prediction. The following comprehensive approach delineates the proposed methodology for attaining a highly effective and precise implementation of solar radiation prediction. By manipulating [6] the power of machine learning techniques, this approach strives to fulfill the prescribed objectives for accurate solar radiation prediction. The integration of machine learning approaches stands as a pivotal concept in significantly enhancing the accuracy of the proposed methodology. Machine learning algorithms possess the ability to gather knowledge from historical data and discern patterns which can subsequently be utilized to predict future solar radiation values. This approach diverges from traditional methods of solar radiation prediction, which often rely on physical models prone to inaccuracy. To assess the accuracy of predictions, the root mean squared error (RMSE), Mean Squared error (MSE), Mean Absolute Error (MAE) and R-squared values are calculated. Lower RMSE, MSE, MAE values indicate a more precise prediction. The implementation of the machine learning approach yielded notably reduced RMSE, MSE and MAE signifying substantial enhancements in the accuracy of solar radiation prediction.

The proposed research represents a highly promising and innovative approach to solar radiation prediction. It surpasses traditional methods in terms of accuracy and exhibits the versatility to predict solar radiation across varying time scales. Consequently, this model can serve as a valuable tool with a multitude of applications, such as solar power forecasting and climate modeling.

FUTURE WORK

While this project successfully explores the application of machine learning algorithms for solar radiation prediction, there are still a number of possible areas for further research and development. These directions can enhance the accuracy, robustness, and applicability of the predictive models. Following are a few potential future work plans:

We will expand the dataset. The more data we have, the better our model will be able to predict solar radiation. We will expand the dataset by collecting data from more locations and collecting data for a longer period of time.

In this project, we focused on traditional machine learning algorithms [7] such as Linear Regression, Gradient Boosting, K-Nearest Neighbor, SVM, Random Forests etc. However, exploring advanced techniques like [8] deep learning models, such as neural networks or convolutional neural networks (CNNs), can potentially capture complex patterns and nonlinear relationships in solar radiation prediction.

We will try to [9] incorporate real-time data for the project in future. In my current model, we used historical data to make predictions. However, we will also try to incorporate real-time data into my model. This would allow our model to make more accurate predictions in the present day.

In order to address potential limitations of limited data availability, we will implement data augmentation. we will also try to fine-tune the hyperparameters of the machine learning models which will lead to improved performance. Advanced hyperparameter optimization techniques can automate the search process and help identify the best set of hyperparameters for the models.

We will build a web application using the best machine learning. This system will provide its users, a user-friendly interface, allow the users to predict solar radiation in advance which will lead to a better decision making. Most importantly, they will get it without any cost.

REFERENCES

[1] [Online]. Available: https://assets.researchsquare.com/files/rs-3152311/v1_covered_c78960b4-c1f3-41f2-823d-ec8d5f17dea9.pdf?c=1689218373.

[2] J. K. M. W. L. F. C. Z. Z. Z. Liexing Huang, "Solar Radiation Prediction Using Different Machine Learning Algorithms and Implications for Extreme Climate Events," *Frontiers in Earth Science*, vol. 9, 2021.

[3] [Online]. Available: https://link.springer.com/book/10.1007/978-3-031-26254-8.

[4] Ü. A. H. B. A. E. G. Y. Ali Etem Gürel, "A state of art review on estimation of solar radiation with various models," *https://www.sciencedirect.com/science/article/pii/S2405844023003742#sec1*, pp. 1-20, 2023.

[5] P. D. S. N. A.-S. Mr. Gabriel de Freitas Viscondi, "Solar Irradiance Prediction with Machine Learning Algorithms: A Brazilian Case Study on Photovoltaic Electricity Generation," *MDPI Journals*, vol. 14, no. 18, pp. 1-30, 2021.

[6] S. C. J. P. A. Edgar Dario Obando, "Solar Radiation Prediction Using Machine Learning Techniques: A

Review," *https://www.researchgate.net/publication/337043958_Solar_Radiation_Prediction_Using_Machine_Learning_Techniques_A_Review,* vol. 14, no. 4, pp. 2-27, 2019.

[7] J. K. L. H. Mengxue Wan, "Solar Radiation Prediction Using Different Machine Learning Algorithms and Implications for Extreme Climate Events," *Frontiers in Earth Science,* vol. 9, pp. 1-20, 2021.

[8] D. O. F. J. C. C. P. N. V. Vinícius Leonardo Gadioli da Silva, "An Approach to Solar Radiation Prediction Using ARX and ARMAX Models," *Frontiers in Energy Research,* vol. 10, 2022.

[9] A. M. A. R. A. W. S. S. C. U. P. S. A. S. Divyansh Divyansh, "Stock Closing Price Prediction of Indian Tech Companies using Deep Neural Networks," *Research Square,* 2023.

[10] [Online]. Available: https://dokumen.pub/advanced-computational-techniques-for-renewable-energy-systems-3031212150-9783031212154.html.

[11] [Online]. Available: https://webthesis.biblio.polito.it/25400/1/tesi.pdf.

[12] H.-B. L. W. W. D.-T. X. Ji-Long Chen, "Estimation of monthly solar radiation from measured temperatures using support vector machines – A case study," *Researchgate,* 2011.

[13] [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0275022.

[14] Y. W. Muhammad Zarar, "Early Stage Diabetes Prediction by Approach Using Machine Learning Techniques," *Research Square,* 2023.

[15] J. B. H.-A. T. F. G. A.-S. View ORCID ProfileMichael Owusu-Adjei, "A systematic review of prediction accuracy as an evaluation measure for determining machine learning model performance in healthcare systems," *medRxiv,* 2023.

[16] D. J. B. Priyanka .R, "A Semi-Supervised Learning Approach for Tackling Twitter Spam Drift," *International Journal for Research in Applied Science and Engineering Technology (IJRASET),* 2022.

[17] H. Tierney, "Forecasting and tracking real-time data revisions in inflation persistence," *Munich Personal RePEc Archive,* 2011.