SOLAR RADIATION PREDICTION USING MACHINE LEARNING

M.RAM KUMAR [1]
ANDE NAVITHA[2], M M JAGAN[2], MATTIGALLA NANDHA KUMAR[2],
AYYAPPARAJU RENUSREE[2], DONAPATI VAMSHIDHARA REDDY[2]

[1] *Assistant Professor, Department of Computer Science & Information Technology, Siddharth Institute of Engineering & Technology, Andhra Pradesh, India*
[2] *Research Scholar, Department of Computer Science & Information Technology, Siddharth Institute of Engineering & Technology, Andhra Pradesh, India*

ABSTRACT

Solar energy is an abundant and sustainable source of power, making it a critical component of the global transition towards clean energy solutions. Accurate prediction of solar radiation is essential for optimizing the performance of solar energy systems, such as photovoltaic panels and solar thermal plants. This study presents a comprehensive exploration of machilearning techniques for the prediction of solar radiation.Machine learning models have been developed and trained on historical solar radiation data, incorporating various meteorological parameters, geographical factors, and time-related features. The predictive accuracy of these models has been evaluated using real-world datasets from diverse geographic locations and climates.The results demonstrate the effectiveness of machine learning algorithms in accurately forecasting solar radiation levels. These predictions can empower energy stakeholders, grid operators, and solar energy system operators to make informed decisions regarding energy generation, distribution, and consumption. Additionally, the study highlights the significance of feature engineering, model selection, and hyper parameter tuning in enhancing prediction performance.The methodology involves data preprocessing, feature engineering, model selection, and evaluation using metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Additionally, model interpretability techniques, such as feature importance analysis, will be employed to enhance the transparency of predictions.

Keyword: - Decision Tree, Random Forest, AdaBoost, Linear Regression, KNN, SVR.

1. INTRODUCTION

Solar energy has emerged as a pivotal component of the world's transition to sustainable and clean energy sources. The harnessing of solar power heavily relies on the accurate prediction of solar radiation, which is vital for optimizing the efficiency and reliability of solar energy systems. As climate change concerns and the quest for greener energy solutions continue to grow, the development of effective solar radiation prediction models becomes increasingly imperative.

This project aims to address this pressing need by leveraging the power of machine learning to predict solar radiation levels. Machine learning, with its ability to discern complex patterns from large datasets, offers a promising approach to enhance the accuracy and precision of solar radiation forecasting. By integrating historical solar radiation data with various meteorological, geographical, and time-related factors, machine learning models can provide real-time and future predictions of solar radiation levels.The significance of this project lies in its potential to revolutionize the renewable energy sector. Accurate solar radiation predictions can assist energy grid operators, solar power plant managers, and energy policymakers in making informed decisions about energy generation, storage, and distribution. Moreover, these predictions can contribute to reducing the dependency on non-renewable energy sources and mitigating greenhouse gas emissions.This project's methodology encompasses data preprocessing, feature engineering, model

selection, and performance evaluation using metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).In this endeavor, we will explore various machine learning techniques, conduct rigorous data analysis, and develop predictive models to optimize solar radiation forecasting. By doing so, we aim to support the transition towards a more sustainable and environmentally friendly energy future.

2. LITERATURE SURVEY

[1] Hossain, M.R., O. A.M.T., Shawkat Ali, A.B.M.: Hybrid prediction method for solar power using different computational intelligence algorithms. Smart Grid and Renewable Energy 4(1), 76-87 (2013).

[2] E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann, and C. Kurz, Regional PV power prediction for improved grid integration, Prog. Photovoltaics Res. Appl., vol. 19, no. 7, pp. 757–771, 2011.
[3] M. Abuella and B. Chowdhury, Solar Power Probabilistic Forecasting by Using Multiple Linear Regression Analysis, in IEEE Southeastcon Proceedings, Ft. Lauderdale, FL, 2015.
[4] Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Annals of Statistics pp. 1189{1232 (2001).

[5] Faizan Jawiad, Khurum Nazir Junejo Predicting Daily Mean Solar Power Using Machine Learning Regression Techniques .

[6] Paulescu, E. Paulescu, P. Gravila, V. Badescu, Weather Modeling and Forecasting of PV Systems Operation, Springer London, London, 2013.

3.METHODOLOGY

3.1EXISTING SYSTEM

In the existing system, implementation of machine learning algorithms is bit complex to build due to the lack of information about the data visualization. Mathematical calculations are used in existing system for K-Meanse, Adaboost and Linear Regression model building this may take the lot of time and complexity. To overcome all this, we use machine learning packages available in the scikit-learn library.

3.1.1DISADVANTAGES OF EXISTING SYSTEM

- High complexity: High complexity in data can lead to increased processing time, resource requirements, and potential challenges in data analysis and interpretation.
- Time consuming: Time-consuming data processing can delay decision-making, hinder real-time insights, and impede the efficiency of data-driven operations and analytics.
- Loss of Trust: Users and stakeholders may lose trust in the machine learning system if it consistently delivers inaccurate results.
- Model Complexity: Sometimes, to improve accuracy, machine learning models become overly complex, which can lead to overfitting and reduced generalization performance. Complex models can also be harder to interpret and maintain.

3.2 PROPOSED METHODOLOGY

Proposed several machine learning models to Predict the Solar radiation, but none have adequately addressed this misdiagnosis problem. Also, similar studies that have proposed models for evaluation of such performance classification mostly do not consider the heterogeneity and the size of the data Therefore, we propose a KNN, Decision Tree, Random forest and AdaBoost Regression techniques.contrast enhancement, and normalization to ensure uniform input quality. A feature extraction module, powered by a DCNN, then analyzes spatial and temporal characteristics of objects within the frames. This network is trained on a diverse dataset containing both original and manipulated videos, allowing it to learn distinctive features associated with tampering.

4. SYSTEM DESIGN

In an information system, input is the raw data that is processed to produce output. During the input design, the developers must consider the input devices such as PC, MICR, OMR, etc.

Therefore, the quality of system input determines the quality of system output. Well-designed input forms and screens have following properties −

- It should serve specific purpose effectively such as storing, recording, and retrieving the information.
- It ensures proper completion with accuracy.
- It should be easy to fill and straightforward.
- It should focus on user's attention, consistency, and simplicity.
- All these objectives are obtained using the knowledge of basic design principles regarding −

4.1 SYSTEM ARCHITECTURE

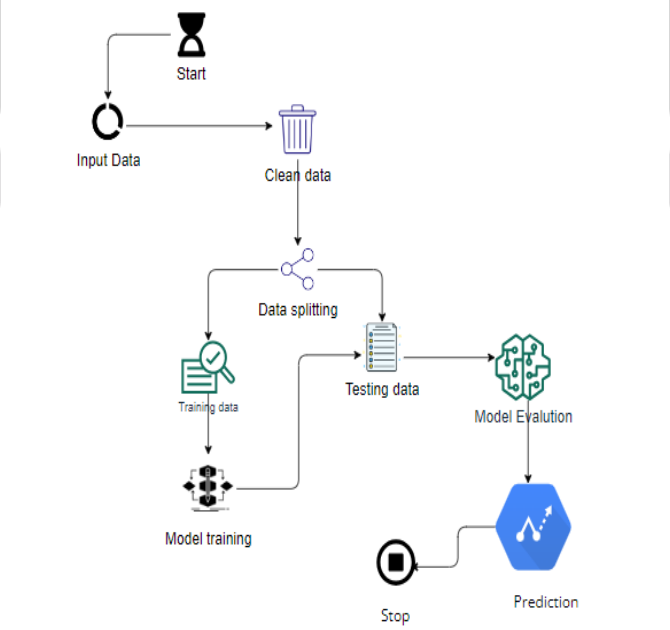


Fig. System Architecture

4.2 MODULES

1. Users can upload a dataset, which is a crucial initial step for the system to work with relevant data. This dataset likely contains historical information or examples that the system will use for its predictions.

2. Users have the capability to view the dataset they've uploaded. This feature helps users confirm the data they've provided and ensures transparency in the process.

3. Users need to input specific values or parameters into the system to request predictions or results. These input values likely correspond to the variables or features in the dataset.

4.2.1 MODULES DESCRIPTION

1. Take the Dataset: The system accepts and processes the dataset provided by the user. This dataset forms the foundation for building the predictive model.

2. Preprocessing: Before training a predictive model, the system preprocesses the dataset. This includes handling missing data, data cleaning, and feature extraction. Preprocessing ensures that the data is in a suitable format for modeling.

3. Training: The system uses machine learning techniques and Python modules to train a model based on the preprocessed dataset. The model learns patterns and relationships within the data, allowing it to make predictions.

3. Generate Results: Once the model is trained, the system can generate results based on user input values. These results typically indicate whether the input data corresponds to a specific condition, event, or prediction, such as Medical Insurance Cost

1. AdaBoost: AdaBoost algorithm, short for Adaptive Boosting, is a Boosting technique used as an Ensemble Method in Machine Learning. It is called Adaptive Boosting as the weights are re-assigned to each instance, with higher weights assigned to incorrectly classified instances. Boosting is used to reduce bias as well as variance for supervised learning. It works on the principle of learners growing sequentially. Except for the first, each subsequent learner is grown from previously grown learners. In simple words, weak learners are converted into strong ones. The AdaBoost algorithm works on the same principle as boosting with a slight difference. Let's discuss this difference in detail.

First, let us discuss how boosting works. It makes 'n' number of decision trees during the data training period. As the first decision tree/model is made, the incorrectly classified record in the first model is given priority. Only these records are sent as input for the second model. The process goes on until we specify a number of base learners we want to create. Remember, repetition of records

is allowed with all boosting techniques.

This figure shows how the first model is made and errors from the first model are noted by the algorithm. The record which is incorrectly classified is used as input for the next model. This process is repeated until the specified condition is met. As you can see in the figure, there are 'n' number of models made by taking the errors from the previous model. This is how boosting works.

The models 1,2, 3,…, N are individual models that can be known as decision trees. All types of boosting models work on the same principle.

Since we now know the boosting principle, it will be easy to understand the AdaBoost algorithm. Let's dive into AdaBoost's working. When the random forest is used, the algorithm makes an 'n' number of trees. It makes proper trees that consist of a start node with several leaf nodes. Some trees might be bigger than others, but there is no fixed depth in a random forest. With AdaBoost, however, the algorithm only makes a node with two leaves, known as Stump.

The figure here represents the stump. It can be seen clearly that it has only one node with two leaves. These stumps are weak learners and boosting techniques prefer this. The order of stumps is very important in AdaBoost. The error of the first stump influences how other stumps are made. Let's understand this with an example.

Here's a sample dataset consisting of only three features where the output is in categorical form. The image shows the actual representation of the dataset. As the output is in binary/categorical form, it becomes a classification problem. In real life, the dataset can have any number of records and features in it. Let us consider 5 datasets for explanation purposes. The output is in categorical form, here in the form of Yes or No. All these records will be assigned a sample weight. The formula used for this is 'W=1/N' where N is the number of records. In this dataset, there are only 5 records, so the sample weight becomes 1/5 initially. Every record gets the same weight. In this case, it's 1/5.

Learn AdaBoost Model from Data

Ada Boosting is best used to boost the performance of decision trees and this is based on binary classification problems.

AdaBoost was originally called AdaBoost.M1 by the author. More recently it may be referred to as discrete Ada Boost. As because it is used for classification rather than regression.

AdaBoost can be used to boost the performance of any machine learning algorithm. It is best used with weak learners.

2. Decision Tree: A tree has many analogies in real life, and turns out that it has influenced a wide area of machine learning, covering both classification and regression. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a tree-like model of decisions. Though a commonly used tool in data mining for deriving a strategy to reach a particular goal.

A decision tree is drawn upside down with its root at the top. In the image on the left, the bold text in black represents a condition/internal node, based on which the tree splits into branches/ edges. The end of the branch that doesn't split anymore is the decision/leaf, in this case, whether the passenger died or survived, represented as red and green text respectively.

Although, a real dataset will have a lot more features and this will just be a branch in a much bigger tree, but you can't ignore the simplicity of this algorithm. The feature importance is clear and relations can be viewed easily. This methodology is more commonly known as learning decision tree from data and above tree is called Classification tree as the target is to classify passenger as survived or died. Regression trees are represented in the same manner, just they predict continuous values like price of a house. In general, Decision Tree algorithms are referred to as CART or Classification and Regression Trees.

So, what is actually going on in the background? Growing a tree involves deciding on which features to choose and what conditions to use for splitting, along with knowing when to stop. As a tree generally grows arbitrarily, you will need to trim it down for it to look beautiful. Let's start with a common technique used for splitting.

3. Random Forest: A random forest is a machine learning technique that's used to solve regression and classification problems. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems.

A random forest algorithm consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.

The (random forest) algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome.

A random forest eradicates the limitations of a decision tree algorithm. It reduces the over fitting of datasets and increases precision. It generates predictions without requiring many configurations in packages (like Scikit-learn).

Features of a Random Forest Algorithm:

5. RESULTS AND DISCUSSION

EXECUTION PROCEDURE

The Execution procedure is as follows :

1. In this research work with data with attributes are observable and then all of them are floating data. And there's a decision class/class variable. This data was collected from Kaggle machine learning repository.
2. In this research 70% data use for train model and 30% data use for testing purpose.
3. Logistic Regressionis used as Classifier .
4. In the classification report we were able to find out the desired result
5. In this analysis the result depends on some part of this research. However, which algorithm gives the best true positive, false positive, true negative, and false negative are the best algorithms in this analysis.

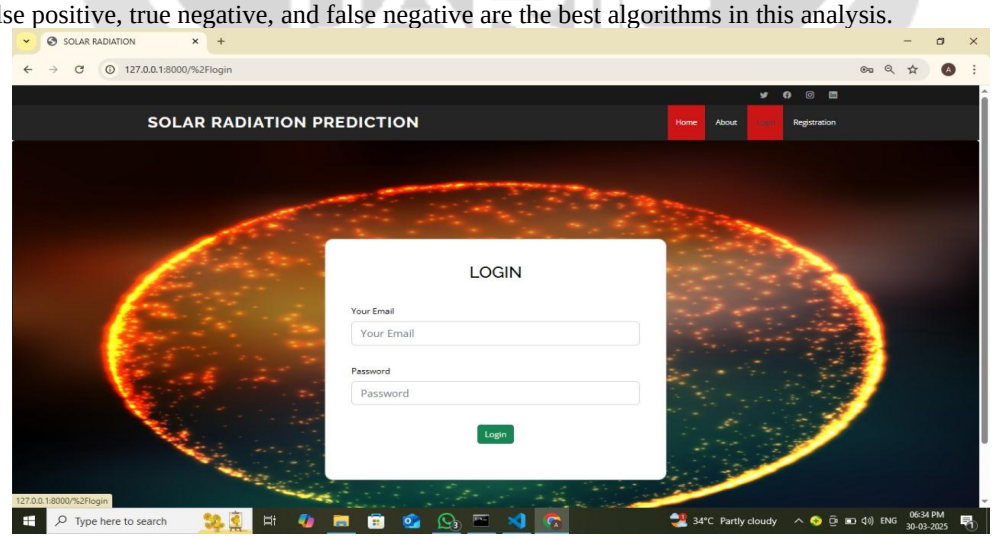


Fig. Login

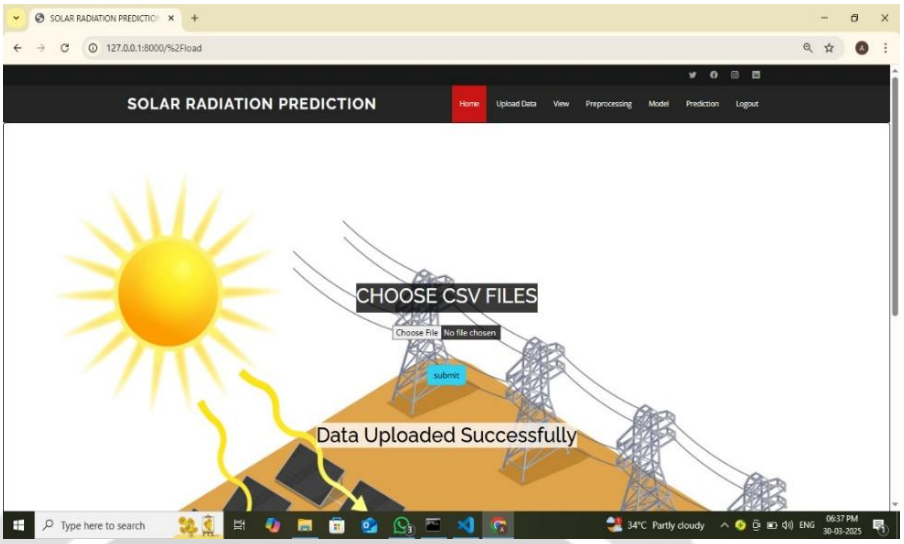
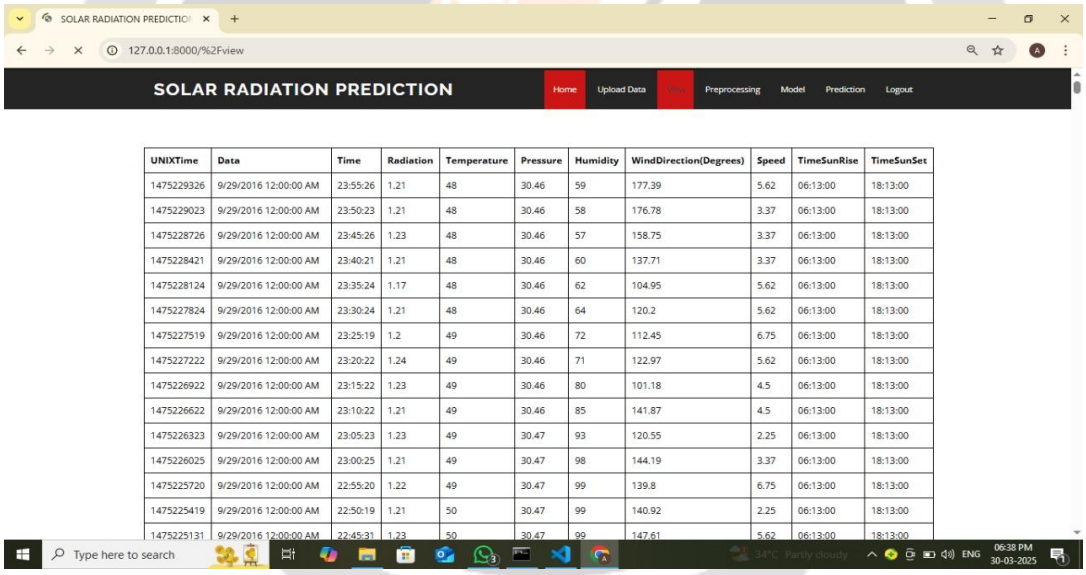


Fig. Data Upload



UNIXTime	Data	Time	Radiation	Temperature	Pressure	Humidity	WindDirection(Degrees)	Speed	TimeSunRise	TimeSunSet
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
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
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
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
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
1475227824	9/29/2016 12:00:00 AM	23:30:24	1.21	48	30.46	64	120.2	5.62	06:13:00	18:13:00
1475227519	9/29/2016 12:00:00 AM	23:25:19	1.2	49	30.46	72	112.45	6.75	06:13:00	18:13:00
1475227222	9/29/2016 12:00:00 AM	23:20:22	1.24	49	30.46	71	122.97	5.62	06:13:00	18:13:00
1475226922	9/29/2016 12:00:00 AM	23:15:22	1.23	49	30.46	80	101.18	4.5	06:13:00	18:13:00
1475226622	9/29/2016 12:00:00 AM	23:10:22	1.21	49	30.46	85	141.87	4.5	06:13:00	18:13:00
1475226323	9/29/2016 12:00:00 AM	23:05:23	1.23	49	30.47	93	120.55	2.25	06:13:00	18:13:00
1475226025	9/29/2016 12:00:00 AM	23:00:25	1.21	49	30.47	98	144.19	3.37	06:13:00	18:13:00
1475225720	9/29/2016 12:00:00 AM	22:55:20	1.22	49	30.47	99	139.8	6.75	06:13:00	18:13:00
1475225419	9/29/2016 12:00:00 AM	22:50:19	1.21	50	30.47	99	140.92	2.25	06:13:00	18:13:00
1475225131	9/29/2016 12:00:00 AM	22:45:31	1.23	50	30.47	99	147.61	5.62	06:13:00	18:13:00

Fig. View all data

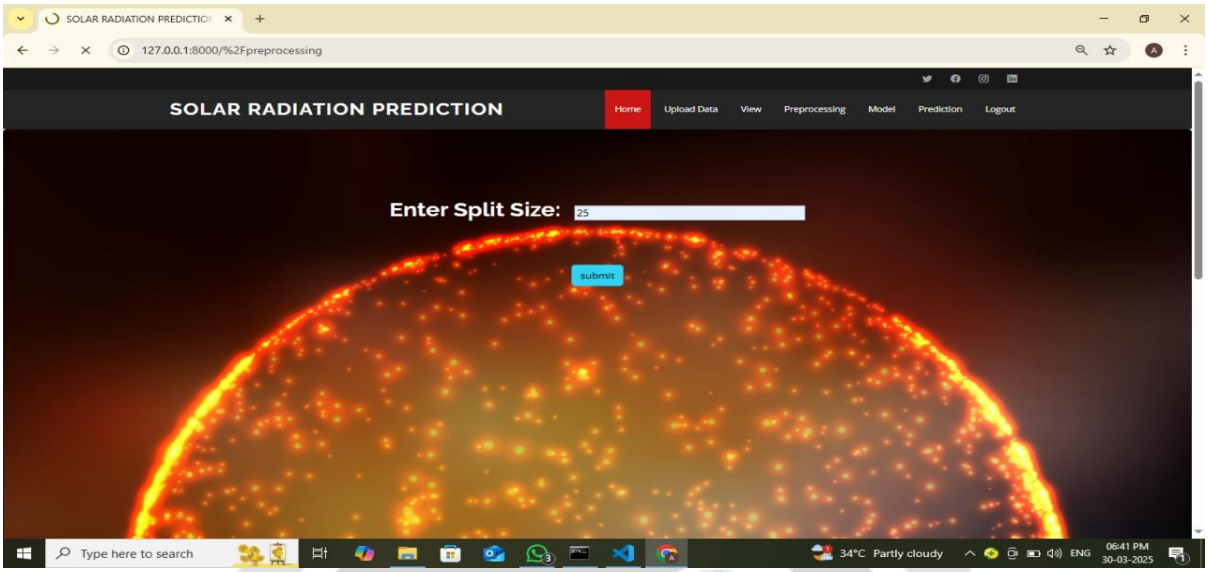
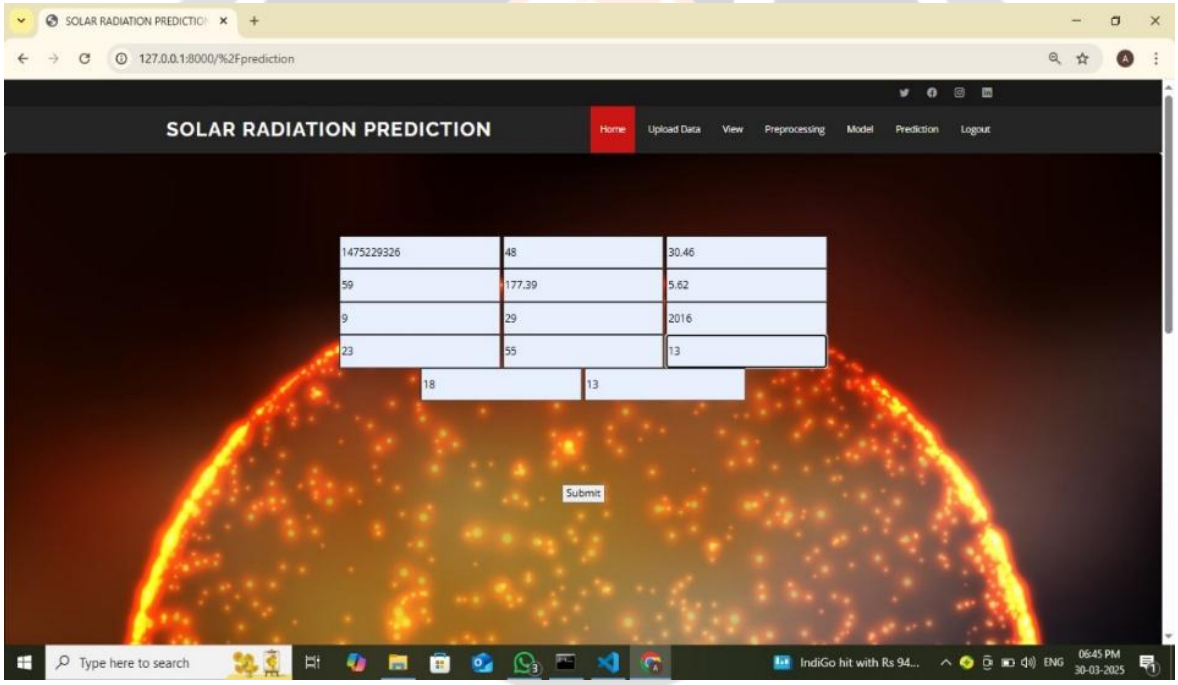


Fig. Pre Processing



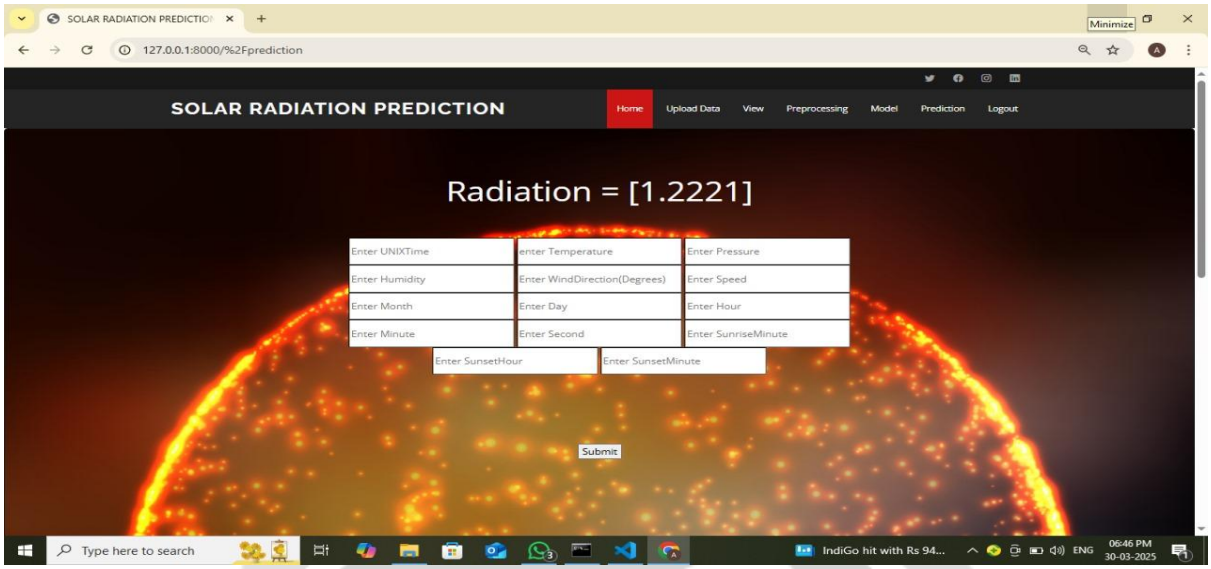


Fig.Result

6. CONCLUSION

In conclusion, the application of machine learning techniques for solar radiation prediction represents a crucial advancement in the field of renewable energy. This technology has demonstrated its potential to significantly improve the efficiency and reliability of solar energy systems. Accurate predictions empower energy stakeholders to optimize energy generation, enhance grid integration, reduce costs, and minimize environmental impact. As we strive for a more sustainable and eco-friendly energy future, machine learning-based solar radiation forecasting emerges as a vital tool, paving the way for the widespread adoption of solar power and contributing to our global efforts to combat climate change and achieve energy sustainability.

7. REFERENCE

[1] Hossain, M.R., O. A.M.T., Shawkat Ali, A.B.M.: Hybrid prediction method for solar power using different computational intelligence algorithms. Smart Grid and Renewable Energy 4(1), 76-87 (2013).

[2] E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann, and C. Kurz, "Regional PV power prediction for improved grid integration," Prog. Photovoltaics Res. Appl., vol. 19, no. 7, pp. 757–771, 2011.

[3] M. Abuella and B. Chowdhury, "Solar Power Probabilistic Forecasting by Using Multiple Linear Regression Analysis," in IEEE Southeastcon Proceedings, Ft. Lauderdale, FL, 2015.

[4] Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Annals of Statistics pp. 1189{1232 (2001).

[5] Faizan Jawiad, Khurum Nazir Junejo "Predicting Daily Mean Solar Power Using Machine Learning Regression Techniques ".

[6] Paulescu, E. Paulescu, P. Gravila, V. Badescu, Weather Modeling and Forecasting of PV Systems Operation, Springer London, London, 2013.