PROJECT PRESENTATION

SOLAR RADIATION PREDICTION USING MACHINE LEARNING

SUPERVISOR: DR. HAILIYA AHSAN(Assistant Professor)
CO-SUPERVISOR: DR. SHOEB HUSSAIN(Assistant Professor**)**

PRESENTED BY:
DHRUV SHARMA (2020BELE024)
ANURUDRA YADAV (2020BELE077)
KAUSHINDER (2020BELE099)

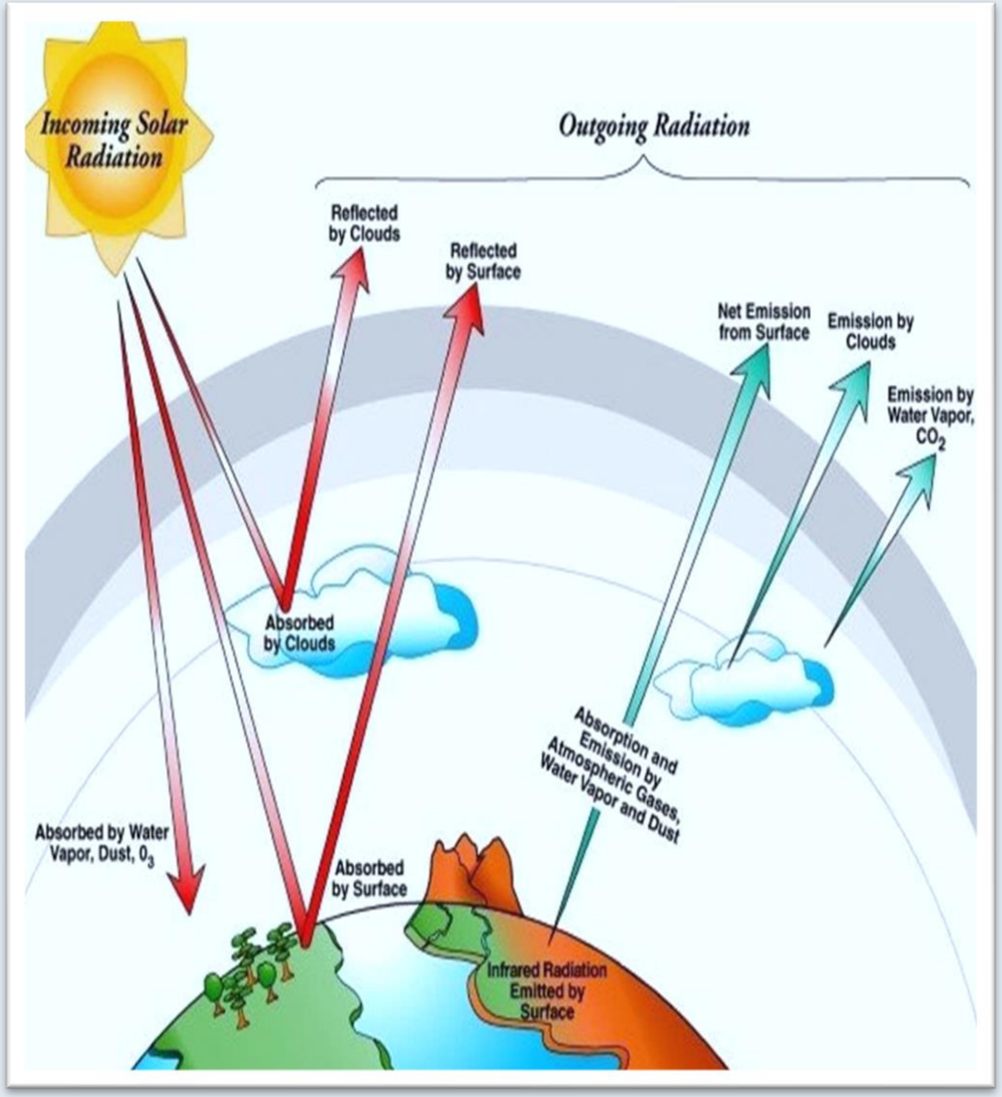
SOLAR RADIATION

- Solar radiation plays a crucial role in various aspects of our lives, from powering renewable energy systems to influencing **agricultural productivity** and **climate patterns.**

- However, accurately predicting solar radiation levels is challenging due to the complex interplay of **weather conditions**, geographical factors, and **atmospheric dynamics**.

- In this project, we aim to develop a model that can predict solar radiation levels with high accuracy.

- We hope to contribute to the advancement of renewable energy technologies and enhance our understanding of earth's energy balance."



IMPORTANCE OF SOLAR RADIATION PREDICTION

1.

- **GRID MANAGEMENT**
- Grid operators to manage the integration of solar energy into the grid. Accurate predictions enable them to balance supply and demand, reduce the need for backup power sources, and prevent grid instability.

2.

- **COST REDUCTION**
- It can help optimize the design of buildings for energy efficiency, as well as assist farmers in planning crop planting and harvesting times based on sunlight availability

3.

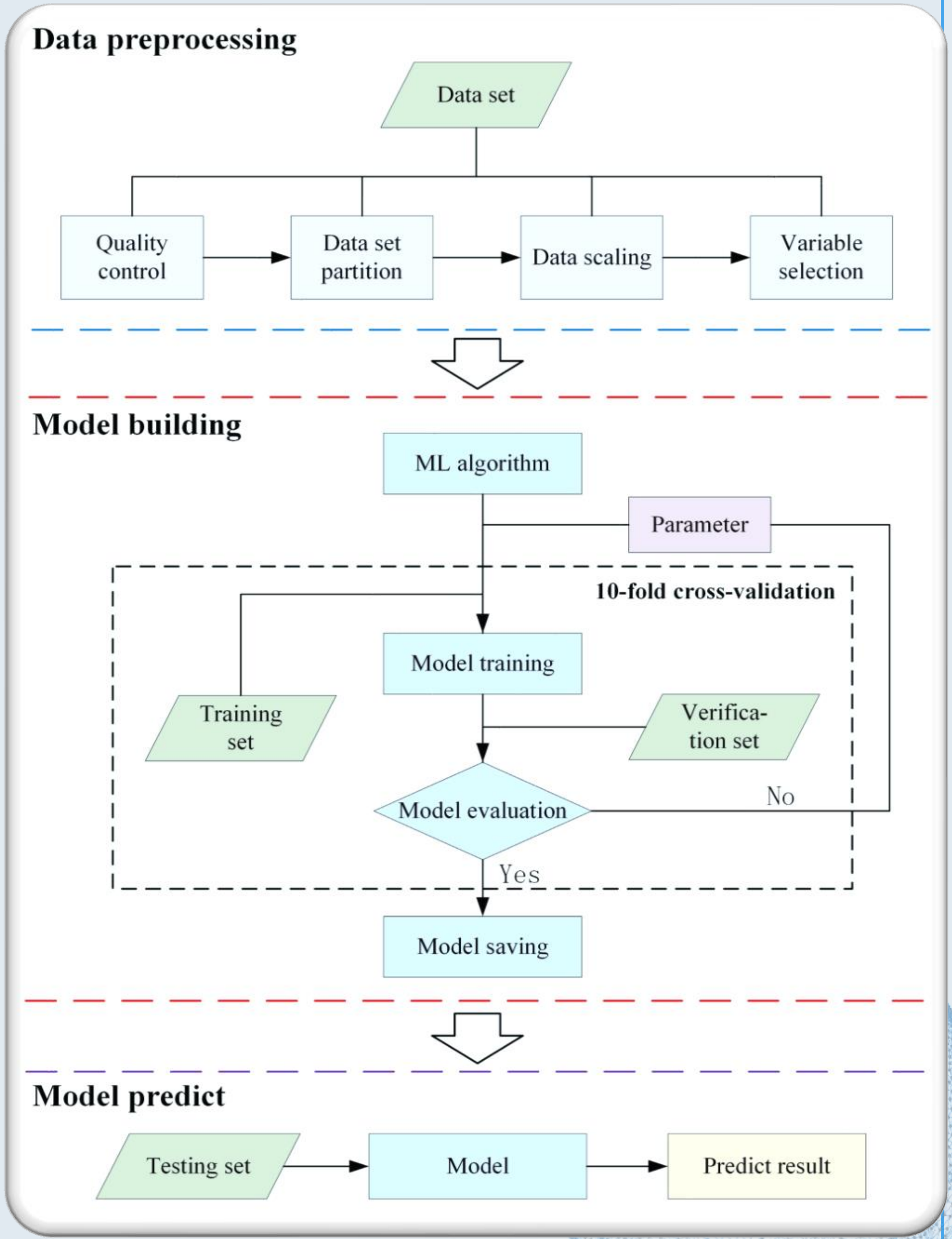
- **URBAN PLANNING AND AGRICULTURE**
- Predicting solar radiation more accurately, energy producers can reduce costs associated with inefficient energy production and grid management.

OUR APPROACH

- Solar radiation relies on weather patterns.
- The dataset has a total of 32,686 samples.
- 11 attributes are corresponding to each sample.
- Radiation is the response variable and the rest of the 10 attributes are the predictor variable.
- We plan to perform an 80-20 split to generate the train and test set.

Standard models:

1. Linear Regression 2. Decision Tree Regressor 3. Random Forest Regressor



DATA PROCESSING

→Data preprocessing is crucial in data science to clean, transform, and prepare data for analysis.

→It ensures data quality, improves model accuracy, and helps in extracting meaningful insights from the data.

→A real world data generally contains noise, missing value and may be in unusable format.

Fig: Solar Data Set after Pre Processing

	Date	Time	Radiation	Temperature	Pressure	Humidity	WindDirection	Speed	TimeSunRise	TimeSunSet
0	273	86126	1.21	48	30.46	59	177.39	5.62	22380	65580
1	273	85823	1.21	48	30.46	58	176.78	3.37	22380	65580
2	273	85526	1.23	48	30.46	57	158.75	3.37	22380	65580
3	273	85221	1.21	48	30.46	60	137.71	3.37	22380	65580
4	273	84924	1.17	48	30.46	62	104.95	5.62	22380	65580
...	...	...	...	...	...	...	...	...	...	...
32681	336	1204	1.22	44	30.43	102	145.42	6.75	24060	63720
32682	336	901	1.17	44	30.42	102	117.78	6.75	24060	63720
32683	336	601	1.20	44	30.42	102	145.19	9.00	24060	63720
32684	336	302	1.23	44	30.42	101	164.19	7.87	24060	63720
32685	336	2	1.20	44	30.43	101	83.59	3.37	24060	63720

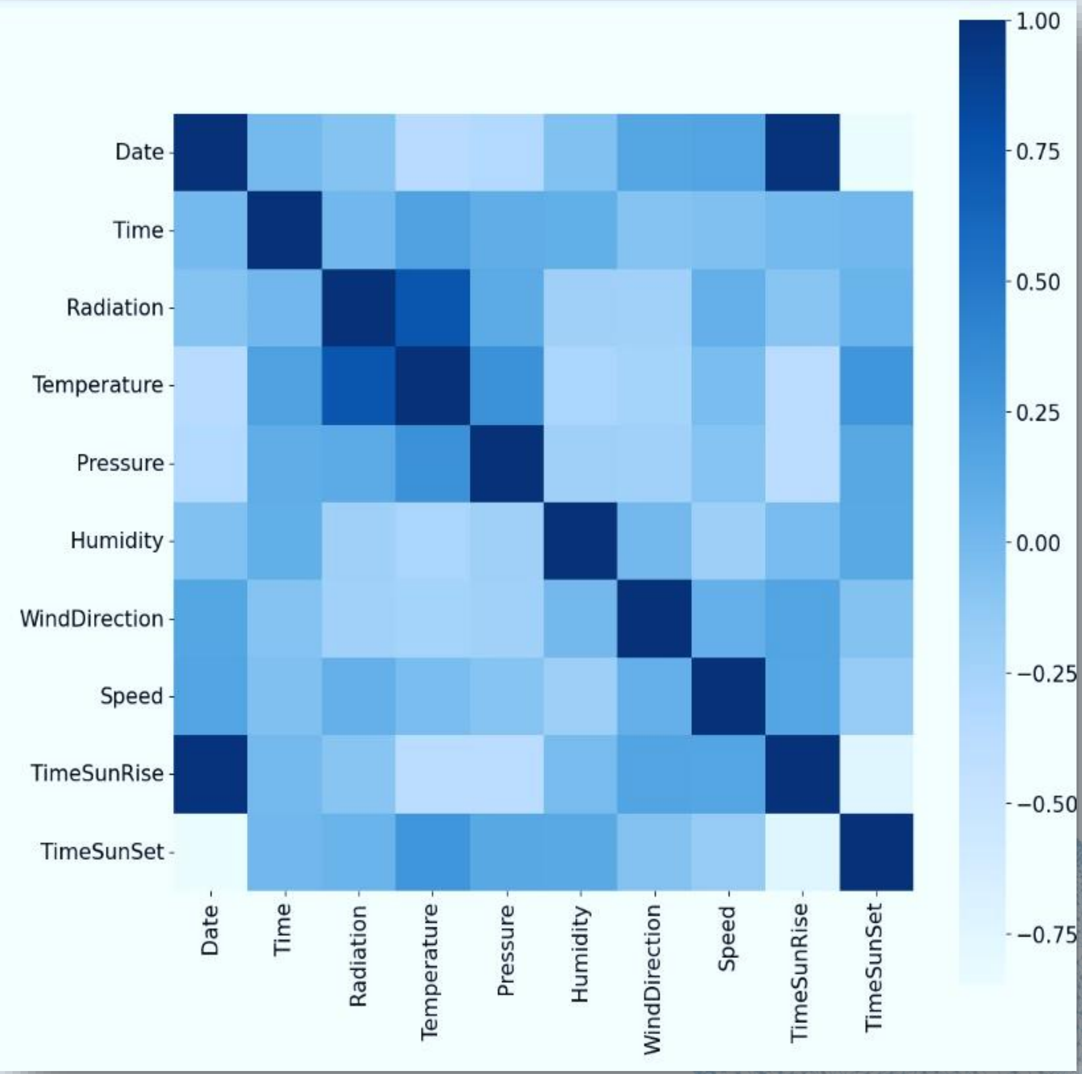
32686 rows × 10 columns

FEATURE SELECTION

- high correlation between temperature, pressure, humidity, and Wind Direction with radiation.
- The sunrise and sunset dataset does not correlate with the radiation prediction so therefore we drop both these attributes

Scatter Plot

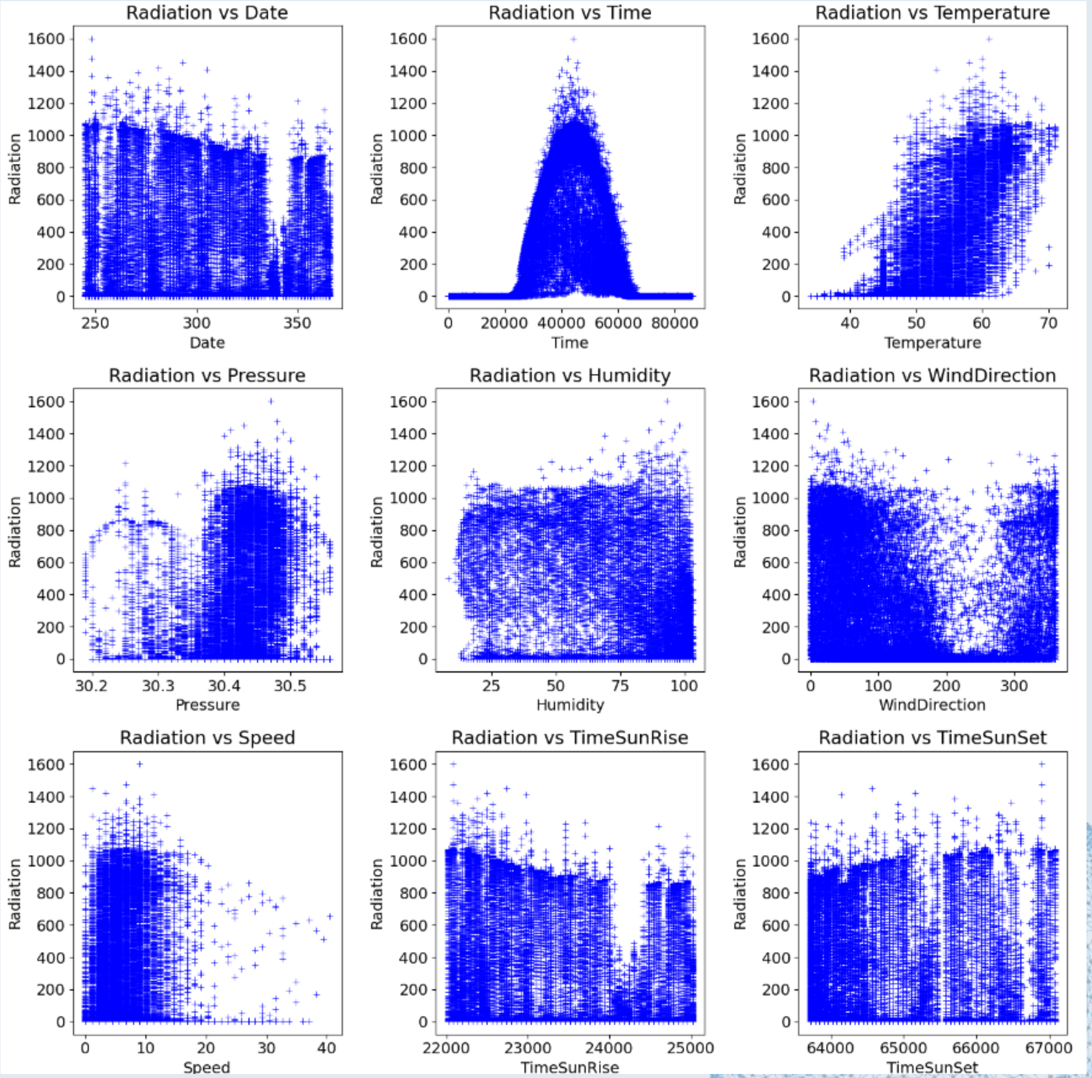
```
In [10]: plt.rcParams["figure.figsize"] = (14, 14)
         plt.rcParams["font.size"] = 14

         ylabel = 'Radiation'
         columns = ['Date', 'Time', 'Temperature', 'Pressure', 'Humidity', 'WindDirectio

         for index, xlabel in enumerate(columns):
           plt.subplot(3, 3, index+1)
           plt.scatter(df[xlabel], df[ylabel],color='blue',marker='+',linewidth=0.5)
           plt.xlabel(xlabel)
           plt.ylabel(ylabel)
           plt.title(ylabel + ' vs ' + xlabel)

         plt.tight_layout()
```



RADIATION VS ATTRIBUTES

- High correlation between temperature, pressure, humidity, and Wind Direction with radiation.
- The plot indicates that radiation depends on the time of the day.
- Low →early morning,
- Maximum → noon, and
- Gradually decreases → end of the day.

Fig Scatterplot of different features of dataset against Radiation



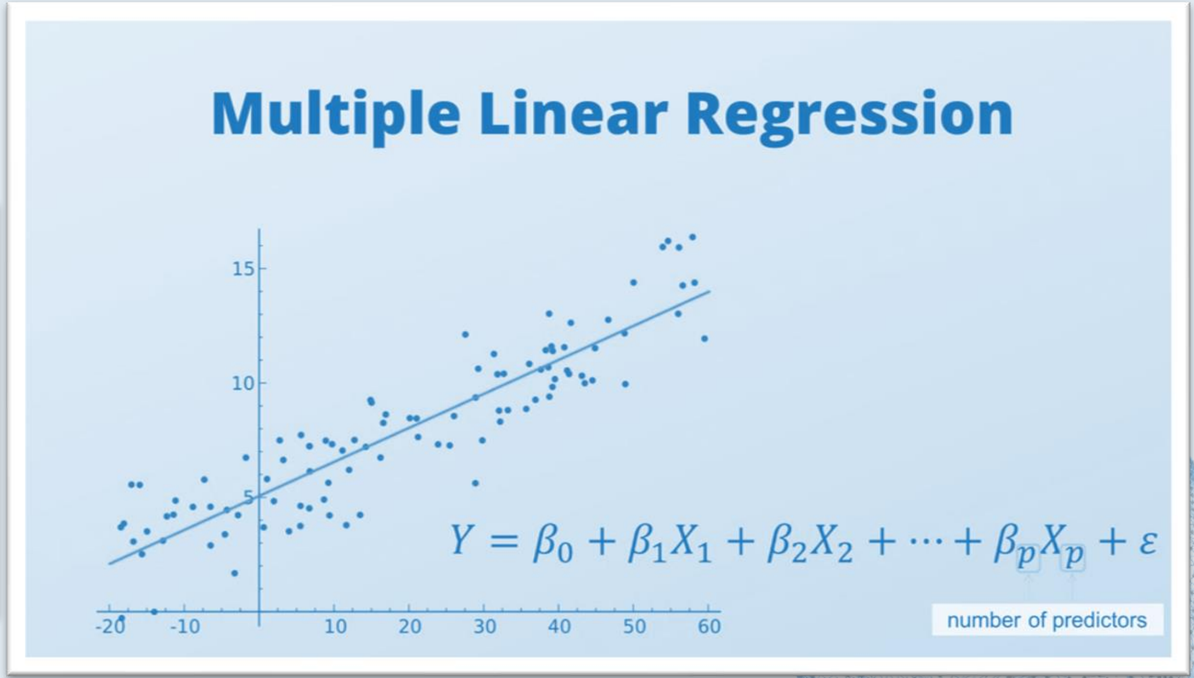
MULTI LINEAR REGRESSION APPROACH

Multiple linear regression is a statistical method used to model the relationship between multiple independent variables and a dependent variable.

Independent variable: Temp, speed. Pressure, humidity, wind direction etc.....

Dependent variable: Radiation

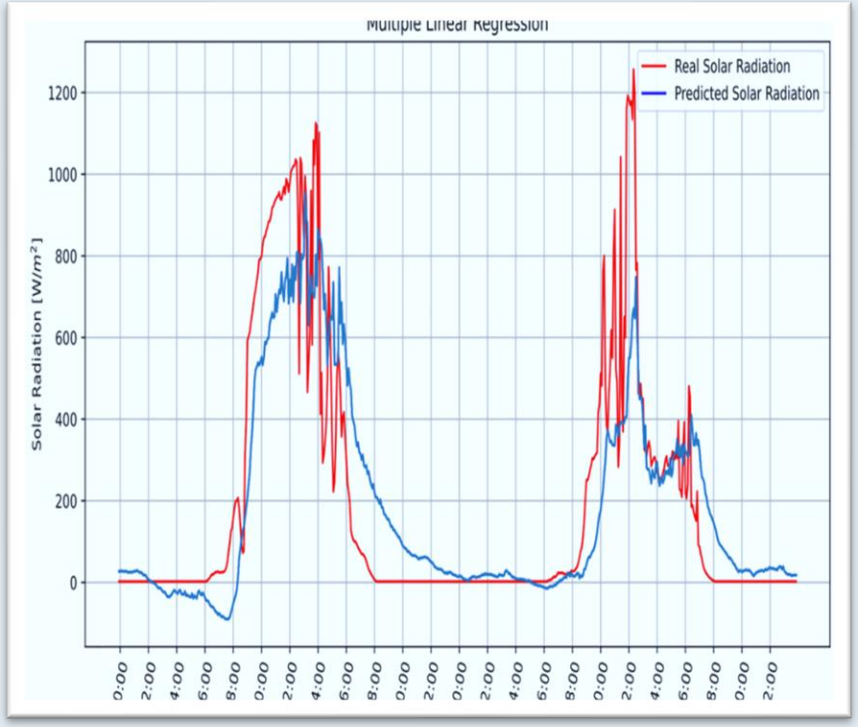
```
In [18]: pred=lr.predict(X_test)

         m=lr.coef_
         m

Out[18]: array([ 1.98104301e+00, -2.07962649e-03,  4.38422669e+01, -3.59256338e+02,
                 5.47659755e-01, -2.74759504e-01,  5.44636736e+00])

In [19]: c=lr.intercept_
         c

Out[19]: 8343.381813532798
```



RESULT OF LINEAR REGRESSION

- After performing different tests and using sensitivity analysis for this forecasting technique,
  it is deduced that the best parameters for tuning are those described in Table

```
In [86]: print("Accuracy for Train data sets")
         print( lr.score(X_train, Y_train)*100,'%')

         Accuracy for Train data sets
         61.90414543792242 %

In [87]: print("Accuracy for Test data sets")
         print( lr.score(X_test, Y_test)*100,'%')

         Accuracy for Test data sets
         62.5918994079995 %
```



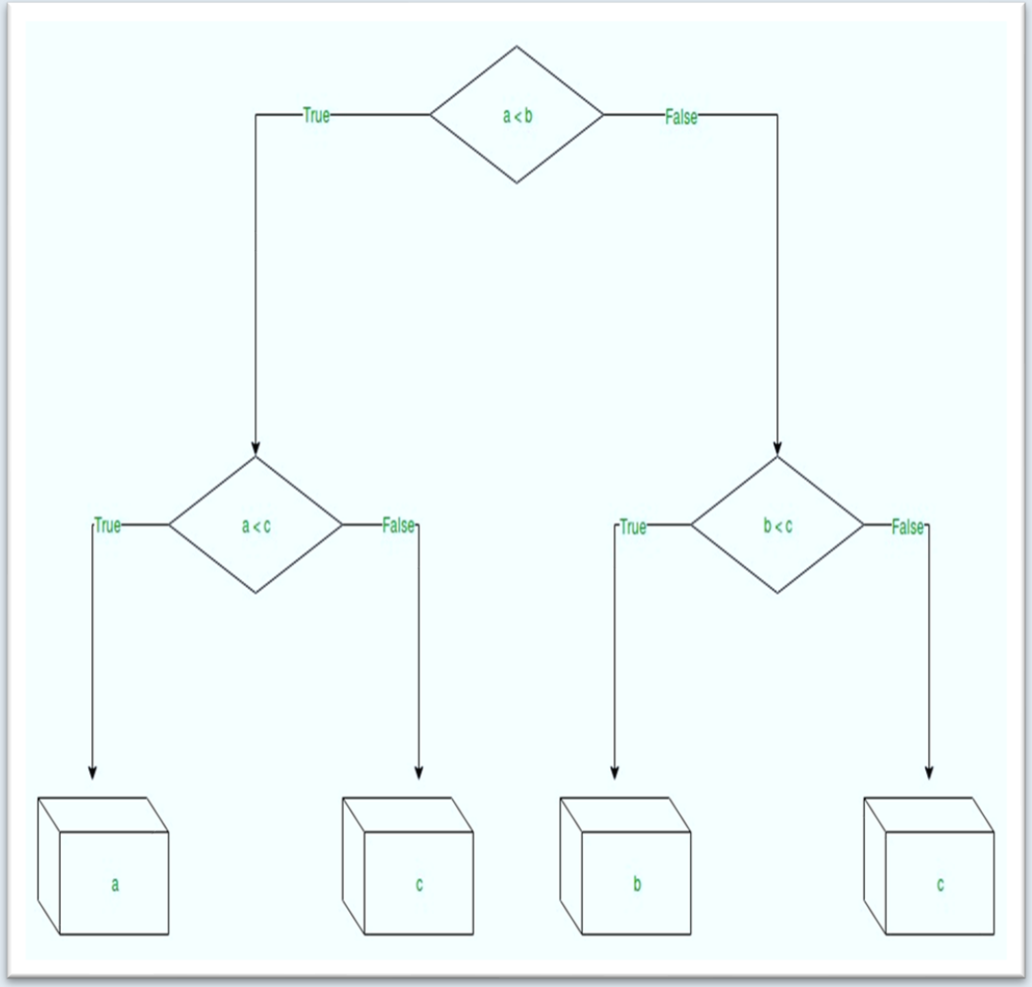
S.No	Date	Time	Temp(◦C)	Pressure(bar)	Humidity	WindDirection	Speed(m/s)	Actual(W/m^2)	Predicted(LR)	Error(%)
1	302	45920	50	30.4	101	69.67	4.5	516.1	377.56	26.74
2	350	73517	36	30.26	103	326.67	5.62	3.24	5.76	77.7
3	323	27019	47	30.44	65	145.42	5.62	21.79	28.14	24.13
4	350	24002	38	30.25	95	239.63	14.62	5.27	6.64	35.64
5	280	77120	40	30.46	102	128.94	3.37	4.22	7.64	81.23

DECISION TREE REGRESSION

- Decision tree regression observes the features of an object and trains a model in the structure of a tree to predict data in the future to produce meaningful continuous output.

Decision Tree Regressor

```
In [49]: from sklearn.tree import DecisionTreeRegressor
         dtr = DecisionTreeRegressor(random_state=RS)
         dtr.fit(X_train, Y_train)

Out[49]: ▾ DecisionTreeRegressor
         DecisionTreeRegressor(random_state=1811)
```



DECISION TREE APPROACH

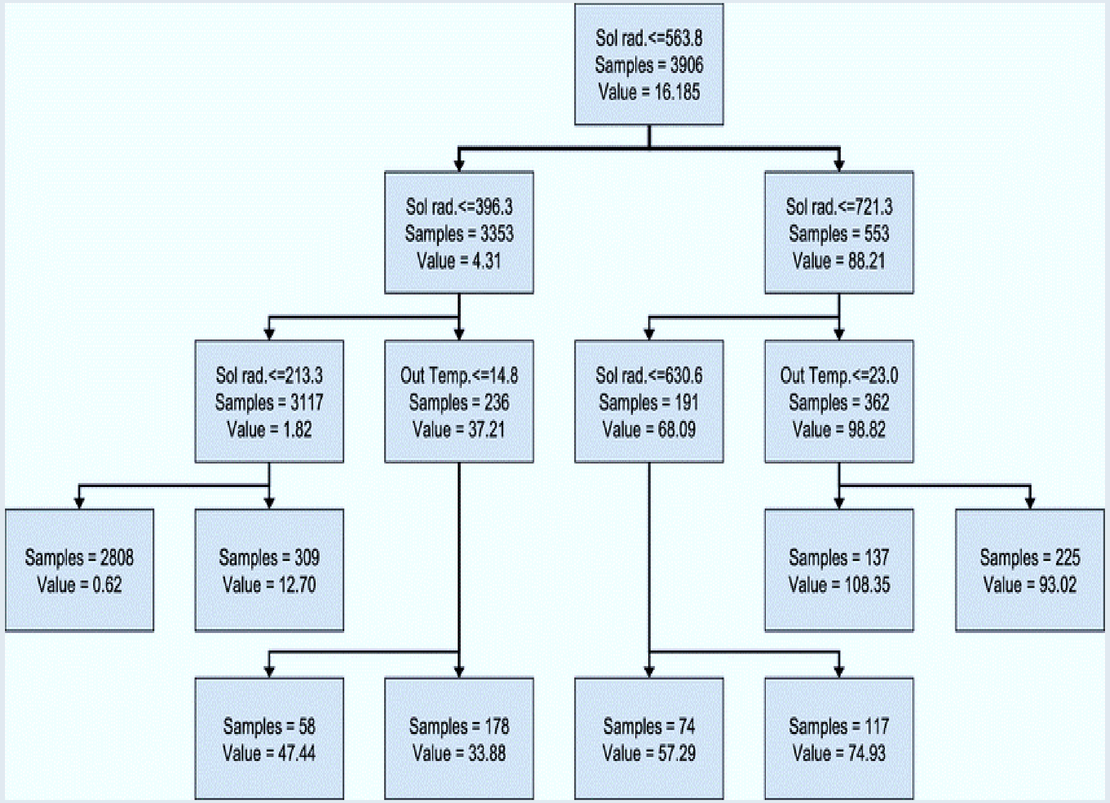
Step-1: Begin the tree with the root node, says S, which contains the complete dataset.

Step-2: Find the best attribute in the dataset using Attribute Selection Measure (ASM).

Step-3: Divide the S into subsets that contains possible values for the best attributes.

Step-4: Recursively make new decision trees using the subsets of the dataset created in step 3. Continue this process until a stage is reached where you cannot further classify the nodes

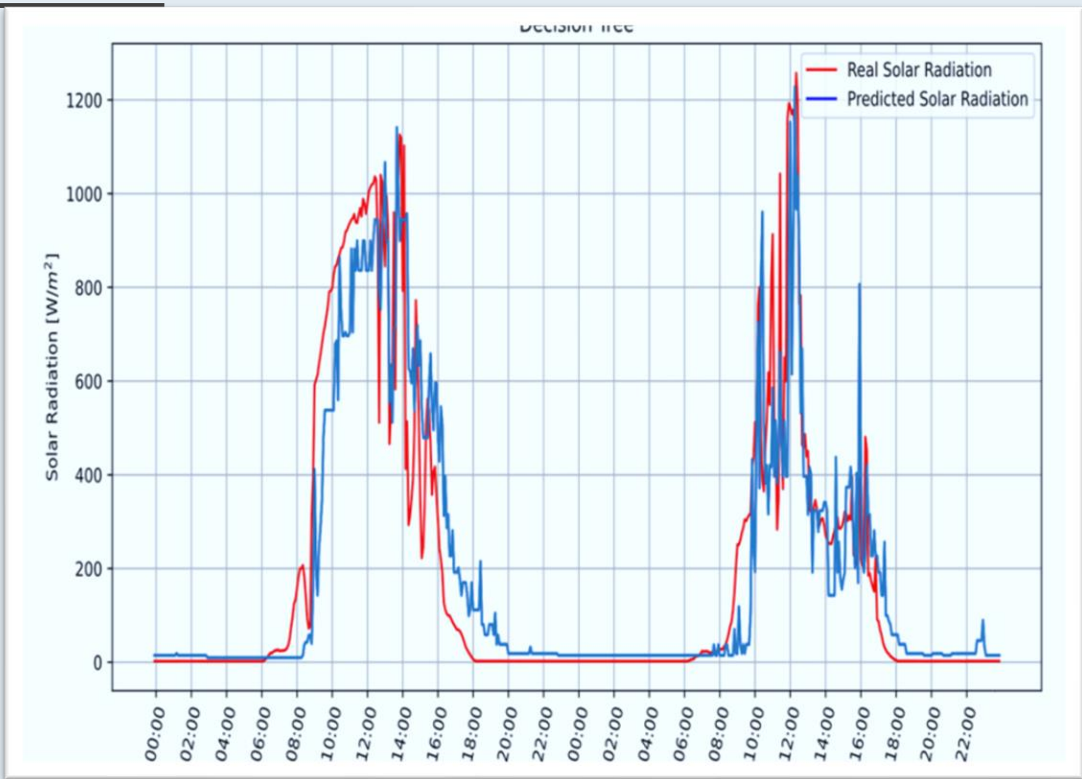
the final node represent output.



RESULT OF DECISION TREE

- After performing different tests and using sensitivity analysis for this forecasting technique,
  it is deduced that the best parameters for tuning are those described in Table

```
In [28]: print("Accuracy for Train data sets")
         print('train', dtr.score(X_train, Y_train))

         Accuracy for Train data sets
         train 1.0

In [29]: print("Accuracy for Test data sets")
         print('train', dtr.score(X_test, Y_test))

         Accuracy for Test data sets
         train 0.8748831378034979
```



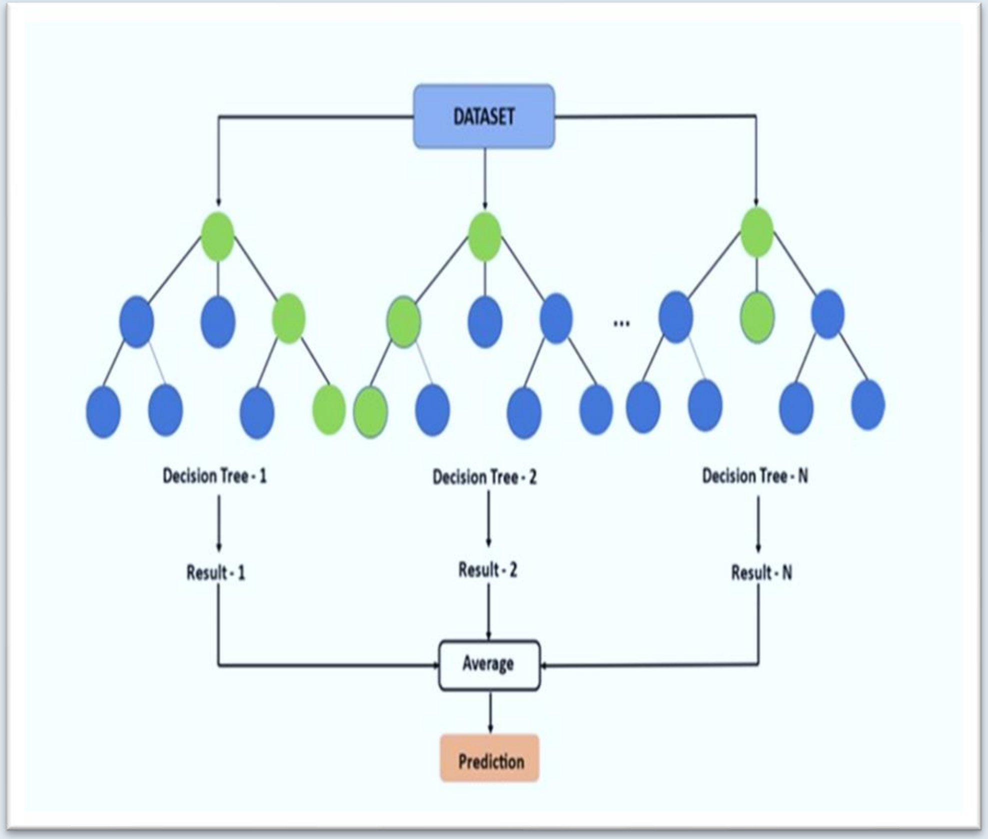
S.No	Date	Time	Temp(◦C)	Pressure(bar)	Humidity	WindDirection	Speed(m/s)	Actual(W/m^2)	Predicted(DT)	Error(%)
1	302	45920	50	30.4	101	69.67	4.5	516.1	549.12	6.39
2	350	73517	36	30.26	103	326.67	5.62	1.24	1.52	22.58
3	323	27019	47	30.44	65	145.42	5.62	21.79	22.09	1.35
4	350	24002	38	30.25	95	239.63	14.62	2.27	1.71	23.81
5	280	77120	40	30.46	102	128.94	3.37	1.22	1.25	1.2

RANDOM FOREST REGRESSOR

- Method that constructs multiple decision trees during training and combines their outputs for improved accuracy and robustness.
- It is widely used for classification and regression tasks by averaging predictions or majority voting

Random Forest

```
In [32]: from sklearn.ensemble import RandomForestRegressor

In [34]: rfr = RandomForestRegressor(random_state=RS)
         rfr.fit(X_train, Y_train)

Out[34]: RandomForestRegressor
         RandomForestRegressor(random_state=1811)
```



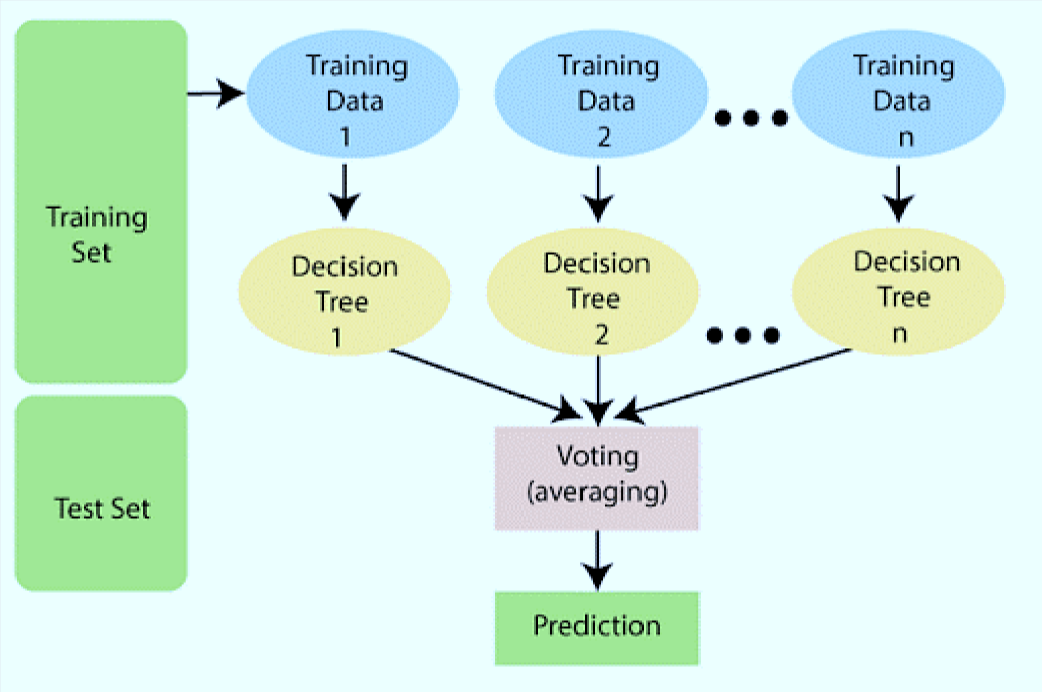
RANDOM FOREST APPROACH

Step 1: Select random samples from a given data or training dataset.

Step 2: This algorithm will construct a decision tree for every training data.

Step 3: Voting will take place by averaging the decision.

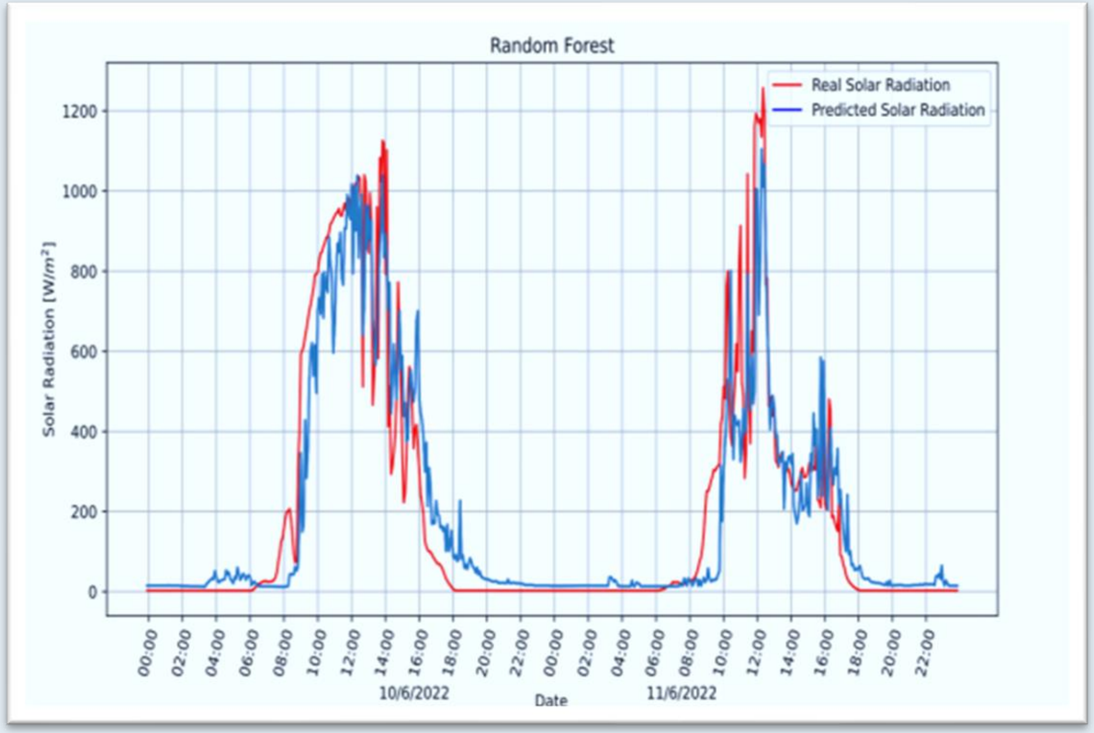
Step 4: Finally, select the average value as the final prediction result.



RANDOM FOREST RESULT

- After performing different tests and using sensitivity analysis for this forecasting technique, it is deduced that the best parameters for tuning are those described in Table

```
In [35]: print('train', rfr.score(X_train, Y_train))
         print('test', rfr.score(X_test, Y_test))

train 0.990636239140551
test 0.938671525993316
```



S.No	Date	Time	Temp(◦C)	Pressure(bar)	Humidity	WindDirection	Speed(m/s)	Actual(W/m^2)	Predicted(RFR)	Error(%)
1	302	45920	50	30.4	101	69.67	4.5	516.1	528.55	2.39
2	350	73517	36	30.26	103	326.67	5.62	1.24	1.24	0
3	323	27019	47	30.44	65	145.42	5.62	21.79	23.54	7.35
4	350	24002	38	30.25	95	239.63	14.62	2.27	2.63	3.81
5	280	77120	40	30.46	102	128.94	3.37	1.22	1.23	1.2

RESULTS AND COMPARISON

S.NO	MODEL-NAME	TRAIN DATASET ACCURACY	TEST DATASET ACCURACY
1	Multiple Linear Regression	61.90414	62.591899
2	Decision Tree Regressor	100	87.488313
3	Random Forest Regressor	99.0623	93.867152

S.No	Date	Time	Temp(◦C)	Pressure(bar)	Humidity	WindDirection	Speed(m/s)	Actual(W/m^2)	Predicted(LR)	Predicted(DT)	Predicted(RFR)
1	302	45920	50	30.4	101	69.67	4.5	516.1	377.56	549.12	528.55
2	350	73517	36	30.26	103	326.67	5.62	1.24	5.76	1.52	1.24
3	323	27019	47	30.44	65	145.42	5.62	21.79	28.14	22.09	23.54
4	350	24002	38	30.25	95	239.63	14.62	2.27	6.64	1.71	2.63
5	280	77120	40	30.46	102	128.94	3.37	1.22	7.64	1.25	1.23

FUTURE OF SOLAR RADIATION PREDICTION

1. **Enhanced Data Collection**: Continuously improve our data collection methods to incorporate new sources of data and enhance the quality and reliability of our predictions.

2. **Model Refinement**: Refine our machine learning algorithms to improve prediction accuracy, particularly in challenging weather conditions.

3. **Global Scalability**: Further refine our model to make it more scalable and adaptable to different regions and climates around the world.

4. **Integration with Energy Systems**: Explore ways to integrate our solar radiation prediction model with energy systems to optimize energy production and consumption.

FUTURE ASPECTS IN SOLAR ENERGY

- Advanced Photovoltaic (PV) Technologies
- Bifacial Solar Panels
- Solar Tracking Systems)
- Energy Storage Solutions
- Floating Solar Farms
- Digital solutions and Smart Grid Solutions

CONCLUSION

- A significant advancement in the field of renewable energy planning and grid management.
- By machine learning algorithms and high-resolution satellite data, we have developed a model that can accurately forecast solar radiation levels, even in complex weather conditions.
- These model has numerous practical applications, including optimizing energy production, improving grid stability, and advancing climate research.
- Through our project, we have demonstrated the potential of data-driven approaches to enhance the reliability and efficiency of solar energy systems.
- Future efforts will focus on enhancing data collection methods, refining machine learning algorithms, and integrating our model with energy systems to maximize its impact.

REFERENCES

- Soomin, H.; Lumi, A.; Okorie, E.C.; Uredo, L.R. Investigation of Solar Energy: The Case Study in Malaysia, Indonesia, Colombia and Nigeria. Int. J. Renew. Energy Res. 2019, 9, 86–95.
- Abdullah, W.S.W.; Osman, M.; Kadir, M.Z.A.A.; Varaiya, R. The potential and status of renewable energy development in Malaysia. Energies 2019, 12, 2437.2) https://www.kaggle.com/datasets/anikannal/solar-power-generation-data
- Solar irradiance forecasting models using machine learning techniques and digital twin: A case study with comparison. https://www.sciencedirect.com/science/article/pii/S2666603023000064
- https://www.researchgate.net/publication/343735894_Prediction_of_daily_global_solar_radiation_using_different_machine_learning_algorithms_Evaluation_and_comparison
- https://www.kaggle.com/datasets/dronio/SolarEnergy

THANK YOU