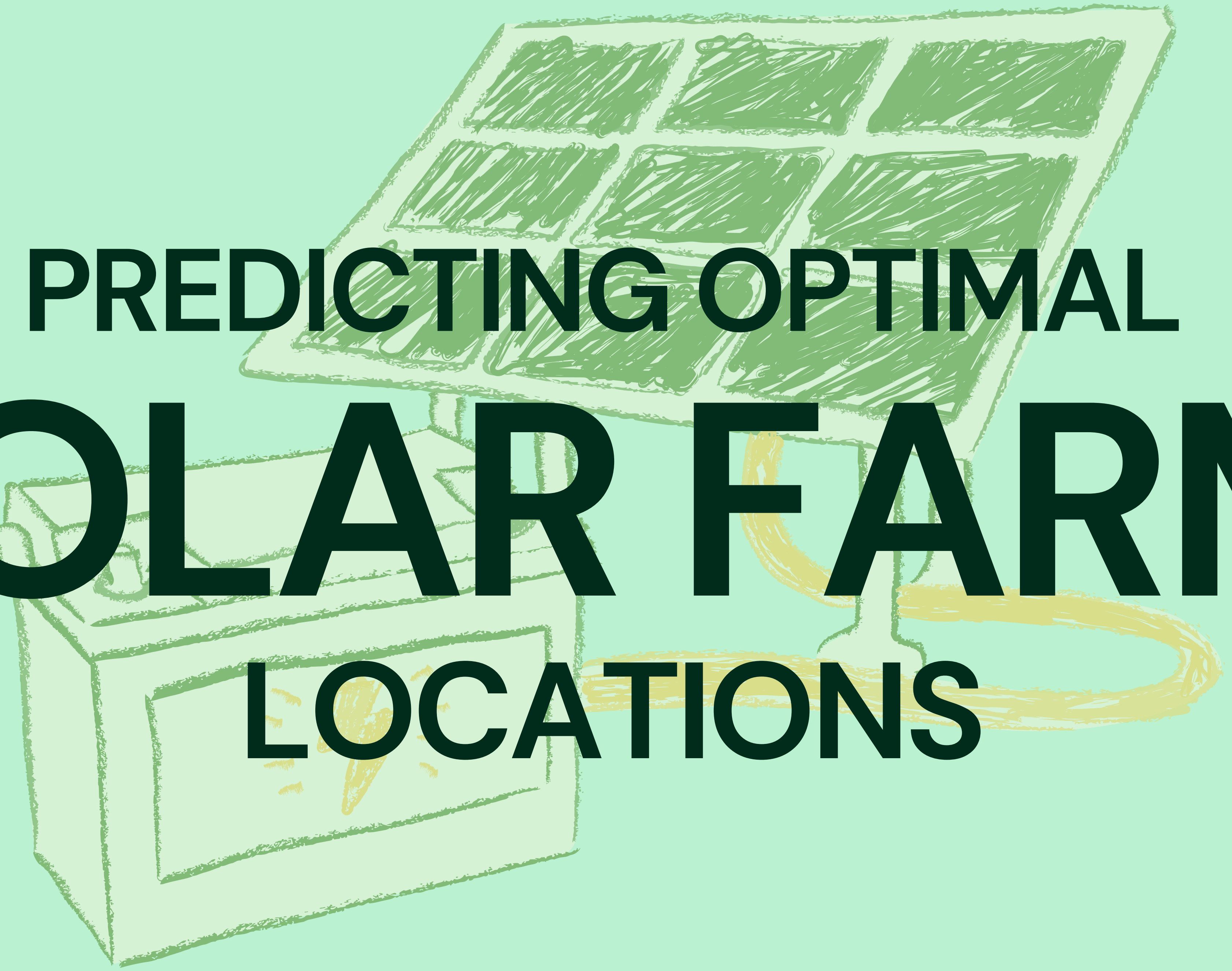
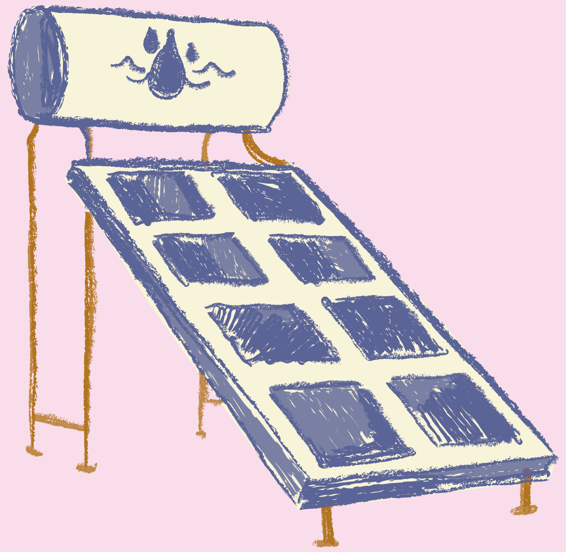


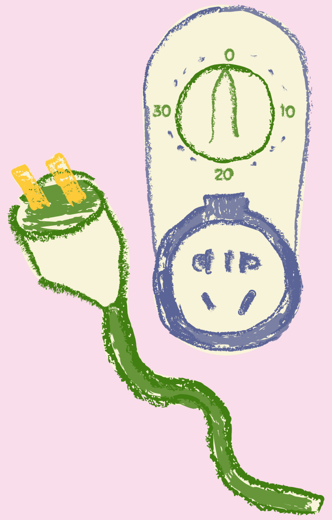
PREDICTING OPTIMAL SOLAR FARM LOCATIONS

RENEWABLE ENERGY SOURCE



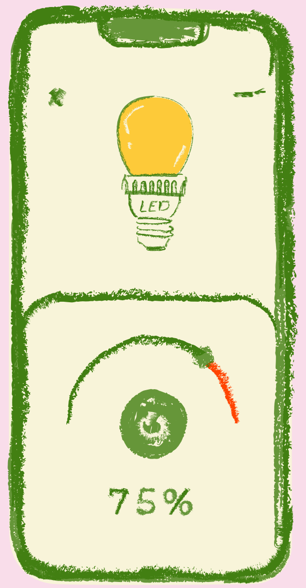
Solar farms harness the power of the sun, an abundant and inexhaustible energy source, making them a sustainable alternative to fossil fuels.

REDUCES CARBON EMISSIONS



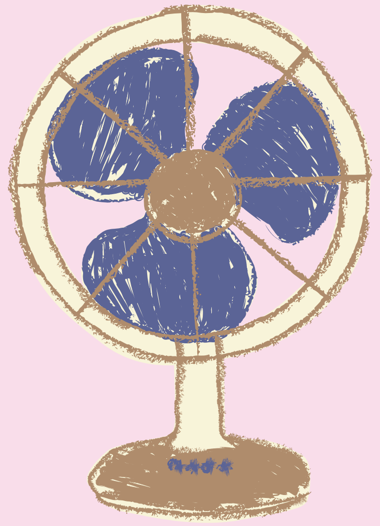
Solar farms significantly reduce greenhouse gas emissions, helping to combat climate change and improve air quality.

ENERGY INDEPENDENCE



Large-scale solar farms contribute to national and regional energy security by reducing dependence on imported fossil fuels and diversifying the energy supply.

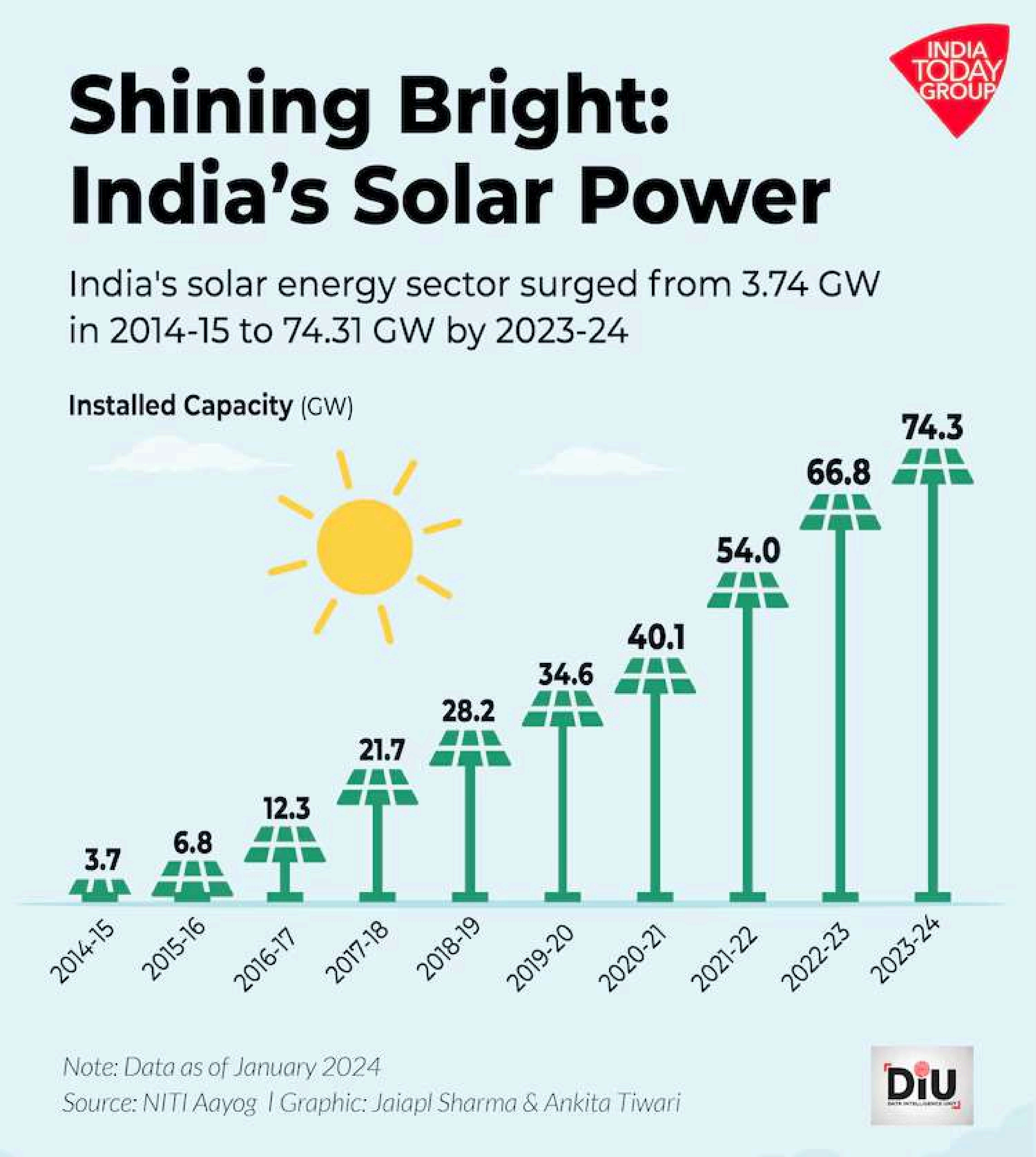
LOW OPERATING COST



Once installed, solar farms require minimal maintenance and have low operational costs compared to traditional power plants.

https://en.wikipedia.org/wiki/Photovoltaic_power_station

THE POTENTIAL OF SOLAR POWER



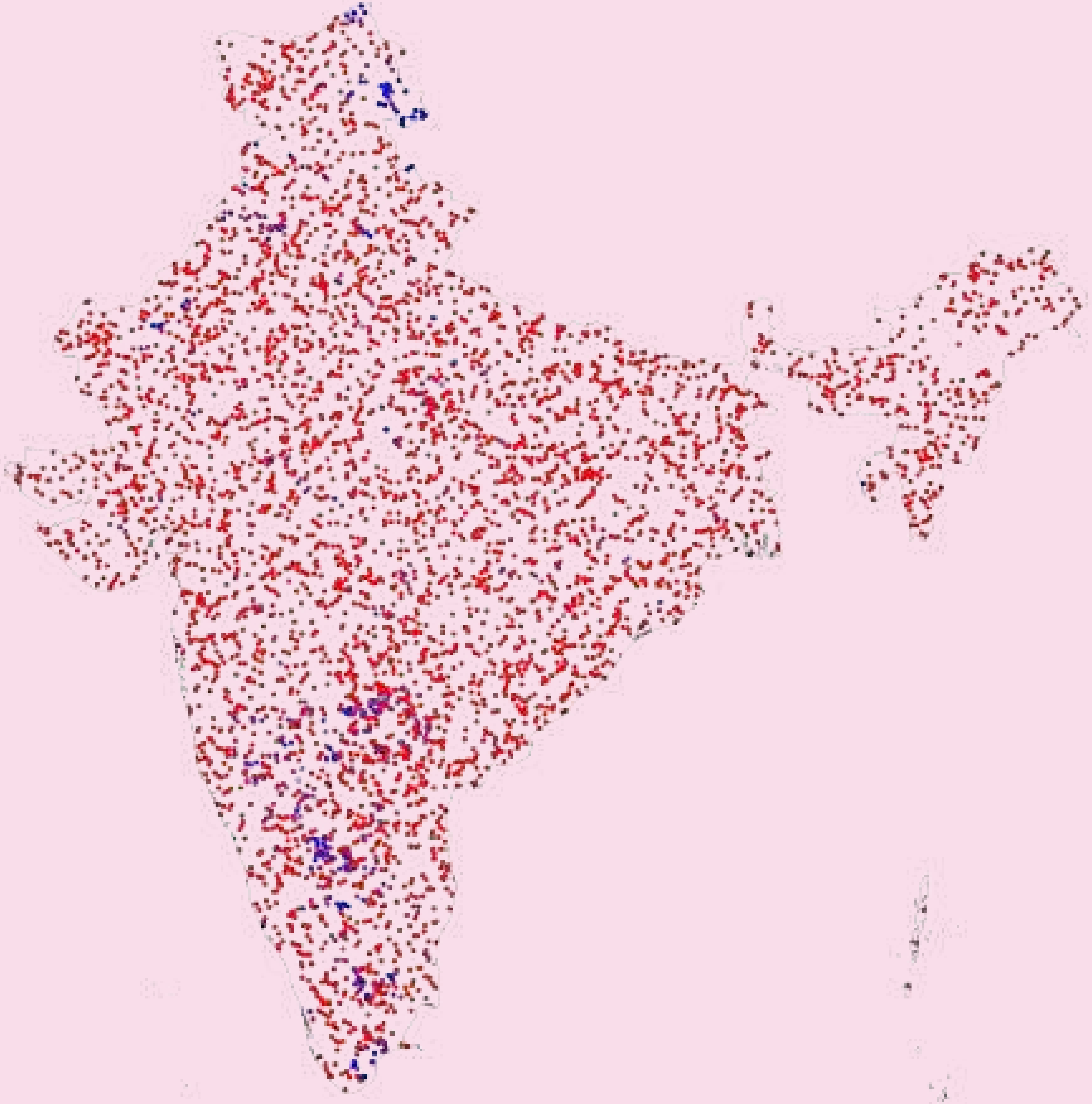
A SOLAR FARM IS A LARGE-SCALE INSTALLATION OF SOLAR PANELS THAT HARNESS SUNLIGHT TO GENERATE ELECTRICITY.

PROBLEM STATEMENT

PREDICTING OPTIMAL SOLAR FARM LOCATIONS IN INDIA!

PREDICTING OPTIMAL SOLAR FARM LOCATIONS IN INDIA!

WITH OUR MODEL, WE PLAN ON UNLOCKING INDIA'S SOLAR POTENTIAL!



WHY IS THIS RELEVANT?

REDUCES FINANCIAL RISK AND ATTRACTS CONSERVATIVE INVESTORS

ENABLES SMARTER LAND USE PLANNING AND ZONING

PROMOTES NATIONAL ENERGY INDEPENDENCE

LITERATURE REVIEW

TRADITIONAL APPROACHES?

- For the past two decades, earliest traceable mention in 2008, **GIS-MCDM** Models have dominated the market.
- A 2020 literature review published by Nabila Tabassum Suprova et al cites 46 studies over 20 different countries using MCDM-GIS methods in solar farm optimization from 2010 to 2020. (Suprova, Zidan, & Rashid, 2022, pg. 3-4)

GIS
(Geographic Information System)

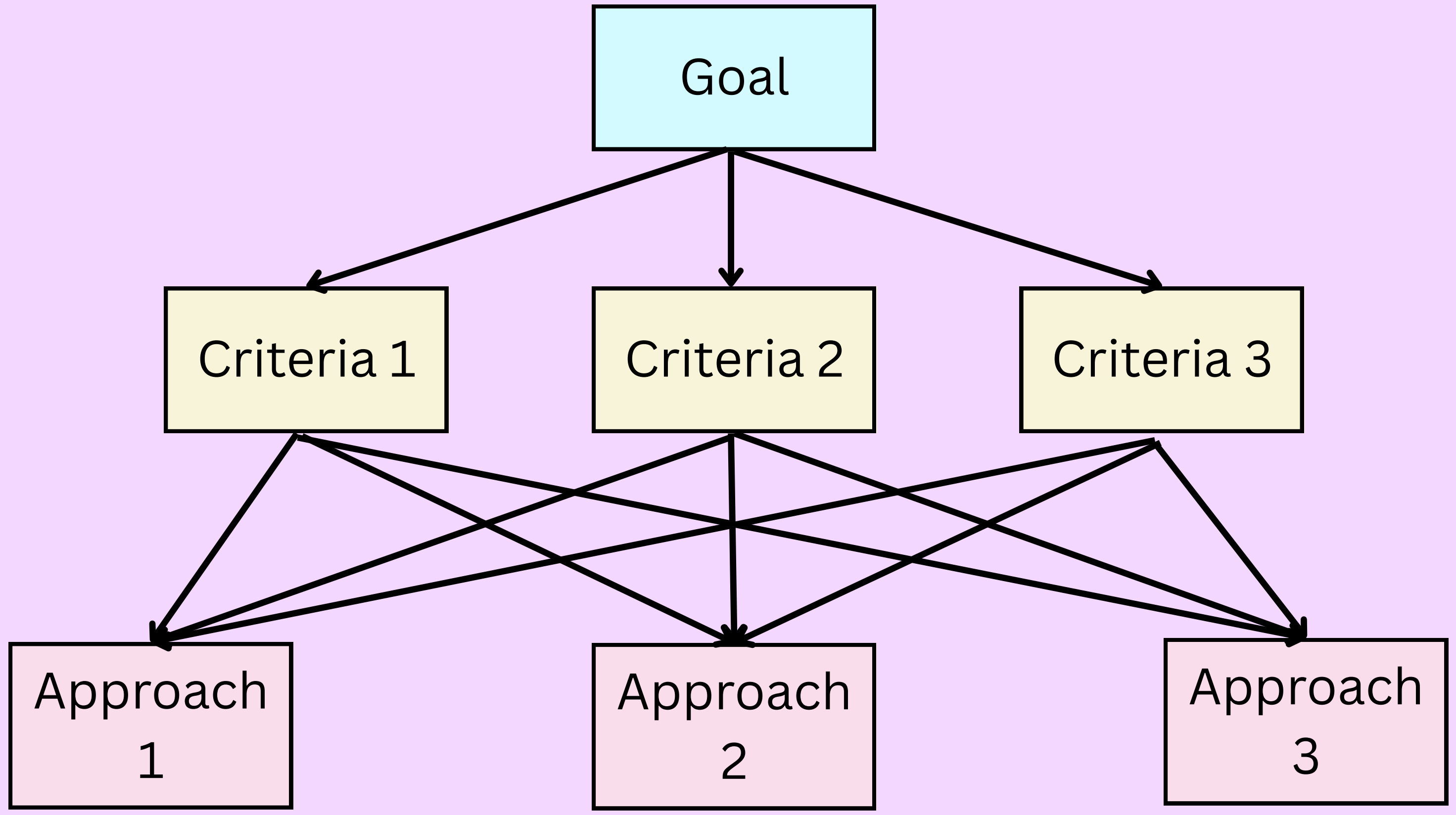
- create extensive databases
- use spatial data sources (maps, aerial photographs and quantitative and qualitative data)

MCDM
(Multi-Criteria Desision Making)

Consider multiple factors to identify the best possible solution according to the decision-maker's preferences.

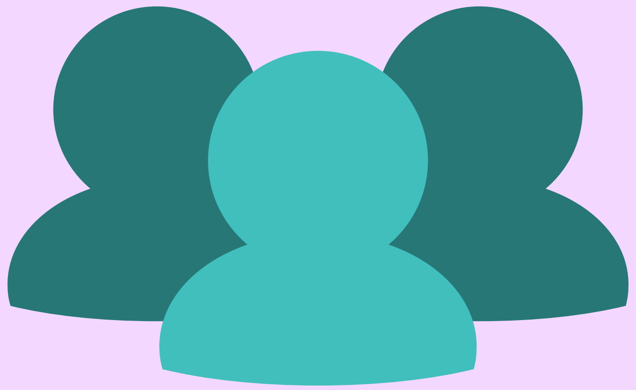
TRADITIONAL APPROACHES?

- AHP, a common MCDM approach in this domain , breaks the problem into a hierarchy structure, and pairwise comparisons are used to determine the relative importance of criteria and alternatives.(Mierzwiak & Całka, 2017, pg.24)
- Through AHP, the weight of particular criteria can be assigned through consultation of experts in the field thus making the criteria flexible and credible.



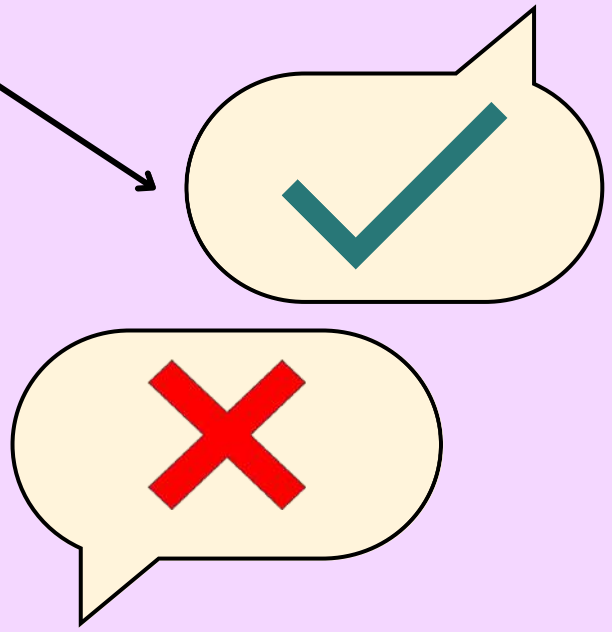
	Criteria 1	Criteria 2	Criteria 3
Criteria 1			
Criteria 2			
Criteria 3			

SHORTCOMINGS



Dependent on expert

Transparency

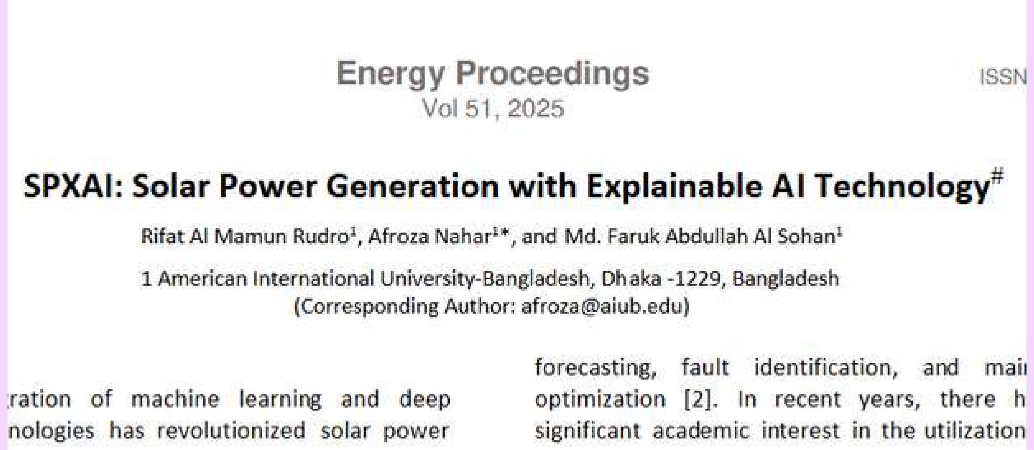


Subjectivity

Cognitive overload for large scale desision problems

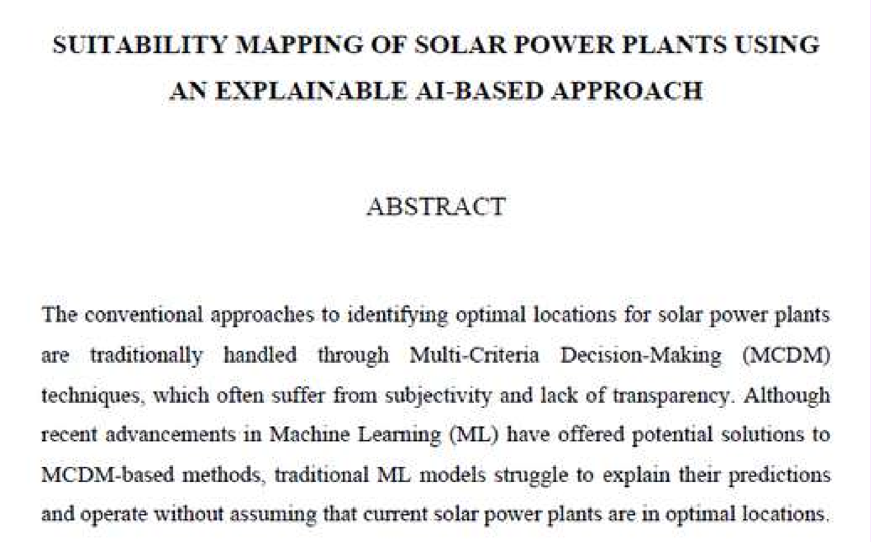
Alves, M. A., Meneghini, I. R., Gaspar-Cunha, A., & Guimarães, F. G. (2023). Machine learning-driven approach for large scale decision making with the analytic hierarchy process. Mathematics, 11(3), 627. https://doi.org/10.3390/math11030627
De Silva, M. H. R. M. (2024). Suitability mapping of solar power plants using an explainable AI-based approach (Master's thesis). NOVA Information Management School, Universidade Nova de Lisboa.
\url{https://run.unl.pt/bitstream/10362/165508/1/TGEO293.pdf}

AI ML BASED APPROACHES?


Energy Proceedings
Vol 51, 2025
ISSN
SPXAI: Solar Power Generation with Explainable AI Technology#
Rifat Al Mamun Rudro[1], Afroza Nahar[1]*, and Md. Faruk Abdullah Al Sohan[1]
1 American International University-Bangladesh, Dhaka -1229, Bangladesh
(Corresponding Author: afroza@aiub.edu)
ration of machine learning and deep nologies has revolutionized solar power
forecasting, fault identification, and mai optimization [2]. In recent years, there h significant academic interest in the utilization

(Rudro et al, 2025)

Rudro, R. A. M., Nahar, A., \& Al Sohan, M. F. A. (2025). SPXAI: Solar power generation with explainable AI technology. \textit{Energy Proceedings}, 51. American International University--Bangladesh.

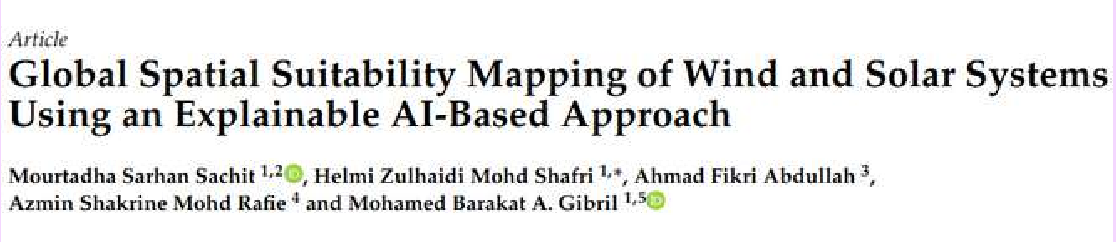

SUITABILITY MAPPING OF SOLAR POWER PLANTS USING AN EXPLAINABLE AI-BASED APPROACH

ABSTRACT

The conventional approaches to identifying optimal locations for solar power plants are traditionally handled through Multi-Criteria Decision-Making (MCDM) techniques, which often suffer from subjectivity and lack of transparency. Although recent advancements in Machine Learning (ML) have offered potential solutions to MCDM-based methods, traditional ML models struggle to explain their predictions and operate without assuming that current solar power plants are in optimal locations.

(De Silva, 2024)

De Silva, M. H. R. M. (2024). Suitability mapping of solar power plants using an explain-
able AI-based approach (Master's thesis). NOVA Information Management School, Uni-
versidade Nova de Lisboa.
https://run.unl.pt/bitstream/10362/165508/1/TGEO293.pdf


Article
Global Spatial Suitability Mapping of Wind and Solar Systems Using an Explainable AI-Based Approach
Mourtadha Sarhan Sachit [1,2], Helmi Zulhaidi Mohd Shafri [1,*], Ahmad Fikri Abdullah [3], Azmin Shakrine Mohd Rafie [4] and Mohamed Barakat A. Gibril [1,5]

(Sachit et al., 2022)

Sachit, M. S., Shafri, H. Z. M., Abdullah, A. F., Rafie, A. S. M., & Gibril, M. B. A. (2022). Global spatial suitability mapping of wind and solar systems using an explainable AI-based approach. ISPRS International Journal of Geo-Information, 11(8), 422. https://doi.org/10.3390/ijgi11080422


Spatial modelling the location choice of large-scale solar photovoltaic power plants: Application of interpretable machine learning techniques and the national inventory
June 2023 · Energy Conversion and Management 289(9)
DOI:10.1016/j.enconman.2023.117198
Authors:
Yanwei Sun
Ningbo University

(Sun, 2023)

Sun, Y. (2023). Spatial modelling the location choice of large-scale solar photovoltaic power plants: Application of interpretable machine learning techniques and the national inventory. Energy Conversion and Management, 289, 117198. https://doi.org/10.1016/j.enconman.2023.117198

SHORTCOMINGS

Paper	Approach of the paper	Shortcomings
(Rudro et al, 2025)	• Proposed SPXAI model collects power production data from solar farms and makes use of ML and DL models to analyze this data on an hourly basis.	• Provided a framework but did not implement the model to prove concept
(De Silva, 2024)	• De Silva integrated Explainable AI (XAI) and implemented 5 AI-ML approaches: RF, DT, MLP, KNN, SVM • Considered ground truth values by calculating solar efficiency of the solar powerplants near the coordinates	• Based on only US Data • Not scalable for Indian context
(Sachit et al., 2022)	• Perform binary classification • Existing sites=1, others=0 • Used GIS to extract environmental, social, technical and economic features • Used SVM, RF, MLP with SHAP.	• Had no validation method • Existing sites=1, others=0
(Sun, 2023)	• Implement Binary Classification and sets solar farm locations as one and randomly samples data to set for zero • Used XGBoost, RF and MLP. • Factors cover topography, climate, environment, infrastructure, and socioeconomics	• Existing sites=1, others=0 • Database specific to China

ABH MODEL BANANE KE LIYE, ITNA SAARA DATA KAAHAN SE AAYEGA?

OUR DATASET

WHERE DID WE GET OUR DATA POINTS FROM?

SOLAR FARM POINTS

Global-Solar-Power-Tracker-February-2025 by Global Energy Monitor organization

OTHER POINTS

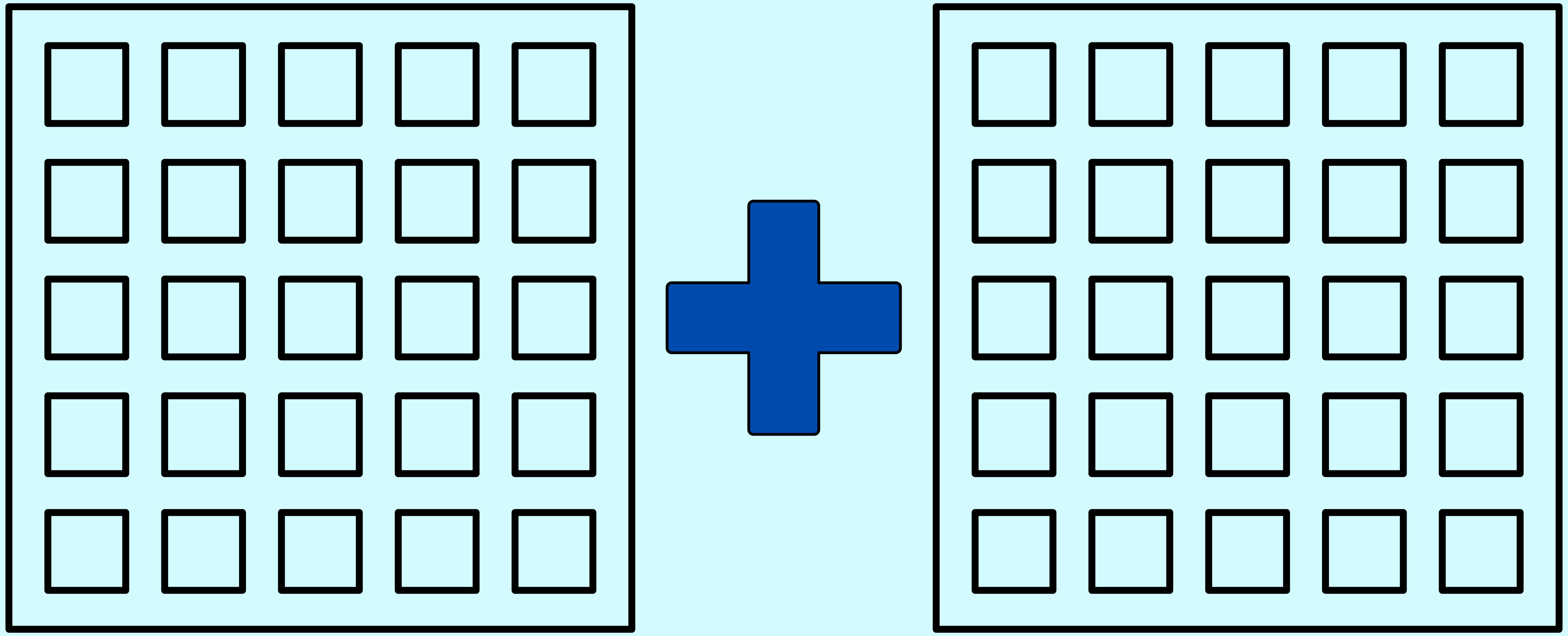
Randomly Sampled coordinates within India’s boundaries, at least 10m apart from each other

FEATURES EXTRACTION

S.No	Features	Sources
1	Average Annual Cloudy Days (Cloud Cover), Elevation, Slope, LULC, Vegetation, Aspect, LST (Land Surface Temperature), Solar Radiation	Earth Engine
2	Distance From Major Roads	Open Street Map
3	Solar Radiation, Average Annual temperature, Sunshine Duration, Wind speed, Dew point, UV index, Snowfall, dust, NO2, SO2, pm25, Water Vapour, Shortwave Radiation, Precipitation, Surface Pressure, Evapo Transpiration, Average Annual Humidity	Open Meteo
4	Lattitude, Longitude, Year comissioned	Global-Solar-Power-Tracker

TOTAL NUMBER OF ROWS: 6525

GROUND TRUTH VALUES BEING COMPUTED USING AHP!



AVERAGED VALUES FROM 2 DIFFERENT AHP MATRICES!

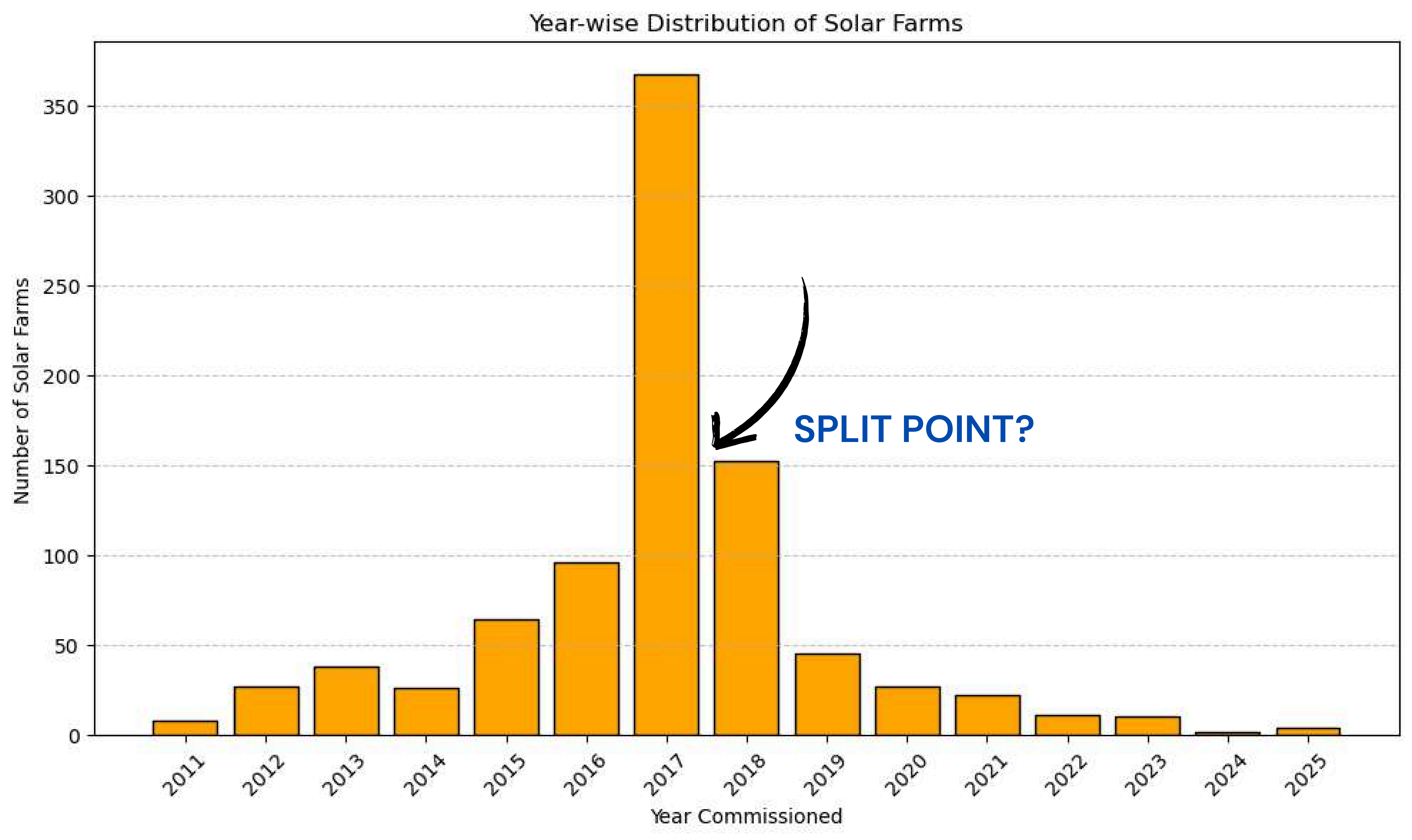
PREPROCESSING

1. LABEL ENCODING (LULC/ASPECT)
2. NORMALISATION
3. DIRECT INVERSE RELATIONSHIPS
4. CLASS IMBALANCE (UNDERSAMPLING MAJORITY CLASS)

```
CUSTOM_LULC_CLASSES = {
    5: "Built-up Areas", 4: "Water Bodies", 3: "Forest",
    2: "Vegetation", 1: "Wasteland/Barren Areas"
}
```

```
if aspect == -1:
    return 0, "Flat"
elif 0 <= aspect <= 22.5 or 337.5 < aspect <= 360:
    return 0, "North"
elif 22.5 < aspect <= 67.5:
    return 0.25, "Northeast"
elif 67.5 < aspect <= 112.5:
    return 0.5, "East"
elif 112.5 < aspect <= 157.5:
    return 0.75, "Southeast"
elif 157.5 < aspect <= 202.5:
    return 1, "South"
elif 202.5 < aspect <= 247.5:
    return 0.75, "Southwest"
elif 247.5 < aspect <= 292.5:
    return 0.5, "West"
elif 292.5 < aspect <= 337.5:
    return 0.25, "Northwest"
else:
    return None, "Unknown"
```

TEMPORAL VALIDATION?



ML METHODOLOGY

RANDOM FOREST

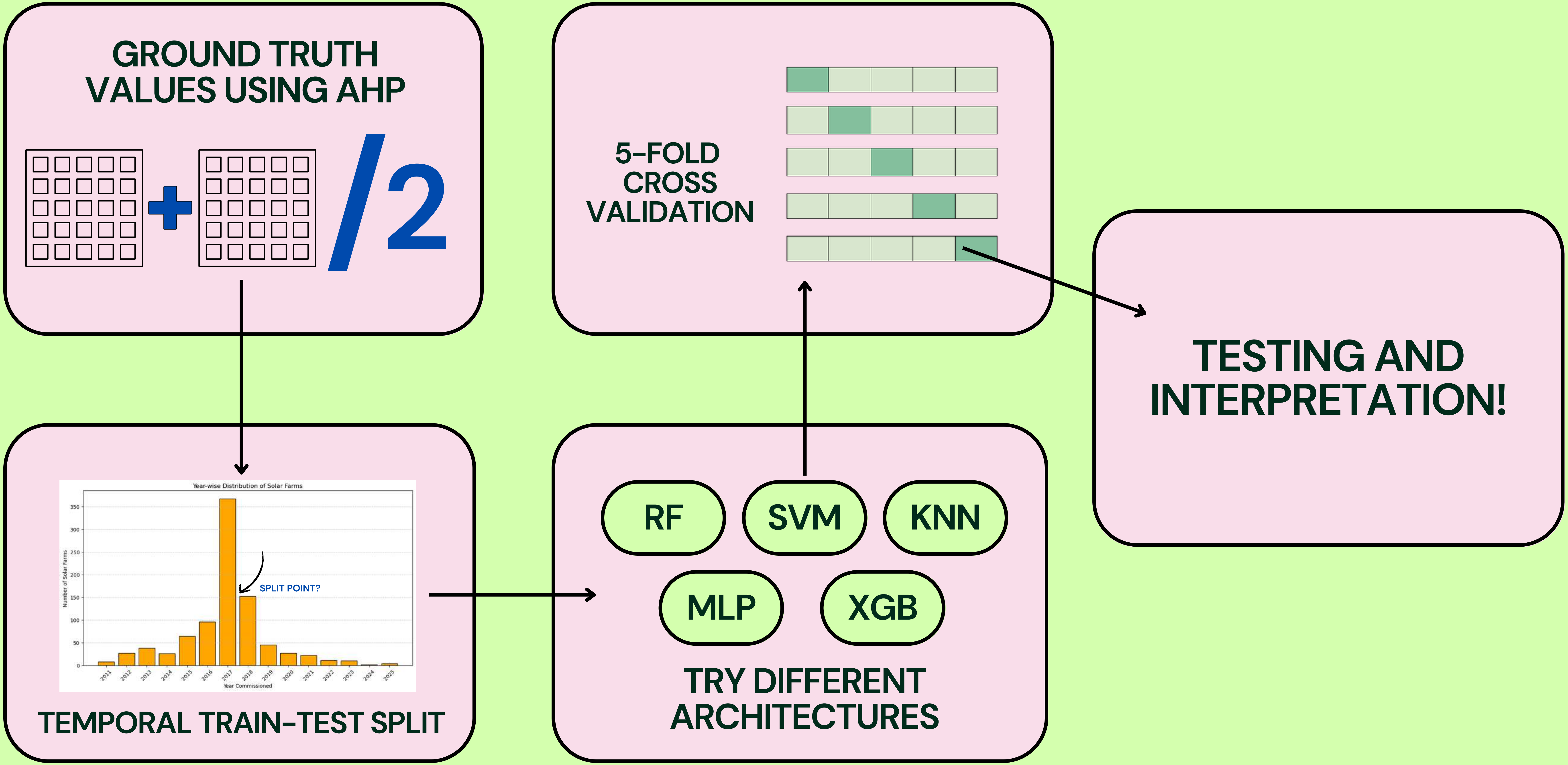
MULTI LAYER PERCEPTRON

SVM

XGBOOST

KNN

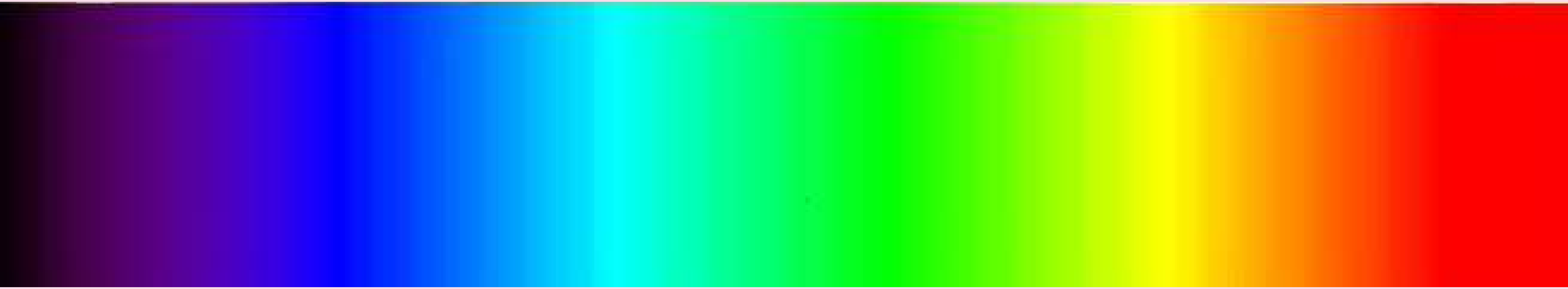
DIFFERENT REGRESSORS WE LOOKED AT


GROUND TRUTH VALUES USING AHP
/2
5-FOLD CROSS VALIDATION
TESTING AND INTERPRETATION!
Year-wise Distribution of Solar Farms
Number of Solar Farms
Year Commissioned
SPLIT POINT?
TEMPORAL TRAIN-TEST SPLIT
RF
SVM
KNN
MLP
XGB
TRY DIFFERENT ARCHITECTURES

PERFORMANCE METRICS AND VALIDATION

UNSUITABLE
0.2-0.4

SUITABLE
0.6-0.8



HIGHLY
UNSUITABLE
0-0.2

NEUTRAL
0.4-0.6

HIGHLY
SUITABLE
0.8-1

OUTPUT INTERPRETATION

R SQUARE

MSE

RMSE

PERFORMANCE METRICS

THE K-NEAREST NEIGHBORS (KNN) ALGORITHM IS A NON-PARAMETRIC, SUPERVISED LEARNING CLASSIFIER, WHICH USES PROXIMITY TO MAKE CLASSIFICATIONS OR PREDICTIONS ABOUT THE GROUPING OF AN INDIVIDUAL DATA POINT.

```
Average CV Results:
Average MSE: 0.0110
Average R²:  0.6804
```

METRIC	RANGE	OBTAINED VALUE
Mean Squared Error(MSE)	**[0,∞)**, The smaller the better	0.0201
Root Mean Squared Error(RMSE)	**[0,∞)**, The smaller the better	0.1417
R^2 Score	**(-∞,1]**, The greater the better	0.6890

https://www.ibm.com/think/topics/knn#:~:text=The%20k%2Dnearest%20neighbors%20(KNN)%20algorithm%20is%20a%20non,used%20in%20machine%20learning%20today.

A SUPPORT VECTOR MACHINE (SVM) IS A SUPERVISED MACHINE LEARNING ALGORITHM THAT CLASSIFIES DATA BY FINDING AN OPTIMAL LINE OR HYPERPLANE THAT MAXIMIZES THE DISTANCE BETWEEN EACH CLASS IN AN N-DIMENSIONAL SPACE.

```
Average CV Results:
Average MSE: 0.0073
Average R²:  0.7868
```

METRIC	RANGE	OBTAINED VALUE
Mean Squared Error(MSE)	**[0,∞)**, The smaller the better	0.0231
Root Mean Squared Error(RMSE)	**[0,∞)**, The smaller the better	0.1519
R^2 Score	**(-∞,1]**, The greater the better	0.6421

https://www.ibm.com/think/topics/support-vector-machine

AN MLP IS A TYPE OF FEEDFORWARD ARTIFICIAL NEURAL NETWORK WITH MULTIPLE LAYERS, INCLUDING AN INPUT LAYER, ONE OR MORE HIDDEN LAYERS, AND AN OUTPUT LAYER. EACH LAYER IS FULLY CONNECTED TO THE NEXT.

```
Average CV Results:
Average MSE: 0.0072
Average R²:  0.7913
```

METRIC	RANGE	OBTAINED VALUE
Mean Squared Error(MSE)	**[0,∞)**, The smaller the better	0.0187
Root Mean Squared Error(RMSE)	**[0,∞)**, The smaller the better	0.1367
R^2 Score	**(-∞,1]**, The greater the better	0.7095

https://www.shiksha.com/online-courses/articles/understanding-multilayer-perceptron-mlp-neural-networks/

RANDOM FOREST COMBINES THE OUTPUT OF MULTIPLE DECISION TREES TO REACH A SINGLE RESULT.


```
Average CV Results:
Average MSE: 0.0058
Average R²:  0.8306
```

METRIC	RANGE	OBTAINED VALUE
Mean Squared Error(MSE)	**[0,∞)**, The smaller the better	0.0147
Root Mean Squared Error(RMSE)	**[0,∞)**, The smaller the better	0.1212
R^2 Score	**(-∞,1]**, The greater the better	0.7729

https://www.ibm.com/think/topics/random-forest#:~:text=Random%20forest%20is%20a%20commonly,both%20classification%20and%20regression%20problems.

XGBOOST (EXTREME GRADIENT BOOSTING) IS A DISTRIBUTED, OPEN-SOURCE MACHINE LEARNING LIBRARY THAT USES GRADIENT BOOSTED DECISION TREES, A SUPERVISED LEARNING BOOSTING ALGORITHM THAT MAKES USE OF GRADIENT DESCENT.

```
Average CV Results:
Average MSE: 0.0048
Average R²:  0.8614
```

METRIC	RANGE	OBTAINED VALUE
Mean Squared Error(MSE)	**[0,∞)**, The smaller the better	0.0135
Root Mean Squared Error(RMSE)	**[0,∞)**, The smaller the better	0.1161
R^2 Score	**(-∞,1]**, The greater the better	0.7915

https://www.ibm.com/think/topics/xgboost

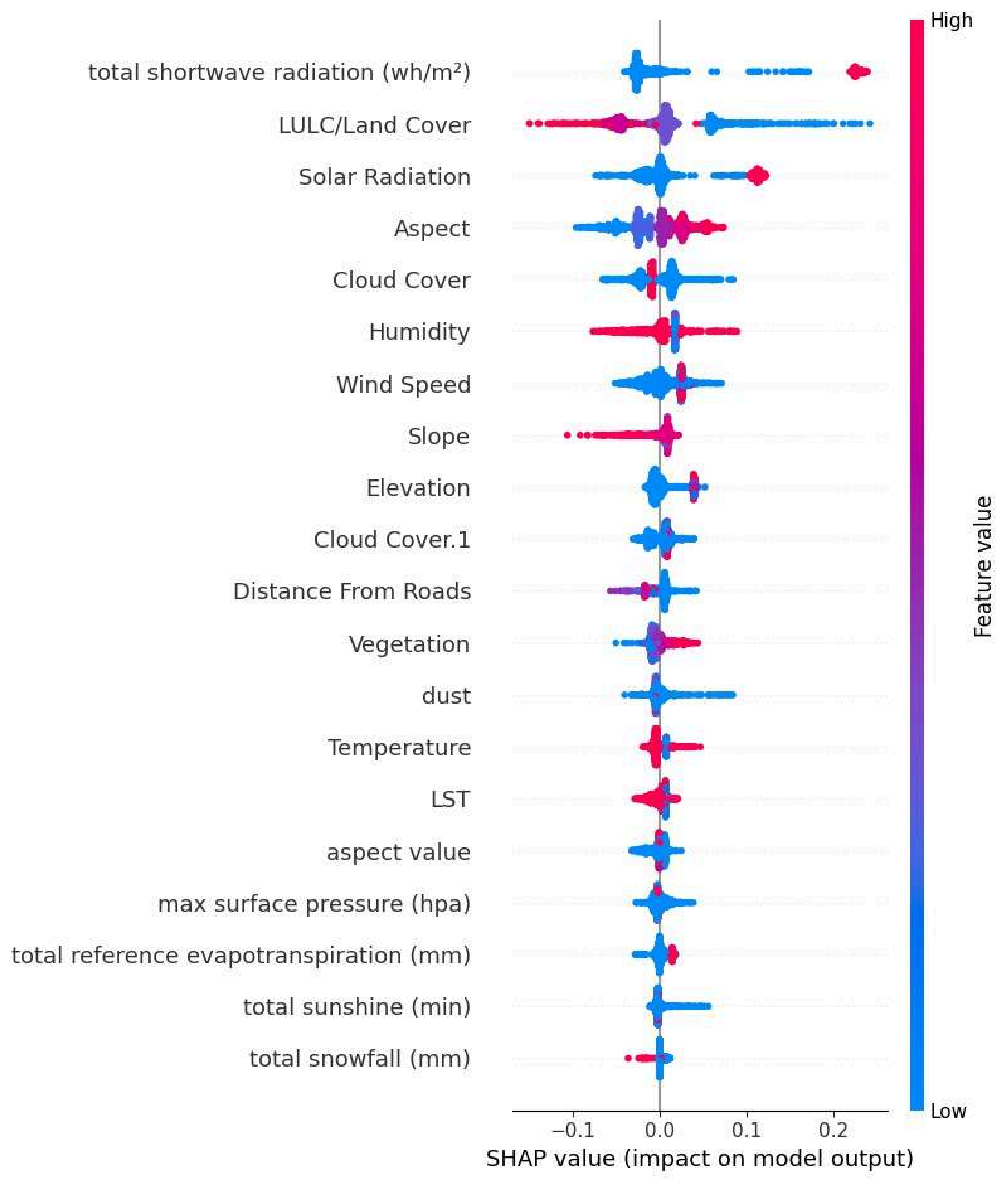
CONCLUSION:
XGBOOST WINS!

VALIDATING OUR RESULTS

POST TRAINING AND TESTING, WHEN TAKING A THRESHOLD OF >0.5, WE HAD 20 DATA POINTS OF MISCLASSIFICATIONS. WE TRIED TO INTERPRET THIS THROUGH SHAP

SHAP RESULTS

Rank	Feature	Importance (%)
1	total shortwave radiation (wh/mÂ²)	17.6
2	LULC/Land Cover	14.3
3	Solar Radiation	9.07
4	Aspect	8.85
5	Cloud Cover	6.06
6	Humidity	5.44
7	Wind Speed	5.34
8	Slope	3.99
9	Elevation	3.57
10	Cloud Cover.1	3.33
11	Distance From Roads	2.91
12	Vegetation	2.77
13	dust	2.39
14	Temperature	2.25
15	LST	2.17
16	aspect value	1.89
17	max surface pressure (hpa)	1.88
18	total reference evapotranspiration (r	1.59
19	total sunshine (min)	1.27
20	total snowfall (mm)	0.6
21	max dew point (Â°c)	0.58
22	max precipitation (mm)	0.58
23	pm2.5	0.55
24	so2	0.53
25	no2	0.5



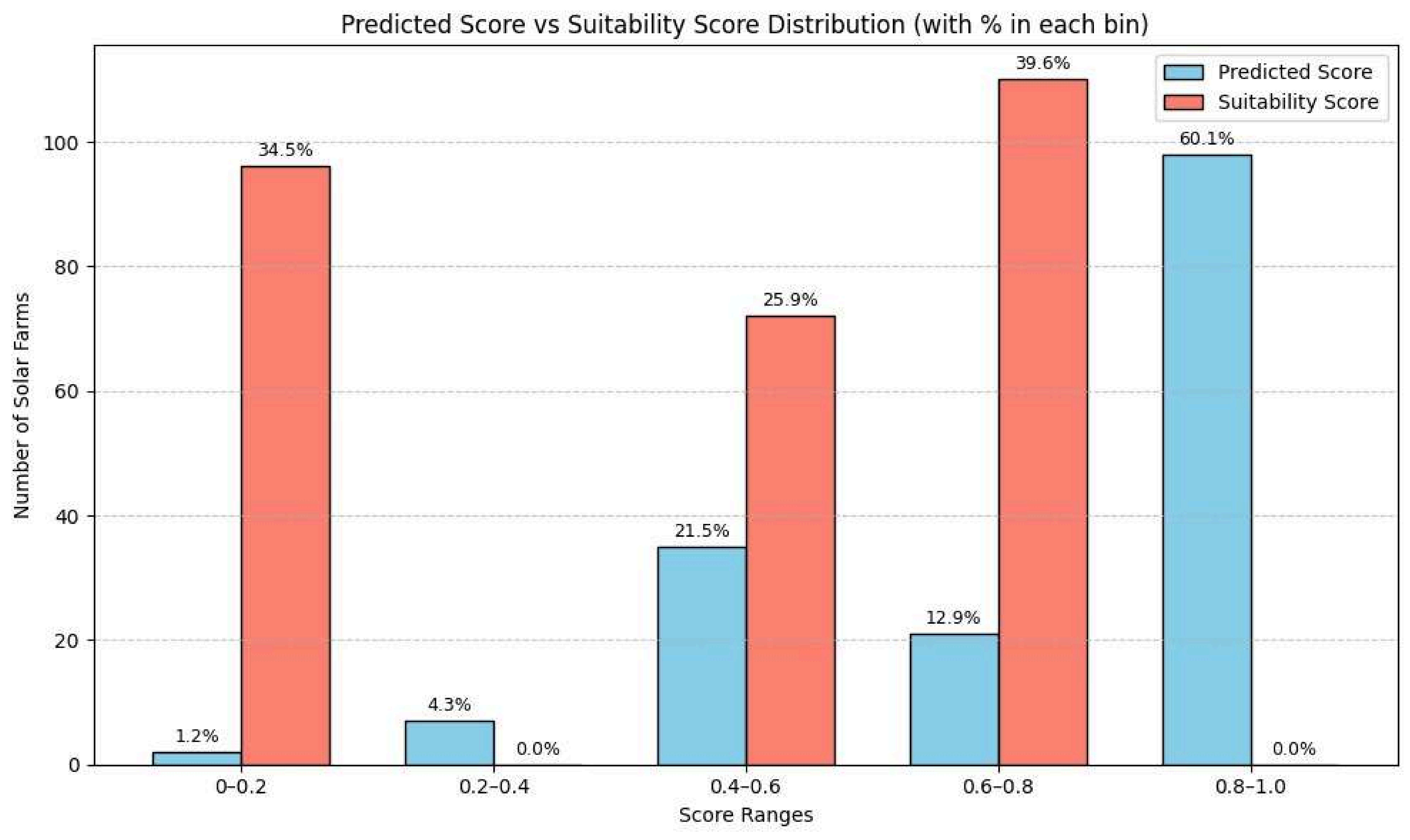
- SHAP PROVIDES CONTEXT OF WHICH FEATURES DICTATED OUR MODEL’S PREDICTIONS FOR A LOCATION
- IT CAN’T HELP IDENTIFY WHETHER THE MODEL’S ASSIGNED WEIGHTS, FEATURES AND INTERPRETATIONS ARE CORRECT

THUS EXPLAINING MISSCLASSIFICATIONS USING SHAP WAS NOT AN APPROPRIATE VALIDATION METHOD.

FINAL VALIDATION METHOD

RUN AHP ON ALL SOLAR FARM COORDINATES TO OBTAIN SUITABILITY SCORES

COMPARE AHP SUITABILITY SCORES WITH MODEL PREDICTIONS AND IDENTIFY WHICH CLASSIFIES THE SOLAR FARM LOCATIONS MORE ACCURATELY


Predicted Score vs Suitability Score Distribution (with % in each bin)
Predicted Score
Suitability Score
Number of Solar Farms
0
20
40
60
80
100
1.2%
34.5%
4.3%
0.0%
21.5%
25.9%
12.9%
39.6%
60.1%
0.0%
0–0.2
0.2–0.4
0.4–0.6
0.6–0.8
0.8–1.0
Score Ranges

CONCLUSION: AI-ML WINS!

EXPERT OPINION

BROOKINGS

EXPERT

Rahul Tongia

Nonresident Senior Fellow – Foreign Policy, Energy Security and Climate Initiative

DEPLOYABILITY
(ALSO, LIMITATIONS)

LIGHTWEIGHT

INTERPRETABLE!

SCALABLE!

FAST!

DEPLOYABILITY!

LIMITATIONS

NEED BETTER WAYS OF OBTAINING GROUND TRUTHS (AS PER EXPERT'S SUGGESTION)

NEED BETTER AND MORE DETAILED VALIDATION BY DOMAIN EXPERTS

NEED A BROADER RANGE OF FEATURES (SOCIAL, ECONOMIC AND POLITICAL ASPECTS) (EXPERT'S SUGGESTION)

MAJOR ASSUMPTION: PLACES WHERE SOLAR FARMS EXIST RIGHT NOW ARE OPTIMAL LOCATIONS

FINAL THOUGHTS:

1) AI-ML IS A WORTHY PLAYER IN THE MARKET

(BEWARE GIS-MCDM)

2) WE KNOW TOO MUCH ABOUT SOLAR FARMS FOR A BUNCH OF 20 YEARS OLDS

THANK YOU! QUESTIONS?