EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI)

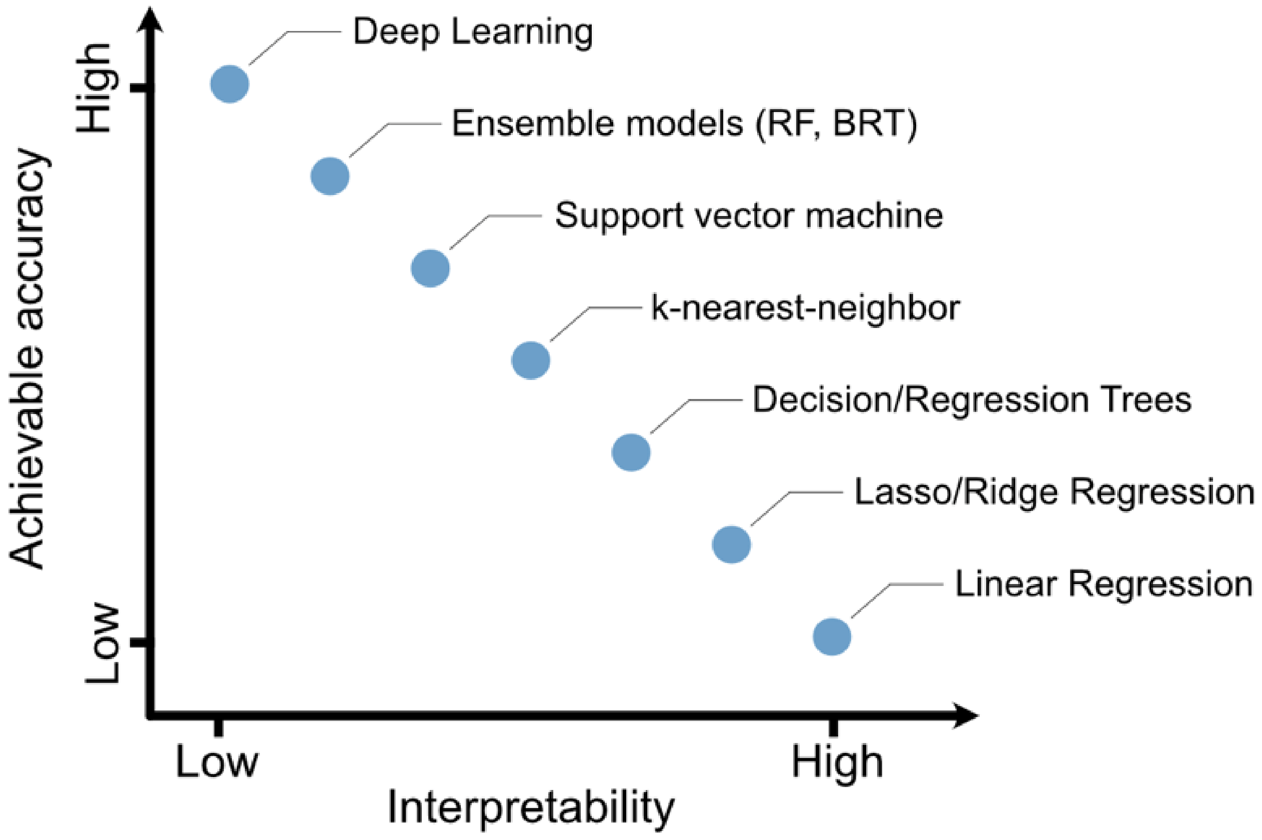
Jonnalagadda
Naveen Reddy

INTRODUCTION

Explainable AI (XAI) refers to the set of techniques, methods, and practices used to make the decision making processes of artificial intelligence systems more **understandable**, **interpretable**, and **transparent to humans**.

The primary goal of XAI is to provide insights into why AI models make specific predictions or decisions, especially in situations where AI is used in critical applications, such as healthcare, finance, autonomous vehicles, and law enforcement.

ACCURACY VS INTERPRETABILITY



EXAMPLE

Understanding the relationship between land use/land cover (LULC) and land surface temperature (LST) :

Independent variables :

1. Landscape features : NDBI, NDWI, GNDVI

2. Topographic features : Elevation, Slope

3. LULC features : Urbanized area, Land Cover Map,
Agricultural area, Forest area,
Grassland area, Wetland area, Bareland area,
Water area.

XGBOOST

We developed the LST prediction model and estimated the LST reduction effects after specific LULC changes.

Results showed that the prediction accuracy of LST was maximized when landscape, topographic, and LULC features within a 150 m buffer radius were adopted as independent variables.

Do all the variables show equal importance in final prediction ?

BLACK BOX

SHAP

Shaply Additive exPlanations

These values assign an importance score to each feature in a prediction.

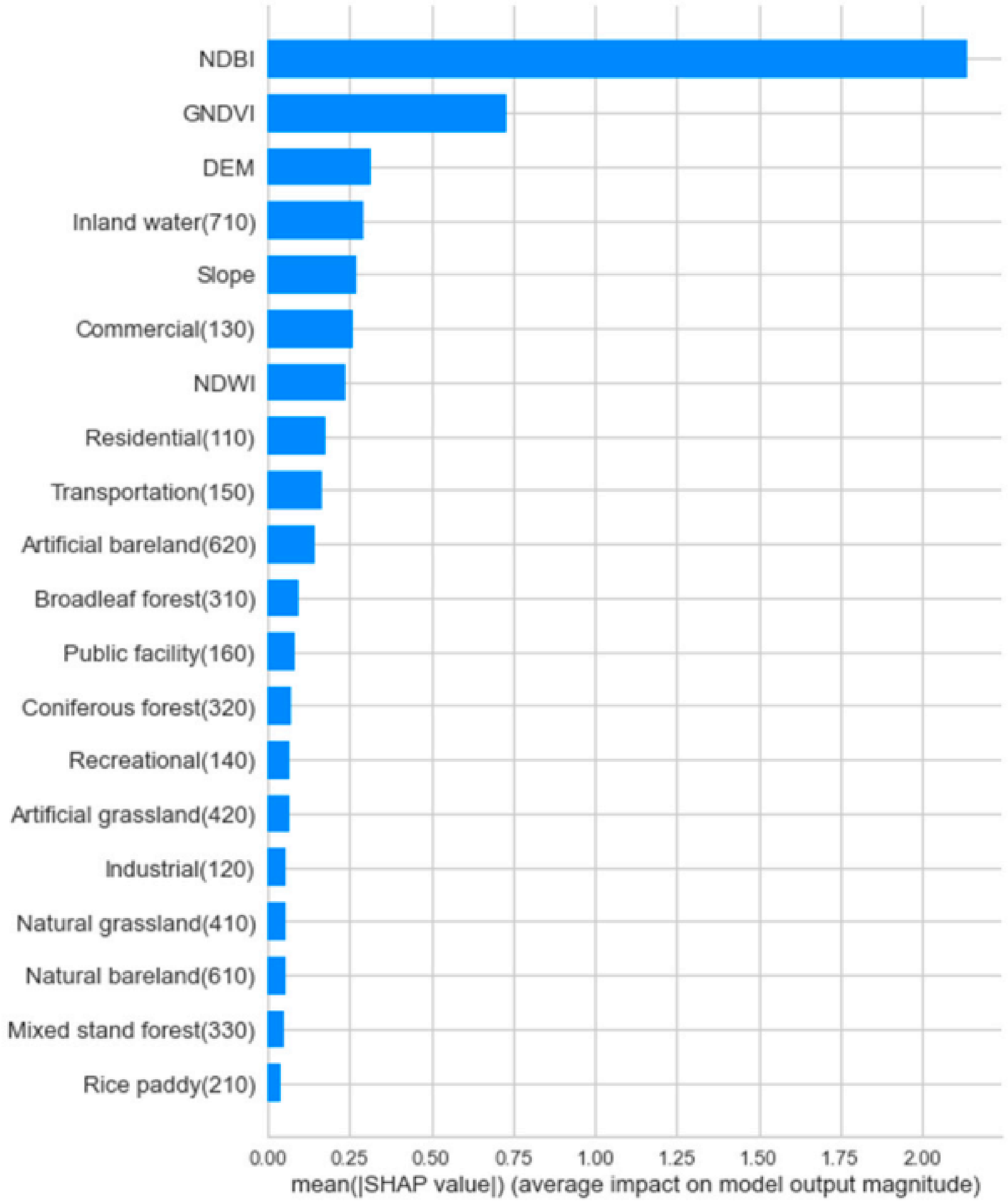
These scores indicate how much each feature contribute to the final prediction. If a feature is irrelevant (i.e., its value doesn't affect the prediction), its SHAP value will be close to zero.

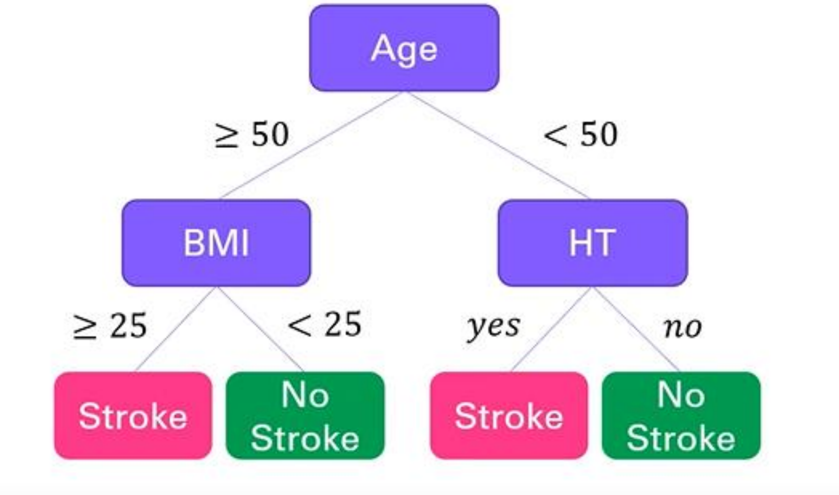
SHAP

Shaply Additive exPlanations

These values assign an importance score to each feature in a prediction.

These scores indicate how much each feature contribute to the final prediction. If a feature is irrelevant (i.e., its value doesn't affect the prediction), its SHAP value will be close to zero.


NDBI
GNDVI
DEM
Inland water(710)
Slope
Commercial(130)
NDWI
Residential(110)
Transportation(150)
Artificial bareland(620)
Broadleaf forest(310)
Public facility(160)
Coniferous forest(320)
Recreational(140)
Artificial grassland(420)
Industrial(120)
Natural grassland(410)
Natural bareland(610)
Mixed stand forest(330)
Rice paddy(210)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
mean(|SHAP value|) (average impact on model output magnitude)


Age
≥ 50
< 50
BMI
HT
≥ 25
< 25
yes
no
Stroke
No Stroke
Stroke
No Stroke


age <= 56.50
Obs: 3889, 3889
age <= 47.50
Obs: 2778, 509
age <= 67.50
Obs: 1111, 3380
bmi <= 5.15
Obs: 2218, 161
Obs: 560, 348
Obs: 559, 900
Obs: 552, 2480
Impurity: 0.26

OBJECTIVE

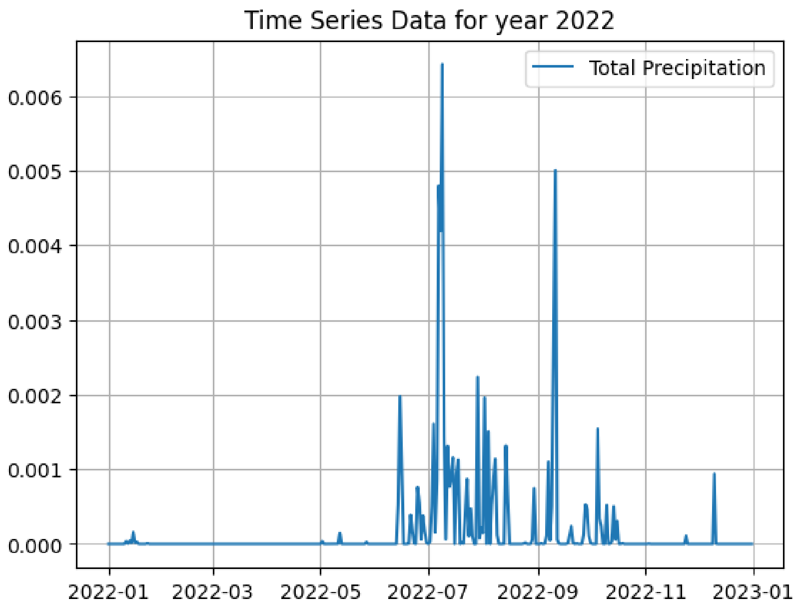
To develop a Long term Precipitation Prediction Model with Explainable Artificial Intelligence (XAI) for transparent decision-making in weather forecasting.

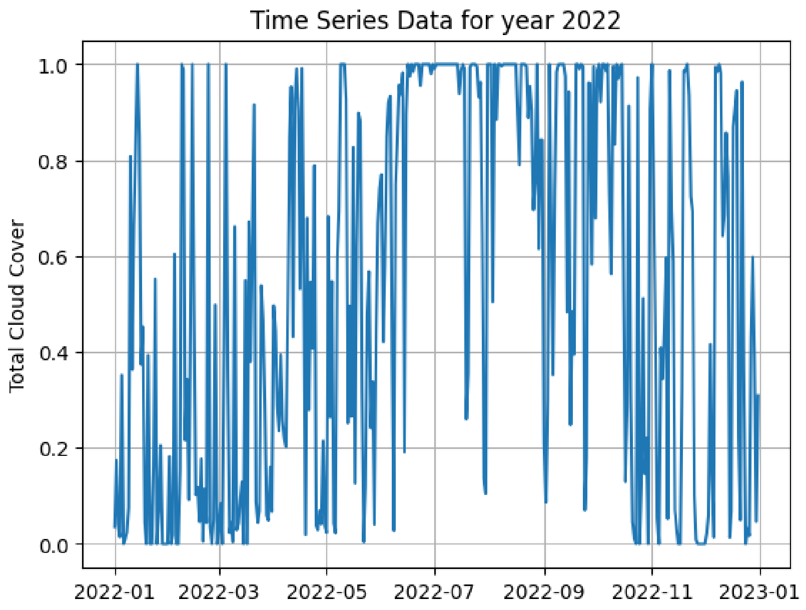
Approach

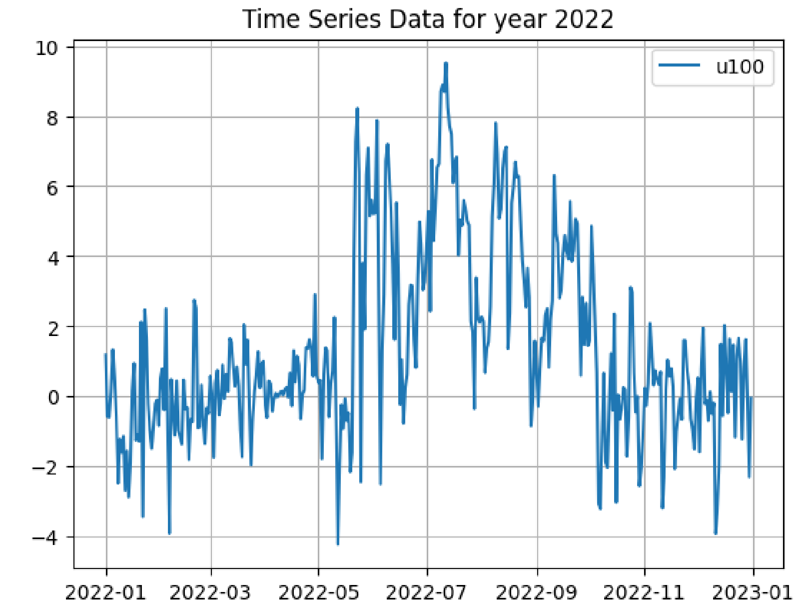
- Data Collection
- Data Preprocessing
- Model Evaluation
- Precipitation prediction
- Explanations
- Visualization of results and explanations

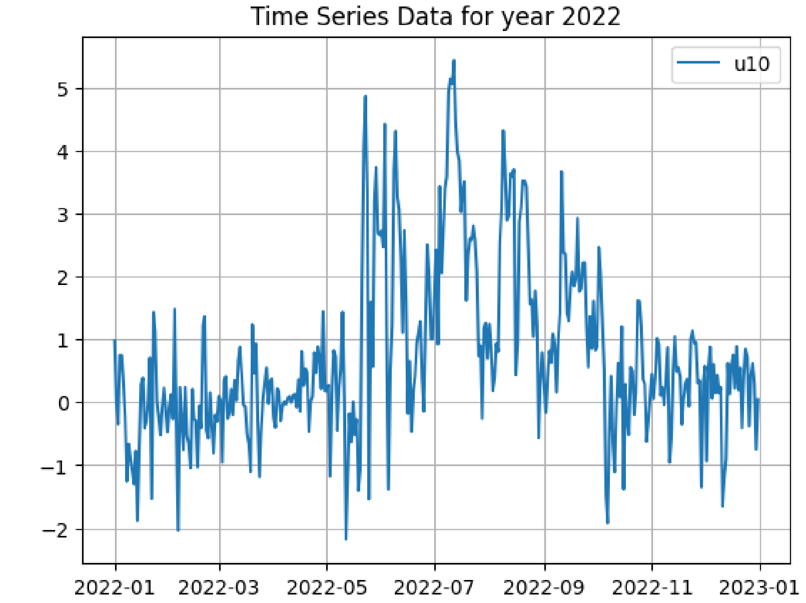
DATA COLLECTION

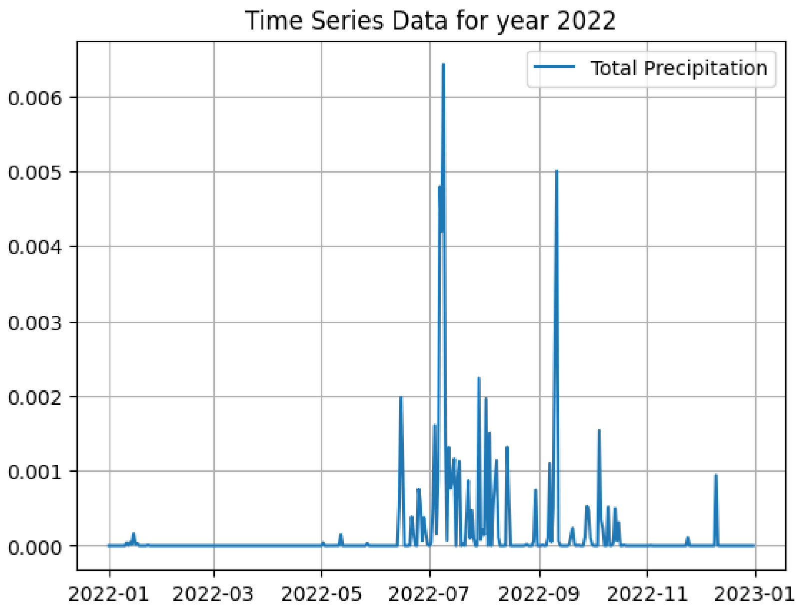
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	time	longitude	latitude	u100	v100	u10	v10	d2m	t2m	sst	ssr	sp	sshf	tcc	tp
2	2000-01-01 00:00:00	80	18	-0.04604	3.880129	0.0641	1.617788	290.3942	291.9719	0	3.6E-12	98190.48	44414.41	0.172509	1.77E-06
3	2000-01-02 00:00:00	80	18	-0.17101	3.917302	0.026193	1.443589	288.931	290.5139	0	3.6E-12	98176.26	45220.3	0.096104	0
4	2000-01-03 00:00:00	80	18	-0.14958	-2.38253	0.013973	-0.83511	284.0286	290.2008	0	3.6E-12	98122.05	36574.59	0.099248	0
5	2000-01-04 00:00:00	80	18	-0.99501	-0.30082	-0.61872	0.493489	281.1429	290.0224	0	3.6E-12	98047.63	16701.22	0.004456	0
6	2000-01-05 00:00:00	80	18	0.360212	3.362156	0.537934	1.630868	284.8356	287.5365	0	3.6E-12	98167.65	28570.48	0.369936	0
7	2000-01-06 00:00:00	80	18	0.365766	-3.13441	0.63719	-1.36934	281.9846	287.7187	0	3.6E-12	98171.42	50626.8	0.004517	0
8	2000-01-07 00:00:00	80	18	-0.68358	-5.91119	0.104999	-2.41951	278.815	287.7191	0	3.6E-12	98150.71	100936.2	0.038881	0
9	2000-01-08 00:00:00	80	18	0.351087	2.215744	0.766123	1.27375	280.9271	286.8779	0	3.6E-12	98211.39	28406.17	0	0
10	2000-01-09 00:00:00	80	18	0.479627	5.196276	0.273585	1.86777	287.538	288.8269	0	3.6E-12	98127.28	84333.3	0.073551	0
11	2000-01-10 00:00:00	80	18	-0.08611	5.586947	0.019959	2.416735	287.6329	289.1188	0	3.6E-12	98036.88	89262.52	0.057132	0
12	2000-01-11 00:00:00	80	18	1.380599	4.337782	0.658886	1.777245	289.4125	290.1498	0	3.6E-12	98108.9	58537.03	0	0
13	2000-01-12 00:00:00	80	18	0.003552	5.408094	0.122955	2.46096	288.7827	289.8275	0	3.6E-12	98030.59	65070.21	0	0
14	2000-01-13 00:00:00	80	18	0.201917	5.2843	0.053626	2.433969	291.447	292.5409	0	3.6E-12	98042.98	60180.1	0.03	0
15	2000-01-14 00:00:00	80	18	1.070357	3.911691	0.390796	1.747347	291.5427	292.5215	0	3.6E-12	97842.92	40533.63	0.154518	0
16	2000-01-15 00:00:00	80	18	0.150739	5.417212	-0.0788	2.689142	291.9762	292.8744	0	3.6E-12	97835.47	53208.77	0.004456	0
17	2000-01-16 00:00:00	80	18	-0.45784	4.253966	-0.22793	2.16177	292.7446	293.6537	0	3.6E-12	98023.82	26927.41	0.497505	0
18	2000-01-17 00:00:00	80	18	0.62126	4.104221	0.252636	1.761466	290.7041	291.9996	0	3.6E-12	98177.71	37677.81	0.03	0
19	2000-01-18 00:00:00	80	18	1.895553	1.466315	1.014761	1.19589	289.5742	292.393	0	3.6E-12	98352.32	15245.92	0	0
20	2000-01-19 00:00:00	80	18	-0.92202	-3.4304	-0.56784	-1.22919	287.3499	293.9274	0	3.6E-12	98239.27	26207.58	0	0
21	2000-01-20 00:00:00	80	18	0.967604	-2.65081	1.302303	-0.99063	284.208	291.8271	0	3.6E-12	97970.39	20198.63	0	0
22	2000-01-21 00:00:00	80	18	0.941817	5.841549	0.045147	2.466358	284.2791	292.1309	0	3.6E-12	97943.58	106616.5	0.004456	0
23	2000-01-22 00:00:00	80	18	0.036084	6.06003	-0.08778	3.259907	292.9136	293.2955	0	3.6E-12	97999.23	62120.5	0.4719	0
24	2000-01-23 00:00:00	80	18	-0.5122	4.386528	-0.40749	2.047991	292.2089	292.7146	0	3.6E-12	98017.72	34853.28	0.376345	0
25	2000-01-24 00:00:00	80	18	0.468122	0.934315	0.735448	0.894209	287.645	293.817	0	3.6E-12	98137.06	6521.984	0	0
26	2000-01-25 00:00:00	80	18	0.766859	4.452458	0.438429	1.717656	288.8885	291.0909	0	3.6E-12	98287.56	41308.22	0.428715	0

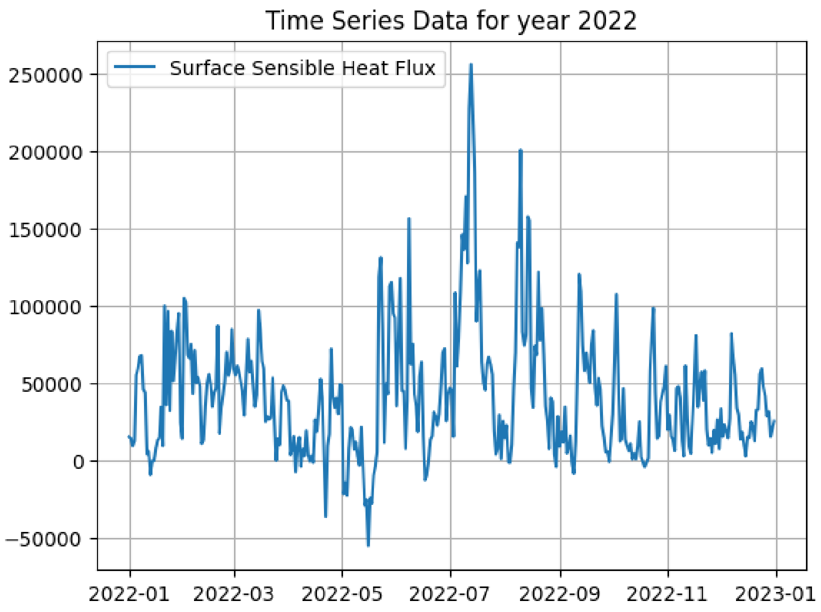

Time Series Data for year 2022
Total Precipitation
0.006
0.005
0.004
0.003
0.002
0.001
0.000
2022-01
2022-03
2022-05
2022-07
2022-09
2022-11
2023-01

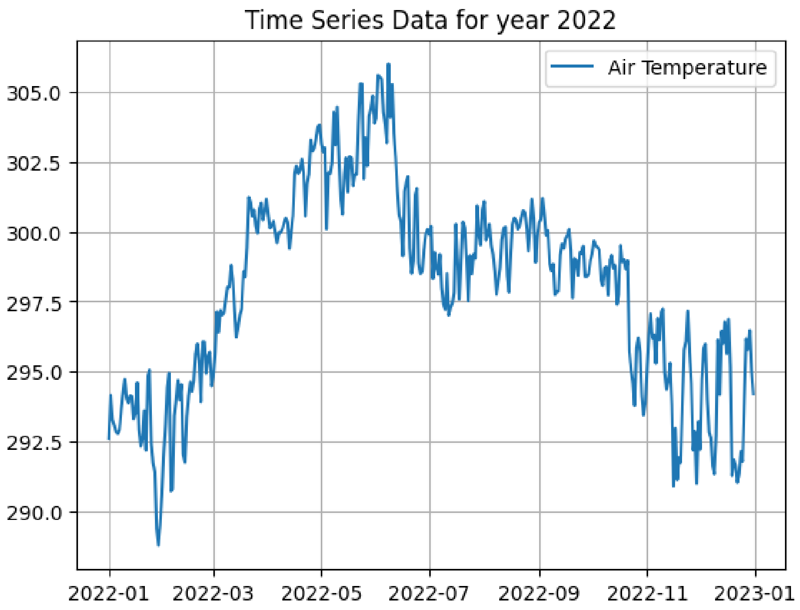

Time Series Data for year 2022
Total Cloud Cover
1.0
0.8
0.6
0.4
0.2
0.0
2022-01
2022-03
2022-05
2022-07
2022-09
2022-11
2023-01


Time Series Data for year 2022
u100
10
8
6
4
2
0
−2
−4
2022-01
2022-03
2022-05
2022-07
2022-09
2022-11
2023-01


Time Series Data for year 2022
u10
5
4
3
2
1
0
−1
−2
2022-01
2022-03
2022-05
2022-07
2022-09
2022-11
2023-01


Time Series Data for year 2022
Total Precipitation
0.006
0.005
0.004
0.003
0.002
0.001
0.000
2022-01
2022-03
2022-05
2022-07
2022-09
2022-11
2023-01


Time Series Data for year 2022
Surface Sensible Heat Flux
250000
200000
150000
100000
50000
0
−50000
2022-01
2022-03
2022-05
2022-07
2022-09
2022-11
2023-01


Time Series Data for year 2022
Air Temperature
305.0
302.5
300.0
297.5
295.0
292.5
290.0
2022-01
2022-03
2022-05
2022-07
2022-09
2022-11
2023-01

Shapley Additive Explanations (SHAP)

The SHAP is one of the representative explainable artificial intelligence (XAI) methods that have been widely used to increase the interpretability of machine learning models.

The SHAP value quantifies the relative importance of each independent variable for the model outcome based on its marginal contribution.

For n features, the SHAP value ϕ_i assigned to each feature *i* is represented as below:

$$\phi_i = \sum_{s \in F} \frac{|S|!\,(|F| - |s| - 1)!}{|\{\mathrm{i}\}|F|!} \left[f_{s \cup \{i\}}\left(x_{s \cup \{i\}}\right) - f_s(x_s) \right]$$

where F represents all the combinable features in S, and $f_{s \cup \{i\}}\left(x_{s \cup \{i\}}\right) - f_s(x_s)$ calculates the difference between the contributions when the feature *i* is used and when it is not used.

THANK YOU