READ ME File()

  This project , Solar Energy generation and Air Quality Prediction, Sustainability Investment using Data Analytics and AI Algorithms

Has used the below ML, AI, DL Methodologies. To generate solar PV energy production , predict the Air Quality prediction and Sustainability ESG analysis using NLP and data analytics/

The code was built on, Google colab with Python 3.8

tensorflow 1.x in 2019

**LIME & SHAP with XGBOOST ,ML ALGORITHMS**

Repository for the DeepSHAP experiments.

**Prerequisites**

* Python, NumPy, Tensorflow, Keras, XGBoost.

**Experiments**

Experiments for evaluating baseline distributions are in:

* 1\_multiple\_references/

Experiments for evaluating series of models are in:

* 3\_loss\_explanation/
* 4\_feature\_extraction/
* 5\_model\_stack/

**Code**

Code underlying the experiments and implementations of DeepSHAP for our specific applications is found in deepshap/.

**Dataset availability**

Datasets are highlighted in the project document

* UNISOLAR is a large-scale open dataset on University data

[Link Source [1]](https://github.com/CDAC-lab/UNISOLAR)

* Power Plant Datasets

Link Source (Git hub link)-- [Solar Power Generation Data](https://www.kaggle.com/datasets/anikannal/solar-power-generation-data?select=Plant_1_Generation_Data.csv)

* Solcast Dataset

Link Source:

<https://solcast.com/blog/bankability-report-and-validation-of-solcast-data>)

* G7 countries RET dataset
* [countrygroups/data/g7.csv at main · openclimatedata/countrygroups](https://github.com/openclimatedata/countrygroups/blob/main/data/g7.csv)
* Link Source(Github link)
* ESG Bank Datasets

Source Link::: [Sovereign ESG Data Portal](https://esgdata.worldbank.org/?lang=en)

* Air Quality Dataset

[Air Quality Forecasting Models: Real-time and Historical Analysis](https://aqicn.org/forecast/models/)

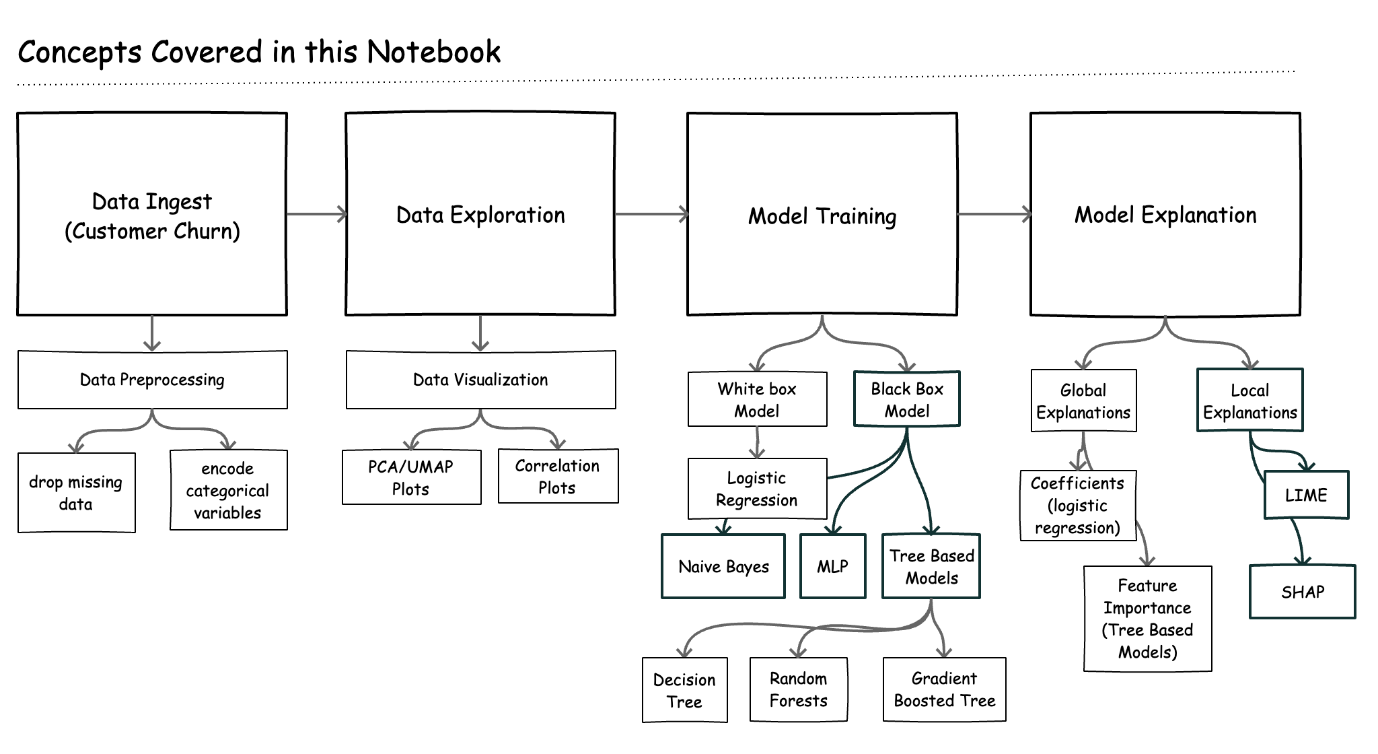
[India NCAP – Air Quality Index (AQI) in Indian Cities 2015-2023 - UrbanEmissions.Info](https://urbanemissions.info/india-air-quality/india-ncap-aqi-indian-cities-2015-2023/)

**Explaining Models with LIME and SHAP**

Logistic Regression, Decision Tree, Random Forest, Gradient Boosted Tree, Multilayer Perceptron) using LIME and SHAP.

The model interpretability as an important step in the data science workflow. Being able to explain how a model works serves many purposes, including building trust in the model's output, satisfying regulatory requirements, model debugging, and verifying model safety, amongst other things. In this article, we revisit two industry standard algorithms for interpretability - LIME and SHAP. We discuss how these two algorithms work, and show some code examples of how to implement them in Python. An overview of model interpretability

* Interpreting white box models, such as Linear/Logistic Regression (using model coefficients) and Tree models (using feature importance scores)
* Interpreting black box models with LIME and SHAP (Kernel Explainer, Tree Explainer) and how to implement this in Python
* Good practices for "debugging" LIME and SHAP explanations
* Limitations of LIME/SHAP (a.k.a., when to choose LIME over SHAP)
* Limitations of LIME/SHAP

[](https://github.com/fastforwardlabs/cml_explainability/blob/master/docs/images/limeshap.png)

**UNISOLAR Dataset**

**About UNISOLAR Dataset**

UNISOLAR is a large-scale open dataset on UNIversity SOLAR energy generation through PV systems.This dataset is publicly released as part of La Trobe University’s commitment to Net Zero Carbon Emissions by 2029, for which we are building the La Trobe Energy AI/Analytics Platform (LEAP) that leverages Artificial Intelligence (AI) and Data Analytics to analyse, predict and optimize the consumption, generation and utilization of electricity, renewables, gas and water resources. UNISOLAR contains solar generation data for solar sites across La Trobe’s five campuses in geographically distributed regions, apprximately across two and a half years, 2020 January-2022 April. The solar generation data is at 15-minute granularity and solar irradiance at hourly intervals. UNISOLAR also contains weather data from the closest weather station to each campus, averaged at 15-minute time intervals along with monthly summarized data and meta data of PV solar sites. A high granularity data dictionary and technical validation of the dataset for forecasting are further contributions of this article that will enable interested research scientists, academics, industry practitioners, sustainability and energy consultants to experiment and evaluate their AI algorithms, models, forecasts, as well as inform the development of energy benchmarks, guidelines and much needed data-driven energy policies.

**Dataset location:** <https://www.kaggle.com/datasets/cdaclab/unisolar>

**Contacts:**

If you want to have more information on the dataset, please feel free to contact us

GEETA BACHASPATI Email: [sgeeta1719@gmail.com](mailto:sgeeta1719@gmail.com)