# library

November 10, 2024

[1]:
```
pip list
```

```
Package                       Version
----------------------------- ---------------
absl-py                       1.4.0
alabaster                     0.7.12
altgraph                      0.17.4
anaconda-client               1.11.2
anaconda-navigator            2.4.2
anaconda-project              0.11.1
anyio                         3.5.0
appdirs                       1.4.4
argon2-cffi                   21.3.0
argon2-cffi-bindings          21.2.0
arrow                         1.2.3
astroid                       2.14.2
astropy                       5.1
asttokens                     2.0.5
astunparse                    1.6.3
atomicwrites                  1.4.0
attrs                         22.1.0
Automat                       20.2.0
autopep8                      1.6.0
Babel                         2.11.0
backcall                      0.2.0
backports.functools-lru-cache 1.6.4
backports.tempfile            1.0
backports.weakref             1.0.post1
bcrypt                        3.2.0
beautifulsoup4                4.11.1
binaryornot                   0.4.4
black                         22.6.0
bleach                        4.1.0
bokeh                         2.4.3
boltons                       23.0.0
Bottleneck                    1.3.5
brotlipy                      0.7.0
cachetools                    5.3.1
```

| | |
|---|---|
| certifi | 2023.5.7 |
| cffi | 1.15.1 |
| chardet | 4.0.0 |
| charset-normalizer | 2.0.4 |
| click | 8.0.4 |
| cloudpickle | 2.0.0 |
| clyent | 1.2.2 |
| cocos2d | 0.6.10 |
| colorama | 0.4.6 |
| colorcet | 3.0.1 |
| comm | 0.1.2 |
| conda | 23.3.1 |
| conda-build | 3.24.0 |
| conda-content-trust | 0.1.3 |
| conda-pack | 0.6.0 |
| conda-package-handling | 2.0.2 |
| conda_package_streaming | 0.7.0 |
| conda-repo-cli | 1.0.41 |
| conda-token | 0.4.0 |
| conda-verify | 3.4.2 |
| constantly | 15.1.0 |
| contourpy | 1.0.5 |
| cookiecutter | 1.7.3 |
| cpgames | 0.1.2 |
| cryptography | 39.0.1 |
| cssselect | 1.1.0 |
| cycler | 0.11.0 |
| cytoolz | 0.12.0 |
| daal4py | 2023.0.2 |
| dask | 2022.7.0 |
| datashader | 0.14.4 |
| datashape | 0.5.4 |
| debugpy | 1.5.1 |
| decorator | 5.1.1 |
| defusedxml | 0.7.1 |
| diff-match-patch | 20200713 |
| dill | 0.3.6 |
| distributed | 2022.7.0 |
| docstring-to-markdown | 0.11 |
| docutils | 0.18.1 |
| entrypoints | 0.4 |
| et-xmlfile | 1.1.0 |
| executing | 0.8.3 |
| fastjsonschema | 2.16.2 |
| filelock | 3.9.0 |
| flake8 | 6.0.0 |
| Flask | 2.2.2 |
| flatbuffers | 23.5.26 |

| | |
|---|---|
| flit_core | 3.6.0 |
| fonttools | 4.25.0 |
| fsspec | 2022.11.0 |
| future | 0.18.3 |
| gast | 0.4.0 |
| gensim | 4.3.0 |
| glob2 | 0.7 |
| google-auth | 2.20.0 |
| google-auth-oauthlib | 1.0.0 |
| google-pasta | 0.2.0 |
| greenlet | 2.0.1 |
| grpcio | 1.56.0 |
| h5py | 3.7.0 |
| HeapDict | 1.0.1 |
| holoviews | 1.15.4 |
| huggingface-hub | 0.10.1 |
| hvplot | 0.8.2 |
| hyperlink | 21.0.0 |
| idna | 3.4 |
| imagecodecs | 2021.8.26 |
| imageio | 2.26.0 |
| imagesize | 1.4.1 |
| imbalanced-learn | 0.10.1 |
| importlib-metadata | 4.11.3 |
| incremental | 21.3.0 |
| inflection | 0.5.1 |
| iniconfig | 1.1.1 |
| intake | 0.6.7 |
| intervaltree | 3.1.0 |
| ipykernel | 6.19.2 |
| ipython | 8.10.0 |
| ipython-genutils | 0.2.0 |
| ipywidgets | 7.6.5 |
| isort | 5.9.3 |
| itemadapter | 0.3.0 |
| itemloaders | 1.0.4 |
| itsdangerous | 2.0.1 |
| jax | 0.4.13 |
| jedi | 0.18.1 |
| jellyfish | 0.9.0 |
| Jinja2 | 3.1.2 |
| jinja2-time | 0.2.0 |
| jmespath | 0.10.0 |
| joblib | 1.1.1 |
| json5 | 0.9.6 |
| jsonpatch | 1.32 |
| jsonpointer | 2.1 |
| jsonschema | 4.17.3 |

| | |
|---|---|
| jupyter | 1.0.0 |
| jupyter_client | 7.3.4 |
| jupyter-console | 6.6.2 |
| jupyter_core | 5.2.0 |
| jupyter-server | 1.23.4 |
| jupyterlab | 3.5.3 |
| jupyterlab-pygments | 0.1.2 |
| jupyterlab_server | 2.19.0 |
| jupyterlab-widgets | 1.0.0 |
| keras | 2.12.0 |
| keyring | 23.4.0 |
| kiwisolver | 1.4.4 |
| lazy-object-proxy | 1.6.0 |
| libarchive-c | 2.9 |
| libclang | 16.0.0 |
| llvmlite | 0.39.1 |
| locket | 1.0.0 |
| lxml | 4.9.1 |
| lz4 | 3.1.3 |
| Markdown | 3.4.1 |
| MarkupSafe | 2.1.1 |
| matplotlib | 3.7.0 |
| matplotlib-inline | 0.1.6 |
| mccabe | 0.7.0 |
| menuinst | 1.4.19 |
| mistune | 0.8.4 |
| mkl-fft | 1.3.1 |
| mkl-random | 1.2.2 |
| mkl-service | 2.4.0 |
| ml-dtypes | 0.2.0 |
| mock | 4.0.3 |
| mpmath | 1.2.1 |
| msgpack | 1.0.3 |
| multipledispatch | 0.6.0 |
| munkres | 1.1.4 |
| mypy-extensions | 0.4.3 |
| mysql-connector-python | 8.1.0 |
| navigator-updater | 0.3.0 |
| nbclassic | 0.5.2 |
| nbclient | 0.5.13 |
| nbconvert | 6.5.4 |
| nbformat | 5.7.0 |
| nest-asyncio | 1.5.6 |
| networkx | 2.8.4 |
| nltk | 3.7 |
| notebook | 6.5.2 |
| notebook-as-pdf | 0.5.0 |
| notebook_shim | 0.2.2 |

| | |
|---|---|
| numba | 0.56.4 |
| numexpr | 2.8.4 |
| numpy | 1.23.5 |
| numpydoc | 1.5.0 |
| oauthlib | 3.2.2 |
| opencv-python | 4.7.0.72 |
| openpyxl | 3.0.10 |
| opt-einsum | 3.3.0 |
| packaging | 22.0 |
| pandas | 1.5.3 |
| pandocfilters | 1.5.0 |
| panel | 0.14.3 |
| param | 1.12.3 |
| paramiko | 2.8.1 |
| parsel | 1.6.0 |
| parso | 0.8.3 |
| partd | 1.2.0 |
| pathspec | 0.10.3 |
| patsy | 0.5.3 |
| pefile | 2023.2.7 |
| pep8 | 1.7.1 |
| pexpect | 4.8.0 |
| pickleshare | 0.7.5 |
| Pillow | 9.4.0 |
| pip | 22.3.1 |
| pkginfo | 1.9.6 |
| platformdirs | 2.5.2 |
| plotly | 5.9.0 |
| pluggy | 1.0.0 |
| ply | 3.11 |
| pooch | 1.4.0 |
| poyo | 0.5.0 |
| prometheus-client | 0.14.1 |
| prompt-toolkit | 3.0.36 |
| Protego | 0.1.16 |
| protobuf | 4.21.12 |
| psutil | 5.9.0 |
| ptyprocess | 0.7.0 |
| pure-eval | 0.2.2 |
| py | 1.11.0 |
| pyasn1 | 0.4.8 |
| pyasn1-modules | 0.2.8 |
| PyAudio | 0.2.13 |
| pycodestyle | 2.10.0 |
| pycosat | 0.6.4 |
| pycparser | 2.21 |
| pyct | 0.5.0 |
| pycurl | 7.45.1 |

| | |
|---|---|
| PyDispatcher | 2.0.5 |
| pydocstyle | 6.3.0 |
| pyee | 8.2.2 |
| pyerfa | 2.0.0 |
| pyflakes | 3.0.1 |
| pygame | 2.5.0 |
| pyglet | 1.5.27 |
| Pygments | 2.11.2 |
| PyHamcrest | 2.0.2 |
| pyinstaller | 6.1.0 |
| pyinstaller-hooks-contrib | 2023.10 |
| PyJWT | 2.4.0 |
| pylint | 2.16.2 |
| pylint-venv | 2.3.0 |
| pyls-spyder | 0.4.0 |
| PyNaCl | 1.5.0 |
| pyodbc | 4.0.34 |
| pyOpenSSL | 23.0.0 |
| pyparsing | 3.0.9 |
| PyPDF2 | 3.0.1 |
| pyppeteer | 1.0.0 |
| PyQt5 | 5.15.7 |
| PyQt5-sip | 12.11.0 |
| PyQtWebEngine | 5.15.4 |
| pyrsistent | 0.18.0 |
| PySocks | 1.7.1 |
| pytest | 7.1.2 |
| python-dateutil | 2.8.2 |
| python-lsp-black | 1.2.1 |
| python-lsp-jsonrpc | 1.0.0 |
| python-lsp-server | 1.7.1 |
| python-slugify | 5.0.2 |
| python-snappy | 0.6.1 |
| pytoolconfig | 1.2.5 |
| pytz | 2022.7 |
| pyviz-comms | 2.0.2 |
| PyWavelets | 1.4.1 |
| pywin32 | 305.1 |
| pywin32-ctypes | 0.2.2 |
| pywinpty | 2.0.10 |
| pyxel | 1.9.18 |
| PyYAML | 6.0 |
| pyzmq | 23.2.0 |
| QDarkStyle | 3.0.2 |
| qstylizer | 0.2.2 |
| QtAwesome | 1.2.2 |
| qtconsole | 5.4.0 |
| QtPy | 2.2.0 |

| | |
|---|---|
| queuelib | 1.5.0 |
| regex | 2022.7.9 |
| requests | 2.28.1 |
| requests-file | 1.5.1 |
| requests-oauthlib | 1.3.1 |
| requests-toolbelt | 0.9.1 |
| rope | 1.7.0 |
| rsa | 4.9 |
| Rtree | 1.0.1 |
| ruamel.yaml | 0.17.21 |
| ruamel.yaml.clib | 0.2.6 |
| ruamel-yaml-conda | 0.17.21 |
| scikit-image | 0.19.3 |
| scikit-learn | 1.2.1 |
| scikit-learn-intelex | 20230228.214818 |
| scipy | 1.10.0 |
| Scrapy | 2.8.0 |
| seaborn | 0.12.2 |
| Send2Trash | 1.8.0 |
| service-identity | 18.1.0 |
| setuptools | 65.6.3 |
| sip | 6.6.2 |
| six | 1.16.0 |
| smart-open | 5.2.1 |
| sniffio | 1.2.0 |
| snowballstemmer | 2.2.0 |
| sortedcontainers | 2.4.0 |
| sounddevice | 0.4.6 |
| soupsieve | 2.3.2.post1 |
| SpeechRecognition | 3.10.0 |
| Sphinx | 5.0.2 |
| sphinxcontrib-applehelp | 1.0.2 |
| sphinxcontrib-devhelp | 1.0.2 |
| sphinxcontrib-htmlhelp | 2.0.0 |
| sphinxcontrib-jsmath | 1.0.1 |
| sphinxcontrib-qthelp | 1.0.3 |
| sphinxcontrib-serializinghtml | 1.1.5 |
| spyder | 5.4.1 |
| spyder-kernels | 2.4.1 |
| SQLAlchemy | 1.4.39 |
| stack-data | 0.2.0 |
| statsmodels | 0.13.5 |
| sympy | 1.11.1 |
| tables | 3.7.0 |
| tabulate | 0.8.10 |
| TBB | 0.2 |
| tblib | 1.7.0 |
| tenacity | 8.0.1 |

| | |
|---|---|
| tensorboard | 2.12.3 |
| tensorboard-data-server | 0.7.1 |
| tensorflow | 2.12.0 |
| tensorflow-estimator | 2.12.0 |
| tensorflow-intel | 2.12.0 |
| tensorflow-io-gcs-filesystem | 0.31.0 |
| termcolor | 2.3.0 |
| terminado | 0.17.1 |
| text-unidecode | 1.3 |
| textdistance | 4.2.1 |
| threadpoolctl | 2.2.0 |
| three-merge | 0.1.1 |
| tifffile | 2021.7.2 |
| tinycss2 | 1.2.1 |
| tldextract | 3.2.0 |
| tokenizers | 0.11.4 |
| toml | 0.10.2 |
| tomli | 2.0.1 |
| tomlkit | 0.11.1 |
| toolz | 0.12.0 |
| torch | 1.12.1 |
| tornado | 6.1 |
| tqdm | 4.64.1 |
| traitlets | 5.7.1 |
| transformers | 4.24.0 |
| Twisted | 22.2.0 |
| twisted-iocpsupport | 1.0.2 |
| typing_extensions | 4.4.0 |
| ujson | 5.4.0 |
| Unidecode | 1.2.0 |
| urllib3 | 1.26.14 |
| w3lib | 1.21.0 |
| watchdog | 2.1.6 |
| wcwidth | 0.2.5 |
| webencodings | 0.5.1 |
| websocket-client | 0.58.0 |
| websockets | 10.4 |
| Werkzeug | 2.2.2 |
| whatthepatch | 1.0.2 |
| wheel | 0.38.4 |
| widgetsnbextension | 3.5.2 |
| win-inet-pton | 1.1.0 |
| wincertstore | 0.2 |
| wrapt | 1.14.1 |
| xarray | 2022.11.0 |
| xlwings | 0.29.1 |
| yapf | 0.31.0 |
| zict | 2.1.0 |

```
zipp                          3.11.0
zope.interface                5.4.0
zstandard                     0.19.0
Note: you may need to restart the kernel to use updated packages.
```

[2]:
```
pip install numpy scipy scikit-learn pandas
```

```
Requirement already satisfied: numpy in c:\users\vaibh\anaconda3\lib\site-
packages (1.23.5)
Requirement already satisfied: scipy in c:\users\vaibh\anaconda3\lib\site-
packages (1.10.0)
Requirement already satisfied: scikit-learn in
c:\users\vaibh\anaconda3\lib\site-packages (1.2.1)
Requirement already satisfied: pandas in c:\users\vaibh\anaconda3\lib\site-
packages (1.5.3)
Requirement already satisfied: threadpoolctl>=2.0.0 in
c:\users\vaibh\anaconda3\lib\site-packages (from scikit-learn) (2.2.0)
Requirement already satisfied: joblib>=1.1.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from scikit-learn) (1.1.1)
Requirement already satisfied: pytz>=2020.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from pandas) (2022.7)
Requirement already satisfied: python-dateutil>=2.8.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\vaibh\anaconda3\lib\site-
packages (from python-dateutil>=2.8.1->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

[3]:
```
pip install matplotlib
```

```
Requirement already satisfied: matplotlib in c:\users\vaibh\anaconda3\lib\site-
packages (3.7.0)
Requirement already satisfied: kiwisolver>=1.0.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: fonttools>=4.22.0 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: numpy>=1.20 in c:\users\vaibh\anaconda3\lib\site-
packages (from matplotlib) (1.23.5)
Requirement already satisfied: cycler>=0.10 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: contourpy>=1.0.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (1.0.5)
Requirement already satisfied: pillow>=6.2.0 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: packaging>=20.0 in
```

```
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (22.0)
Requirement already satisfied: six>=1.5 in c:\users\vaibh\anaconda3\lib\site-
packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

[ ]:

```

```

# solar-data-preparation

November 10, 2024

[1]:
```python
import os
import numpy as np
import pandas as pd
```

[2]:
```python
BASE_PATH = 'datasets'
```

[3]:
```python
sites = os.listdir(BASE_PATH)
print('Total number of sites:', len(sites))
```

```
Total number of sites: 20
```

[4]:
```python
df = None
for site in sites:
    if df is not None:
        df = pd.concat((df,pd.read_csv(os.path.join(BASE_PATH, site),␣
↪skiprows=10)))
    else:
        df = pd.read_csv(os.path.join(BASE_PATH, site), skiprows=10)
df
```

[4]:
```
              Site  Latitude  Longitude                   Date  \
0   Layla - TVTC  22.27948   46.73319   7/1/2013 12:00:00 AM
1   Layla - TVTC  22.27948   46.73319   8/1/2013 12:00:00 AM
2   Layla - TVTC  22.27948   46.73319   9/1/2013 12:00:00 AM
3   Layla - TVTC  22.27948   46.73319  10/1/2013 12:00:00 AM
4   Layla - TVTC  22.27948   46.73319  11/1/2013 12:00:00 AM
..           …         …         …                     …
16  Riyadh - KSU  24.72359   46.61639   3/1/2016 12:00:00 AM
17  Riyadh - KSU  24.72359   46.61639   4/1/2016 12:00:00 AM
18  Riyadh - KSU  24.72359   46.61639   5/1/2016 12:00:00 AM
19  Riyadh - KSU  24.72359   46.61639   6/1/2016 12:00:00 AM
20  Riyadh - KSU  24.72359   46.61639   7/1/2016 12:00:00 AM

    Air Temperature (C°)  Air Temperature Uncertainty (C°)  \
0                   38.4                               0.5
1                   37.1                               0.5
2                   34.4                               0.5
```

```
3                   28.3                                  0.5
4                   23.0                                  0.5
..                   …                                 …
16                  23.4                                  0.5
17                  26.6                                  0.5
18                  33.8                                  0.5
19                  36.7                                  0.5
20                  38.5                                  0.5

    Wind Direction at 3m (°N)  Wind Direction at 3m Uncertainty (°N)  \
0                        22.0                                    4.0
1                        25.0                                    4.0
2                        36.0                                    4.0
3                        56.0                                    4.0
4                        79.0                                    4.0
..                        …                                   …
16                       88.0                                    4.0
17                      172.0                                    4.0
18                       16.0                                    4.0
19                       26.0                                    4.0
20                       21.0                                    4.0

    Wind Speed at 3m (m/s)  Wind Speed at 3m Uncertainty (m/s)  …  \
0                      2.7                                 0.1  …
1                      2.6                                 0.0  …
2                      2.3                                 0.0  …
3                      2.1                                 0.0  …
4                      2.2                                 0.0  …
..                     …                                 …  …
16                     2.3                                 0.0  …
17                     1.8                                 0.0  …
18                     2.1                                 0.0  …
19                     1.8                                 0.0  …
20                     2.0                                 0.0  …

    Standard Deviation DNI (Wh/m2)  GHI (Wh/m2)  GHI Uncertainty (Wh/m2)  \
0                              NaN       7236.2                    638.7
1                              NaN       7266.6                    657.1
2                              NaN       6899.4                    463.1
3                              NaN       6116.5                    386.5
4                              NaN       4978.0                    359.8
..                             …        …                      …
16                          2545.4       5715.6                    401.4
17                          2549.7       6880.5                    461.1
18                          2018.2       7507.5                    473.6
19                          1041.9       7997.9                    474.5
20                          1295.6       7922.3                    494.0
```

| | Standard Deviation GHI (Wh/m2) | Peak Wind Speed at 3m (m/s) |
|---|---|---|
| 0 | NaN | 17.1 |
| 1 | NaN | 14.1 |
| 2 | NaN | 10.7 |
| 3 | NaN | 11.2 |
| 4 | NaN | 12.0 |
| .. | … | … |
| 16 | 1125.6 | 13.3 |
| 17 | 1241.5 | 12.5 |
| 18 | 848.5 | 13.6 |
| 19 | 212.1 | 11.5 |
| 20 | 380.0 | 12.5 |

| | Peak Wind Speed at 3m Uncertainty (m/s) | Relative Humidity (%) |
|---|---|---|
| 0 | 0.1 | 10.7 |
| 1 | 0.1 | 12.8 |
| 2 | 0.1 | 14.5 |
| 3 | 0.1 | 18.0 |
| 4 | 0.1 | 43.5 |
| .. | … | … |
| 16 | 0.1 | 34.2 |
| 17 | 0.1 | 34.0 |
| 18 | 0.1 | 18.1 |
| 19 | 0.1 | 12.2 |
| 20 | 0.1 | 13.2 |

| | Relative Humidity Uncertainty (%) | Barometric Pressure (mB (hPa equiv)) |
|---|---|---|
| 0 | 3.0 | 938.5 |
| 1 | 3.0 | 940.6 |
| 2 | 3.0 | 945.2 |
| 3 | 3.0 | 950.5 |
| 4 | 3.0 | 952.6 |
| .. | … | … |
| 16 | 3.0 | 939.4 |
| 17 | 3.0 | 937.9 |
| 18 | 3.0 | 935.2 |
| 19 | 3.0 | 932.6 |
| 20 | 3.0 | 929.0 |

| | Barometric Pressure Uncertainty (mB (hPa equiv)) |
|---|---|
| 0 | 4.7 |
| 1 | 4.7 |
| 2 | 4.7 |
| 3 | 4.8 |
| 4 | 4.8 |
| .. | … |

```
16                                                  4.7
17                                                  4.7
18                                                  4.7
19                                                  4.7
20                                                  4.6

[642 rows x 27 columns]
```

[5]:
```
#Making target col as the last col
y=df.pop('GHI (Wh/m2)')
df['GHI (Wh/m2)']=y
```

[6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 642 entries, 0 to 20
Data columns (total 27 columns):
 #   Column                                          Non-Null Count  Dtype
---  ------                                          --------------  -----
 0   Site                                            642 non-null    object
 1   Latitude                                        642 non-null    float64
 2   Longitude                                       642 non-null    float64
 3   Date                                            642 non-null    object
 4   Air Temperature (C°)                            642 non-null    float64
 5   Air Temperature Uncertainty (C°)                642 non-null    float64
 6   Wind Direction at 3m (°N)                       609 non-null    float64
 7   Wind Direction at 3m Uncertainty (°N)           609 non-null    float64
 8   Wind Speed at 3m (m/s)                          609 non-null    float64
 9   Wind Speed at 3m Uncertainty (m/s)              609 non-null    float64
 10  Wind Speed at 3m (std dev) (m/s)                609 non-null    float64
 11  Wind Speed at 3m (std dev) Uncertainty (m/s)    0 non-null      float64
 12  DHI (Wh/m2)                                     640 non-null    float64
 13  DHI Uncertainty (Wh/m2)                         640 non-null    float64
 14  Standard Deviation DHI (Wh/m2)                  499 non-null    float64
 15  DNI (Wh/m2)                                     640 non-null    float64
 16  DNI Uncertainty (Wh/m2)                         640 non-null    float64
 17  Standard Deviation DNI (Wh/m2)                  499 non-null    float64
 18  GHI Uncertainty (Wh/m2)                         641 non-null    float64
 19  Standard Deviation GHI (Wh/m2)                  500 non-null    float64
 20  Peak Wind Speed at 3m (m/s)                     609 non-null    float64
 21  Peak Wind Speed at 3m Uncertainty (m/s)         609 non-null    float64
 22  Relative Humidity (%)                           642 non-null    float64
 23  Relative Humidity Uncertainty (%)               642 non-null    float64
 24  Barometric Pressure (mB (hPa equiv))            642 non-null    float64
 25  Barometric Pressure Uncertainty (mB (hPa equiv))  642 non-null    float64
 26  GHI (Wh/m2)                                     641 non-null    float64
dtypes: float64(25), object(2)
```

```
memory usage: 140.4+ KB
```

[7]:
```
#drop wind speed at 3m std dev uncertanity (m/s) as all values are missing
df.drop(columns='Wind Speed at 3m (std dev) Uncertainty (m/s)',inplace=True)
df
```

[7]:
```
            Site  Latitude  Longitude                   Date  \
0   Layla - TVTC  22.27948   46.73319   7/1/2013 12:00:00 AM
1   Layla - TVTC  22.27948   46.73319   8/1/2013 12:00:00 AM
2   Layla - TVTC  22.27948   46.73319   9/1/2013 12:00:00 AM
3   Layla - TVTC  22.27948   46.73319  10/1/2013 12:00:00 AM
4   Layla - TVTC  22.27948   46.73319  11/1/2013 12:00:00 AM
..           …        …         …                    …
16  Riyadh - KSU  24.72359   46.61639   3/1/2016 12:00:00 AM
17  Riyadh - KSU  24.72359   46.61639   4/1/2016 12:00:00 AM
18  Riyadh - KSU  24.72359   46.61639   5/1/2016 12:00:00 AM
19  Riyadh - KSU  24.72359   46.61639   6/1/2016 12:00:00 AM
20  Riyadh - KSU  24.72359   46.61639   7/1/2016 12:00:00 AM

    Air Temperature (C°)  Air Temperature Uncertainty (C°)  \
0                   38.4                               0.5
1                   37.1                               0.5
2                   34.4                               0.5
3                   28.3                               0.5
4                   23.0                               0.5
..                   …                                …
16                  23.4                               0.5
17                  26.6                               0.5
18                  33.8                               0.5
19                  36.7                               0.5
20                  38.5                               0.5

    Wind Direction at 3m (°N)  Wind Direction at 3m Uncertainty (°N)  \
0                        22.0                                    4.0
1                        25.0                                    4.0
2                        36.0                                    4.0
3                        56.0                                    4.0
4                        79.0                                    4.0
..                        …                                     …
16                       88.0                                    4.0
17                      172.0                                    4.0
18                       16.0                                    4.0
19                       26.0                                    4.0
20                       21.0                                    4.0

    Wind Speed at 3m (m/s)  Wind Speed at 3m Uncertainty (m/s)  …  \
0                      2.7                                 0.1  …
```

```
1                     2.6                                0.0  …
2                     2.3                                0.0  …
3                     2.1                                0.0  …
4                     2.2                                0.0  …
..                    …                                …  …
16                    2.3                                0.0  …
17                    1.8                                0.0  …
18                    2.1                                0.0  …
19                    1.8                                0.0  …
20                    2.0                                0.0  …

    Standard Deviation DNI (Wh/m2)  GHI Uncertainty (Wh/m2)  \
0                              NaN                    638.7
1                              NaN                    657.1
2                              NaN                    463.1
3                              NaN                    386.5
4                              NaN                    359.8
..                             …                      …
16                          2545.4                    401.4
17                          2549.7                    461.1
18                          2018.2                    473.6
19                          1041.9                    474.5
20                          1295.6                    494.0

    Standard Deviation GHI (Wh/m2)  Peak Wind Speed at 3m (m/s)  \
0                              NaN                         17.1
1                              NaN                         14.1
2                              NaN                         10.7
3                              NaN                         11.2
4                              NaN                         12.0
..                             …                         …
16                          1125.6                         13.3
17                          1241.5                         12.5
18                           848.5                         13.6
19                           212.1                         11.5
20                           380.0                         12.5

    Peak Wind Speed at 3m Uncertainty (m/s)  Relative Humidity (%)  \
0                                       0.1                   10.7
1                                       0.1                   12.8
2                                       0.1                   14.5
3                                       0.1                   18.0
4                                       0.1                   43.5
..                                      …                   …
16                                      0.1                   34.2
17                                      0.1                   34.0
18                                      0.1                   18.1
```

```
19                                         0.1                     12.2
20                                         0.1                     13.2

    Relative Humidity Uncertainty (%)  Barometric Pressure (mB (hPa equiv))  \
0                                 3.0                                 938.5
1                                 3.0                                 940.6
2                                 3.0                                 945.2
3                                 3.0                                 950.5
4                                 3.0                                 952.6
..                                …                                   …
16                                3.0                                 939.4
17                                3.0                                 937.9
18                                3.0                                 935.2
19                                3.0                                 932.6
20                                3.0                                 929.0

    Barometric Pressure Uncertainty (mB (hPa equiv))  GHI (Wh/m2)
0                                                4.7       7236.2
1                                                4.7       7266.6
2                                                4.7       6899.4
3                                                4.8       6116.5
4                                                4.8       4978.0
..                                               …            …
16                                               4.7       5715.6
17                                               4.7       6880.5
18                                               4.7       7507.5
19                                               4.7       7997.9
20                                               4.6       7922.3

[642 rows x 26 columns]
```

[8]:
```
#Drop any exam with missing target value
df.dropna(how='all', axis='index',subset=['GHI (Wh/m2)'],inplace =True)
df
```

[8]:
```
             Site  Latitude  Longitude                   Date  \
0   Layla - TVTC  22.27948   46.73319   7/1/2013 12:00:00 AM
1   Layla - TVTC  22.27948   46.73319   8/1/2013 12:00:00 AM
2   Layla - TVTC  22.27948   46.73319   9/1/2013 12:00:00 AM
3   Layla - TVTC  22.27948   46.73319  10/1/2013 12:00:00 AM
4   Layla - TVTC  22.27948   46.73319  11/1/2013 12:00:00 AM
..          …        …         …                    …
16  Riyadh - KSU  24.72359   46.61639   3/1/2016 12:00:00 AM
17  Riyadh - KSU  24.72359   46.61639   4/1/2016 12:00:00 AM
18  Riyadh - KSU  24.72359   46.61639   5/1/2016 12:00:00 AM
19  Riyadh - KSU  24.72359   46.61639   6/1/2016 12:00:00 AM
20  Riyadh - KSU  24.72359   46.61639   7/1/2016 12:00:00 AM
```

```
    Air Temperature (C°)  Air Temperature Uncertainty (C°)  \
0                   38.4                               0.5
1                   37.1                               0.5
2                   34.4                               0.5
3                   28.3                               0.5
4                   23.0                               0.5
..                   …                               …
16                  23.4                               0.5
17                  26.6                               0.5
18                  33.8                               0.5
19                  36.7                               0.5
20                  38.5                               0.5

    Wind Direction at 3m (°N)  Wind Direction at 3m Uncertainty (°N)  \
0                        22.0                                    4.0
1                        25.0                                    4.0
2                        36.0                                    4.0
3                        56.0                                    4.0
4                        79.0                                    4.0
..                        …                                    …
16                       88.0                                    4.0
17                      172.0                                    4.0
18                       16.0                                    4.0
19                       26.0                                    4.0
20                       21.0                                    4.0

    Wind Speed at 3m (m/s)  Wind Speed at 3m Uncertainty (m/s)  …  \
0                      2.7                                 0.1  …
1                      2.6                                 0.0  …
2                      2.3                                 0.0  …
3                      2.1                                 0.0  …
4                      2.2                                 0.0  …
..                     …                                 …  …
16                     2.3                                 0.0  …
17                     1.8                                 0.0  …
18                     2.1                                 0.0  …
19                     1.8                                 0.0  …
20                     2.0                                 0.0  …

    Standard Deviation DNI (Wh/m2)  GHI Uncertainty (Wh/m2)  \
0                              NaN                    638.7
1                              NaN                    657.1
2                              NaN                    463.1
3                              NaN                    386.5
4                              NaN                    359.8
..                             …                    …
```

```
16                        2545.4                   401.4
17                        2549.7                   461.1
18                        2018.2                   473.6
19                        1041.9                   474.5
20                        1295.6                   494.0

    Standard Deviation GHI (Wh/m2)  Peak Wind Speed at 3m (m/s)  \
0                              NaN                         17.1
1                              NaN                         14.1
2                              NaN                         10.7
3                              NaN                         11.2
4                              NaN                         12.0
..                             …                          …
16                          1125.6                         13.3
17                          1241.5                         12.5
18                           848.5                         13.6
19                           212.1                         11.5
20                           380.0                         12.5

    Peak Wind Speed at 3m Uncertainty (m/s)  Relative Humidity (%)  \
0                                       0.1                   10.7
1                                       0.1                   12.8
2                                       0.1                   14.5
3                                       0.1                   18.0
4                                       0.1                   43.5
..                                     …                    …
16                                      0.1                   34.2
17                                      0.1                   34.0
18                                      0.1                   18.1
19                                      0.1                   12.2
20                                      0.1                   13.2

    Relative Humidity Uncertainty (%)  Barometric Pressure (mB (hPa equiv))  \
0                                 3.0                                 938.5
1                                 3.0                                 940.6
2                                 3.0                                 945.2
3                                 3.0                                 950.5
4                                 3.0                                 952.6
..                               …                                   …
16                                3.0                                 939.4
17                                3.0                                 937.9
18                                3.0                                 935.2
19                                3.0                                 932.6
20                                3.0                                 929.0

    Barometric Pressure Uncertainty (mB (hPa equiv))  GHI (Wh/m2)
0                                                4.7       7236.2
```

```
1                                                   4.7       7266.6
2                                                   4.7       6899.4
3                                                   4.8       6116.5
4                                                   4.8       4978.0
..                                                  …          …
16                                                  4.7       5715.6
17                                                  4.7       6880.5
18                                                  4.7       7507.5
19                                                  4.7       7997.9
20                                                  4.6       7922.3

[641 rows x 26 columns]
```

[9]:

```
#checking missiong values
for col in df.columns:
    if sum(pd.isnull(df[col]).values)>0:
        print('{0:s} has {1:d} missing values.' .format(col, sum(pd.
 ↪isnull(df[col]).values)))
```

```
Wind Direction at 3m (°N) has 33 missing values.
Wind Direction at 3m Uncertainty (°N) has 33 missing values.
Wind Speed at 3m (m/s) has 33 missing values.
Wind Speed at 3m Uncertainty (m/s) has 33 missing values.
Wind Speed at 3m (std dev) (m/s) has 33 missing values.
DHI (Wh/m2) has 1 missing values.
DHI Uncertainty (Wh/m2) has 1 missing values.
Standard Deviation DHI (Wh/m2) has 142 missing values.
DNI (Wh/m2) has 1 missing values.
DNI Uncertainty (Wh/m2) has 1 missing values.
Standard Deviation DNI (Wh/m2) has 142 missing values.
Standard Deviation GHI (Wh/m2) has 141 missing values.
Peak Wind Speed at 3m (m/s) has 33 missing values.
Peak Wind Speed at 3m Uncertainty (m/s) has 33 missing values.
```

## 0.1 option 1:remove all records with any column value missing

[10]:

```
# drop records with missing values
df_dropped = df.dropna(how='any' , axis ='index')
df_dropped
```

[10]:

```
            Site  Latitude  Longitude                   Date  \
10  Layla - TVTC  22.27948   46.73319  5/1/2014 12:00:00 AM
11  Layla - TVTC  22.27948   46.73319  6/1/2014 12:00:00 AM
12  Layla - TVTC  22.27948   46.73319  7/1/2014 12:00:00 AM
13  Layla - TVTC  22.27948   46.73319  8/1/2014 12:00:00 AM
14  Layla - TVTC  22.27948   46.73319  9/1/2014 12:00:00 AM
..           …        …         …                      …
```

```
16  Riyadh - KSU  24.72359   46.61639  3/1/2016 12:00:00 AM
17  Riyadh - KSU  24.72359   46.61639  4/1/2016 12:00:00 AM
18  Riyadh - KSU  24.72359   46.61639  5/1/2016 12:00:00 AM
19  Riyadh - KSU  24.72359   46.61639  6/1/2016 12:00:00 AM
20  Riyadh - KSU  24.72359   46.61639  7/1/2016 12:00:00 AM

    Air Temperature (C°)  Air Temperature Uncertainty (C°)  \
10                  34.4                               0.5
11                  36.8                               0.5
12                  38.5                               0.5
13                  38.1                               0.5
14                  35.1                               0.5
..                   …                                …
16                  23.4                               0.5
17                  26.6                               0.5
18                  33.8                               0.5
19                  36.7                               0.5
20                  38.5                               0.5

    Wind Direction at 3m (°N)  Wind Direction at 3m Uncertainty (°N)  \
10                       76.0                                    4.0
11                       18.0                                    4.0
12                      351.0                                    4.0
13                       20.0                                    4.0
14                       42.0                                    4.0
..                        …                                    …
16                       88.0                                    4.0
17                      172.0                                    4.0
18                       16.0                                    4.0
19                       26.0                                    4.0
20                       21.0                                    4.0

    Wind Speed at 3m (m/s)  Wind Speed at 3m Uncertainty (m/s)  …  \
10                     2.3                                 0.0  …
11                     2.4                                 0.0  …
12                     2.3                                 0.0  …
13                     2.5                                 0.0  …
14                     2.4                                 0.0  …
..                     …                                 …   …
16                     2.3                                 0.0  …
17                     1.8                                 0.0  …
18                     2.1                                 0.0  …
19                     1.8                                 0.0  …
20                     2.0                                 0.0  …

    Standard Deviation DNI (Wh/m2)  GHI Uncertainty (Wh/m2)  \
10                          2084.9                    475.3
```

```
11                 2053.9                484.0
12                 1579.2                479.8
13                 1130.4                451.7
14                 1665.5                430.2
..                   …                   …
16                 2545.4                401.4
17                 2549.7                461.1
18                 2018.2                473.6
19                 1041.9                474.5
20                 1295.6                494.0

   Standard Deviation GHI (Wh/m2)  Peak Wind Speed at 3m (m/s)  \
10                          503.0                         15.7
11                          585.2                         13.1
12                          294.9                         11.7
13                          351.2                         20.8
14                          543.0                         12.5
..                            …                          …
16                         1125.6                         13.3
17                         1241.5                         12.5
18                          848.5                         13.6
19                          212.1                         11.5
20                          380.0                         12.5

   Peak Wind Speed at 3m Uncertainty (m/s)  Relative Humidity (%)  \
10                                     0.1                   13.5
11                                     0.1                    9.9
12                                     0.1                    9.6
13                                     0.1                   12.6
14                                     0.1                   15.0
..                                     …                     …
16                                     0.1                   34.2
17                                     0.1                   34.0
18                                     0.1                   18.1
19                                     0.1                   12.2
20                                     0.1                   13.2

   Relative Humidity Uncertainty (%)  Barometric Pressure (mB (hPa equiv))  \
10                               3.0                                 946.3
11                               3.0                                 942.9
12                               3.0                                 940.8
13                               3.0                                 941.1
14                               3.0                                 944.9
..                               …                                   …
16                               3.0                                 939.4
17                               3.0                                 937.9
18                               3.0                                 935.2
```

```
19                                  3.0                                  932.6
20                                  3.0                                  929.0

    Barometric Pressure Uncertainty (mB (hPa equiv))  GHI (Wh/m2)
10                                               4.7       7721.0
11                                               4.7       7765.4
12                                               4.7       7823.9
13                                               4.7       7463.4
14                                               4.7       6833.3
..                                               …           …
16                                               4.7       5715.6
17                                               4.7       6880.5
18                                               4.7       7507.5
19                                               4.7       7997.9
20                                               4.6       7922.3

[470 rows x 26 columns]
```

## 0.2 option 2 :replace missing values

[11]:
```
for col in df.columns:
    if sum(pd.isnull(df[col]).values) >0:
        #Calculate mean of the column
        value = df[col].mean()
        #Replace missing value with calculated value
        df[col].fillna(value = value, inplace=True)
```

## 0.3 Calculated value can be mean, median, mode, etc

[12]:
```
# Checking missing values
for col in df_dropped.columns:
    if sum(pd.isnull(df_dropped[col]).values) > 0:
        print('{0:s} has {1:d} missing values.'.format(col, sum(pd.
 ↪isnull(df_dropped[col]).values)))
```

[13]:
```
import sklearn
```

## 0.4 Data Normalization

Apply normalization on dataframe with missing values removed or replaced

[14]:
```
df.columns
```

[14]:
```
Index(['Site', 'Latitude', 'Longitude', 'Date', 'Air Temperature (C°)',
       'Air Temperature Uncertainty (C°)', 'Wind Direction at 3m (°N)',
       'Wind Direction at 3m Uncertainty (°N)', 'Wind Speed at 3m (m/s)',
```

```
       'Wind Speed at 3m Uncertainty (m/s)',
       'Wind Speed at 3m (std dev) (m/s)', 'DHI (Wh/m2)',
       'DHI Uncertainty (Wh/m2)', 'Standard Deviation DHI (Wh/m2)',
       'DNI (Wh/m2)', 'DNI Uncertainty (Wh/m2)',
       'Standard Deviation DNI (Wh/m2)', 'GHI Uncertainty (Wh/m2)',
       'Standard Deviation GHI (Wh/m2)', 'Peak Wind Speed at 3m (m/s)',
       'Peak Wind Speed at 3m Uncertainty (m/s)', 'Relative Humidity (%)',
       'Relative Humidity Uncertainty (%)',
       'Barometric Pressure (mB (hPa equiv))',
       'Barometric Pressure Uncertainty (mB (hPa equiv))', 'GHI (Wh/m2)'],
      dtype='object')
```

[15]:
```python
cols = list(df.columns[4:])
cols
```

[15]:
```
['Air Temperature (C°)',
 'Air Temperature Uncertainty (C°)',
 'Wind Direction at 3m (°N)',
 'Wind Direction at 3m Uncertainty (°N)',
 'Wind Speed at 3m (m/s)',
 'Wind Speed at 3m Uncertainty (m/s)',
 'Wind Speed at 3m (std dev) (m/s)',
 'DHI (Wh/m2)',
 'DHI Uncertainty (Wh/m2)',
 'Standard Deviation DHI (Wh/m2)',
 'DNI (Wh/m2)',
 'DNI Uncertainty (Wh/m2)',
 'Standard Deviation DNI (Wh/m2)',
 'GHI Uncertainty (Wh/m2)',
 'Standard Deviation GHI (Wh/m2)',
 'Peak Wind Speed at 3m (m/s)',
 'Peak Wind Speed at 3m Uncertainty (m/s)',
 'Relative Humidity (%)',
 'Relative Humidity Uncertainty (%)',
 'Barometric Pressure (mB (hPa equiv))',
 'Barometric Pressure Uncertainty (mB (hPa equiv))',
 'GHI (Wh/m2)']
```

## 0.5 Scale all columns except the target

[16]:
```python
from sklearn.preprocessing import StandardScaler
scalar = StandardScaler()
scaled_data = scalar.fit_transform(df[cols[:-1]].to_numpy())
scaled_df = pd.DataFrame(scaled_data , columns= cols[:-1])
scaled_df
```

[16]:
```
      Air Temperature (C°)  Air Temperature Uncertainty (C°)  \
0                 1.546567                               0.0
1                 1.366193                               0.0
2                 0.991570                               0.0
3                 0.145200                               0.0
4                -0.590171                               0.0
..                     …                              …
636              -0.534672                               0.0
637              -0.090674                               0.0
638               0.908321                               0.0
639               1.310693                               0.0
640               1.560442                               0.0

      Wind Direction at 3m (°N)  Wind Direction at 3m Uncertainty (°N)  \
0                     -1.467370                                0.18201
1                     -1.442493                                0.18201
2                     -1.351279                                0.18201
3                     -1.185436                                0.18201
4                     -0.994717                                0.18201
..                         …                                     …
636                   -0.920088                                0.18201
637                   -0.223547                                0.18201
638                   -1.517122                                0.18201
639                   -1.434201                                0.18201
640                   -1.475662                                0.18201

      Wind Speed at 3m (m/s)  Wind Speed at 3m Uncertainty (m/s)  \
0                  -0.331420                            0.710869
1                  -0.459034                           -1.483082
2                  -0.841878                           -1.483082
3                  -1.097107                           -1.483082
4                  -0.969492                           -1.483082
..                       …                                  …
636                -0.841878                           -1.483082
637                -1.479950                           -1.483082
638                -1.097107                           -1.483082
639                -1.479950                           -1.483082
640                -1.224721                           -1.483082

      Wind Speed at 3m (std dev) (m/s)  DHI (Wh/m2)  DHI Uncertainty (Wh/m2)  \
0                            -0.239125     2.033152                 2.543475
1                            -0.709637     0.726671                 1.321083
2                            -1.180148     0.040304                 0.242664
3                            -1.180148    -0.687110                -0.125577
4                            -0.709637    -1.045514                -0.243248
..                                 …          …                      …
636                          -1.180148     0.825019                 0.710578
```

```
637                        -1.415404     0.544577                  0.763184
638                        -0.709637     1.575850                  1.660252
639                        -0.944892     1.397198                  1.233868
640                        -0.474381     0.397882                  0.136068

     Standard Deviation DHI (Wh/m2)  …  DNI Uncertainty (Wh/m2)  \
0                          0.000000  …                 0.286709
1                          0.000000  …                 1.222345
2                          0.000000  …                 0.147899
3                          0.000000  …                 0.272695
4                          0.000000  …                 0.281370
..                              …  …                     …
636                        1.135155  …                 0.545644
637                        0.859923  …                 0.777884
638                        1.241456  …                 0.207294
639                       -0.461654  …                -0.094352
640                       -0.620243  …                 0.390817

     Standard Deviation DNI (Wh/m2)  GHI Uncertainty (Wh/m2)  \
0                          0.000000                 0.450382
1                          0.000000                 0.486600
2                          0.000000                 0.104739
3                          0.000000                -0.046037
4                          0.000000                -0.098592
..                              …                      …
636                        1.447571                -0.016709
637                        1.456262                 0.100802
638                        0.382034                 0.125407
639                       -1.591192                 0.127178
640                       -1.078432                 0.165561

     Standard Deviation GHI (Wh/m2)  Peak Wind Speed at 3m (m/s)  \
0                      4.131493e-16                     0.335854
1                      4.131493e-16                    -0.359885
2                      4.131493e-16                    -1.148389
3                      4.131493e-16                    -1.032433
4                      4.131493e-16                    -0.846902
..                              …                          …
636                    1.710223e+00                    -0.545415
637                    2.131415e+00                    -0.730946
638                    7.032136e-01                    -0.475841
639                   -1.609528e+00                    -0.962859
640                   -9.993622e-01                    -0.730946

     Peak Wind Speed at 3m Uncertainty (m/s)  Relative Humidity (%)  \
0                                   0.182479              -1.336438
1                                   0.182479              -1.234901
```

```
2                                   0.182479             -1.152704
3                                   0.182479             -0.983476
4                                   0.182479              0.249471
..                                       …                    …
636                                 0.182479             -0.200192
637                                 0.182479             -0.209862
638                                 0.182479             -0.978641
639                                 0.182479             -1.263911
640                                 0.182479             -1.215561

     Relative Humidity Uncertainty (%)  Barometric Pressure (mB (hPa equiv))  \
0                                  0.0                             -0.449612
1                                  0.0                             -0.402440
2                                  0.0                             -0.299112
3                                  0.0                             -0.180059
4                                  0.0                             -0.132888
..                                 …                                    …
636                                0.0                             -0.429396
637                                0.0                             -0.463090
638                                0.0                             -0.523739
639                                0.0                             -0.582142
640                                0.0                             -0.663008

     Barometric Pressure Uncertainty (mB (hPa equiv))
0                                           -0.469979
1                                           -0.469979
2                                           -0.469979
3                                           -0.021681
4                                           -0.021681
..                                                …
636                                         -0.469979
637                                         -0.469979
638                                         -0.469979
639                                         -0.469979
640                                         -0.918278

[641 rows x 21 columns]
```

### 0.6 Min-Max scaling for target variable

[17]:
```
from sklearn.preprocessing import MinMaxScaler
scalar2 = MinMaxScaler(feature_range=(-1,1))
scaled_df[cols[-1]] = scalar2.fit_transform(df[cols[-1]].to_numpy().
 ↪reshape(-1,1))
scaled_df
```

[17]:

```
      Air Temperature (C°)  Air Temperature Uncertainty (C°)  \
0                 1.546567                               0.0
1                 1.366193                               0.0
2                 0.991570                               0.0
3                 0.145200                               0.0
4                -0.590171                               0.0
..                     …                                …
636              -0.534672                               0.0
637              -0.090674                               0.0
638               0.908321                               0.0
639               1.310693                               0.0
640               1.560442                               0.0

     Wind Direction at 3m (°N)  Wind Direction at 3m Uncertainty (°N)  \
0                    -1.467370                                0.18201
1                    -1.442493                                0.18201
2                    -1.351279                                0.18201
3                    -1.185436                                0.18201
4                    -0.994717                                0.18201
..                         …                                    …
636                  -0.920088                                0.18201
637                  -0.223547                                0.18201
638                  -1.517122                                0.18201
639                  -1.434201                                0.18201
640                  -1.475662                                0.18201

     Wind Speed at 3m (m/s)  Wind Speed at 3m Uncertainty (m/s)  \
0                 -0.331420                            0.710869
1                 -0.459034                           -1.483082
2                 -0.841878                           -1.483082
3                 -1.097107                           -1.483082
4                 -0.969492                           -1.483082
..                      …                                 …
636               -0.841878                           -1.483082
637               -1.479950                           -1.483082
638               -1.097107                           -1.483082
639               -1.479950                           -1.483082
640               -1.224721                           -1.483082

     Wind Speed at 3m (std dev) (m/s)  DHI (Wh/m2)  DHI Uncertainty (Wh/m2)  \
0                           -0.239125     2.033152                 2.543475
1                           -0.709637     0.726671                 1.321083
2                           -1.180148     0.040304                 0.242664
3                           -1.180148    -0.687110                -0.125577
4                           -0.709637    -1.045514                -0.243248
..                                …          …                      …
636                         -1.180148     0.825019                 0.710578
```

```
637                         -1.415404     0.544577                  0.763184
638                         -0.709637     1.575850                  1.660252
639                         -0.944892     1.397198                  1.233868
640                         -0.474381     0.397882                  0.136068

     Standard Deviation DHI (Wh/m2)  …  Standard Deviation DNI (Wh/m2)  \
0                          0.000000  …                        0.000000
1                          0.000000  …                        0.000000
2                          0.000000  …                        0.000000
3                          0.000000  …                        0.000000
4                          0.000000  …                        0.000000
..                              …  …                             …
636                        1.135155  …                        1.447571
637                        0.859923  …                        1.456262
638                        1.241456  …                        0.382034
639                       -0.461654  …                       -1.591192
640                       -0.620243  …                       -1.078432

     GHI Uncertainty (Wh/m2)  Standard Deviation GHI (Wh/m2)  \
0                   0.450382                    4.131493e-16
1                   0.486600                    4.131493e-16
2                   0.104739                    4.131493e-16
3                  -0.046037                    4.131493e-16
4                  -0.098592                    4.131493e-16
..                       …                           …
636                -0.016709                    1.710223e+00
637                 0.100802                    2.131415e+00
638                 0.125407                    7.032136e-01
639                 0.127178                   -1.609528e+00
640                 0.165561                   -9.993622e-01

     Peak Wind Speed at 3m (m/s)  Peak Wind Speed at 3m Uncertainty (m/s)  \
0                       0.335854                                 0.182479
1                      -0.359885                                 0.182479
2                      -1.148389                                 0.182479
3                      -1.032433                                 0.182479
4                      -0.846902                                 0.182479
..                           …                                      …
636                    -0.545415                                 0.182479
637                    -0.730946                                 0.182479
638                    -0.475841                                 0.182479
639                    -0.962859                                 0.182479
640                    -0.730946                                 0.182479

     Relative Humidity (%)  Relative Humidity Uncertainty (%)  \
0                -1.336438                                0.0
1                -1.234901                                0.0
```

```
2              -1.152704                               0.0
3              -0.983476                               0.0
4               0.249471                               0.0
..                   …                               …
636            -0.200192                               0.0
637            -0.209862                               0.0
638            -0.978641                               0.0
639            -1.263911                               0.0
640            -1.215561                               0.0

     Barometric Pressure (mB (hPa equiv))  \
0                                -0.449612
1                                -0.402440
2                                -0.299112
3                                -0.180059
4                                -0.132888
..                                     …
636                              -0.429396
637                              -0.463090
638                              -0.523739
639                              -0.582142
640                              -0.663008

     Barometric Pressure Uncertainty (mB (hPa equiv))  GHI (Wh/m2)
0                                           -0.469979     0.509240
1                                           -0.469979     0.520716
2                                           -0.469979     0.382095
3                                           -0.021681     0.086544
4                                           -0.021681    -0.343249
..                                                …          …
636                                         -0.469979    -0.064799
637                                         -0.469979     0.374960
638                                         -0.469979     0.611657
639                                         -0.469979     0.796787
640                                         -0.918278     0.768248

[641 rows x 22 columns]
```

[18]:
```python
print(y.shape)
```

```
(642,)
```

[19]:
```python
def get_class(ghi):
    if ghi>0.4:
        return 'running'
    elif ghi>=-0.4:
        return 'Monitoring'
```

```
    elif ghi>=-1:
        return 'Inspecting'
```

## 0.7 Generate classes

3 classes: **Running :** GHI> +0.4, **Moniotoring :** 0.4<= GHI >=-0.4 **Inspecting:** -0.4< GHI >= -0.1

[20]:
```
scaled_df['Class']=scaled_df['GHI (Wh/m2)'].apply(get_class)
scaled_df
```

[20]:
```
     Air Temperature (C°)  Air Temperature Uncertainty (C°)  \
0                1.546567                               0.0
1                1.366193                               0.0
2                0.991570                               0.0
3                0.145200                               0.0
4               -0.590171                               0.0
..                    …                                …
636             -0.534672                               0.0
637             -0.090674                               0.0
638              0.908321                               0.0
639              1.310693                               0.0
640              1.560442                               0.0

     Wind Direction at 3m (°N)  Wind Direction at 3m Uncertainty (°N)  \
0                    -1.467370                                0.18201
1                    -1.442493                                0.18201
2                    -1.351279                                0.18201
3                    -1.185436                                0.18201
4                    -0.994717                                0.18201
..                         …                                  …
636                  -0.920088                                0.18201
637                  -0.223547                                0.18201
638                  -1.517122                                0.18201
639                  -1.434201                                0.18201
640                  -1.475662                                0.18201

     Wind Speed at 3m (m/s)  Wind Speed at 3m Uncertainty (m/s)  \
0                 -0.331420                            0.710869
1                 -0.459034                           -1.483082
2                 -0.841878                           -1.483082
3                 -1.097107                           -1.483082
4                 -0.969492                           -1.483082
..                      …                                  …
636               -0.841878                           -1.483082
637               -1.479950                           -1.483082
638               -1.097107                           -1.483082
```

```
639             -1.479950                           -1.483082
640             -1.224721                           -1.483082

     Wind Speed at 3m (std dev) (m/s)  DHI (Wh/m2)  DHI Uncertainty (Wh/m2)  \
0                           -0.239125     2.033152                 2.543475
1                           -0.709637     0.726671                 1.321083
2                           -1.180148     0.040304                 0.242664
3                           -1.180148    -0.687110                -0.125577
4                           -0.709637    -1.045514                -0.243248
..                                ...          ...                      ...
636                         -1.180148     0.825019                 0.710578
637                         -1.415404     0.544577                 0.763184
638                         -0.709637     1.575850                 1.660252
639                         -0.944892     1.397198                 1.233868
640                         -0.474381     0.397882                 0.136068

     Standard Deviation DHI (Wh/m2)  ...  GHI Uncertainty (Wh/m2)  \
0                          0.000000  ...                 0.450382
1                          0.000000  ...                 0.486600
2                          0.000000  ...                 0.104739
3                          0.000000  ...                -0.046037
4                          0.000000  ...                -0.098592
..                              ...  ...                      ...
636                        1.135155  ...                -0.016709
637                        0.859923  ...                 0.100802
638                        1.241456  ...                 0.125407
639                       -0.461654  ...                 0.127178
640                       -0.620243  ...                 0.165561

     Standard Deviation GHI (Wh/m2)  Peak Wind Speed at 3m (m/s)  \
0                      4.131493e-16                     0.335854
1                      4.131493e-16                    -0.359885
2                      4.131493e-16                    -1.148389
3                      4.131493e-16                    -1.032433
4                      4.131493e-16                    -0.846902
..                              ...                          ...
636                    1.710223e+00                    -0.545415
637                    2.131415e+00                    -0.730946
638                    7.032136e-01                    -0.475841
639                   -1.609528e+00                    -0.962859
640                   -9.993622e-01                    -0.730946

     Peak Wind Speed at 3m Uncertainty (m/s)  Relative Humidity (%)  \
0                                   0.182479              -1.336438
1                                   0.182479              -1.234901
2                                   0.182479              -1.152704
3                                   0.182479              -0.983476
```

```
4                                  0.182479              0.249471
..                                      …                     …
636                                0.182479             -0.200192
637                                0.182479             -0.209862
638                                0.182479             -0.978641
639                                0.182479             -1.263911
640                                0.182479             -1.215561

     Relative Humidity Uncertainty (%)  Barometric Pressure (mB (hPa equiv))  \
0                                 0.0                             -0.449612
1                                 0.0                             -0.402440
2                                 0.0                             -0.299112
3                                 0.0                             -0.180059
4                                 0.0                             -0.132888
..                                  …                                     …
636                               0.0                             -0.429396
637                               0.0                             -0.463090
638                               0.0                             -0.523739
639                               0.0                             -0.582142
640                               0.0                             -0.663008

     Barometric Pressure Uncertainty (mB (hPa equiv))  GHI (Wh/m2)       Class
0                                           -0.469979     0.509240     running
1                                           -0.469979     0.520716     running
2                                           -0.469979     0.382095  Monitoring
3                                           -0.021681     0.086544  Monitoring
4                                           -0.021681    -0.343249  Monitoring
..                                                …           …           …
636                                         -0.469979    -0.064799  Monitoring
637                                         -0.469979     0.374960  Monitoring
638                                         -0.469979     0.611657     running
639                                         -0.469979     0.796787     running
640                                         -0.918278     0.768248     running

[641 rows x 23 columns]
```

[21]:
```
scaled_df['Class'].value_counts()
```

[21]:
```
Monitoring    291
running       201
Inspecting    149
Name: Class, dtype: int64
```

[22]:
```
print(y.shape)  # Should print (442, n_features)
print(y.shape)         # Should print (442,)
```

```
(642,)
(642,)
```

## 0.8 Save data

[23]:
```
scaled_df.to_csv("Solar_radiation_classification.csv",index=False)
```

[24]:
```
print(scaled_df['Class'].isna().sum())  # Should output 0 if cleaned properly
```

```
0
```

[ ]:
```
```

# linear-regression-study-model

November 10, 2024

[2]:
```
pip install numpy scipy scikit-learn pandas matplotlib
```

```
Requirement already satisfied: numpy in c:\users\vaibh\anaconda3\lib\site-
packages (1.23.5)
Requirement already satisfied: scipy in c:\users\vaibh\anaconda3\lib\site-
packages (1.10.0)
Requirement already satisfied: scikit-learn in
c:\users\vaibh\anaconda3\lib\site-packages (1.2.1)
Requirement already satisfied: pandas in c:\users\vaibh\anaconda3\lib\site-
packages (1.5.3)
Requirement already satisfied: matplotlib in c:\users\vaibh\anaconda3\lib\site-
packages (3.7.0)
Requirement already satisfied: joblib>=1.1.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from scikit-learn) (1.1.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in
c:\users\vaibh\anaconda3\lib\site-packages (from scikit-learn) (2.2.0)
Requirement already satisfied: pytz>=2020.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from pandas) (2022.7)
Requirement already satisfied: python-dateutil>=2.8.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: packaging>=20.0 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (22.0)
Requirement already satisfied: cycler>=0.10 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: pillow>=6.2.0 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (9.4.0)
Requirement already satisfied: contourpy>=1.0.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (1.0.5)
Requirement already satisfied: fonttools>=4.22.0 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: pyparsing>=2.3.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: kiwisolver>=1.0.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: six>=1.5 in c:\users\vaibh\anaconda3\lib\site-
packages (from python-dateutil>=2.8.1->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

[3]:
```
%matplotlib inline
```

[4]:
```
import numpy as np #for numerical cal and matrix handling
import matplotlib.pyplot as plt # for plotting

from sklearn.linear_model import LinearRegression # for linear regression
from sklearn.model_selection import train_test_split # Divide dat as training
from sklearn.metrics import mean_squared_error # for evaluation

np.random.seed(0) # to control the randopm num generator
```

## 0.1 Load dataset

[5]:
```
from sklearn import datasets
X,y = datasets.load_diabetes(return_X_y=True)
```

## 0.2 Generate your own data

[6]:
```
def gen_target(X):
    return np.cos(1.5* X)+2
```

[7]:
```
n_records =300
X= np.sort(np.random.rand(n_records)) #Randomly generate data points
y= gen_target(X) +np.random.randn(n_records) * 0.1 #Generate regression

X= X.reshape(-1,1) # COnvert input data as 20
# Generate higher-prder features
from sklearn.preprocessing import PolynomialFeatures
poly_feat = PolynomialFeatures(degree=2)
X=poly_feat.fit_transform(X)
```

[8]:
```
print(y.shape)
```

```
(300,)
```

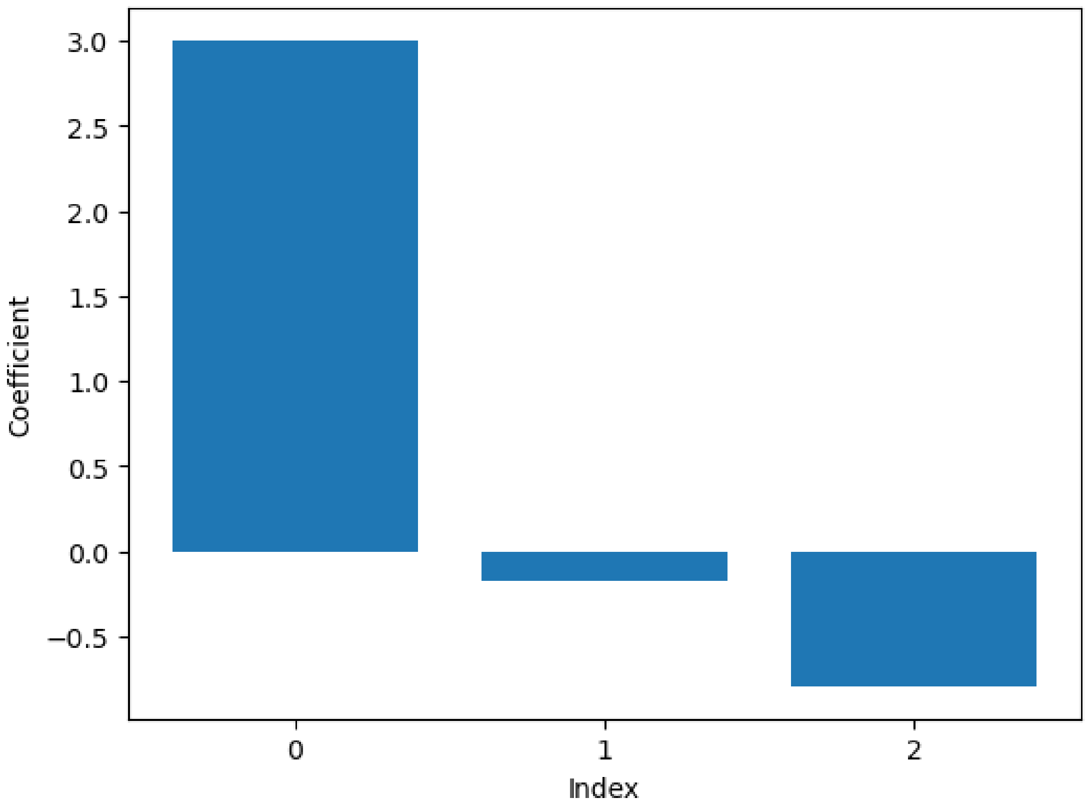
[9]:
```
print('Number of training examples:' , X.shape[0])
print('Number of predictors: ',y.shape[1] if len(y.shape)>1 else 1)
```

```
Number of training examples: 300
Number of predictors:  1
```

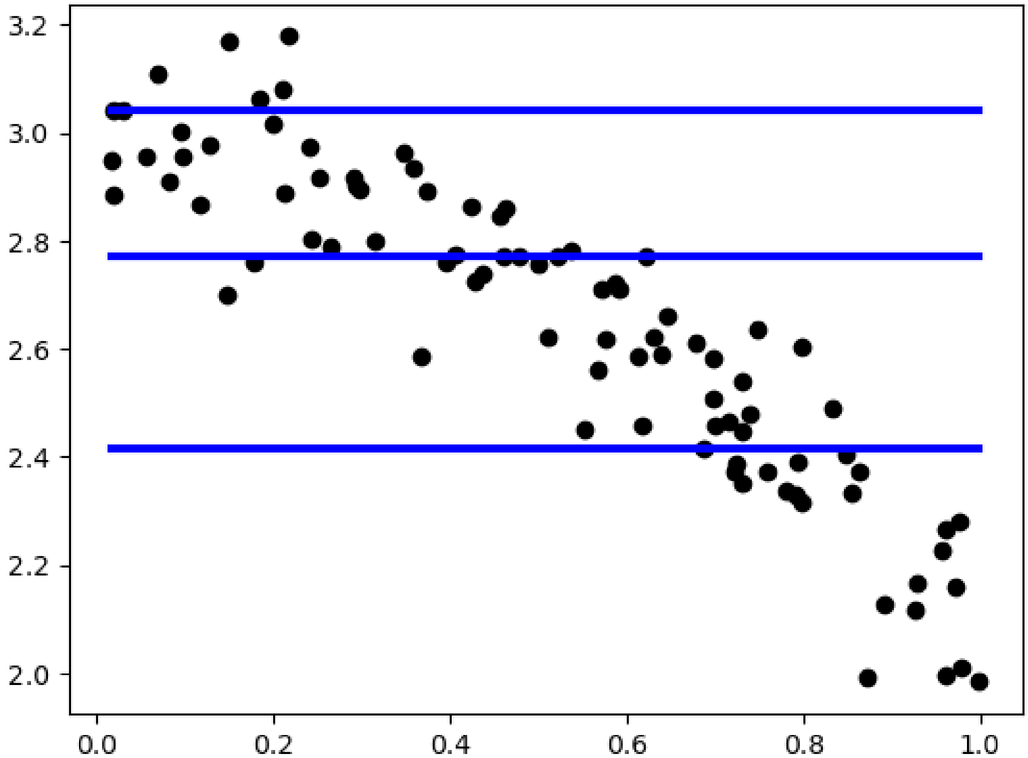
[17]:
```
#Split the data into training/testing sets
X_train , X_test , y_train,y_test= train_test_split(X,y,test_size = 0.
 ↪3,random_state=0)

lr= LinearRegression(fit_intercept=False) #INITIALIZE linear
lr.fit(X_train, y_train) #Train the model using training dta
```

```
y_pred = lr.predict(X_test) #Make predictiobns using the testing data

#Display coefficients
print ("Coefficients: \n")
print('Intercept :{0:2.4f}'.format(lr.intercept_))
for ii , coef in enumerate(lr.coef_):
    print('Coeff-{0:2d}:{1:2.4f}'.format(ii, coef))
plt.bar(range(len(lr.coef_)), lr.coef_)
plt.xticks(range(len(lr.coef_)))
plt.xlabel('Index')
plt.ylabel('Coefficient')
plt.show()

print('\nMean squared error: {:2.4f}'.format(mean_squared_error(y_test,␣
 ↪y_pred)))


plt.figure()
plt.scatter(X_test[:, 1], y_test, color="black")
plt.plot(np.sort(X_test[:, 1]), y_test[np.argsort(X_test)], color="blue",␣
 ↪linewidth=3)

plt.show()
```

```
Coefficients:

Intercept :0.0000
Coeff- 0:3.0011
Coeff- 1:-0.1721
Coeff- 2:-0.7934
```

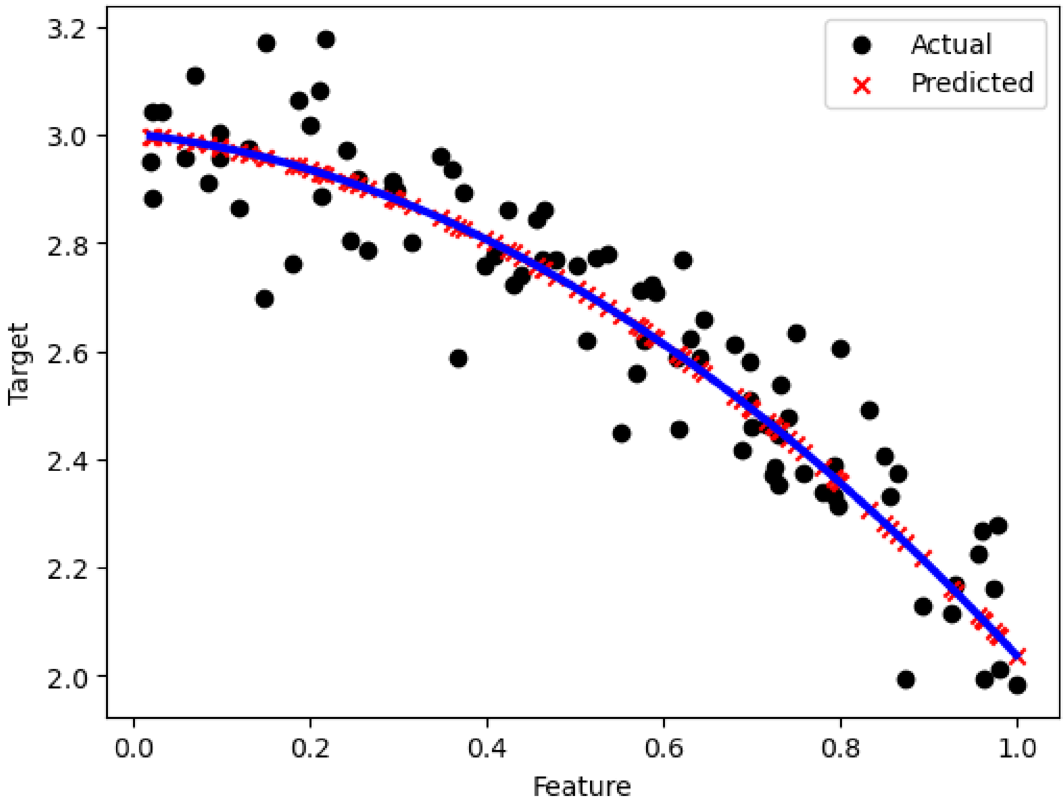



```
Mean squared error: 0.0116
```




[16]:
```python
# The mean squared error
print('\nMean squared error: {:2.4f}'.format(mean_squared_error(y_test,␣
↪y_pred)))

# Plot output
plt.figure()
plt.scatter(X_test[:, 1], y_test, color="black", label="Actual")
plt.scatter(X_test[:, 1], y_pred, color="red", label="Predicted", marker='x')
plt.plot(np.sort(X_test[:, 1]), y_pred[np.argsort(X_test[:, 1])], color="blue",␣
↪linewidth=3)
plt.xlabel('Feature')
plt.ylabel('Target')
plt.legend()
plt.show()
```

```
Mean squared error: 0.0116
```



Actual
Predicted
Target
Feature
3.2
3.0
2.8
2.6
2.4
2.2
2.0
0.0
0.2
0.4
0.6
0.8
1.0


[ ]:

# linear-regression-grad-descent

November 10, 2024

## 0.1 Import lib

[3]:
```
pip install numpy scipy scikit-learn pandas matplotlib
```

```
Requirement already satisfied: numpy in c:\users\vaibh\anaconda3\lib\site-
packages (1.23.5)
Requirement already satisfied: scipy in c:\users\vaibh\anaconda3\lib\site-
packages (1.10.0)
Requirement already satisfied: scikit-learn in
c:\users\vaibh\anaconda3\lib\site-packages (1.2.1)
Requirement already satisfied: pandas in c:\users\vaibh\anaconda3\lib\site-
packages (1.5.3)
Requirement already satisfied: matplotlib in c:\users\vaibh\anaconda3\lib\site-
packages (3.7.0)
Requirement already satisfied: joblib>=1.1.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from scikit-learn) (1.1.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in
c:\users\vaibh\anaconda3\lib\site-packages (from scikit-learn) (2.2.0)
Requirement already satisfied: python-dateutil>=2.8.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from pandas) (2022.7)
Requirement already satisfied: pyparsing>=2.3.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: cycler>=0.10 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: packaging>=20.0 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (22.0)
Requirement already satisfied: contourpy>=1.0.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (1.0.5)
Requirement already satisfied: fonttools>=4.22.0 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: pillow>=6.2.0 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (9.4.0)
Requirement already satisfied: six>=1.5 in c:\users\vaibh\anaconda3\lib\site-
packages (from python-dateutil>=2.8.1->pandas) (1.16.0)
```

```
Note: you may need to restart the kernel to use updated packages.
```

[4]:
```python
import numpy as np
import matplotlib.pyplot as plt
```

## 0.2 Load data

[5]:
```python
x=[[60,67,71,75,78],
  [22,24,15,20,16]]
y = [140,159,192,200,212]
```

[6]:
```python
## Add bias
x.insert(0, [1 for ii in x[0]]);
```

[7]:
```python
x
```

[7]:
```
[[1, 1, 1, 1, 1], [60, 67, 71, 75, 78], [22, 24, 15, 20, 16]]
```

[8]:
```python
mse = lambda w,x,y: np.mean((x.T@w - y)**2, 0)/2;
```

## 0.3 Constants

[9]:
```python
EPOCHS = 2;   #Max iterations
alpha = 0.0002; #Learning rate
w= [0,1,1]; #initial weights
```

[10]:
```python
w = np.array(w)
x = np.array(x)
y = np.array(y)

for ii in range(EPOCHS+1):
    print('-'*10, f'Iter-{ii}', '-'*10);
    print(f'W_0: {w[0]: 5.4f}');
    print(f'W_1: {w[1]: 5.4f}');
    print(f'W_2: {w[2]: 5.4f}');
    e= mse(w,x,y);
    print(f'MSE:{e: 9.4f}');
    w= w-alpha * np.mean((x.T@w - y)*x, 1)
```

```
---------- Iter-0 ----------
W_0:  0.0000
W_1:  1.0000
W_2:  1.0000
MSE: 4417.3000
---------- Iter-1 ----------
W_0:  0.0182
W_1:  2.3056
```

```
W_2:  1.3394
MSE: 172.0173
---------- Iter-2 ----------
W_0:  0.0167
W_1:  2.2223
W_2:  1.3004
MSE: 151.5906
```

[11]:
```python
EPOCHS = 4;      #Max. iterations
alpha = 0.0005; #Learning rate
w = [0, 1, 1];  #Initial weights
```

[12]:
```python
w = np.array(w)
x = np.array(x)
y = np.array(y)

for ii in range(EPOCHS+1):
    print('-'*10, f'Iter-{ii}','-'*10);
    print(f'W_0: {w[0]: 5.4f}');
    print(f'W_1: {w[1]: 5.4f}');
    print(f'W_2: {w[2]: 5.4f}');
    e = mse(w, x, y);
    print(f'MSE:{e: 9.4f}');
    w = w-alpha * np.mean((x.T@w - y)*x, 1)
```

```
---------- Iter-0 ----------
W_0:  0.0000
W_1:  1.0000
W_2:  1.0000
MSE: 4417.3000
---------- Iter-1 ----------
W_0:  0.0455
W_1:  4.2641
W_2:  1.8486
MSE: 12013.9064
---------- Iter-2 ----------
W_0: -0.0318
W_1: -1.1531
W_2:  0.3317
MSE: 33157.7072
---------- Iter-3 ----------
W_0:  0.0957
W_1:  7.9107
W_2:  2.7618
MSE: 91997.6012
---------- Iter-4 ----------
W_0: -0.1185
W_1: -7.1817
```

```
W_2: -1.3911
MSE: 255729.9348
```

[ ]:

```

```

# pt-2-linear-regression-study-model

November 10, 2024

[2]:
```
pip install numpy scipy scikit-learn pandas matplotlib
```

```
Requirement already satisfied: numpy in c:\users\vaibh\anaconda3\lib\site-
packages (1.23.5)
Requirement already satisfied: scipy in c:\users\vaibh\anaconda3\lib\site-
packages (1.10.0)
Requirement already satisfied: scikit-learn in
c:\users\vaibh\anaconda3\lib\site-packages (1.2.1)
Requirement already satisfied: pandas in c:\users\vaibh\anaconda3\lib\site-
packages (1.5.3)
Requirement already satisfied: matplotlib in c:\users\vaibh\anaconda3\lib\site-
packages (3.7.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in
c:\users\vaibh\anaconda3\lib\site-packages (from scikit-learn) (2.2.0)
Requirement already satisfied: joblib>=1.1.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from scikit-learn) (1.1.1)
Requirement already satisfied: pytz>=2020.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from pandas) (2022.7)
Requirement already satisfied: python-dateutil>=2.8.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pyparsing>=2.3.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: cycler>=0.10 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: kiwisolver>=1.0.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: contourpy>=1.0.1 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (1.0.5)
Requirement already satisfied: fonttools>=4.22.0 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: packaging>=20.0 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (22.0)
Requirement already satisfied: pillow>=6.2.0 in
c:\users\vaibh\anaconda3\lib\site-packages (from matplotlib) (9.4.0)
Requirement already satisfied: six>=1.5 in c:\users\vaibh\anaconda3\lib\site-
packages (from python-dateutil>=2.8.1->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

[3]:
```
import numpy as np #for numerical cal and matrix handling
import matplotlib.pyplot as plt # for plotting

from sklearn.linear_model import LinearRegression # for linear regression
from sklearn.model_selection import train_test_split # Divide dat as training
from sklearn.metrics import mean_squared_error # for evaluation

np.random.seed(0) # to control the randopm num generator
```

[4]:
```
%matplotlib inline
```

[5]:
```
from sklearn import datasets
X,y = datasets.load_diabetes(return_X_y=True)
```

[6]:
```
import pandas as pd
filename = 'Solar_radiation_classification.csv' #Path to external in CSV format
data = pd.read_csv(filename, header=0)
data.drop(columns=['Class'], inplace=True)
X = data.values[:, :-1]
y = data.values[:, -1]
data
```

[6]:
```
     Air Temperature (C°)  Air Temperature Uncertainty (C°)  \
0                1.546567                               0.0
1                1.366193                               0.0
2                0.991570                               0.0
3                0.145200                               0.0
4               -0.590171                               0.0
..                    …                                  …
636             -0.534672                               0.0
637             -0.090674                               0.0
638              0.908321                               0.0
639              1.310693                               0.0
640              1.560442                               0.0

     Wind Direction at 3m (°N)  Wind Direction at 3m Uncertainty (°N)  \
0                    -1.467370                                0.18201
1                    -1.442493                                0.18201
2                    -1.351279                                0.18201
3                    -1.185436                                0.18201
4                    -0.994717                                0.18201
..                         …                                      …
636                  -0.920088                                0.18201
637                  -0.223547                                0.18201
638                  -1.517122                                0.18201
639                  -1.434201                                0.18201
640                  -1.475662                                0.18201
```

```
     Wind Speed at 3m (m/s)  Wind Speed at 3m Uncertainty (m/s)  \
0                 -0.331420                            0.710869
1                 -0.459034                           -1.483082
2                 -0.841878                           -1.483082
3                 -1.097107                           -1.483082
4                 -0.969492                           -1.483082
..                      …                                  …
636               -0.841878                           -1.483082
637               -1.479950                           -1.483082
638               -1.097107                           -1.483082
639               -1.479950                           -1.483082
640               -1.224721                           -1.483082

     Wind Speed at 3m (std dev) (m/s)  DHI (Wh/m2)  DHI Uncertainty (Wh/m2)  \
0                           -0.239125     2.033152                 2.543475
1                           -0.709637     0.726671                 1.321083
2                           -1.180148     0.040304                 0.242664
3                           -1.180148    -0.687110                -0.125577
4                           -0.709637    -1.045514                -0.243248
..                                …            …                        …
636                         -1.180148     0.825019                 0.710578
637                         -1.415404     0.544577                 0.763184
638                         -0.709637     1.575850                 1.660252
639                         -0.944892     1.397198                 1.233868
640                         -0.474381     0.397882                 0.136068

     Standard Deviation DHI (Wh/m2)  …  Standard Deviation DNI (Wh/m2)  \
0                          0.000000  …                        0.000000
1                          0.000000  …                        0.000000
2                          0.000000  …                        0.000000
3                          0.000000  …                        0.000000
4                          0.000000  …                        0.000000
..                              …  …                             …
636                        1.135155  …                        1.447571
637                        0.859923  …                        1.456262
638                        1.241456  …                        0.382034
639                       -0.461654  …                       -1.591192
640                       -0.620243  …                       -1.078432

     GHI Uncertainty (Wh/m2)  Standard Deviation GHI (Wh/m2)  \
0                   0.450382                    4.131493e-16
1                   0.486600                    4.131493e-16
2                   0.104739                    4.131493e-16
3                  -0.046037                    4.131493e-16
4                  -0.098592                    4.131493e-16
..                       …                             …
```

```
636               -0.016709                      1.710223e+00
637                0.100802                      2.131415e+00
638                0.125407                      7.032136e-01
639                0.127178                     -1.609528e+00
640                0.165561                     -9.993622e-01

     Peak Wind Speed at 3m (m/s)  Peak Wind Speed at 3m Uncertainty (m/s)  \
0                       0.335854                                  0.182479
1                      -0.359885                                  0.182479
2                      -1.148389                                  0.182479
3                      -1.032433                                  0.182479
4                      -0.846902                                  0.182479
..                           …                                         …
636                    -0.545415                                  0.182479
637                    -0.730946                                  0.182479
638                    -0.475841                                  0.182479
639                    -0.962859                                  0.182479
640                    -0.730946                                  0.182479

     Relative Humidity (%)  Relative Humidity Uncertainty (%)  \
0                -1.336438                                0.0
1                -1.234901                                0.0
2                -1.152704                                0.0
3                -0.983476                                0.0
4                 0.249471                                0.0
..                     …                                 …
636              -0.200192                                0.0
637              -0.209862                                0.0
638              -0.978641                                0.0
639              -1.263911                                0.0
640              -1.215561                                0.0

     Barometric Pressure (mB (hPa equiv))  \
0                               -0.449612
1                               -0.402440
2                               -0.299112
3                               -0.180059
4                               -0.132888
..                                    …
636                             -0.429396
637                             -0.463090
638                             -0.523739
639                             -0.582142
640                             -0.663008

     Barometric Pressure Uncertainty (mB (hPa equiv))  GHI (Wh/m2)
0                                           -0.469979     0.509240
```

```
1                                                           -0.469979    0.520716
2                                                           -0.469979    0.382095
3                                                           -0.021681    0.086544
4                                                           -0.021681   -0.343249
..                                                                 …          …
636                                                         -0.469979   -0.064799
637                                                         -0.469979    0.374960
638                                                         -0.469979    0.611657
639                                                         -0.469979    0.796787
640                                                         -0.918278    0.768248

[641 rows x 22 columns]
```

[7]:

```
print(y.shape)
print(X.shape)
```

```
(641,)
(641, 21)
```

[8]:

```
print('Number of training examples: ', X.shape[0])
print('Number of training examples: ', y.shape[1] if len(y.shape)>1 else 1)
```

```
Number of training examples:  641
Number of training examples:  1
```

#build and evaluate model

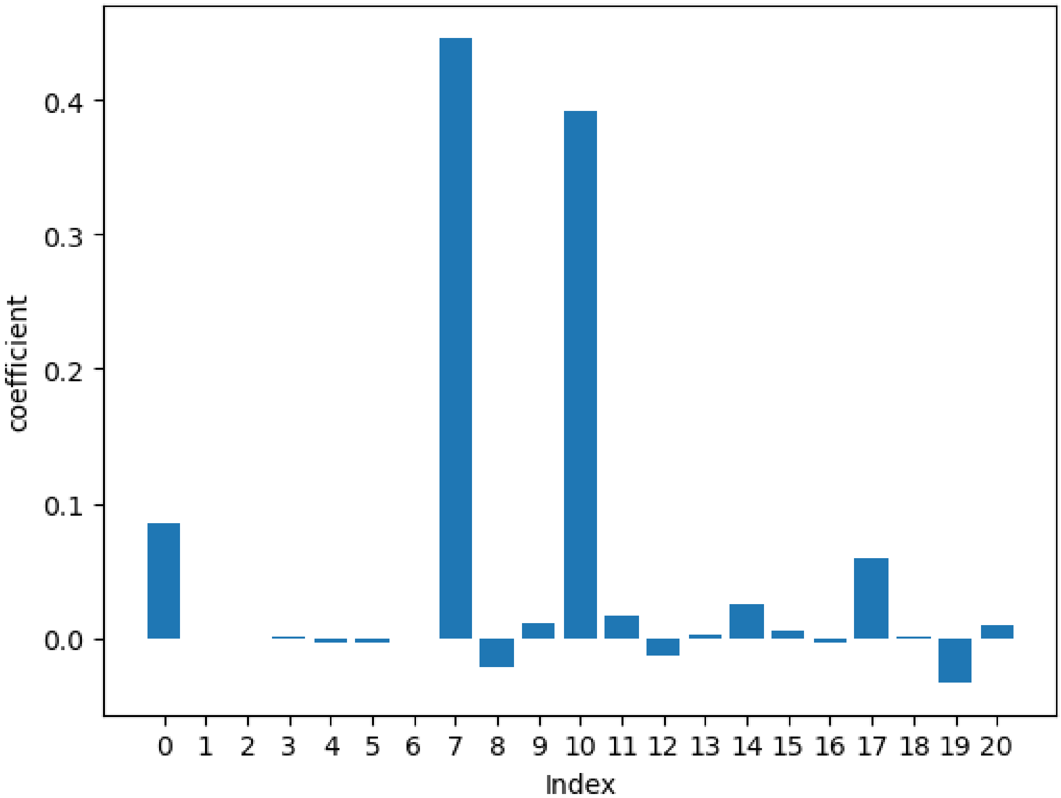
[11]:

```
# Split the data into training/testing sets

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

lr = LinearRegression(fit_intercept=True) # Initialize Linear regression model

lr.fit(X_train, y_train) # Train the model using the training data

y_pred = lr.predict(X_test) # Make predictions using the testing data

# Display coefficients

print("Coefficients: \n")

print('Intercept: {0:2.4f}'.format(lr.intercept_))

for ii, coef in enumerate(lr.coef_):
    print('Coeff-{0:2d}: {1:2.4f}'.format(ii, coef))

plt.bar(range(len(lr.coef_)), lr.coef_)
plt.xticks(range(len(lr.coef_)))
```

```
plt.xlabel('Index')
plt.ylabel('coefficient')
plt.show()
#The mean squared error

print('\nMean squared error: {:2.4f}'.format(mean_squared_error(y_test,␣
↪y_pred)))

#Plot outputs
#plt.figure()
#plt.scatter(X_test[:, 1], y test, color="black")
#plt.plot(np.sort(X_test[:, 1]), y_test[np.argsort(X_test)], color="blue",␣
↪Linewidth=3)
#plt.show()

print('\nMean squared error: {:2.4f}'.format(mean_squared_error(y_test,␣
↪y_pred)))


plt.figure()
plt.scatter(X_test[:, 1], y_test, color="black")
plt.plot(np.sort(X_test[:, 1]), y_test[np.argsort(X_test)], color="blue",␣
↪linewidth=3)

plt.show()
```

```
Coefficients:

Intercept: 0.0880
Coeff- 0: 0.0855
Coeff- 1: -0.0000
Coeff- 2: -0.0011
Coeff- 3: 0.0005
Coeff- 4: -0.0032
Coeff- 5: -0.0033
Coeff- 6: -0.0003
Coeff- 7: 0.4462
Coeff- 8: -0.0220
Coeff- 9: 0.0103
Coeff-10: 0.3916
Coeff-11: 0.0161
Coeff-12: -0.0139
Coeff-13: 0.0020
Coeff-14: 0.0247
Coeff-15: 0.0051
Coeff-16: -0.0032
Coeff-17: 0.0595
```

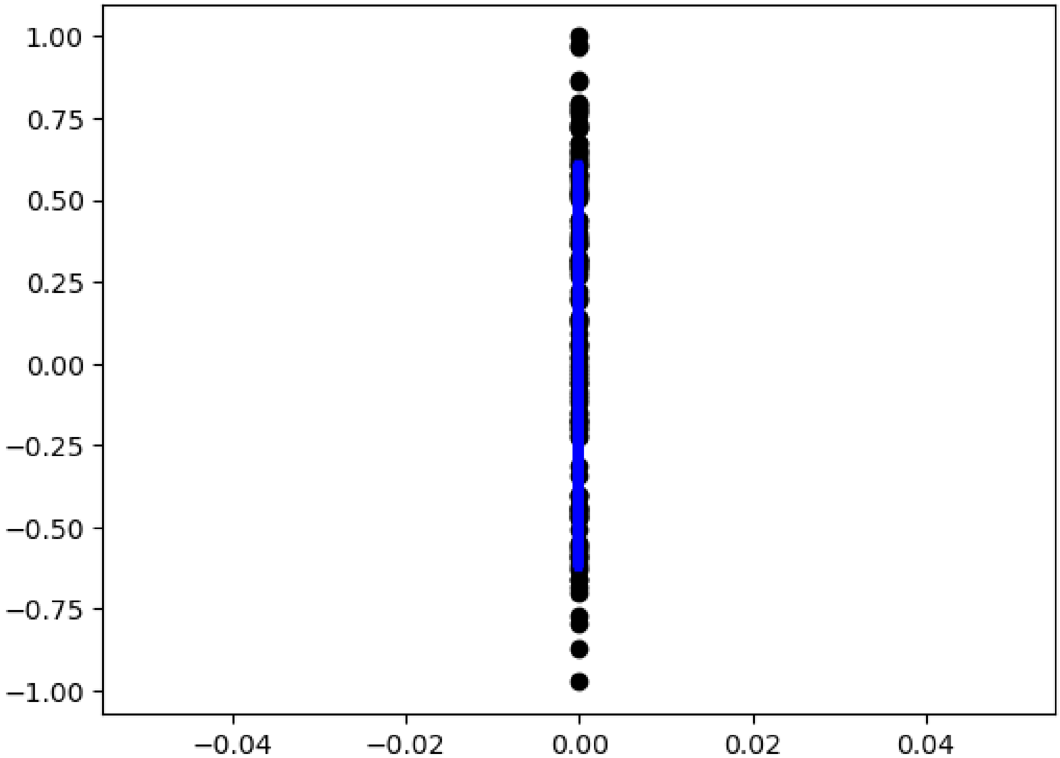
```
Coeff-18: 0.0009
Coeff-19: -0.0332
Coeff-20: 0.0089
```




```
Mean squared error: 0.0025
```



```
Mean squared error: 0.0025
```


1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−0.04
−0.02
0.00
0.02
0.04

[ ]:

# exp-3-logistic-regression

November 10, 2024

[5]:
```
%matplotlib inline
```

## 1 Logistic Regression Example

[6]:
```
# Import necessary libraries
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report,␣
↪confusion_matrix
```

[7]:
```
import pandas as pd
filename = 'Solar_radiation_classification.csv' # Path to external dataset in␣
↪CSV format
data = pd.read_csv(filename, header=0)
```

[8]:
```
# Inspect the data
print(data.head())
```

```
   Air Temperature (C°)  Air Temperature Uncertainty (C°)  \
0              1.598833                          -0.039841
1              1.415552                          -0.039841
2              1.034891                          -0.039841
3              0.174880                          -0.039841
4             -0.572343                          -0.039841

   Wind Direction at 3m (°N)  Wind Direction at 3m Uncertainty (°N)  \
0                  -1.358521                               0.163933
1                  -1.333544                               0.163933
2                  -1.241959                               0.163933
3                  -1.075441                               0.163933
4                  -0.883946                               0.163933

   Wind Speed at 3m (m/s)  Wind Speed at 3m Uncertainty (m/s)  \
0               -0.351941                            0.677977
1               -0.460051                           -1.548603
```

```
2             -0.784380                          -1.548603
3             -1.000600                          -1.548603
4             -0.892490                          -1.548603

   Wind Speed at 3m (std dev) (m/s)  DHI (Wh/m2)  DHI Uncertainty (Wh/m2)  \
0                         -0.309934     2.087472                 2.420683
1                         -0.731123     0.780151                 1.294103
2                         -1.152311     0.093343                 0.300212
3                         -1.152311    -0.634539                -0.039165
4                         -0.731123    -0.993174                -0.147613

   Standard Deviation DHI (Wh/m2)  …  GHI Uncertainty (Wh/m2)  \
0                   -6.322829e-16  …                 0.645691
1                   -6.322829e-16  …                 0.693011
2                   -6.322829e-16  …                 0.194101
3                   -6.322829e-16  …                -0.002891
4                   -6.322829e-16  …                -0.071556

   Standard Deviation GHI (Wh/m2)  Peak Wind Speed at 3m (m/s)  \
0                             0.0                     0.358986
1                             0.0                    -0.401816
2                             0.0                    -1.264058
3                             0.0                    -1.137258
4                             0.0                    -0.934377

   Peak Wind Speed at 3m Uncertainty (m/s)  Relative Humidity (%)  \
0                                 0.116591              -1.346289
1                                 0.116591              -1.239218
2                                 0.116591              -1.152541
3                                 0.116591              -0.974088
4                                 0.116591               0.326067

   Relative Humidity Uncertainty (%)  Barometric Pressure (mB (hPa equiv))  \
0                           -0.04948                             -0.350323
1                           -0.04948                             -0.305136
2                           -0.04948                             -0.206155
3                           -0.04948                             -0.092112
4                           -0.04948                             -0.046925

   Barometric Pressure Uncertainty (mB (hPa equiv))  GHI (Wh/m2)       Class
0                                         -0.364441     0.516210     Running
1                                         -0.364441     0.527461     Running
2                                         -0.364441     0.391562  Monitoring
3                                          0.054434     0.101813  Monitoring
4                                          0.054434    -0.319541  Monitoring

[5 rows x 23 columns]
```

[9]:
```
data['Class'].value_counts()
```

[9]:
```
Class
Monitoring    576
Running       430
Inspecting    256
Name: count, dtype: int64
```

[10]:
```
# # Define feature columns and the target column
# X = data.drop('Class', axis=1)  # Features (all columns except 'Class')
# y = data['Class'].map({'Running': 1, 'Monitoring': 0})  # Convert 'Running'␣
↪to 1 and 'Monitoring' to 0
```

[11]:
```
# Define feature columns and the target column
X = data.drop('Class', axis=1)  # Features (all columns except 'Class')
y = data['Class']  # Assuming 'Class' column has 3 unique classes
```

[12]:
```
# Split the dataset into training and testing sets (80% train, 20% test)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
↪random_state=42)
```

[13]:
```
# Standardize the features
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

[14]:
```
# Initialize the logistic regression model for multiclass classification
#log_reg = LogisticRegression()  for 2 class classification
log_reg =␣
↪LogisticRegression(multi_class='multinomial',solver='lbfgs',max_iter=1000)
```

[15]:
```
# Fit the model to the training data
log_reg.fit(X_train_scaled, y_train)

# Make predictions on the test data
y_pred = log_reg.predict(X_test_scaled)
```

[16]:
```
# Evaluate the model
from sklearn.metrics import␣
↪accuracy_score,recall_score,precision_score,f1_score,precision_recall_curve
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy:.4f}')
recall=recall_score(y_test, y_pred,average='macro')
# macro: Unweighted average of the metrics for each class.All classes are␣
↪treated equally.
print(f'Recall: {recall:.4f}')
precision=precision_score(y_test, y_pred,average='macro')
```

```
print(f'Precision: {precision:.4f}')
f1=f1_score(y_test, y_pred,average='macro')
print(f'F1-Score: {f1:.4f}')
```

```
Accuracy: 0.9763
Recall: 0.9793
Precision: 0.9734
F1-Score: 0.9761
```

[17]:
```
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy:.4f}')
recall=recall_score(y_test, y_pred,average='weighted')
print(f'Recall: {recall:.4f}')
#weighted: Takes class imbalance into account by weighting each class's␣
↪contribution by
#its support (number of true instances).
precision=precision_score(y_test, y_pred,average='weighted')
print(f'Precision: {precision:.4f}')
f1=f1_score(y_test, y_pred,average='weighted')
print(f'F1-Score: {f1:.4f}')
```

```
Accuracy: 0.9763
Recall: 0.9763
Precision: 0.9767
F1-Score: 0.9762
```

[18]:
```
print('Confusion Matrix:')
print(confusion_matrix(y_test, y_pred))
```

```
Confusion Matrix:
[[ 53   1   0]
 [  2 110   3]
 [  0   0  84]]
```

[19]:
```
print('Classification Report:')
print(classification_report(y_test, y_pred))
```

```
Classification Report:
              precision    recall  f1-score   support

  Inspecting       0.96      0.98      0.97        54
  Monitoring       0.99      0.96      0.97       115
     Running       0.97      1.00      0.98        84

    accuracy                           0.98       253
   macro avg       0.97      0.98      0.98       253
weighted avg       0.98      0.98      0.98       253
```

[20]:

```
# # Import necessary libraries
# import numpy as np
# import matplotlib.pyplot as plt
# from sklearn.metrics import precision_recall_curve, average_precision_score
# from sklearn.preprocessing import label_binarize
# from sklearn.metrics import PrecisionRecallDisplay

# # Assuming y_test are the true labels and y_pred are the predicted␣
 ↪probabilities
# # Let's also assume there are 3 classes in the classification task (adjust if␣
 ↪needed)

# n_classes = 3  # Number of classes
# # Binarize the output (one-vs-rest strategy)
# y_test_bin = label_binarize(y_test, classes=[0, 1, 2])
# y_pred_prob = log_reg.predict_proba(X_test_scaled)

# # Initialize variables to store precision, recall, and average precision for␣
 ↪each class
# precision = dict()
# recall = dict()
# average_precision = dict()

# # Calculate precision-recall curve and average precision for each class
# for i in range(n_classes):
#     precision[i], recall[i], _ = precision_recall_curve(y_test_bin[:, i],␣
 ↪y_pred_prob[:, i])
#     average_precision[i] = average_precision_score(y_test_bin[:, i],␣
 ↪y_pred_prob[:, i])

# # Plot the Precision-Recall curve for each class
# plt.figure(figsize=(8, 6))
# for i in range(n_classes):
#     plt.plot(recall[i], precision[i], lw=2, label=f'Class {i} (AP =␣
 ↪{average_precision[i]:0.2f})')

# plt.xlabel('Recall')
# plt.ylabel('Precision')
# plt.title('Precision-Recall Curve for Multiclass Classification')
# plt.legend(loc='best')
# plt.grid()
# plt.show()
```

[ ]:

```

```

# exp-4-decisiontree

November 10, 2024

[1]:
```
# Import necessary libraries
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report,␣
↪confusion_matrix
```

[2]:
```
# Load the dataset
data = pd.read_csv('Solar_radiation_classification.xls')

# Inspect the data
#print(data.head())
data.head()
```

[2]:
```
   Air Temperature (C°)  Air Temperature Uncertainty (C°)  \
0              1.598833                         -0.039841
1              1.415552                         -0.039841
2              1.034891                         -0.039841
3              0.174880                         -0.039841
4             -0.572343                         -0.039841

   Wind Direction at 3m (°N)  Wind Direction at 3m Uncertainty (°N)  \
0                  -1.358521                               0.163933
1                  -1.333544                               0.163933
2                  -1.241959                               0.163933
3                  -1.075441                               0.163933
4                  -0.883946                               0.163933

   Wind Speed at 3m (m/s)  Wind Speed at 3m Uncertainty (m/s)  \
0               -0.351941                            0.677977
1               -0.460051                           -1.548603
2               -0.784380                           -1.548603
3               -1.000600                           -1.548603
4               -0.892490                           -1.548603

   Wind Speed at 3m (std dev) (m/s)  DHI (Wh/m2)  DHI Uncertainty (Wh/m2)  \
```

```
0                          -0.309934     2.087472                 2.420683
1                          -0.731123     0.780151                 1.294103
2                          -1.152311     0.093343                 0.300212
3                          -1.152311    -0.634539                -0.039165
4                          -0.731123    -0.993174                -0.147613

   Standard Deviation DHI (Wh/m2)  …  GHI Uncertainty (Wh/m2)  \
0                   -6.322829e-16  …                 0.645691
1                   -6.322829e-16  …                 0.693011
2                   -6.322829e-16  …                 0.194101
3                   -6.322829e-16  …                -0.002891
4                   -6.322829e-16  …                -0.071556

   Standard Deviation GHI (Wh/m2)  Peak Wind Speed at 3m (m/s)  \
0                             0.0                     0.358986
1                             0.0                    -0.401816
2                             0.0                    -1.264058
3                             0.0                    -1.137258
4                             0.0                    -0.934377

   Peak Wind Speed at 3m Uncertainty (m/s)  Relative Humidity (%)  \
0                                 0.116591              -1.346289
1                                 0.116591              -1.239218
2                                 0.116591              -1.152541
3                                 0.116591              -0.974088
4                                 0.116591               0.326067

   Relative Humidity Uncertainty (%)  Barometric Pressure (mB (hPa equiv))  \
0                           -0.04948                             -0.350323
1                           -0.04948                             -0.305136
2                           -0.04948                             -0.206155
3                           -0.04948                             -0.092112
4                           -0.04948                             -0.046925

   Barometric Pressure Uncertainty (mB (hPa equiv))  GHI (Wh/m2)       Class
0                                         -0.364441     0.516210     Running
1                                         -0.364441     0.527461     Running
2                                         -0.364441     0.391562  Monitoring
3                                          0.054434     0.101813  Monitoring
4                                          0.054434    -0.319541  Monitoring

[5 rows x 23 columns]
```

[3]:
```python
# Define feature columns and the target column
X = data.drop('Class', axis=1)  # Features (all columns except 'Class')
y = data['Class']  # Assuming 'Class' column has 3 unique classes
print(data['Class'].isna().sum())  # Should output 0 if cleaned properly
```

0

[4]:
```python
data['Class'].value_counts()
```

[4]:
```
Monitoring    576
Running       430
Inspecting    256
Name: Class, dtype: int64
```

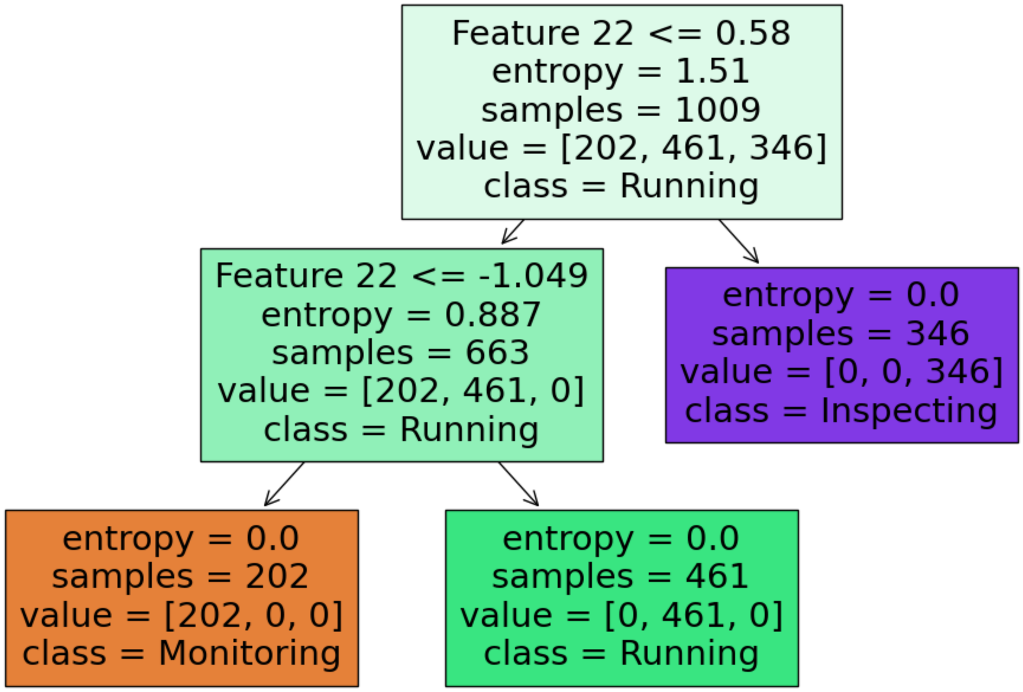
[5]:
```python
X.shape,y.shape
```

[5]:
```
((1262, 22), (1262,))
```

[6]:
```python
# Split the dataset into training and testing sets (80% train, 20% test)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
↪random_state=42)

# Standardize the features (optional, but may improve performance for some␣
↪models)
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

[22]:
```python
# Initialize the Decision Tree classifier
dt_classifier = DecisionTreeClassifier(criterion="entropy", max_depth=3,␣
↪random_state=42)



# Fit the model to the training data
dt_classifier.fit(X_train_scaled, y_train)

# Make predictions on the test data
y_pred = dt_classifier.predict(X_test_scaled)

#vaibhav Chavan

import matplotlib.pyplot as plt
from sklearn import tree
feature_names = [f'Feature {i+1}' for i in range(22)]  # Modify with actual␣
↪feature names if available
class_names = ['Monitoring', 'Running', 'Inspecting']  # Modify with actual␣
↪class names

# Visualize the Decision Tree
plt.figure(figsize=(12,8))
tree.plot_tree(dt_classifier, filled=True, feature_names=feature_names,␣
↪class_names=class_names)
```

```
plt.savefig("Solar_Radiation_dt_log_loss_2.pdf")
plt.show()
```




[9]:
```
# Classification report for detailed metrics
print("\nClassification Report:")
print(classification_report(y_test, y_pred, target_names=class_names))
```

```
Classification Report:
              precision    recall  f1-score   support

  Monitoring       1.00      1.00      1.00        54
     Running       1.00      1.00      1.00       115
  Inspecting       1.00      1.00      1.00        84

    accuracy                           1.00       253
   macro avg       1.00      1.00      1.00       253
weighted avg       1.00      1.00      1.00       253
```

[10]:
```
# Evaluate the model
from sklearn.metrics import␣
↪accuracy_score,recall_score,precision_score,f1_score,precision_recall_curve
```

```python
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy:.4f}')
recall=recall_score(y_test, y_pred,average='macro')
# macro: Unweighted average of the metrics for each class.All classes are␣
↪treated equally.
print(f'Recall: {recall:.4f}')
precision=precision_score(y_test, y_pred,average='macro')
print(f'Precision: {precision:.4f}')
f1=f1_score(y_test, y_pred,average='macro')
print(f'F1-Score: {f1:.4f}')
```

```
Accuracy: 1.0000
Recall: 1.0000
Precision: 1.0000
F1-Score: 1.0000
```

[11]:
```python
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy:.4f}')
recall=recall_score(y_test, y_pred,average='weighted')
print(f'Recall: {recall:.4f}')
#weighted: Takes class imbalance into account by weighting each class's␣
↪contribution by
#its support (number of true instances).
precision=precision_score(y_test, y_pred,average='weighted')
print(f'Precision: {precision:.4f}')
f1=f1_score(y_test, y_pred,average='weighted')
print(f'F1-Score: {f1:.4f}')
```

```
Accuracy: 1.0000
Recall: 1.0000
Precision: 1.0000
F1-Score: 1.0000
```

[12]:
```python
print('Confusion Matrix:')
print(confusion_matrix(y_test, y_pred))
```

```
Confusion Matrix:
[[ 54   0   0]
 [  0 115   0]
 [  0   0  84]]
```

[13]:
```python
print('Classification Report:')
print(classification_report(y_test, y_pred))
```

```
Classification Report:
              precision    recall  f1-score   support

  Inspecting       1.00      1.00      1.00        54
```

```
  Monitoring       1.00      1.00      1.00       115
     Running       1.00      1.00      1.00        84

    accuracy                           1.00       253
   macro avg       1.00      1.00      1.00       253
weighted avg       1.00      1.00      1.00       253
```

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

# exp-5-naivebayes

November 10, 2024

[1]:
```
# Import necessary libraries
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score, classification_report,␣
↪confusion_matrix
import matplotlib.pyplot as plt
```

[8]:
```
# Load the dataset
data = pd.read_csv('Solar_radiation_classification.xls')

# Inspect the data
#print(data.head())
```

[9]:
```
# Define feature columns and the target column
X = data.drop('Class', axis=1)  # Features (all columns except 'Class')
y = data['Class']  # Assuming 'Class' column has 3 unique classes
```

[10]:
```
data['Class'].value_counts()
```

[10]:
```
Monitoring    576
Running       430
Inspecting    256
Name: Class, dtype: int64
```

[11]:
```
X.shape,y.shape
```

[11]:
```
((1262, 22), (1262,))
```

[12]:
```
# Split the dataset into training and testing sets (80% train, 20% test)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
↪random_state=1)

# Standardize the features (optional, but may improve performance for some␣
↪models)
scaler = StandardScaler()
```

```
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

[13]:
```
# Initialize the NaiveBayes classifier
gnb= GaussianNB()

# Fit the model to the training data
gnb.fit(X_train_scaled, y_train)

# Make predictions on the test data
y_pred = gnb.predict(X_test_scaled)
```

[14]:
```
print('Confusion Matrix:')
print(confusion_matrix(y_test, y_pred))
```

```
Confusion Matrix:
[[ 11  38   0]
 [  0 103   4]
 [  0   2  95]]
```

[15]:
```
print('Classification Report:')
print(classification_report(y_test, y_pred))
```

```
Classification Report:
              precision    recall  f1-score   support

  Inspecting       1.00      0.22      0.37        49
  Monitoring       0.72      0.96      0.82       107
     Running       0.96      0.98      0.97        97

    accuracy                           0.83       253
   macro avg       0.89      0.72      0.72       253
weighted avg       0.87      0.83      0.79       253
```

[16]:
```
#comparing actual response values (y_test) with predicted response values␣
↪(y_pred)
from sklearn import metrics
print ("Gaussian Naive Bayes model accuracy(in% ):", metrics.
↪accuracy_score(y_test,y_pred)*100)
```

```
Gaussian Naive Bayes model accuracy(in% ): 82.6086956521739
```

[ ]:

[ ]:

[ ]:

```

```

[ ]:

```

```

[ ]:

```

```

[ ]:

```

```

[ ]:

```

```

[ ]:

```

```

# exp-6-support-vectorclassification

November 10, 2024

[9]:
```
# Import necessary libraries
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report,␣
 ↪confusion_matrix
from sklearn.svm import SVC
```

[10]:
```
# Load the dataset
data = pd.read_csv('Solar_radiation_classification.csv')

# Inspect the data
#print(data.head())
data.head()
```

[10]:
```
   Air Temperature (C°)  Air Temperature Uncertainty (C°)  \
0              1.598833                         -0.039841
1              1.415552                         -0.039841
2              1.034891                         -0.039841
3              0.174880                         -0.039841
4             -0.572343                         -0.039841

   Wind Direction at 3m (°N)  Wind Direction at 3m Uncertainty (°N)  \
0                  -1.358521                               0.163933
1                  -1.333544                               0.163933
2                  -1.241959                               0.163933
3                  -1.075441                               0.163933
4                  -0.883946                               0.163933

   Wind Speed at 3m (m/s)  Wind Speed at 3m Uncertainty (m/s)  \
0               -0.351941                            0.677977
1               -0.460051                           -1.548603
2               -0.784380                           -1.548603
3               -1.000600                           -1.548603
4               -0.892490                           -1.548603
```

```
   Wind Speed at 3m (std dev) (m/s)  DHI (Wh/m2)  DHI Uncertainty (Wh/m2)  \
0                         -0.309934     2.087472                 2.420683
1                         -0.731123     0.780151                 1.294103
2                         -1.152311     0.093343                 0.300212
3                         -1.152311    -0.634539                -0.039165
4                         -0.731123    -0.993174                -0.147613

   Standard Deviation DHI (Wh/m2)  …  GHI Uncertainty (Wh/m2)  \
0                   -6.322829e-16  …                 0.645691
1                   -6.322829e-16  …                 0.693011
2                   -6.322829e-16  …                 0.194101
3                   -6.322829e-16  …                -0.002891
4                   -6.322829e-16  …                -0.071556

   Standard Deviation GHI (Wh/m2)  Peak Wind Speed at 3m (m/s)  \
0                             0.0                     0.358986
1                             0.0                    -0.401816
2                             0.0                    -1.264058
3                             0.0                    -1.137258
4                             0.0                    -0.934377

   Peak Wind Speed at 3m Uncertainty (m/s)  Relative Humidity (%)  \
0                                 0.116591              -1.346289
1                                 0.116591              -1.239218
2                                 0.116591              -1.152541
3                                 0.116591              -0.974088
4                                 0.116591               0.326067

   Relative Humidity Uncertainty (%)  Barometric Pressure (mB (hPa equiv))  \
0                           -0.04948                             -0.350323
1                           -0.04948                             -0.305136
2                           -0.04948                             -0.206155
3                           -0.04948                             -0.092112
4                           -0.04948                             -0.046925

   Barometric Pressure Uncertainty (mB (hPa equiv))  GHI (Wh/m2)       Class
0                                         -0.364441     0.516210     Running
1                                         -0.364441     0.527461     Running
2                                         -0.364441     0.391562  Monitoring
3                                          0.054434     0.101813  Monitoring
4                                          0.054434    -0.319541  Monitoring

[5 rows x 23 columns]
```

[11]:

```
# Define feature columns and the target column
X = data.drop('Class', axis=1)  # Features (all columns except 'Class')
y = data['Class']  # Assuming 'Class' column has 3 unique classes
```

```
print(data['Class'].isna().sum())  # Should output 0 if cleaned properly
```

0

[12]:
```
data['Class'].value_counts()
```

[12]:
```
Class
Monitoring    576
Running       430
Inspecting    256
Name: count, dtype: int64
```

[13]:
```
X.shape,y.shape
```

[13]:
```
((1262, 22), (1262,))
```

[14]:
```
# Split the dataset into training and testing sets (80% train, 20% test)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=42)

# Standardize the features (optional, but may improve performance for some␣
 ↪models)
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

[17]:
```
# Initialize the Decision Tree classifier
sv_classifier = SVC(kernel="linear",  random_state=0)


# Fit the model to the training data
sv_classifier.fit(X_train_scaled, y_train)

# Make predictions on the test data
y_pred = sv_classifier.predict(X_test_scaled)

#vaibhav Chavan
import matplotlib.pyplot as plt
from sklearn import tree
feature_names = [f'Feature {i+1}' for i in range(22)]  # Modify with actual␣
 ↪feature names if available
class_names = ['Monitoring', 'Running', 'Inspecting']  # Modify with actual␣
 ↪class names
```

[18]:
```
# Classification report for detailed metrics
print("\nClassification Report:")
print(classification_report(y_test, y_pred, target_names=class_names))
```

```
Classification Report:
              precision    recall  f1-score   support

  Monitoring       0.96      1.00      0.98        54
     Running       0.98      0.96      0.97       115
  Inspecting       0.96      0.98      0.97        84

    accuracy                           0.97       253
   macro avg       0.97      0.98      0.97       253
weighted avg       0.97      0.97      0.97       253
```

[19]:

```
# Evaluate the model
from sklearn.metrics import␣
↪accuracy_score,recall_score,precision_score,f1_score,precision_recall_curve
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy:.4f}')
recall=recall_score(y_test, y_pred,average='macro')
# macro: Unweighted average of the metrics for each class.All classes are␣
↪treated equally.
print(f'Recall: {recall:.4f}')
precision=precision_score(y_test, y_pred,average='macro')
print(f'Precision: {precision:.4f}')
f1=f1_score(y_test, y_pred,average='macro')
print(f'F1-Score: {f1:.4f}')
```

```
Accuracy: 0.9723
Recall: 0.9776
Precision: 0.9704
F1-Score: 0.9738
```

[20]:

```
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy:.4f}')
recall=recall_score(y_test, y_pred,average='weighted')
print(f'Recall: {recall:.4f}')
#weighted: Takes class imbalance into account by weighting each class's␣
↪contribution by
#its support (number of true instances).
precision=precision_score(y_test, y_pred,average='weighted')
print(f'Precision: {precision:.4f}')
f1=f1_score(y_test, y_pred,average='weighted')
print(f'F1-Score: {f1:.4f}')
```

```
Accuracy: 0.9723
Recall: 0.9723
Precision: 0.9725
F1-Score: 0.9723
```

[21]:
```
print('Confusion Matrix:')
print(confusion_matrix(y_test, y_pred))
```

```
Confusion Matrix:
[[ 54   0   0]
 [  2 110   3]
 [  0   2  82]]
```

[22]:
```
print('Classification Report:')
print(classification_report(y_test, y_pred))
```

```
Classification Report:
              precision    recall  f1-score   support

  Inspecting       0.96      1.00      0.98        54
  Monitoring       0.98      0.96      0.97       115
     Running       0.96      0.98      0.97        84

    accuracy                           0.97       253
   macro avg       0.97      0.98      0.97       253
weighted avg       0.97      0.97      0.97       253
```

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

# exp-7-kmeans

November 10, 2024

[10]:
```
# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
```

[11]:
```
# Load the dataset
df = pd.read_csv('Solar_radiation_classification.xls')

# Inspect the data
#print(data.head())
# Explore the dataset
print("First 5 rows:\n", df.head())
```

```
First 5 rows:
    Air Temperature (C°)  Air Temperature Uncertainty (C°)  \
0              1.598833                         -0.039841
1              1.415552                         -0.039841
2              1.034891                         -0.039841
3              0.174880                         -0.039841
4             -0.572343                         -0.039841

   Wind Direction at 3m (°N)  Wind Direction at 3m Uncertainty (°N)  \
0                  -1.358521                               0.163933
1                  -1.333544                               0.163933
2                  -1.241959                               0.163933
3                  -1.075441                               0.163933
4                  -0.883946                               0.163933

   Wind Speed at 3m (m/s)  Wind Speed at 3m Uncertainty (m/s)  \
0               -0.351941                            0.677977
1               -0.460051                           -1.548603
2               -0.784380                           -1.548603
3               -1.000600                           -1.548603
4               -0.892490                           -1.548603
```

```
   Wind Speed at 3m (std dev) (m/s)  DHI (Wh/m2)  DHI Uncertainty (Wh/m2)  \
0                         -0.309934     2.087472                 2.420683
1                         -0.731123     0.780151                 1.294103
2                         -1.152311     0.093343                 0.300212
3                         -1.152311    -0.634539                -0.039165
4                         -0.731123    -0.993174                -0.147613

   Standard Deviation DHI (Wh/m2)  …  GHI Uncertainty (Wh/m2)  \
0                   -6.322829e-16  …                 0.645691
1                   -6.322829e-16  …                 0.693011
2                   -6.322829e-16  …                 0.194101
3                   -6.322829e-16  …                -0.002891
4                   -6.322829e-16  …                -0.071556

   Standard Deviation GHI (Wh/m2)  Peak Wind Speed at 3m (m/s)  \
0                             0.0                     0.358986
1                             0.0                    -0.401816
2                             0.0                    -1.264058
3                             0.0                    -1.137258
4                             0.0                    -0.934377

   Peak Wind Speed at 3m Uncertainty (m/s)  Relative Humidity (%)  \
0                                 0.116591              -1.346289
1                                 0.116591              -1.239218
2                                 0.116591              -1.152541
3                                 0.116591              -0.974088
4                                 0.116591               0.326067

   Relative Humidity Uncertainty (%)  Barometric Pressure (mB (hPa equiv))  \
0                           -0.04948                             -0.350323
1                           -0.04948                             -0.305136
2                           -0.04948                             -0.206155
3                           -0.04948                             -0.092112
4                           -0.04948                             -0.046925

   Barometric Pressure Uncertainty (mB (hPa equiv))  GHI (Wh/m2)       Class
0                                         -0.364441     0.516210     Running
1                                         -0.364441     0.527461     Running
2                                         -0.364441     0.391562  Monitoring
3                                          0.054434     0.101813  Monitoring
4                                          0.054434    -0.319541  Monitoring

[5 rows x 23 columns]
```

[12]:
```
# Feature selection: Select columns for clustering (Adjust if necessary)
# selected_features = df[['DHI (Wh/m2)', 'Relative Humidity (%)', 'GHI (Wh/
↪m2)']]
```

```
selected_features1 = df[['DHI (Wh/m2)', 'GHI (Wh/m2)']]
```

[13]:
```
# Standardize the features to normalize the data
scaler = StandardScaler()
X_scaled = scaler.fit_transform(selected_features1)
```

[15]:
```
# Determine the optimal number of clusters using the Elbow method
wcss = []  # Within-cluster sum of squares

for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', random_state=42)
    kmeans.fit(X_scaled)
    wcss.append(kmeans.inertia_)
```

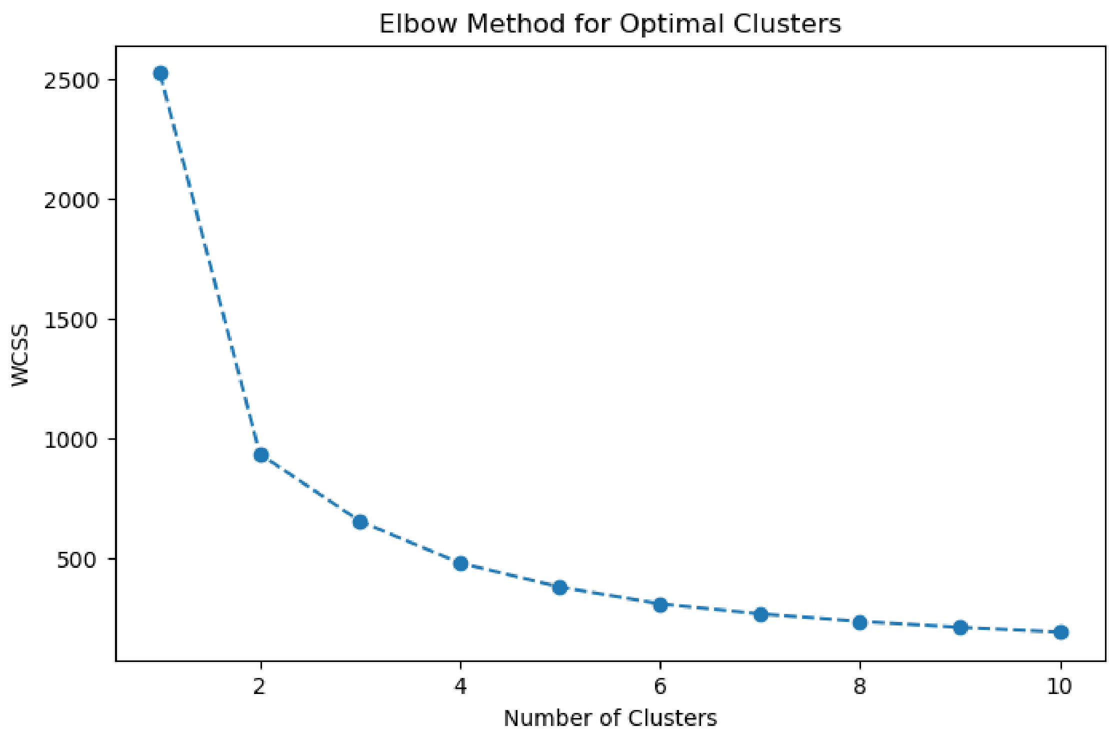
```
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1382:
UserWarning: KMeans is known to have a memory leak on Windows with MKL, when
there are less chunks than available threads. You can avoid it by setting the
environment variable OMP_NUM_THREADS=5.
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1382:
UserWarning: KMeans is known to have a memory leak on Windows with MKL, when
there are less chunks than available threads. You can avoid it by setting the
environment variable OMP_NUM_THREADS=5.
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1382:
UserWarning: KMeans is known to have a memory leak on Windows with MKL, when
there are less chunks than available threads. You can avoid it by setting the
environment variable OMP_NUM_THREADS=5.
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1382:
UserWarning: KMeans is known to have a memory leak on Windows with MKL, when
```

```
there are less chunks than available threads. You can avoid it by setting the
environment variable OMP_NUM_THREADS=5.
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1382:
UserWarning: KMeans is known to have a memory leak on Windows with MKL, when
there are less chunks than available threads. You can avoid it by setting the
environment variable OMP_NUM_THREADS=5.
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1382:
UserWarning: KMeans is known to have a memory leak on Windows with MKL, when
there are less chunks than available threads. You can avoid it by setting the
environment variable OMP_NUM_THREADS=5.
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1382:
UserWarning: KMeans is known to have a memory leak on Windows with MKL, when
there are less chunks than available threads. You can avoid it by setting the
environment variable OMP_NUM_THREADS=5.
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1382:
UserWarning: KMeans is known to have a memory leak on Windows with MKL, when
there are less chunks than available threads. You can avoid it by setting the
environment variable OMP_NUM_THREADS=5.
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1382:
UserWarning: KMeans is known to have a memory leak on Windows with MKL, when
there are less chunks than available threads. You can avoid it by setting the
environment variable OMP_NUM_THREADS=5.
  warnings.warn(
```

```
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1382:
UserWarning: KMeans is known to have a memory leak on Windows with MKL, when
there are less chunks than available threads. You can avoid it by setting the
environment variable OMP_NUM_THREADS=5.
  warnings.warn(
```

[6]:

```
# Plot the Elbow curve
plt.figure(figsize=(8, 5))
plt.plot(range(1, 11), wcss, marker='o', linestyle='--')
plt.title('Elbow Method for Optimal Clusters')
plt.xlabel('Number of Clusters')
plt.ylabel('WCSS')
plt.show()
```




[16]:

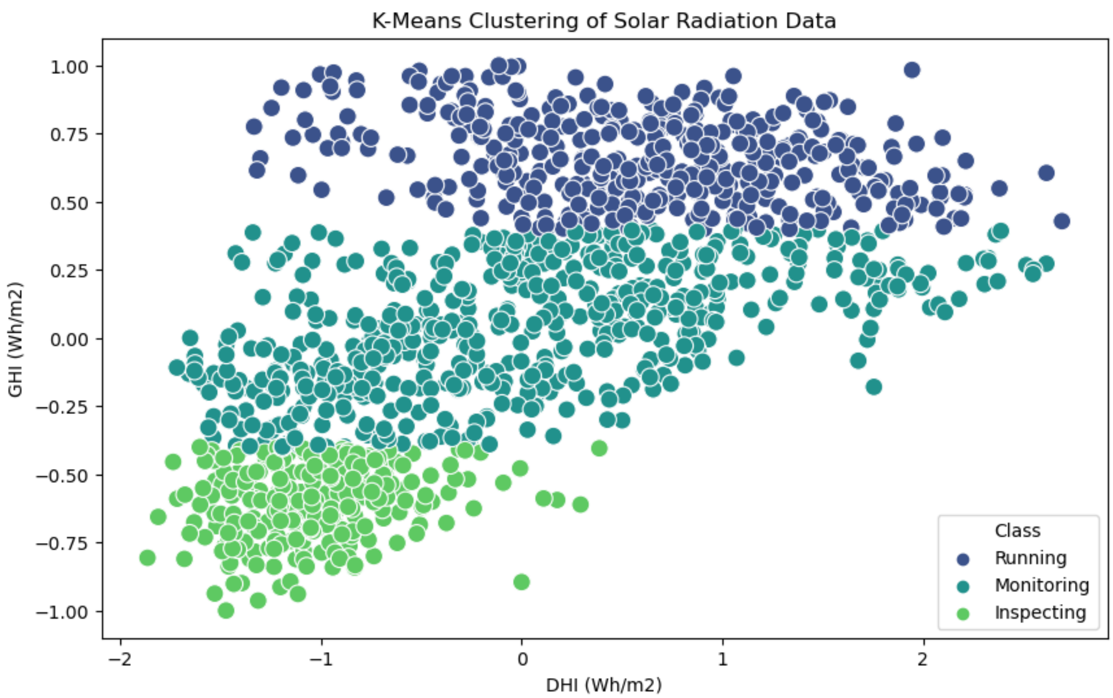
```
# Choose the optimal number of clusters (e.g., from Elbow plot) - say k=3
k =3
kmeans = KMeans(n_clusters=k, init='k-means++', random_state=42)
clusters = kmeans.fit_predict(X_scaled)
```

```
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:870:
```

```
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Admin\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1382:
UserWarning: KMeans is known to have a memory leak on Windows with MKL, when
there are less chunks than available threads. You can avoid it by setting the
environment variable OMP_NUM_THREADS=5.
  warnings.warn(
```

[8]:
```
# Add the cluster labels to the original DataFrame
#df['Cluster'] = clusters
```

[9]:
```
# Visualize the clusters (use scatterplot for two selected features)
plt.figure(figsize=(10, 6))
sns.scatterplot(data=df, x='DHI (Wh/m2)', y='GHI (Wh/m2)', hue='Class',␣
 ↪palette='viridis', s=100)
#sns.scatterplot(data=df, x='DHI (Wh/m2)', y='GHI (Wh/m2)',  palette='viridis',␣
 ↪s=100)
plt.title('K-Means Clustering of Solar Radiation Data')
plt.show()
#df[['DHI (Wh/m2)', 'Relative Humidity (%)', 'GHI (Wh/m2)']]
```




[ ]: