

This is a problem on identifying classification rules for a light sensor that is placed on a outdoor location. Your goal is to compare 2-3 classification algorithms to predict whether the sensor is on or off, depending on a number of features that are described below.

Attached you'll find two datasets. `Mousa_2017-2021.csv` contains four columns: tag name (you can ignore this column), time, sun angle, and value. Value refers to whether the sensor is on or off, and is the target value of this classification task. Sun angle is a feature to help you classify under what circumstances the sensor is on or off.

The second file, `Mousa.csv`, contains several columns that provide additional climate features that can be affecting the sensor's behaviour. Please ignore the latitude and longitude columns, as these refer to the position of the sensor, and thus do not change over time. Please also note that the observations in this file are at hourly intervals, while the time column at the previous file was showing the time when the sensor behaviour is changing from on to off, or vice versa (this can be at multiple points during an hourly period).

Your first task is to merge the two files above, so that you have a complete dataset of 8 features: sun angle, temp 2m, dewpoint 2m, pressure, precipitation, GHI, DHI, and BNI. Please remember that the time columns in the two .csv files do not match, so you'll need to come up with a solution of how to handle the observations from the different times during this merging exercise.

Once the merging is complete, you should move to the second task, which is the classification task. The features in the merged file should be used to train classifiers to predict whether the sensor is on or off. It is up to you to decide which classifiers you will select. Please justify your option in the comments section of your code. Please feel free to perform whatever analysis you like to identify the most appropriate classifier for this task.