Assignment 4 – Data Processing

Ludvig Flodin ludflo-0

April/May 2025

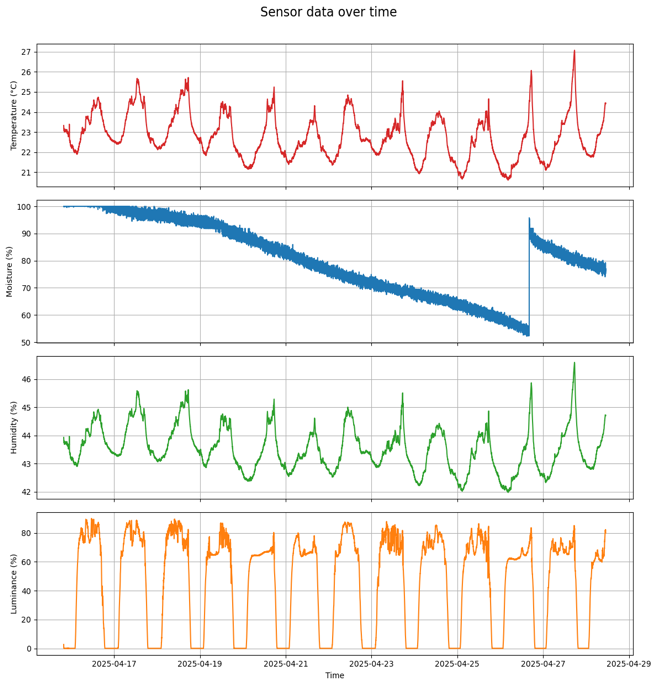
The provided dataset

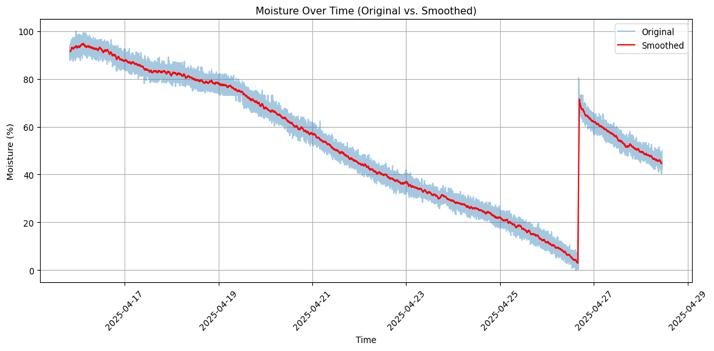
The dataset I got was provided by Anton and Louise. The given metadata was “Metadata Data was captured in Sweden’s time zone, as it is summertime now, it’s CEST (UTC+2). The moisture sensor is about 3cm deep in the pot and the living plant and sensors are located on a windowsill where it can get full sunlight. Maximum and minimum values for calibration of the moisture sensor are 3050 and 1400 respectively and for the light sensor they are 4095 and 0 respectively. Descriptive Data Analysis The data contains 8 features UID, time, temperature, moisture, humidity, and light level.

* UID: A unique ID for this plant, we set ours to 329901.
* Time: Number of seconds since 2000-01-01 00:00 UTC.
* Temperature: The temperature in °C.
* Moisture: Moisture level in % (between 0-100).
* Humidity: Humidity level in % (between 0-100).
* Light: Light level in % (between 0-100).
* Moisture reading raw.
* Light reduction raw.

Working with the dataset

I began by adding a header to each column in the provided order, and it seemed to be correct. Here comes by first thumbs up, good metadata and a good touch by adding a UID, not nescessary but thoughtful. I think that the time format they provided, measuring seconds since Y2K, does work but I added a column for the pleasure of the Gregorian calendar. To do this, I imported a library called datetime, with the function timedelta. I visualized every column to time, and realized that the column for humidity was broken. After speaking with peers, we conclude that the provided code for the sensor was faulty and every group besides one had this problem. When glancing at the graphs, we can see that the relative value moisture has values between 50 and 100% and that gives me the impression that the sensor is either not calibrated, or the contributor chose to water the plant when it was somewhat moist. I think that a detailed desciption on how the sensor was calibrarted should have been stated in the meta data, so that I can draw better conlcusions on the data. The meta data states that the values should be between 0 and 100%, but that is not the case. Thumbs down. I normalized (scaled) the moisture readings to a new column and smooothed out the data with a running average of 70 points.

The light sensor does not peak at 100% which gives me the impression that the light sensor was not calibrated enough. On the other hand, maybe our light sensor was faulty, as it was satrated in direct sunlight. I normalised the light values, scaled the values so that max light value is 100%. I kept the original column if it turns out that the light value is true and not relative.

I think that the dataset is overall very good! As stated before, I wish that the meta data would include a sentence about calibration.

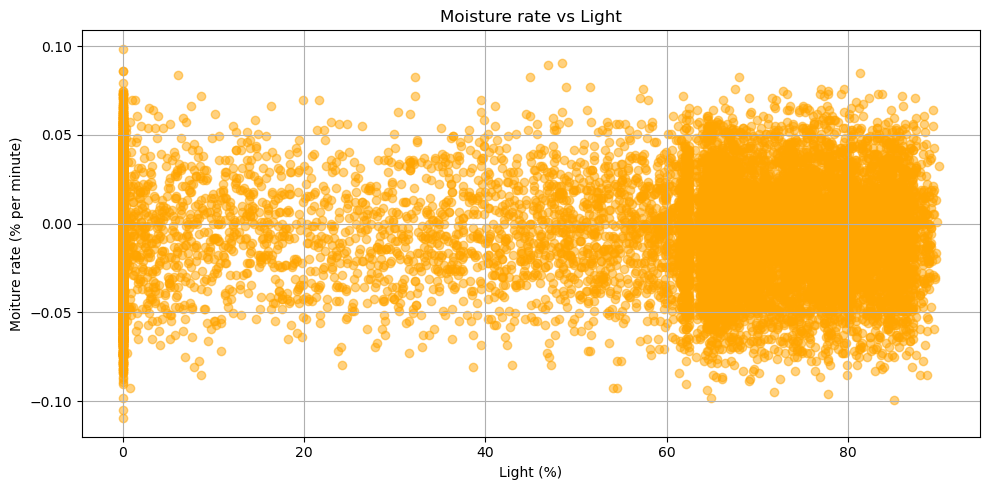
Prognostic analysis

I used past drying behavior to forecast future soil moisture levels. A time-based prediction curve was created, along with a ±1 standard deviation band to indicate uncertainty. The median time to fully deplete the moisure is about 260h (12 days), with the uncertainty std of ±30h. I also plotted the other variables such as light, temperature and time of day to moisture. I cannot see any obvious correlations between any of them which makes me assume that the drying rate is weakly influenced by each variable. Air humidity might be more related to drying, as I assume that drying rate should be related to the delta of air humidity.

A graph of soil moisture forecast

AI-generated content may be incorrect.

A graph showing moisture rate and temperature

AI-generated content may be incorrect.

A graph showing a number of hours

AI-generated content may be incorrect.