Webee Smart Irrigation

Amir Bralin, Zachary Hamilton, Jingjing Tao, Thirawat Bureetes

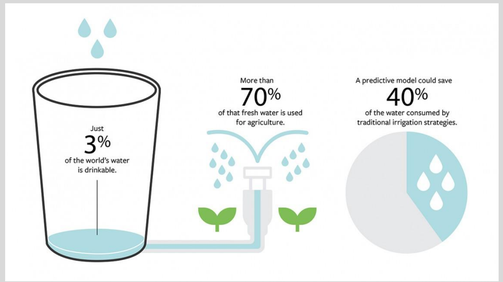
PURDUE UNIVERSITY | The Data Mine

The Data Mine Corporate Partners Symposium 2022

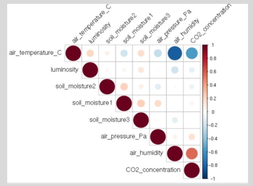
webee

Introduction

Food is a virtual factor to sustain life. Irrigation is a key part that makes farming and food production possible. In average, around 70% of fresh water is used for agriculture. While demand for food keeps increasing, there is a limited water resource available. Smart irrigation is a technology that utilizes sensors to provide insight information for farmers. This technology helps to optimize irrigation strategy. The water will be used at the right time and right amount.



Methodology



Initial Results were not good

1. Low R-squared values in regression reports indicate model is not effective for prediction, even though it is statistically significant
2. Correlation matrix presents weak relationship between variables in dataset

Why the initial soil moisture prediction does not work?

Irrigation is dominated influencer for misleading the correlations

- Regardless of other variable's change (such as air humanity or temperature), when irrigation is on, the soil moisture will increase

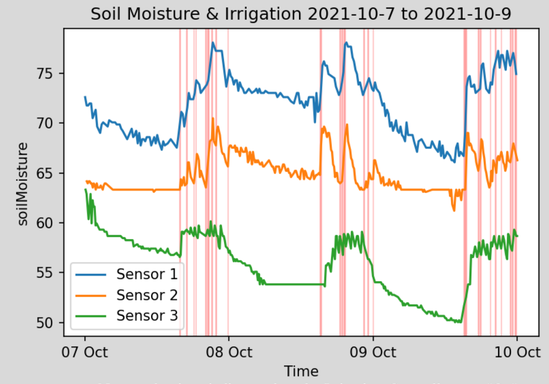
Too much noisy values, such as invalid constant values of soil moisture and void values in our dataset need further cleaning and extracting

How we will switch to soil moisture change prediction?

- Look at data traits and trend day by day to pick out useful dates and times (when the irrigation was "off")

- When irrigation was off, the rate of soil moisture change will depend on environmental condition such temperature at that time.

- Utilize the selected reasonable time range to construct new models



Note: red stripes indicate when the Irrigation Controller was "On".

Data

The data was recorded by a berry farm in CA.
There are 7 main quantities:

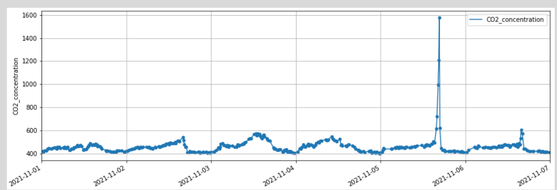
- 1 dependent variable Soil Moisture,
- 5 independent variables (Air Temperature, Pressure, and Humidity, Light Luminosity, and CO^2 Concentration),
- 1 parameter Irrigation Controller.

We have 6 months of data.
It was recorded every 5-15 min.
About 30,000 data points for each variable.
We sample them by 30, 45, and 60 min-long intervals, reducing the size of the dataset.

timestamp	soil_moisture1	soil_moisture2	soil_moisture3	air_humidity	CO2_concentration	air_temperature_F	air_temperature_C	air_pressure_Pa	luminosity
2021-08-01 00:00:00	51.395	72.810000	59.730000	63.595000	412.800000	71.069000	21.705000	101105.800000	870.0
2021-08-01 00:30:00	51.500	72.460000	59.310000	63.753636	406.818182	70.867273	21.592727	101096.545455	130.0
2021-08-01 01:00:00	51.710	71.650000	58.670000	64.725000	412.375000	70.091250	21.162500	101104.250000	120.0

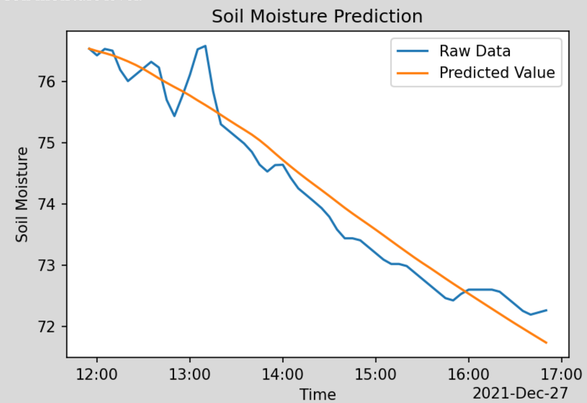
We cleaned all data in Python using Pandas.
Only two variables had outliers needed to be discarded.

1. The allowed range of Soil Moisture is (0, 100). We dropped all points that were > 100.
2. The graph of CO^2 Concentration vs time showed some abrupt spikes that didn't seem natural. So, we dropped all points that were > 600 ppm.



Model Building

- The raw data is noisy. Therefore, the weighted exponential moving average is used to remove noise.
- Selected the time when irrigation was off. So, the changing rate of soil moisture is directed by 5 features.
- The model has R-Square = 0.83 and Root Mean Square Error = 0.8 % per hour.
- The model predicts changing rate of soil moisture which can refer back to soil moisture level.



Conclusion

- Knowing soil moisture level helps preventing overwatering that wastes water and drought that hurts plant growth.
- Predicting soil moisture directly is not feasible due to heavy influence by irrigation which is not a environmental factor.
- By selecting the time when irrigation was off, the rate of moisture change was determined by environmental conditions at that time.
- The model can predict the rate of moisture change from 5 environmental features with R-square = 0.83.

Future Goal

- Integrate the model with sensors and irrigation controller to validate the prediction accuracy in real situation.
- Evaluate the model performance with different soil type, plant species, and location.
- Collect irrigation data and build the model to predict amount of water needed to maintain optimal soil moisture level range.

Thanks to Webee team: Lucas Funes, Davor Margetic, and Yamil Abraham. And Shuennhau Chang from Purdue Data Mine