ArduFarmBot Data Analysis

Garden automation with help of IoT

Marcelo José Rovai

UCSanDiegoX: DSE200x - Python for Data Science

Abstract

The real “crop data” used on this project was generated by an IoT automatic gardening system, the “ArduFarmBot”, which decide automatically when to water and heat a small home tomato plantation, based on captured data from sensors as: temperature, relative air humidity, lightness and soil moisture (humidity).

Using “Scikit” to perform a **Decision Tree based Classification of data** logged on “ThingSpeak” website, was possible to create a model that could predict with very high accuracy, when and how a pump and a lamp would be turned on to help on the crop growth.

With a a very high accuracy, this project confirms that the Pump was planned to be Turned On every time that soil humidity drops bellow 65% and the air temperature in low (at morning for example) and that the Lamp was Turned On every time that the air temperature was low (around 12oC) keeping this way, the crop warm.

Motivation

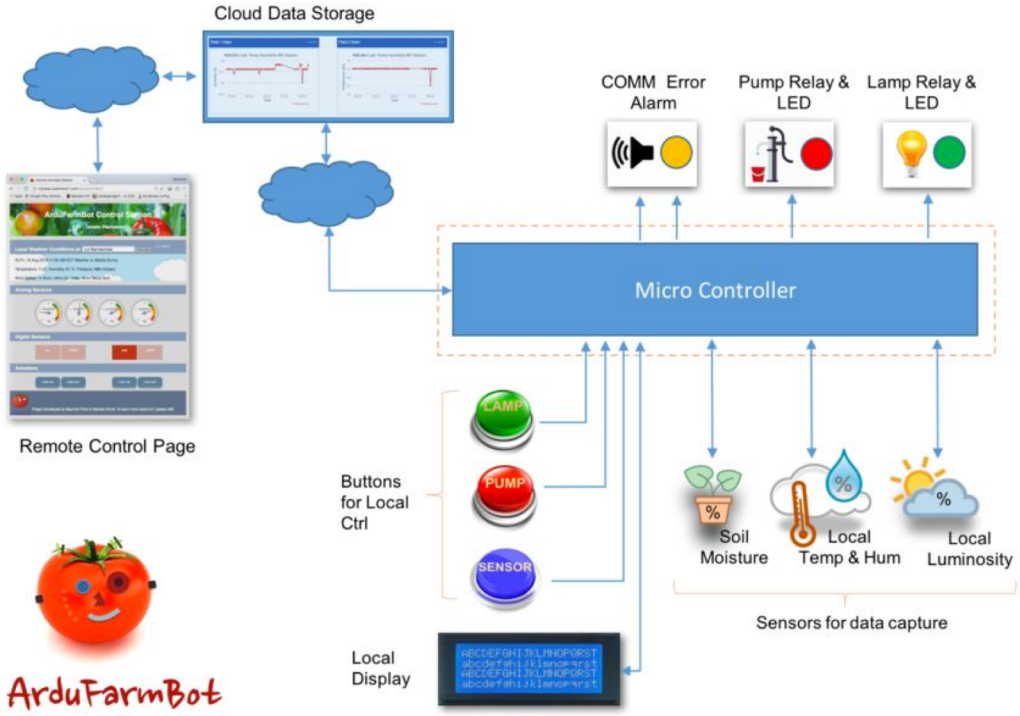
A couple of years ago, I developed the ArduFarmBot, an automatic gardening IoT based system, that automatically controls a pump and a lamp to control water and heat to be delivered to a plantation. The system action was based on real data, which were also logged on ThingSpeak.com (both, sensors and actuators' echo).

So, a question come to my mind: "Can the ArduFarmBot historical data stored on ThingSpeak website be used to predict when the Water Pump and Lamp should be turned on in order to keep a small plantation warm and wet? Using Classification technics, seems to me a good path to follow here. This analysis and prediction is my first attempt to integrate IoT with Data Science.

Dataset

- ✓ ArduFarmBot is a gardening system, that capture data as temperature, relative air humidity, lightness and soil moisture (humidity).
- ✓ ArduFarmBot has 2 actuators, a pump and a lamp
- ✓ ArduFarmBot logs all captured data from sensors and actuators (“echo” of commands) on *ThingSpeak.com*.
- ✓ The Dataset used on this work is the historical data found on ThingSpeak website for an ArduFarmBot operation, that happened from September to December, 2016 (https://thingspeak.com/channels/146159)
- ✓ See, block diagram on next page for details.

ArduFarmBot – General Block Diagram



Data Preparation and Cleaning

The Dataset used in this work is the historical data retrieved from ThingSpeak website from September to December 2016 (*):

On the dataset, there are 47,164 samples divided into 10 columns:

- ✓ “created_at",
- ✓ "entry_id",
- ✓ "Temperature",
- ✓ "Humidity",
- ✓ "Luminosity",
- ✓ "Soil Moisture",
- ✓ "Pump Echo",
- ✓ "Lamp Echo",
- ✓ "Capacitive Soil Moisture" and
- ✓ "Spare".

(*) https://thingspeak.com/channels/146159.

Data Preparation and Cleaning

The data has a lot of samples with NaN values:

	created_at	entry_id	Temperature	Humidity	Luminosity	Soil Moisture	Pump Echo	Lamp Echo	Capacitive Soil Mositure	Spare
47159	2016-12-14 14:26:34 UTC	47160	27.0	17.0	91.0	88.0	0.0	0.0	97.0	0.0
47160	2016-12-14 14:27:18 UTC	47161	NaN	NaN	NaN	NaN	0.0	0.0	NaN	NaN
47161	2016-12-14 14:31:26 UTC	47162	NaN	NaN	NaN	NaN	0.0	0.0	NaN	NaN
47162	2016-12-14 14:35:34 UTC	47163	NaN	NaN	NaN	NaN	0.0	0.0	NaN	NaN
47163	2016-12-14 14:36:41 UTC	47164	28.0	17.0	91.0	89.0	0.0	0.0	98.0	0.0

Any sample with more than 3 NaN will be eliminated

Samples were sensors have "zero" values also must be eliminated.

Data Preparation and Cleaning

Only important instances as: Sensors: Temperature, Humidity, Luminosity, Soil Moisture, and Actuators: Pump and Lamp, are kept and its names changed as bellow:

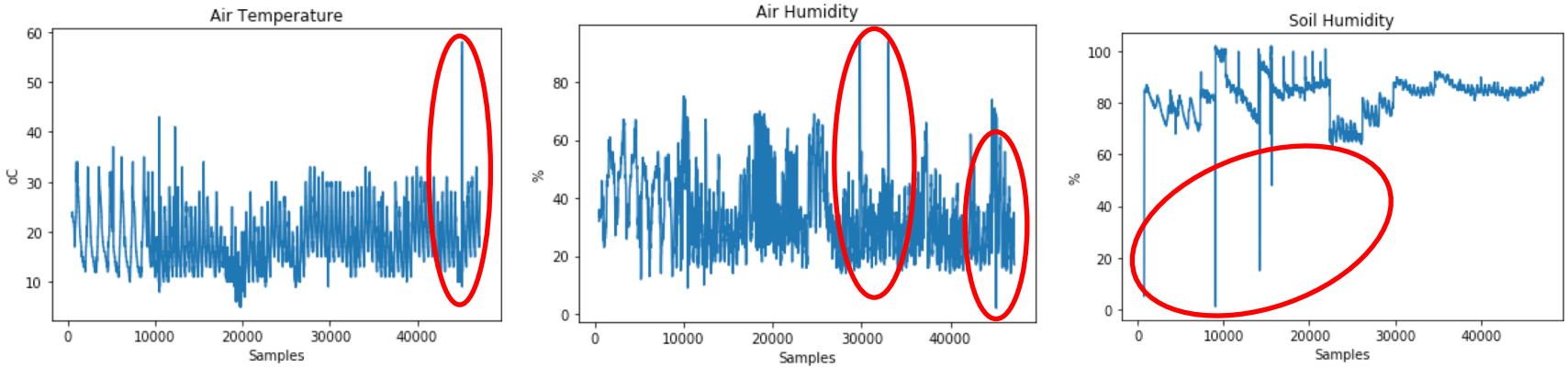
	temp	humi	lumi	soil	pump	lamp
504	23.0	32.0	73.0	8.0	0.0	0.0
505	23.0	32.0	73.0	8.0	0.0	0.0
506	23.0	32.0	73.0	8.0	0.0	0.0
507	23.0	32.0	73.0	8.0	0.0	0.0
508	23.0	32.0	73.0	8.0	0.0	0.0

Sensors Actuators

Data Preparation and Cleaning

Noise on data should be eliminated (all sample). The right range should be:

- ✓ Temperature lower than 45oC
- ✓ Air Humidity between 10% and 80%
- ✓ Soil Moisture with humidity greater than 60%



After total cleaning, around 17,700 samples are available for analyses.

Research Question

- ✓ *Can you use ArduFarmBot historical data saved on ThingSpeak.com to, based on sensors readings and using Classification, preview when the Water Pump and Lamp should be turned on in order to keep a small plantation warm and wet?*

Methods

Pandas is used to retrieve and manipulate data from the dataset. After cleaning, were *MatPlotLib* is also used for data visualization, Machine Learning estimates when Pump and lamp should be turned on.

On this analysis, for ML, we use *scikit-learn* to perform a *decision tree based classification* of data. The sensor data (‘temp’, ‘humi’, ‘lumi’ and ‘soil’) is used to train the model. Two separated analysis are performed:

1. Prediction of Pump operation: Target Variable (y) ➔ ‘pump’
2. Prediction of Lamp operation: Target Variable (y) ➔ ‘lamp’

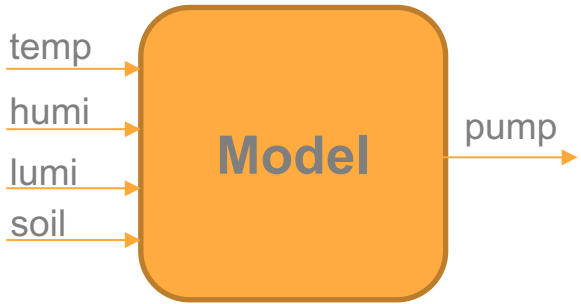
Being for both cases, the Features (X) ➔ 'temp',’ humi', 'lumi’ and 'soil’

Methods

In both cases, the dataset is split in 2 parts:

- ✓ Training data 67%
- ✓ Testing data 33%

1. Prediction of Pump operation

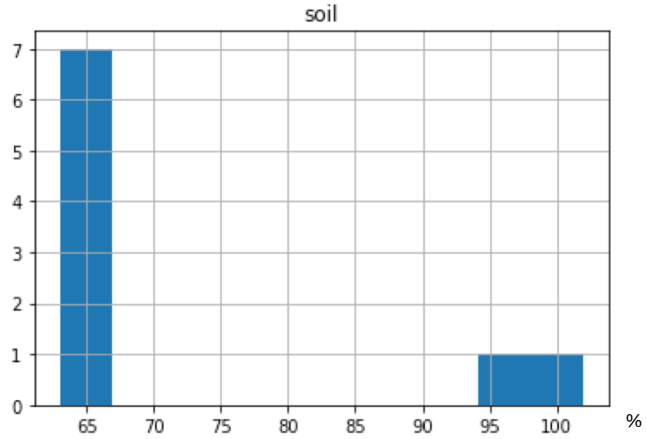


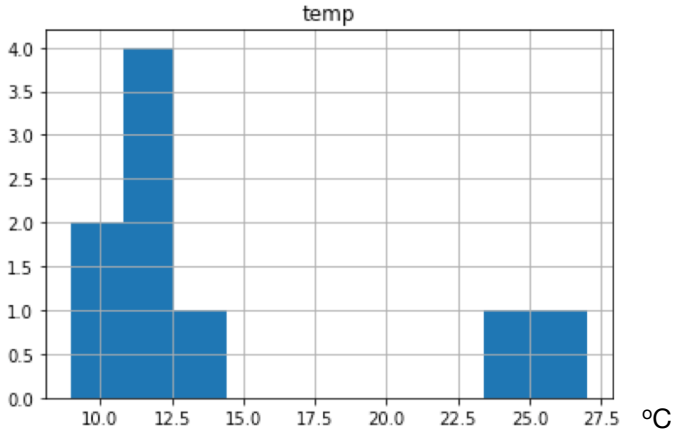
2. Prediction of Lamp operation

Findings:

Prediction on Pump operation:

- ✓ Applying the model on test data we got a 99% of accuracy
- ✓ Looking only on samples were target variable was “1” (Pump Turned ON), we get:



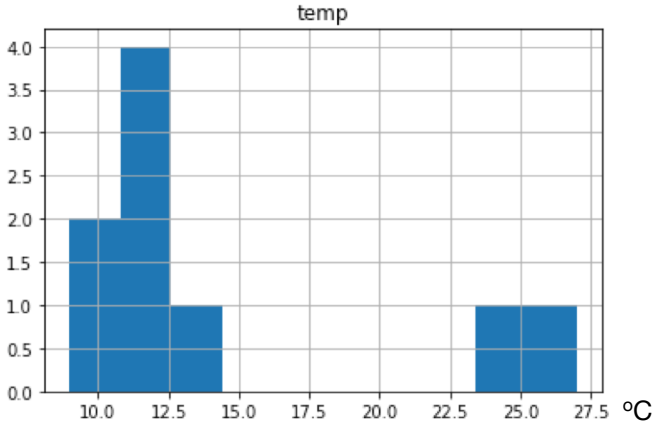


Seems that the Pump is Turned ON automatically every time that soil humidity reachs its lowers at 65% and Temperature is low (10ºC to 15ºC). Some actuations also appears around 100% what should be manual commands. Other sensors conditions were irrelevant.

Findings:

Prediction on Lamp operation:

- ✓ Applying the model on test data we got a 93% of accuracy
- ✓ Looking only on samples were target variable was “1” (Lamp Turned ON), we get:



The lamp is turned ON automatically at lower temperatures (around 12ºC). Some actuations also appear around 25ºC what can be attributed to manual commands. Other sensors conditions were irrelevant.

Conclusions

Using scikit-learn to perform a *decision tree based classification* of data logged on ThingSpeak website, was possible to create a model that could predict with very high accuracy, when and how a pump and a lamp would be turned to help in the crop growth. In short:

The Pump should be Turned On every time that soil humidity drops bellow 65% and the air temperature in low, at morning for example. This make sense and it is really advise not water a plantation when the day is hot.

The Lamp was Turned On every time that the air temperature was low (around 12oC). This is key to help, mainly on the early stages of seed germination phase. Temperatures lower than 10oC for tomatoes (from were the data were got), delay gemination

The data related with Air Humidity and luminosity were not used for actuators control.

Acknowledgements

This dataset is real and got from the *ArduFarmBot running on my experimental home tomato plantation. You can learn more about it on my tutorials at Hackster.io:*

https://www.hackster.io/mjrobot/ardufarmbot-controlling-a-tomato-home-farm-using-arduino-e22c4b

https://www.hackster.io/mjrobot/ardufarmbot-part-2-remote-station-an-iot-implementation-6ccc29

References

- ✓ **“UCSanDiegoX: DSE200x - Python for Data Science”**

  My sincere thanks to professors :

  - ✓ Ilkay Altintas, Chief Data Science Officer, San Diego Supercomputer Center (SDSC)
  - ✓ Leo Porter, Assistant Teaching Professor, Computer Science and Engineering Department

- ✓ **ArduFarmBot: Tomato garden automation with help of "Internet of Things" - IoT (MJRoBot Tutorials Book 1) by Marcelo Rovai and Maurício Pinto**

  English | 6 Apr. 2017 | ASIN: B06Y4CTW23 | 125 Pages | PDF | 29.88 MB

- ✓ **Project Jupyter Notebook**:

  https://github.com/Mjrovai/Python4DS/blob/master/ArduFarmBot_Data_Analysis/MJR_ArduFarmBot.ipynb

Thank you

Marcelo José Rovai

Santiago, Chile **April 24th 2018**