Hackathon

HIVEMQTT FOR IOT MONITORING: DETECT DDOS ATTACK

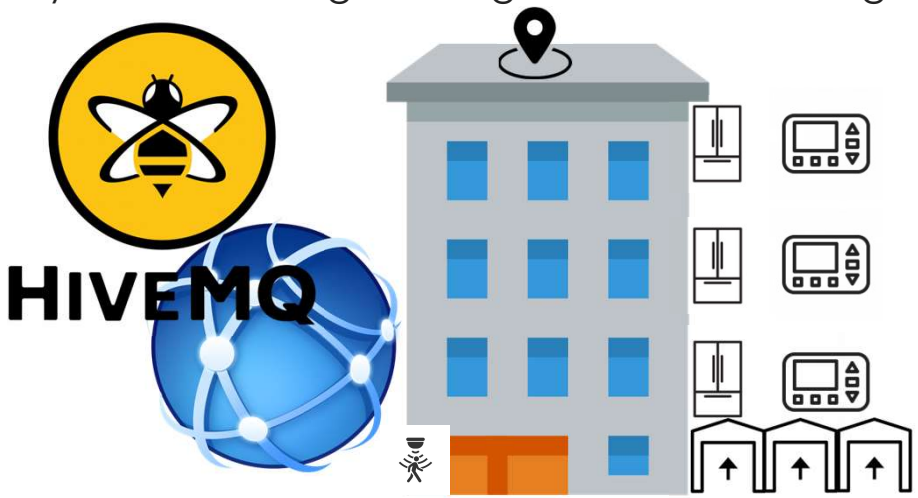
BENAZIR DE LA ROSA

Agenda

- Business Case Scenario: Monitoring IoT information to detect DDOS attack
- Data Sources Description
- Data Architecture Structure Description
- Data Exploratory Analysis
- Artificial Intelligence Implementation: Detect DDos Attack
- Next Steps

Use Case Scenario: Detect DDoS Attack

- This case uses HiveMQ along side with IoT emulations to monitor historical information from 3 different locations. It is common to find hacking activity within IoT systems. The advance analytics application in this demo is going to **detect a DDos Attack**.
- A distributed denial-of-service (DDoS) attack is a malicious attempt to disrupt the normal traffic of a targeted server, service or network by overwhelming the target or its surrounding infrastructure with a flood of Internet traffic.
- The locations consist of:
  1. Garage Door
  2. Fridge
  3. GPS Tracker
  4. Weather
  5. Motion Light



Data Sources

DATA AVAILABILITY COURTESY FROM UNIVERSITY OF NEW SOUTH WALES, SPECIAL THANKS TO DR. NOUR MOUSTAFA

Data Sources

- I got the data from this website IoT_original_normal - Files - CloudStor (aarnet.edu.au)

- Devices to be emulated for the proof of concept:
  - Fridge x3
  - Normal Garage Door x3
  - IoT GPS Tracker x3
  - IoT normal Motion Light x3
  - IoT normal Weather x3






Datasets Description

- **Fridge Datasets**: 3 log files with information in JSON format per Fridge and its corresponding timestamp. It contains ID, TIMESTAMP, FRIDGE_TEMPERATURE and CONDITION. For the proof of concept they were break on per event basis.

- **Garage Door Datasets**: 3 log files with information in JSON format per Garage Door and its corresponding timestamp. It contains ID, DEVICE_TITLE, DOOR_STATE, DOOR_STATE_TEXT and SPHONE SIGNAL.

- **IoT GPS Tracker Datasets**: 3 log files with information in JSON format per GPS Locator and its corresponding timestamp. It contains ID, TIMESTAMP, LAT and LON.






Datasets Description

- **IoT Normal Motion Light Datasets**: 3 log files with information in array format per Motion Light and its corresponding timestamp. It contains DATE, TIME, MOTION_STATUS and LIGHT STATUS.

- **IoT_normal_weather Datasets**: 3 log files with information in JSON format per Wheather station and its corresponding timestamp. It contains ID, TIMESTAMP, PRESSURE, TEMPERATURE and HUMIDITY.



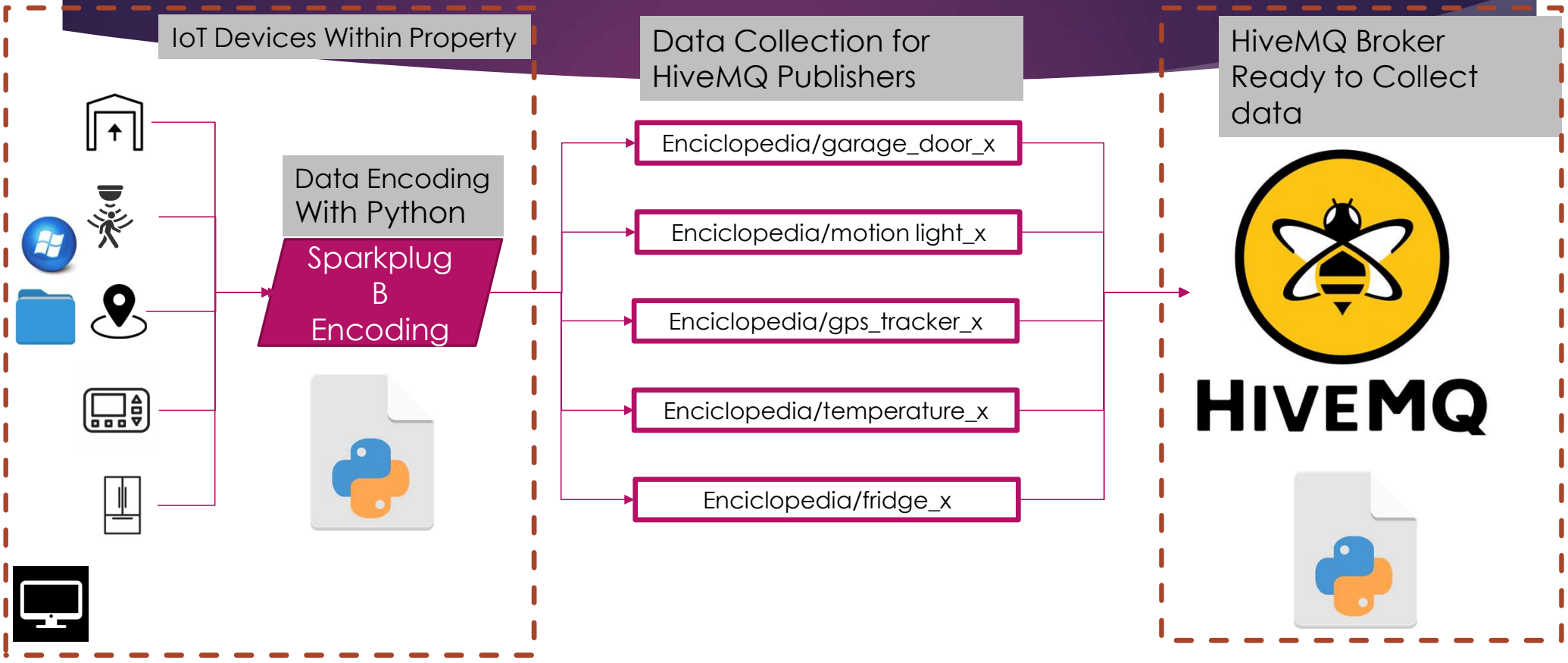



* For demo modbus device was left out. There is a lack in the data meaning.

Data Architecture Structure Description

POC

Local Emulator with HiveMQ: Data Reading



Local Emulator with HiveMQ: Data Transformation

HiveMQ Broker Transmits data with python

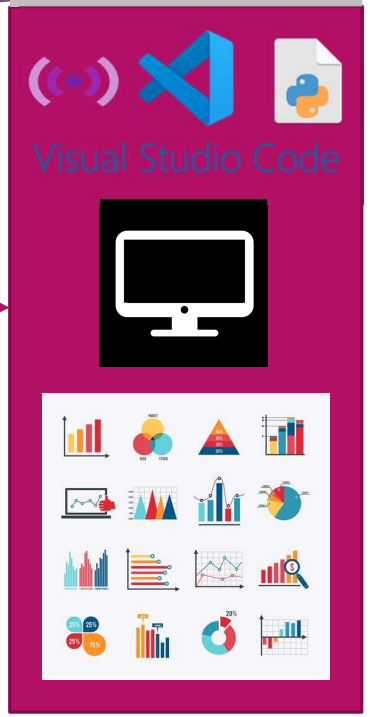
Data Reading from HiveMQ Subscribers

- Enciclopedia/garage_door_x
- Enciclopedia/motion light_x
- Enciclopedia/gps_tracker_x
- Enciclopedia/temperature_x
- Enciclopedia/fridge_x

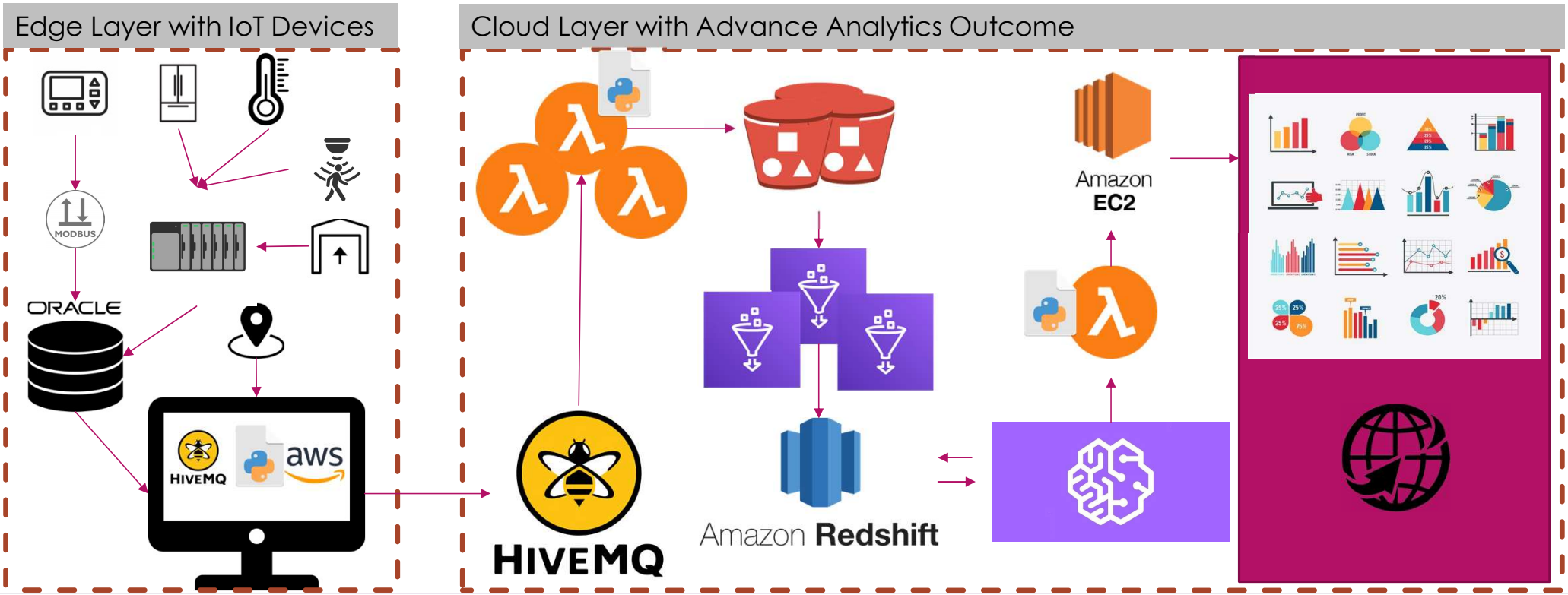
Data Decoding With Python

Sparkplug B Decoding

Data Visualization With Python



Production Data Architecture Infrastructure

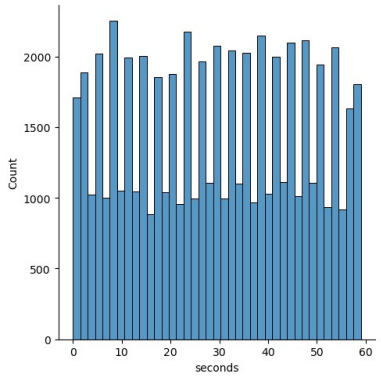


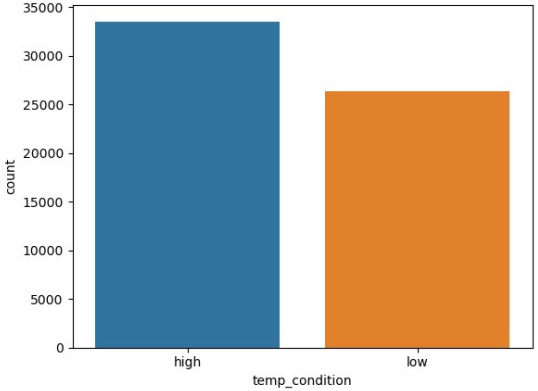
Data Exploratory Analysis

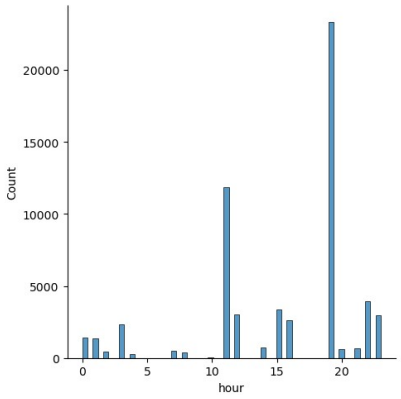
POC

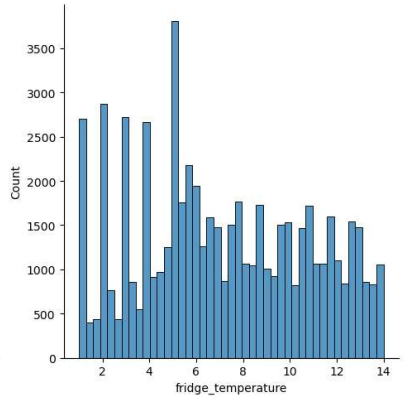
General Information

- Datasets dates are from 31st March 2019 to 25th April 2019.
- Datasets for ML process live in data folder under ML_app_data.
- Code is in tcs_hackaton folder Jupyter Hackaton ML process.
- There are 5000 registers with DDos label and 35000 with normal label for all devices. Proportion of anomalous class is ~15%.
- Fridge findings are the next ones. It looks suspicious what is happening in the 19 hour:



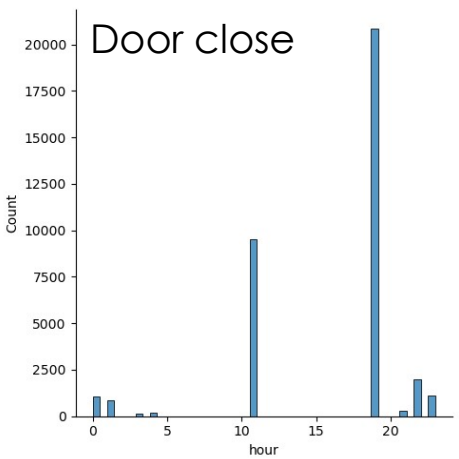


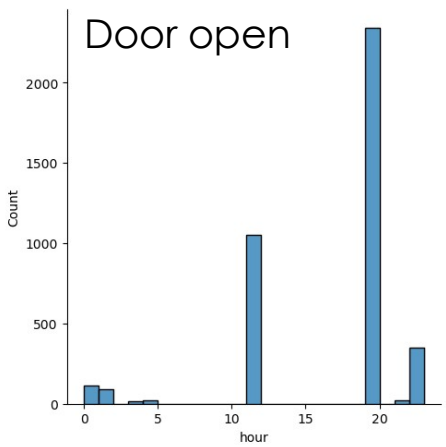


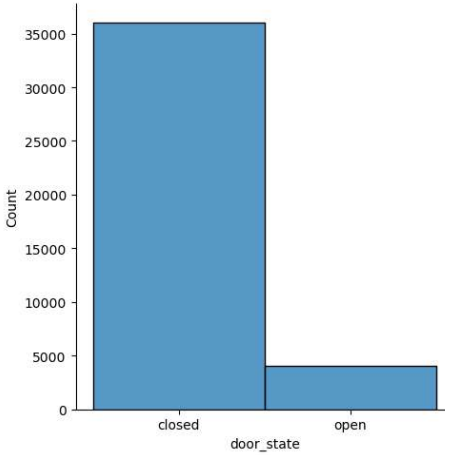


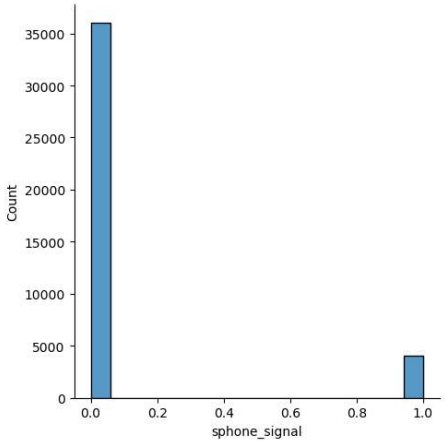
General Information

- **Garage Door** findings are the next ones. Most activity takes place around 19 hour :







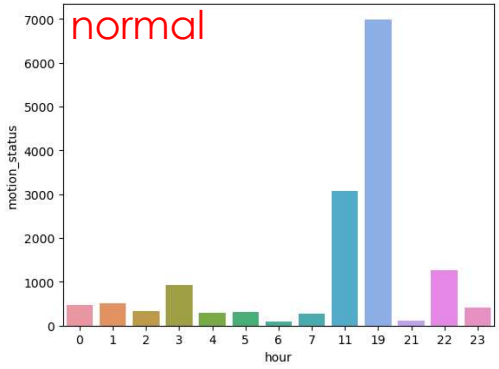


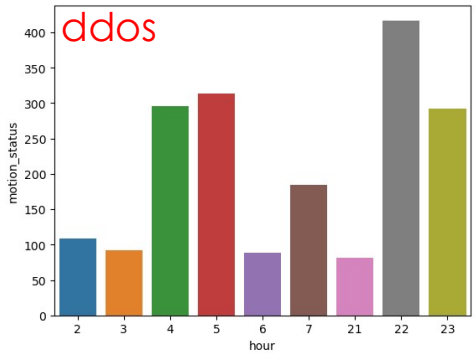
General Information

- The **gps tracker** shows a clear high mean value in latitude and longitude on hacking condition.

	type	latitude	longitude
0	ddos	136.296010	148.977345
1	normal	24.841856	34.720488

- The **motion lights** information displays an unusual behavior during the night. The graph shows active status during all the hours.





General Information

- **Modbus and Thermostate** data could not be use due to the lack of ddos.

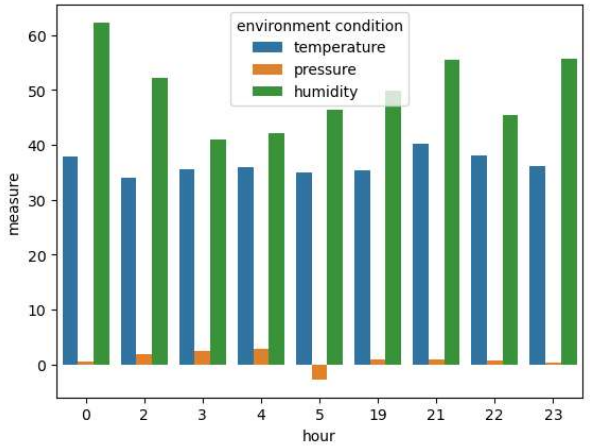
```
#leave out signal because there is no data when it is in ddos attack
modbus_df['type'].unique()
```
✓ 0.0s
```
array(['injection', 'backdoor', 'normal', 'password', 'scanning', 'xss'],
      dtype=object)
```

```
#we have to leave out thermostate because it does not have data under ddos attack.
thermostate_df['type'].unique()
```
✓ 0.0s
```
array(['injection', 'backdoor', 'normal', 'password', 'ransomware',
       'scanning', 'xss'], dtype=object)
```

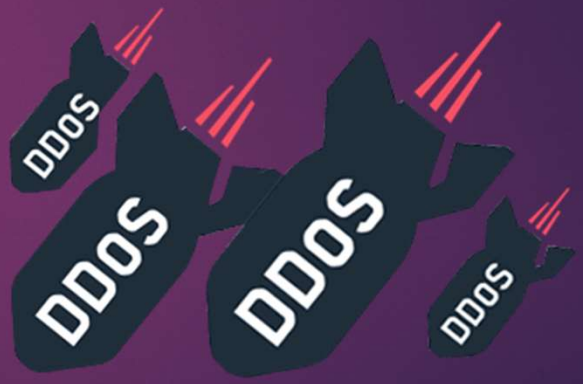
- **Weather** data demonstrate a clear hacking activity during the night and even strange behavior on pressure signal during **hacking activity hour 5.**

Conditions on Hacking Events

	hour	temperature
0	0	41.675220
1	5	35.009764
2	21	41.480547
3	22	40.782752
4	23	39.064633



Artificial Intelligence Implementation: Detect DDos Attack



POC

Feature Engineering

- **time_components**: extract month, day, hour, minute and second per register in electronic devices.
- **ts_delta_fridge**: time lapsus between previous and current register in time. The goal is to detect delay between messages send.
- **temp_delta_fridge**: delta change in temperature between current and previous send measure.
- **high**: fridge on
- **off**: fridge off
- **ts_delta_door**: time lapsus between current and last send message from delta door emulated device.

Feature Engineering

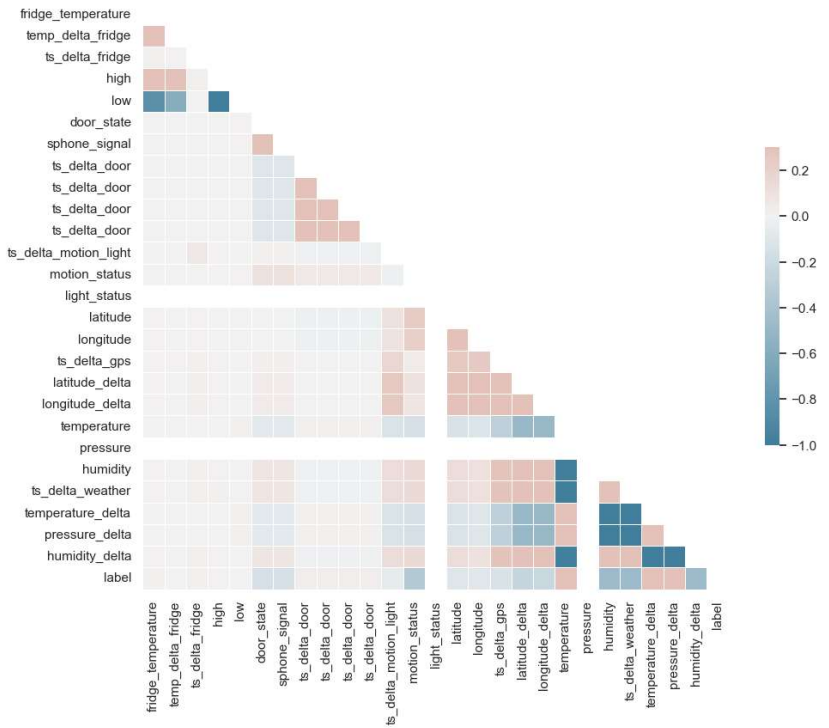
- **ts_delta_gps**: time lapsus between current and last message sent by gps trackers.
- **latitude_delta**: difference in latitude between current position and previous position from gps trackers
- **longitude_delta**: difference in longitude between current position and previous position from gps trackers.
- **Ts_delta_motion_light**: difference in time lapsus between current and last message sent by motion light actuator.
- **ts_delta_weather**: difference in time lapsus between current timestamp and previous timestamp from weather.

Feature Engineering

- **Temperature_delta**: difference in current and previous temperature measure from weather device.
- **Pressure_delta**: difference in current and previous pressure measure from weather device.
- **Humidity_delta**: difference in current and previous humidity measure from weather device.

Correlation Matrix

- As we can see in the correlation matrix whatever happens with the fridge or the garage door. Does not influence directly the result of the label. (The correlation is near to 0.0)
- However, some information influence when combined together with the motion_light. Thus, we erased **fridge_temperature**, **door_state**, **temp_delta_fridge** and **ts_delta_door** from feature engineering.



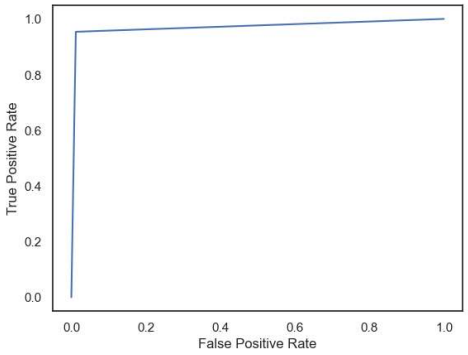
ML Algorithm Training

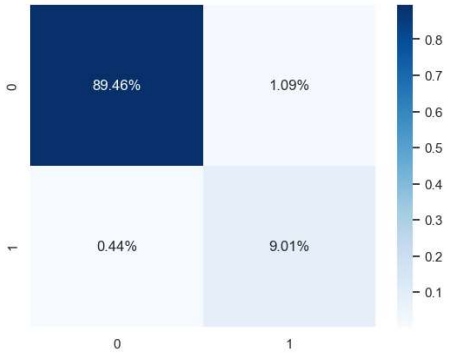
- An Xgboost algorithm was trained to predict the ddos attack. A k-fold (10) cross validation was used to check overall performance.
- The algorithm detected accurately ddos attaks within the network with a 93% accuracy and a very low percentage of false predictions ~1%.

```
k_folds = KFold(n_splits = 10)

scores = cross_val_score(bst, X, y, cv = k_folds, scoring='accuracy')

print("Cross Validation Scores: ", scores)
print("Average CV Score: ", scores.mean())
print("Number of CV Scores used in Average: ", len(scores))
```

✓ 5.9s

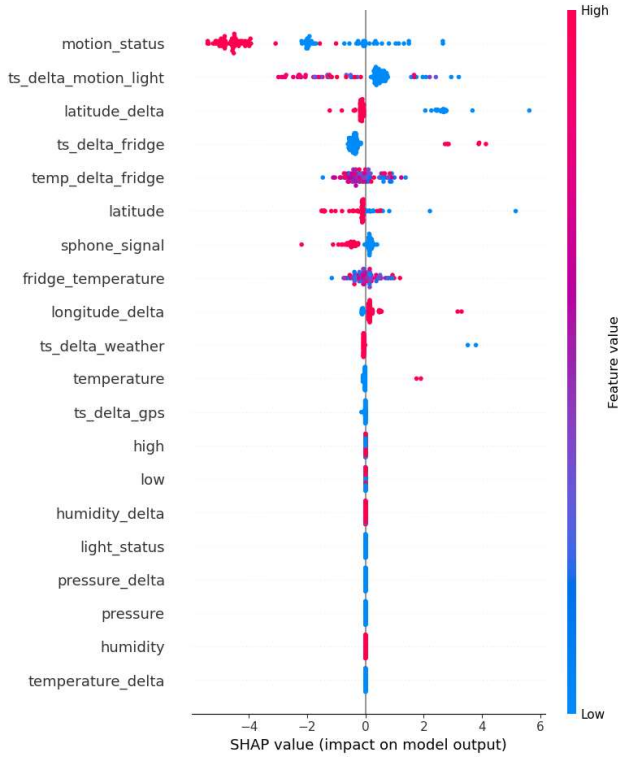
```
Cross Validation Scores:  [0.97546523 0.61794365 0.90300447 0.82087535 0.99729702 1.
 1.         1.         1.         1.        ]
Average CV Score:  0.9314585715770871
Number of CV Scores used in Average:  10
```





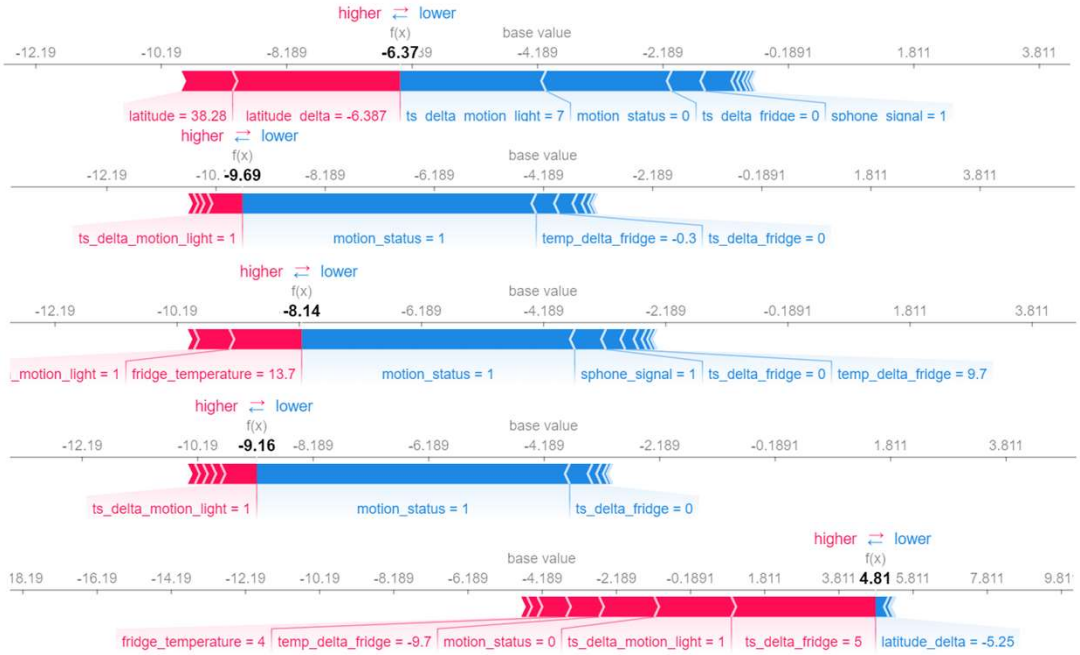
ML Algorithm Training

- In the image can be appreciate which are the elements that influence the most in the algorithm detection been **motion_status**, the change in the motion status signal and the **latitude from the gps** tracker the features that influence the most to detect intrusion.



ML Algorithm Training

- In here we can appreciate 5 observations detected as not dangerous (mostly blue, first 4) and one dangerous (mostly red, last). The time lapsus between signals from the fridge (**ts_delta_fridge**) seems to be a strong indicator to detect it as dangerous.



Conclusions

- We hope and the audience can understand better the exploitation of IoT advance analytics applications.

- We appreciate the support from the USWN for the datasets.

Thank You!

BENAZIR DE LA ROSA