



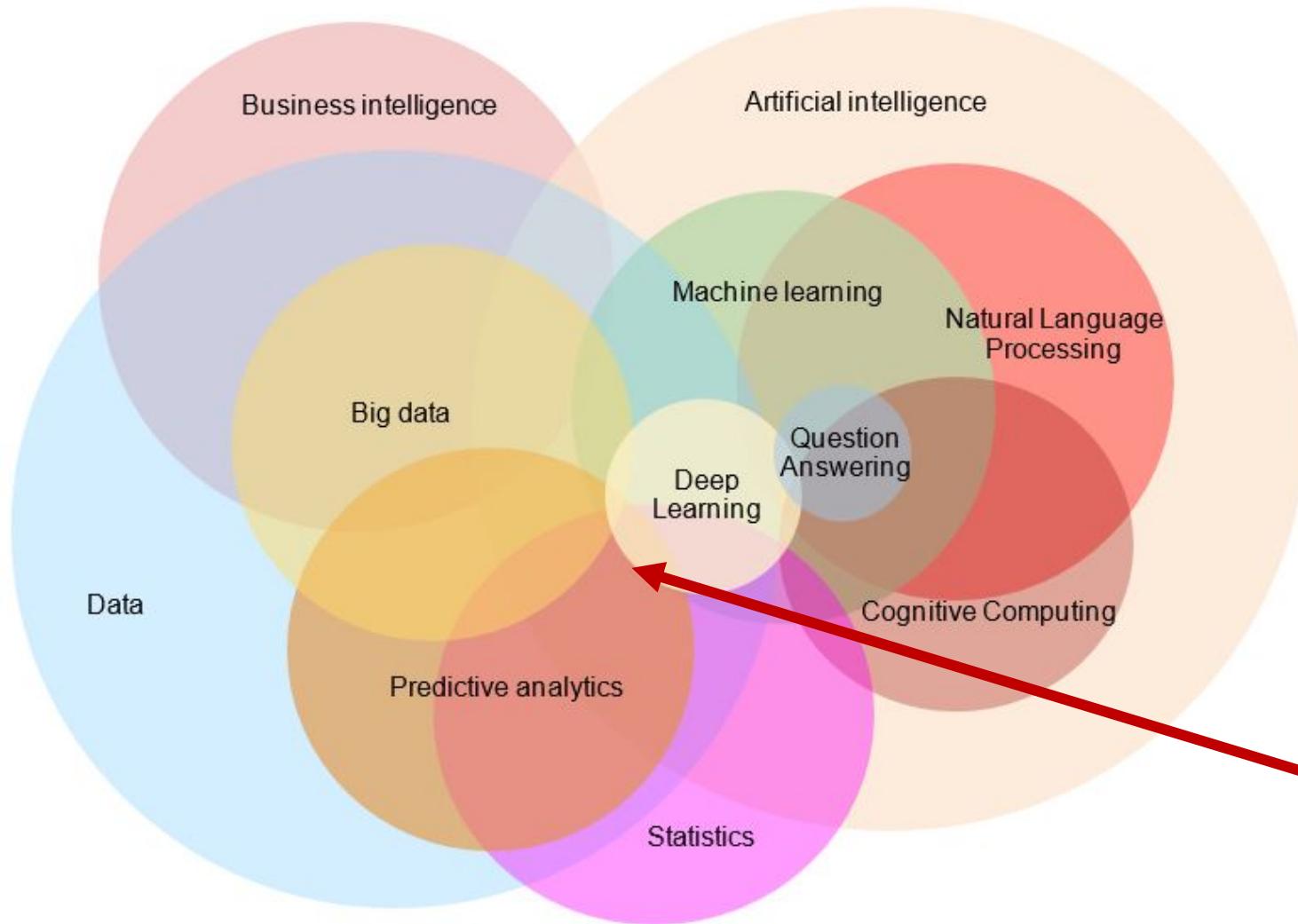
# vSight Machine Learning Algorithm to Enhance Virtual Infrastructure Efficiency

Amir Oghbaee, Ali Syed  
Marat Khandros





# The what is “ML” slide



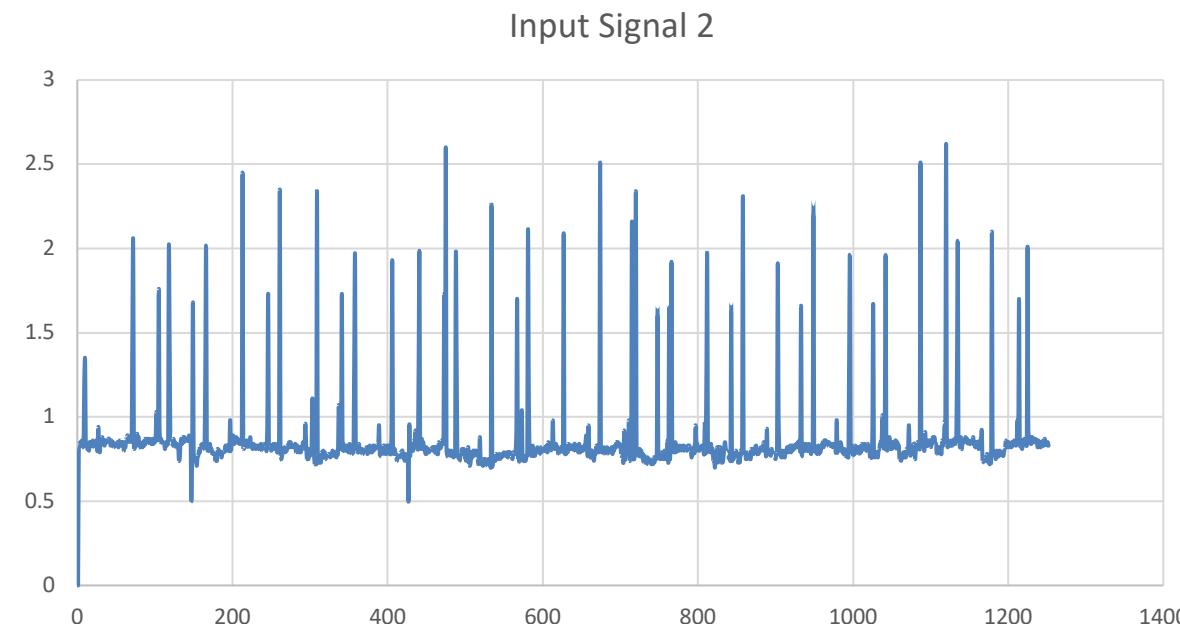
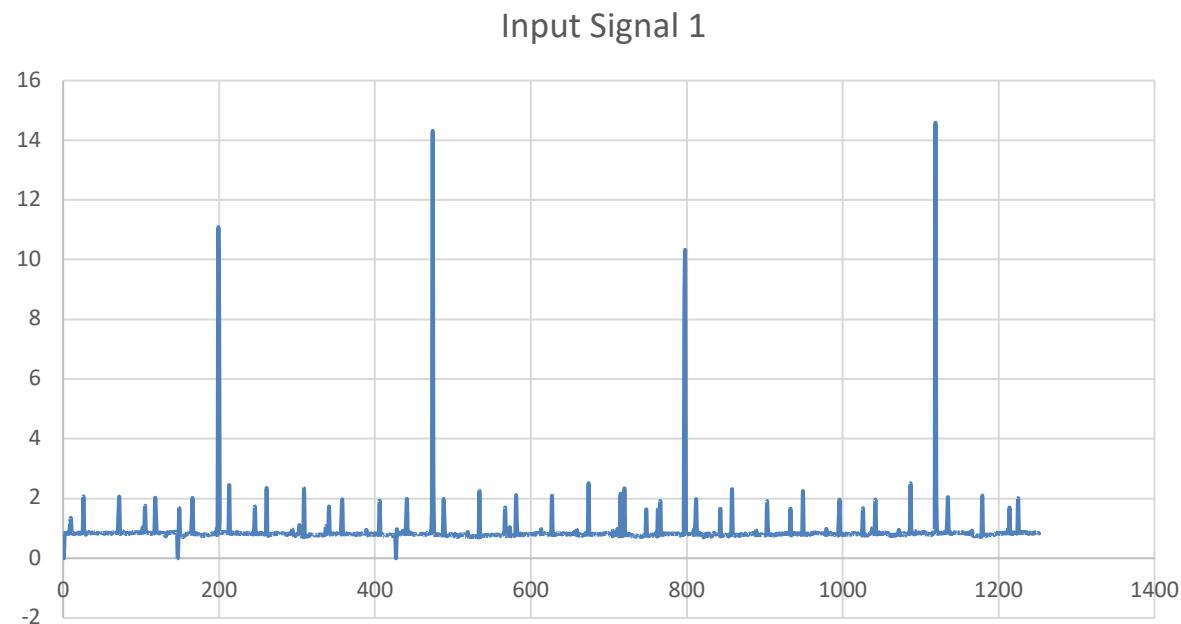
vSight ML algorithm  
lies on the verge of  
Predictive Analysis,  
and Statistics

# Here is a ‘theoretical’ problem:



Classify ~38,000 time-series datasets (e.g. input signals 1 or 2), into two different categories:

- Signal-1 look-alikes
- Signal-2 look-alikes





# Here is a ‘theoretical’ solution:

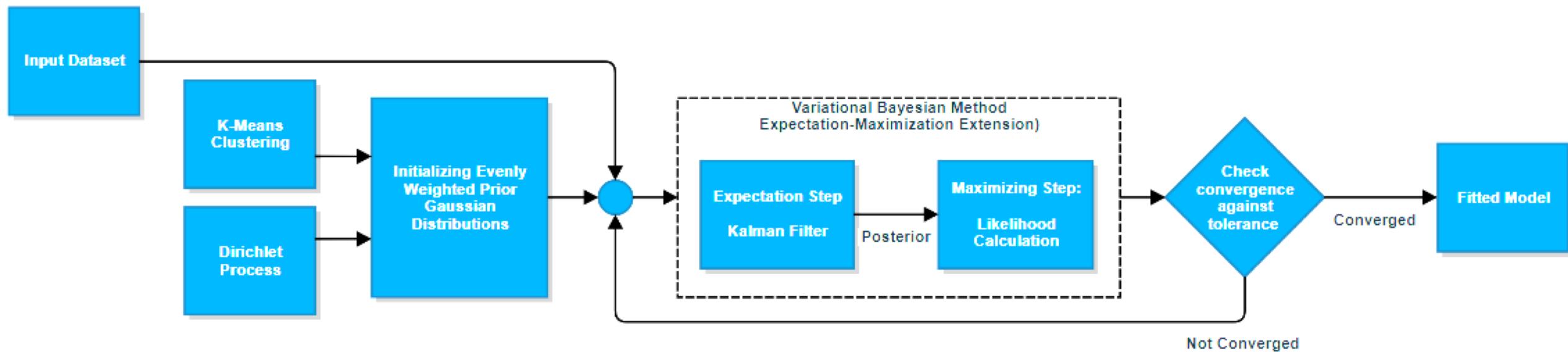
- Implements the expectation-maximization (EM) algorithm for fitting a mixture-of-Gaussian models.
  - EM algorithm: `sklearn.mixture` library
- Evaluate the model prediction to find the errors by computing the logarithm of posterior probabilities of all datapoints
  - Posterior log probability: `numpy` library
- Normalize all the log probabilities to get a single score for each dataset, i.e. anomaly score.
- Set a threshold to categorize the datasets into two sets based on anomaly scores!



# GMM classification - Fit



- Find the model for each dataset



# Our ‘real’ problem

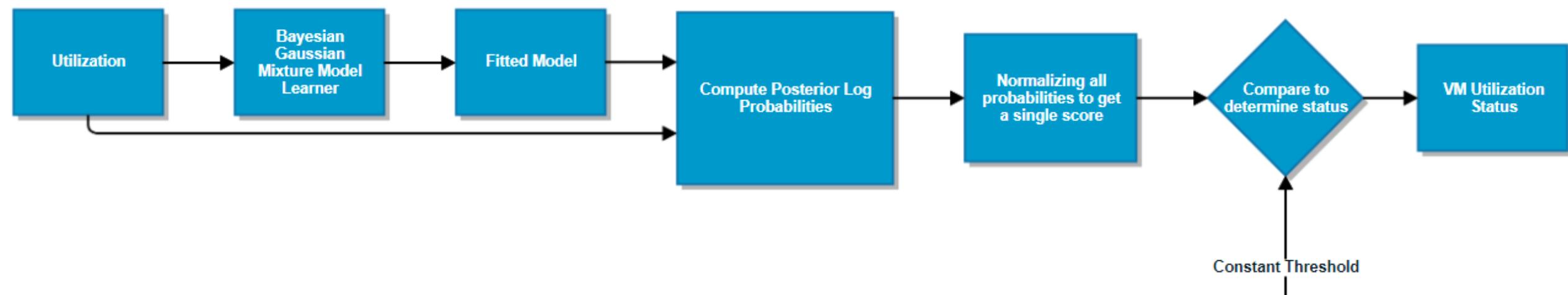


- RBC hosts ~38K Virtual Machines (VMs)
- Some VMs are not utilized to their full capacities
- Rightsizing VMs will enhance the infrastructure and save money
- Two time-series datasets are available per VM
  - cpu, and network utilization
- Goal is to categorizes VMs into two categories; normal and Idle

# Our Solution: GMM classification - Predict

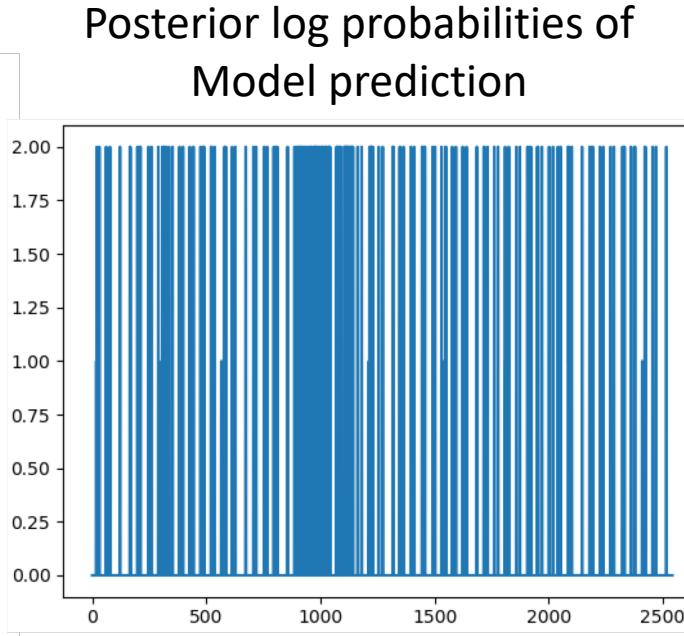
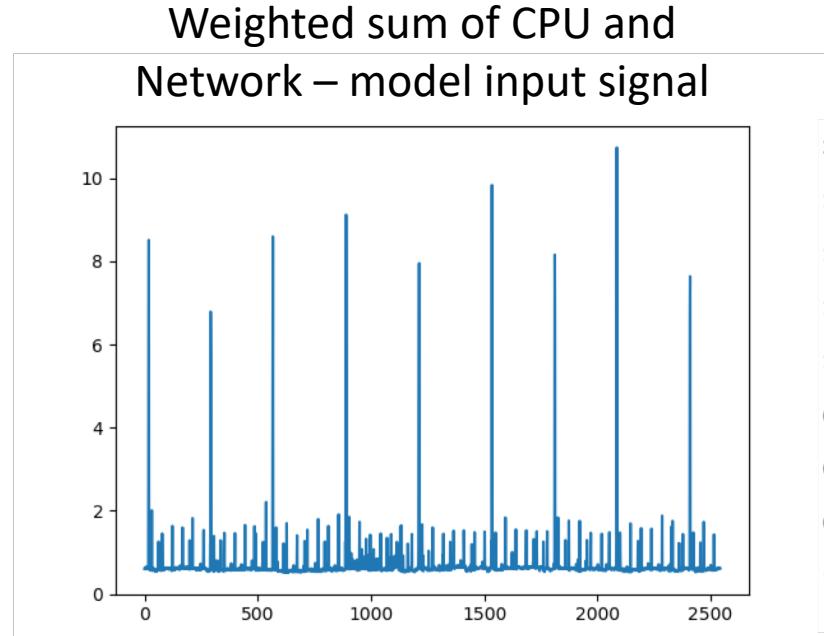
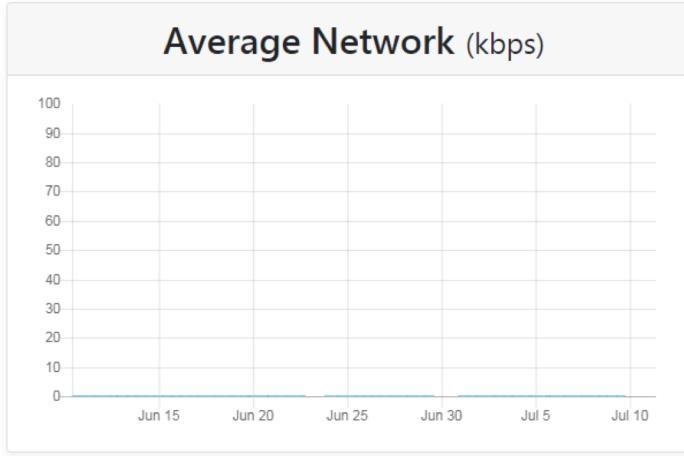
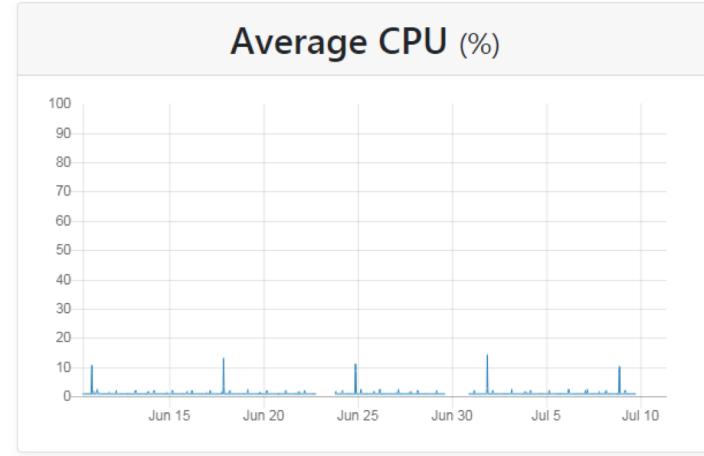


- Using our 'theoretical' solution in action :



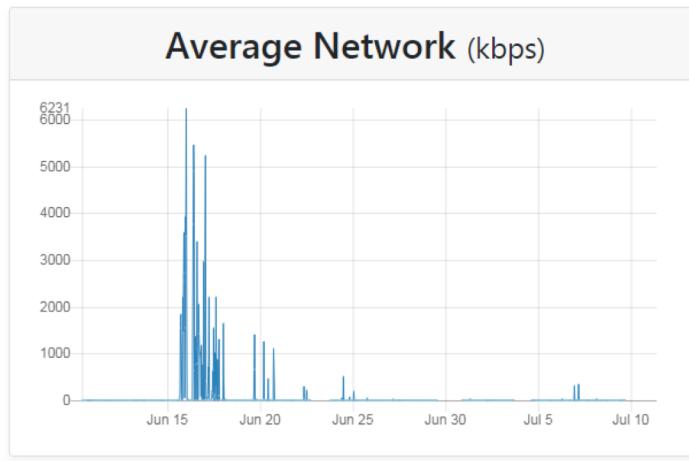
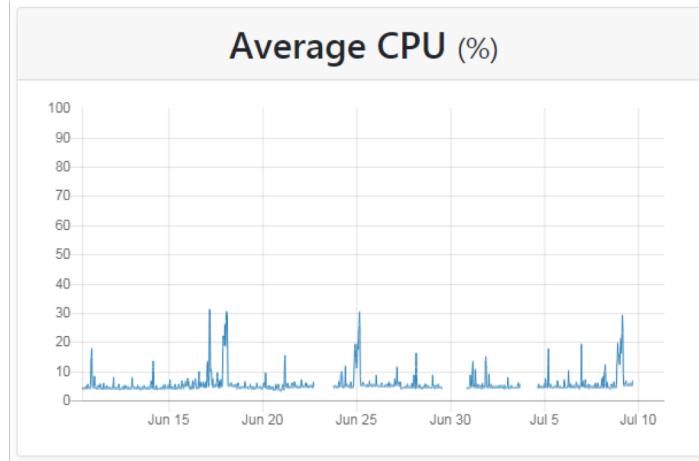


# Result: Idle VM – Signal-1 look-alike

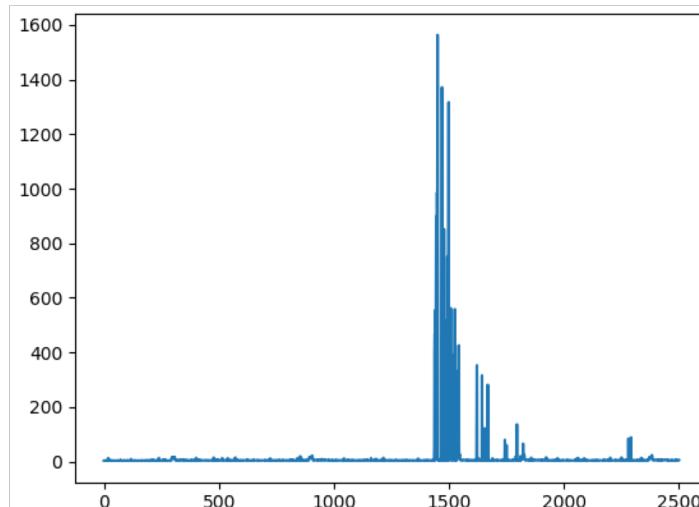


- CPU and network utilization of an idle VM looks like the Signal-1 of the ‘theoretical’ problem!
- We use our ‘theoretical’ solution which leads to the anomaly score of 4.22 for this VM.
- This score is above the set threshold.
- VM is Idle and needs to be decommissioned.

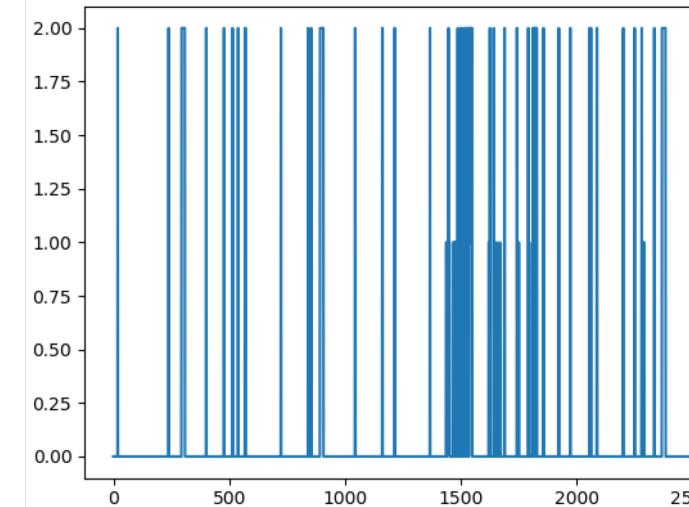
# Result: Normal VM – Signal-2 look-alike



Weighted sum of CPU and Network – model input signal



Posterior log probabilities of Model prediction



- CPU and network utilization of a normal VM looks like the Signal-2 of ‘theoretical’ problem!
- We use our ‘theoretical’ solution which leads to the anomaly score of 0.11 for this VM.
  - This score is below the set threshold.
  - This is a normal VM and no action is needed.

# Fundamental references



- <https://scikit-learn.org/>
  - specifically <https://scikit-learn.org/stable/modules/mixture.html>
- <http://www.numpy.org/>
- <https://www.elastic.co>