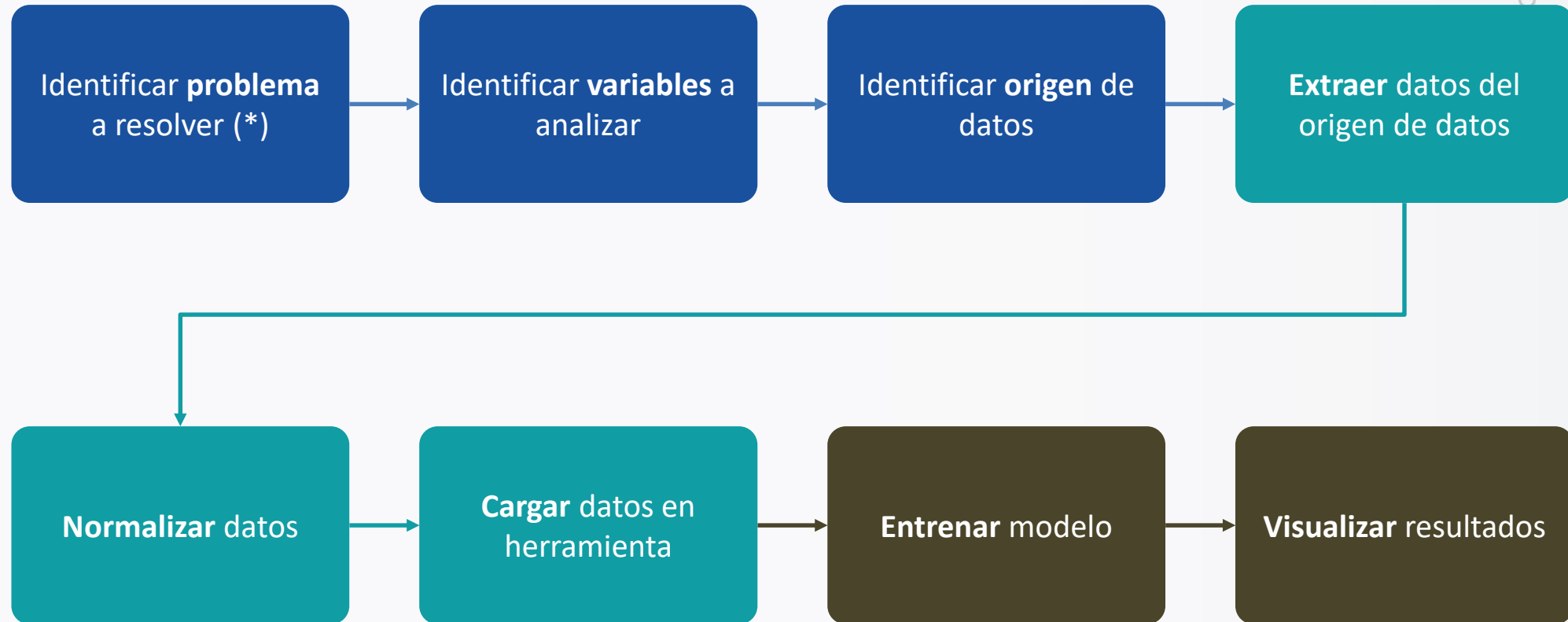




IA/ML en la Cloud de Presidencia

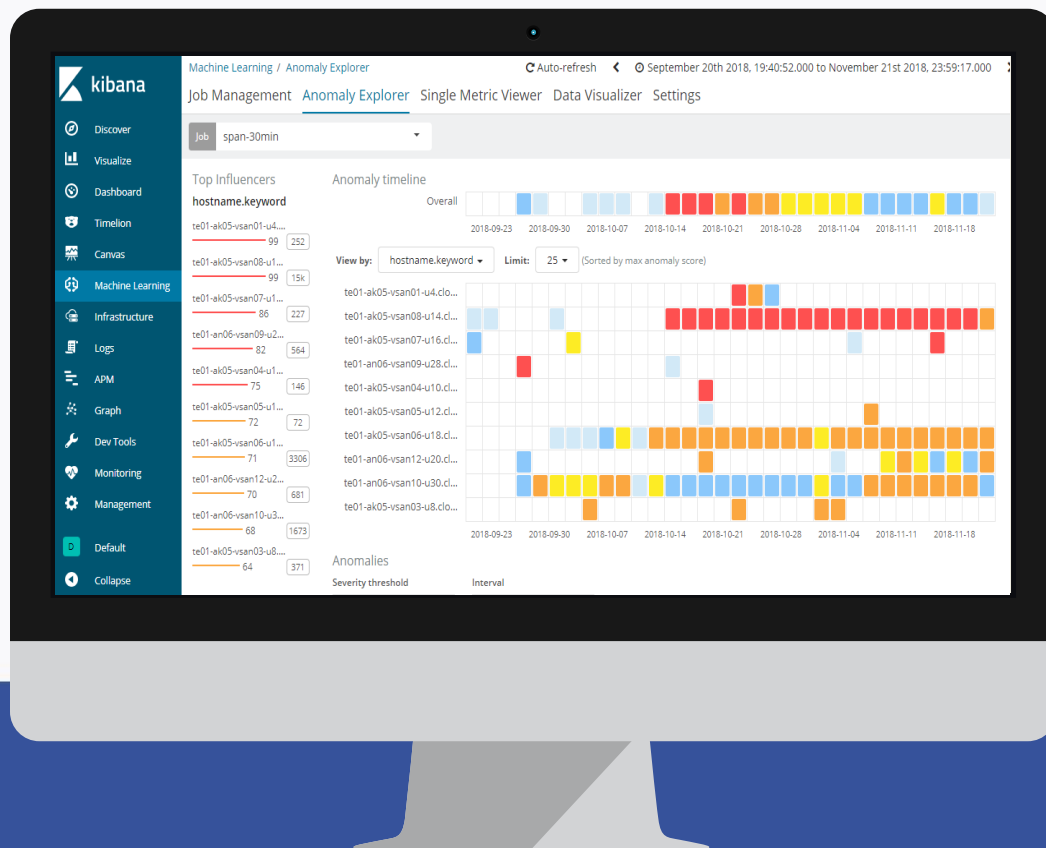
E. Galván / F. Cormenzana

¿Qué hay que hacer para aplicar ML?



Un poco de historia...

En octubre 2018 comenzamos a analizar datos de plataforma, específicamente **scsi-errors** y **scsi-latency** desde LogInsight con **ELK ML**.



Resultados



Identificamos
hosts que se
comportaban
erráticamente

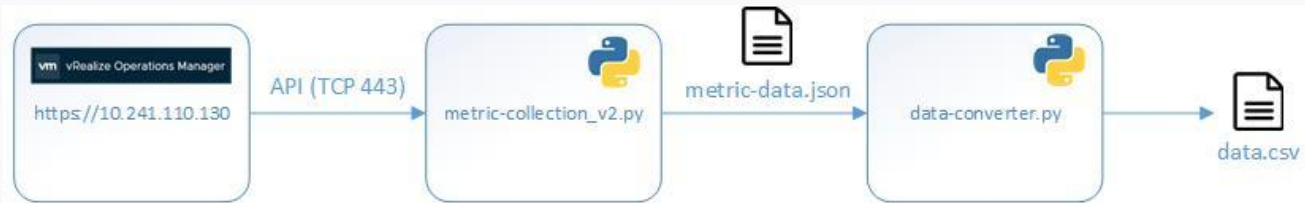


Operaciones
ya contaba
con esta
información

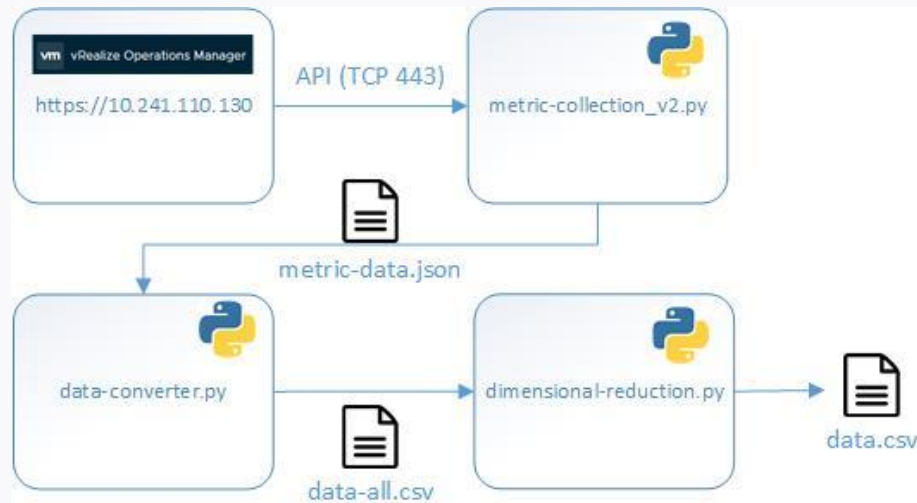


2019

Comenzamos a utilizar **algoritmos de ML** para analizar datos de **VMs** desde VROPs

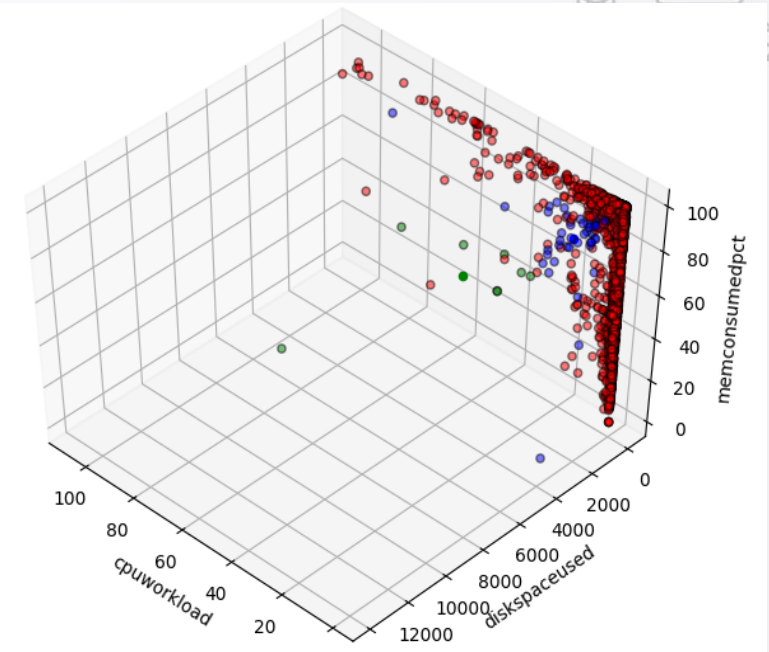
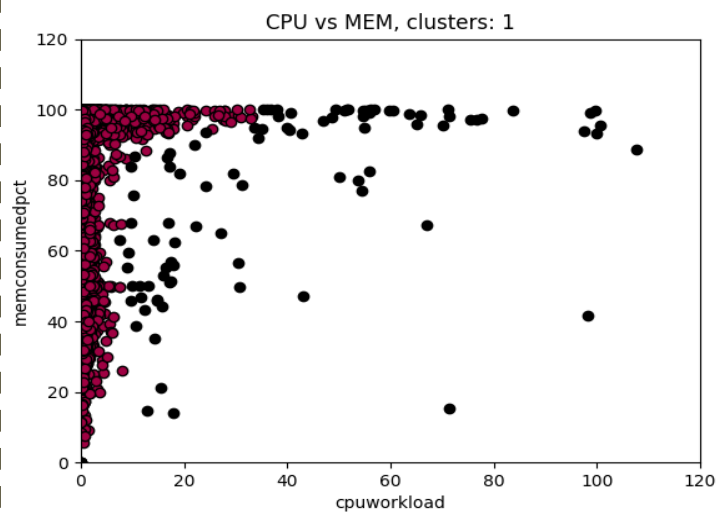
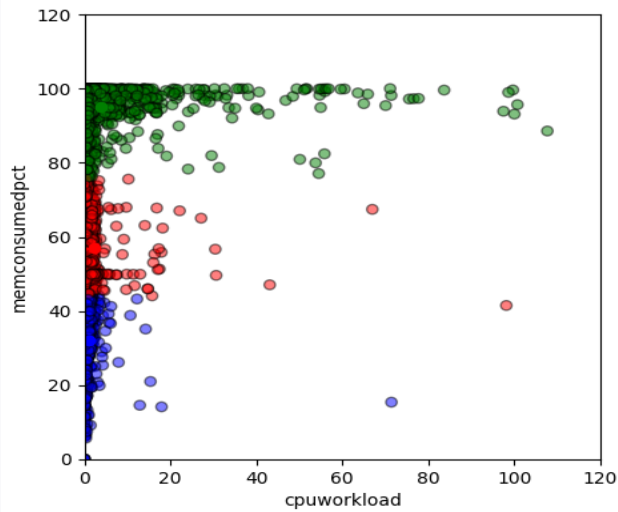


Al principio dataset en csv y
luego analizábamos los datos



Demasiadas features, tuvimos
que eliminar algunas...

Comenzamos implementando algoritmos de **clusterización** (aprendizaje no supervisado) intentando buscar **patrones** en los datos



Desarrollamos una “herramienta”

- » Objetivo: **análisis de datos**.
- » Permite escoger 2 **algoritmos**: KMEANS y DBSCAN.
- » Lee en tiempo **real** info de VROPs.
- » Devuelve una **gráfica en 3D** de los datos analizados.
- » Habilita ejercicios de **Datamining**.
- » Permite **cruzar** datos “*metrics*” y *configuración* (“*properties*”).
- » **Escala** los datos para una correcta implementación de los algoritmos.
- » Se **eliminan** variables (*dimensional reduction*) :
de +6.000 *features* pasamos a **60**.



Desafíos

1

Armado de los
datasets

2

Complejidad
matemática de
los algoritmos

3

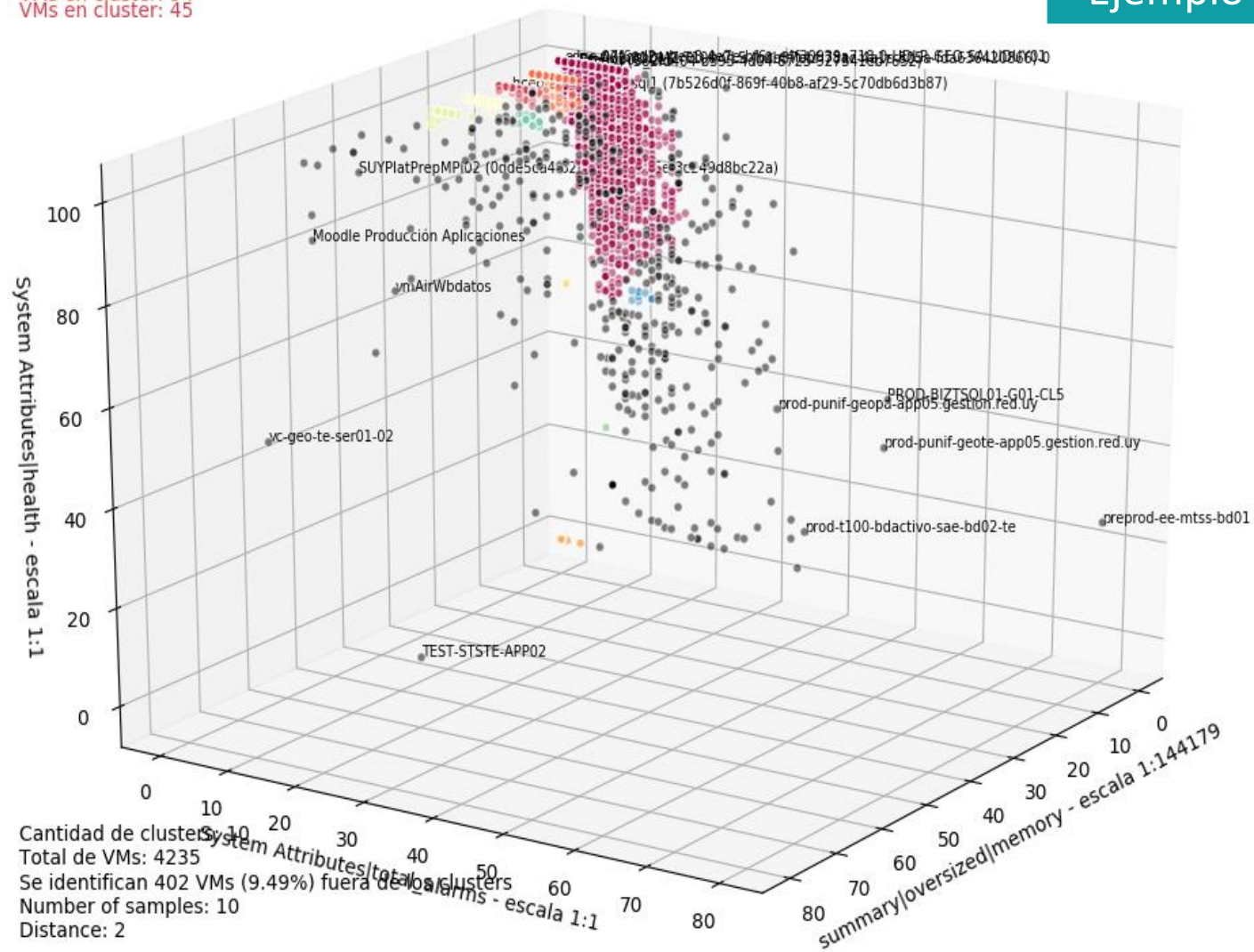
Definir un objetivo
“a priori”, es
necesaria la
iteración



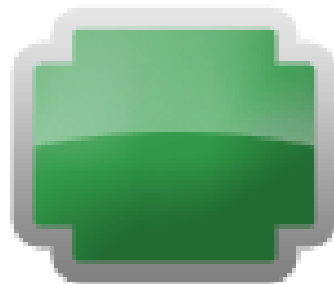
TOP 5 CLUSTERS
 VMs en cluster: 3440
 VMs en cluster: 402
 VMs en cluster: 151
 VMs en cluster: 94
 VMs en cluster: 45

DBSCAN

Ejemplo



Caso de Uso: **SystemHealth**



Health

Immediate issues



¿Qué features elegir?

60+ features = 1.770+ posibilidades



=> utilizamos **feature selection**

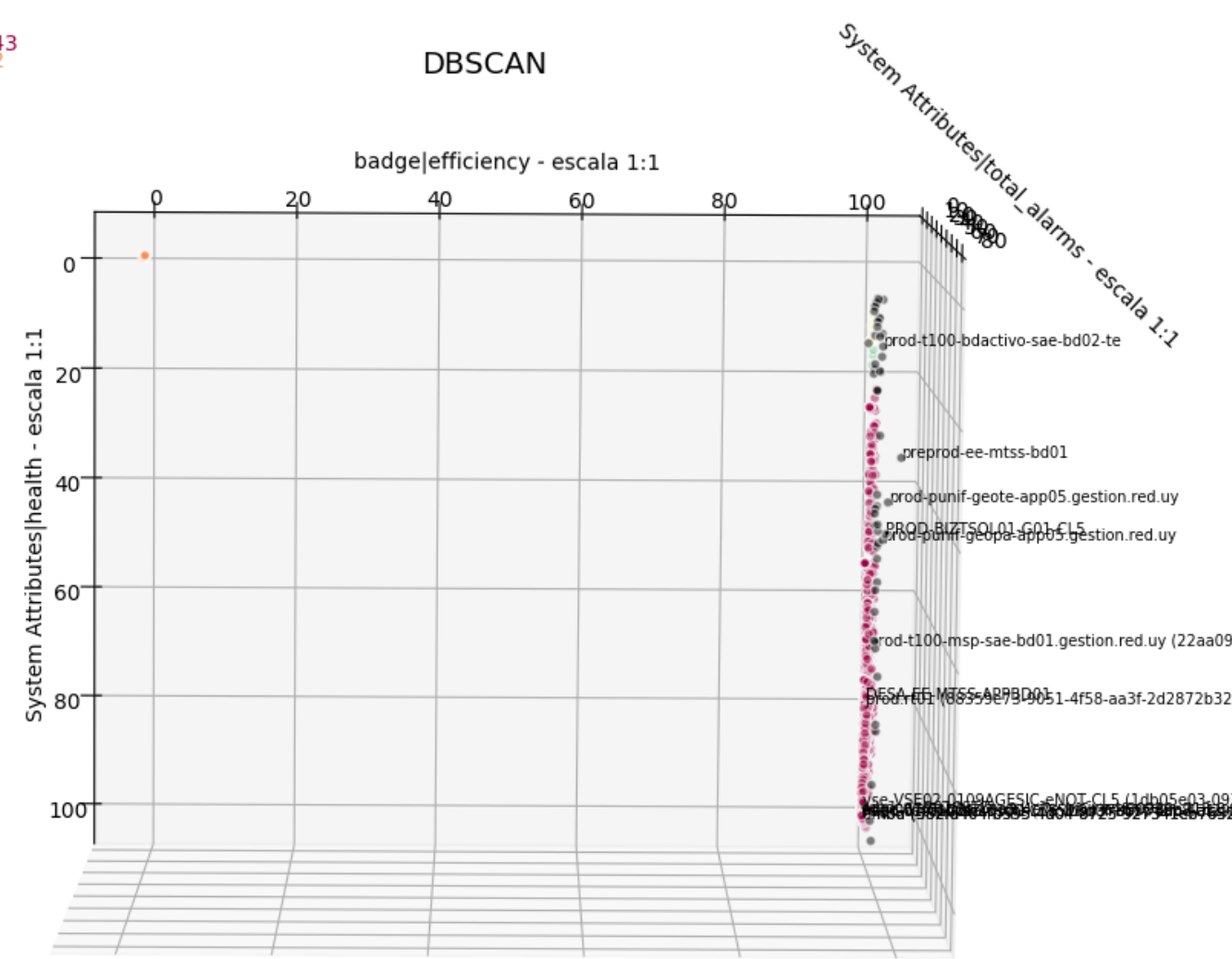
```
emilitus@HGL0113:/mnt/c/Users/egalvan/Desktop$ python feature_selection.py
```

	Specs	Score
8	System Attributes total_alarms	0.431031
9	badge efficiency	0.250032
10	badge health	0.197193
7	System Attributes child_all_metrics	0.166915
54	sys osUptime_latest	0.156419



System Attributes|health
 System Attributes|total_alarms
 badge|efficiency

TOP 5 CLUSTERS
 VMs en cluster: 4043
 VMs en cluster: 152
 VMs en cluster: 54
 VMs en cluster: 10
 VMs en cluster: 3

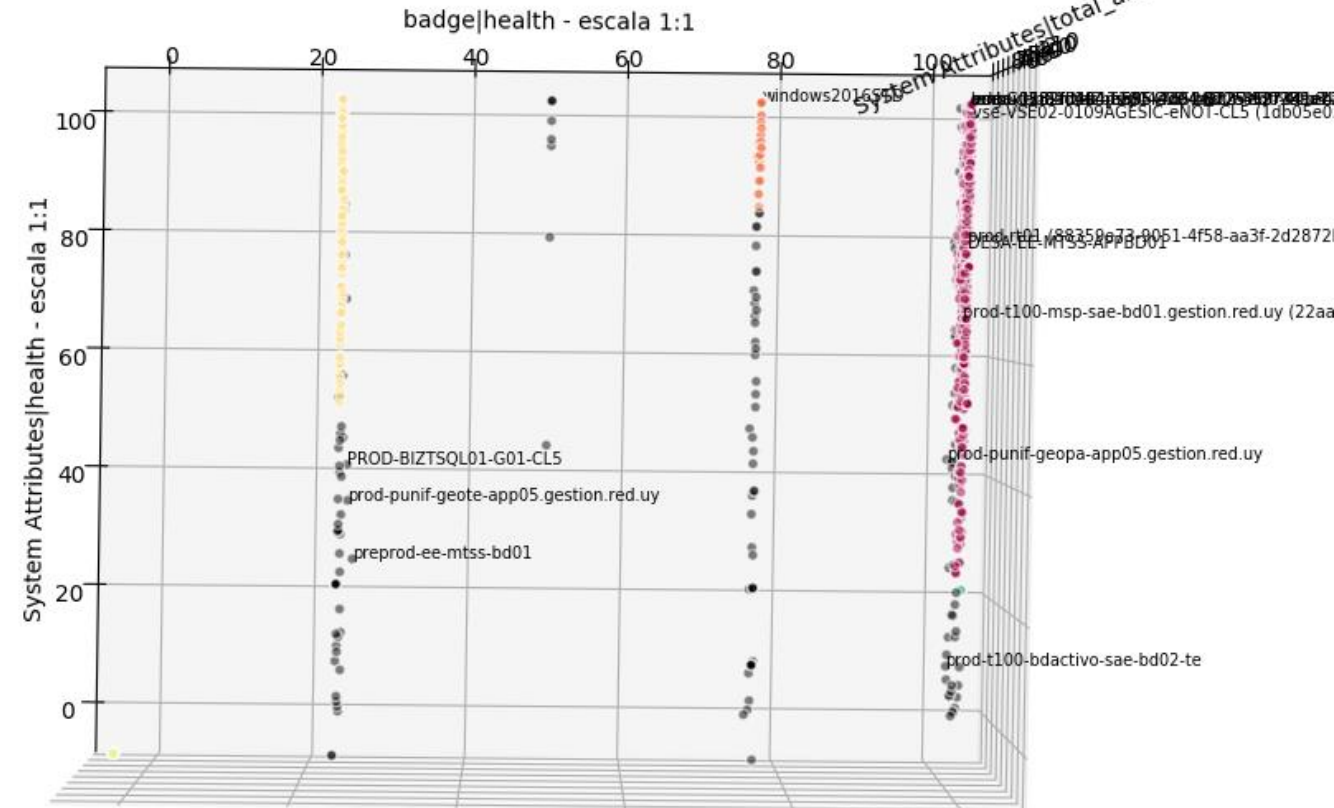


Cantidad de clusters: 4
 Total de VMs: 4262
 Se identifican 54 VMs (1.27%) fuera de los clusters
 Number of samples: 10
 Distance: 5

System Attributes|health
System Attributes|total_alarms
badge|health

TOP 5 CLUSTERS
VMs en cluster: 3594
VMs en cluster: 291
VMs en cluster: 149
VMs en cluster: 144
VMs en cluster: 76

DBSCAN



Cantidad de clusters: 5
Total de VMs: 4262
Se identifican 144 VMs (3.38%) fuera de los clusters
Number of samples: 10
Distance: 5

VMs en cluster: 18

Cantidad de clusters: 5
 Total de VMs: 4157
 Se identifican 227 VMs (5.46%) fuera de los clusters
 Number of samples: 10

Distance: 4

Resultado

- VMs con **total_alarms < 20**
tienden a tener un health entre 50 y 100
- VMs con **child_all_metrics = 0 o
entre 120 y 480**
tienden a tener un health entre 50 y 100





¿preguntas?



