**Split 0, depth 6, size 6**

**Correct 127 slices**

1. max\_no\_maintenance\_streak>=3
   1. Found 6 out of 127 slices correctly
   2. Found another 29 slices incorrectly (out of which 9 in the immediate vicinity of the correct slice)
2. max\_no\_maintenance\_streak>=3 AND total\_no\_maintenance>=3
   1. Redundant same as 1
3. max\_no\_maintenance\_streak>=3 AND utilization\_rate\_\_longest\_strike\_below\_mean: [4.0:5.0]
   1. Found 4 out of 127 slices correctly
   2. Found another 16 slices incorrectly (out of which 6 in the immediate vicinity of the correct slice)
4. max\_no\_maintenance\_streak>=3 AND total\_no\_maintenance>=3 AND utilization\_rate\_\_longest\_strike\_below\_mean: [4.0:5.0]
   1. Found 4 out of 127 slices correctly
   2. Found another 16 slices incorrectly (out of which 6 in the immediate vicinity of the correct slice)
   3. Redundant same as 3
5. max\_no\_maintenance\_streak>=3 AND operating\_temperature\_\_longest\_strike\_above\_mean: [3.0:4.0]
   1. Found 5 out of 127 slices correctly
   2. Found another 16 slices incorrectly (out of which 6 in the immediate vicinity of the correct slice)
6. max\_no\_maintenance\_streak>=3 AND operating\_temperature\_\_longest\_strike\_above\_mean: [3.0:4.0] AND total\_no\_maintenance>=3
   1. Redundant same as 5

=> found max\_no\_maintenance\_streak>=3 AND operating\_temperature\_\_longest\_strike\_above\_mean: [3.0:4.0] AND total\_no\_maintenance>=3 AND utilization\_rate\_\_longest\_strike\_below\_mean: [4.0:5.0]

**Split 25, depth 4 size 6**

**Correct 172 slices**

1. max\_no\_maintenance\_streak>=3

a. found 2 out of 172 slices correctly

b. found another 40 slices incorrectly (out of which 9 in the immediate vicinity of the correct slice)

2. max\_no\_maintenance\_streak>=3 AND total\_no\_maintenance>=3

a.redundant same as 1

3. max\_no\_maintenance\_streak>=3 AND pressure\_\_longest\_strike\_above\_mean: [3.0:4.0[

a. found 2 out of 172 slices correctly

b. found another 24 slices incorrectly (out of which 7 in the immediate vicinity of the correct slice)

4. max\_no\_maintenance\_streak>=3 AND pressure\_\_longest\_strike\_above\_mean: [3.0:4.0[ AND total\_no\_maintenance>=3

a. redundant same as 3

5. max\_no\_maintenance\_streak>=3 AND operating\_temperature\_\_longest\_strike\_below\_mean: [3.0:4.0]

a. found 2 out of 172 slices correctly

b. found another 30 slices incorrectly (out of which 7 in the immediate vicinity of the correct slice)

6. max\_no\_maintenance\_streak>=3 AND operating\_temperature\_\_longest\_strike\_below\_mean: [3.0:4.0[ AND total\_no\_maintenance>=3

a. redundant same as 5

=> . max\_no\_maintenance\_streak>=3 AND operating\_temperature\_\_longest\_strike\_below\_mean: [3.0:4.0[ AND total\_no\_maintenance>=3 AND pressure\_\_longest\_strike\_above\_mean: [3.0:4.0[

**Split 50, depth 6 size 6**

**Correct 251**

1. max\_no\_maintenance\_streak>=3.0 AND power\_consumption\_\_minimum==0.0

a. found 7 out of 251 slices correctly

b. found another 49 slices incorrectly (out of which 13 in the immediate vicinity of the correct slice)

2. max\_no\_maintenance\_streak>=3.0 AND power\_consumption\_\_minimum==0.0 AND total\_no\_maintenance>=3.0

Redundant same as 1

3. load\_\_median==1.0 AND max\_no\_maintenance\_streak>=3.0 AND power\_consumption\_\_minimum==0.0

Redundant same as 1

4. load\_\_mean==1.0 AND max\_no\_maintenance\_streak>=3.0 AND power\_consumption\_\_minimum==0.0

Redundant same as 1

5. load\_\_mean==1.0 AND max\_no\_maintenance\_streak>=3.0 AND power\_consumption\_\_minimum==0.0 AND total\_no\_maintenance>=3.0

Redundant same as 1

6. load\_\_mean==1.0 AND load\_\_median==1.0 AND max\_no\_maintenance\_streak>=3.0 AND power\_consumption\_\_minimum==0.0

Redundant same as 1

=> load\_\_median==1.0 AND max\_no\_maintenance\_streak>=3.0 AND power\_consumption\_\_minimum==0.0 AND load\_\_mean==1.0 AND total\_no\_maintenance>=3.0

**Split 75, depth 4 size 6**

**Correct 503**

1. max\_no\_maintenance\_streak>=3

a. found 9 out of 503 slices correctly

b. found another 124 slices incorrectly (out of which 22 in the immediate vicinity of the correct slice)

2. max\_no\_maintenance\_streak>=3 AND total\_no\_maintenance>=3

Redundant same as 1

3. machine\_id==3 AND max\_no\_maintenance\_streak>=3

a. found 6 out of 503 slices correctly

b. found another 84 slices incorrectly (out of which 15 in the immediate vicinity of the correct slice)

4. machine\_id==3 AND max\_no\_maintenance\_streak>=3 AND total\_no\_maintenance>=3

Redundant same as 3

5. machine\_age==36 AND max\_no\_maintenance\_streak>=3 AND total\_no\_maintenance>=3

Redundant same as 3

6. machine\_age==36 AND machine\_id==3 AND max\_no\_maintenance\_streak>=3 AND total\_no\_maintenance>=3

Redundant same as 3

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Split | Precision | Recall | Rank | Precision extra | Recall extra | Rank extra |
| 0 | 0.1714 | 0.0472 | 1 | 0.4285 | 0.118 | 1 |
| 25 | 0.0476 | 0.0116 | 4 | 0.2619 | 0.0639 | 3 |
| 50 | 0.125 | 0.0278 | 2 | 0.3571 | 0.0796 | 2 |
| 75 | 0.0676 | 0.0178 | 3 | 0.2333 | 0.0616 | 4 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Split | TP | FP | TN | FN | Extra P |
| 0 | 6 | 29 | 1669 | 121 | 9 |
| 25 | 2 | 40 | 2218 | 170 | 9 |
| 50 | 7 | 49 | 3345 | 244 | 13 |
| 75 | 9 | 124 | 6658 | 494 | 22 |

The extra: considers identifying a slice right next to the correct slice (so the previous or the next one) as correct

Discussion Madalina:

* SD algo does not find completely different subgroups, it finds a main condition and adds to it

Discussion Andrea:

- different overlaps seem to have focused on different attributes that characterize some of the patterns which were injected (ie: the power consumption min? operating temp? total no maintenance? All usually only appear within a single slicing overlap)

- maybe our implementation of gp-growth was not the optimal method of recreating the other results, and we recommend using sd-map\*

- inherently during the feature extraction, tsfresh created some unhelpful and unnecessary features that led to it being difficult to tell if our patterns were indeed being discovered + inherent bias of inputting our own features

- inherently during the overlapping, it is actually possible that our patterns were lost and the data was overwritten/we lost the aspect of timeseries or having something “last for a certain while” in the data

- including the proximal entries in calculating the precision/recall really changed the results