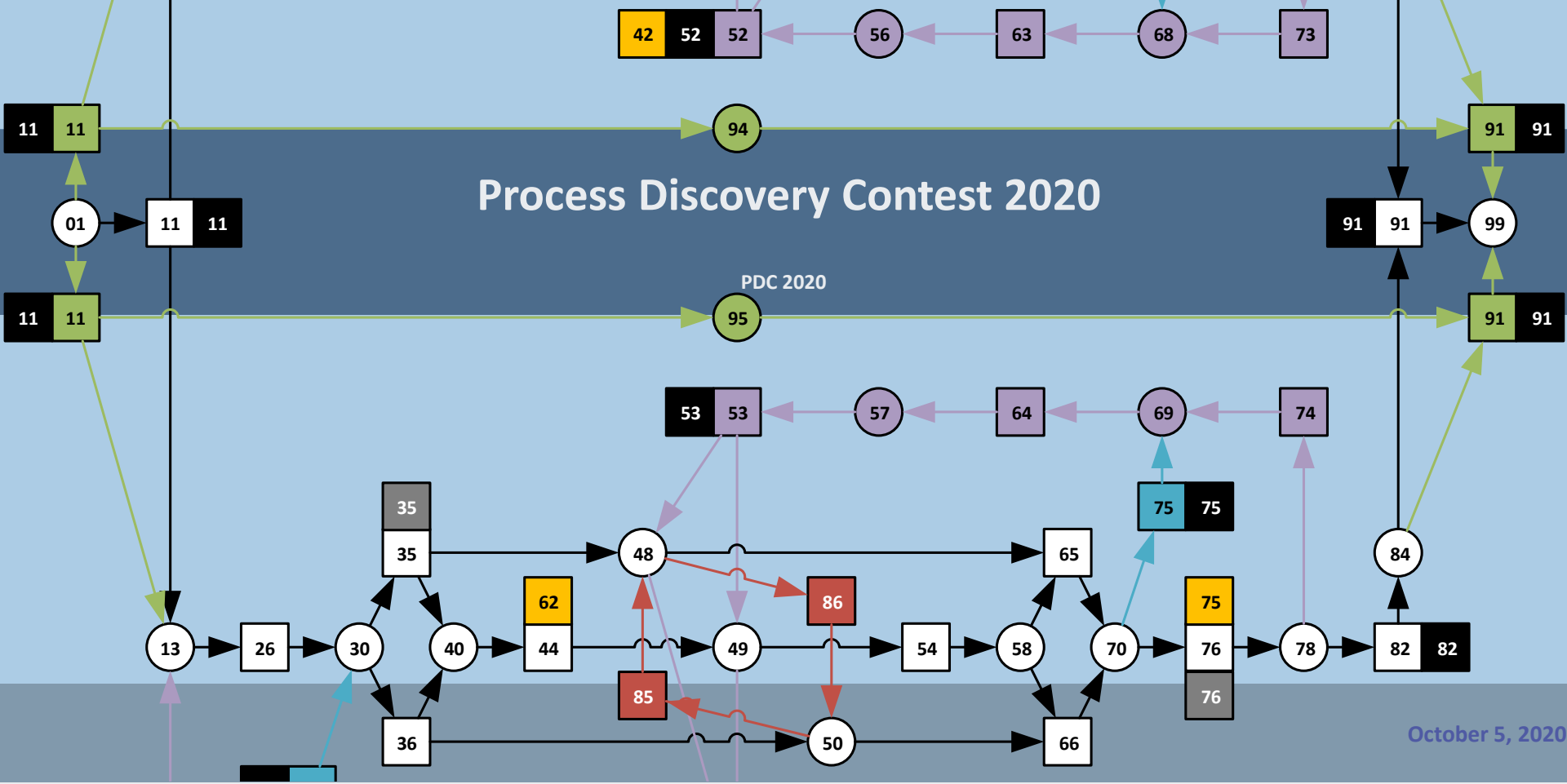
Process Discovery Contest 2020

PDC 2020



October 5, 2020

Josep Carmona, Benoît Depaire, Eric Verbeek, PDC 2020 organizers

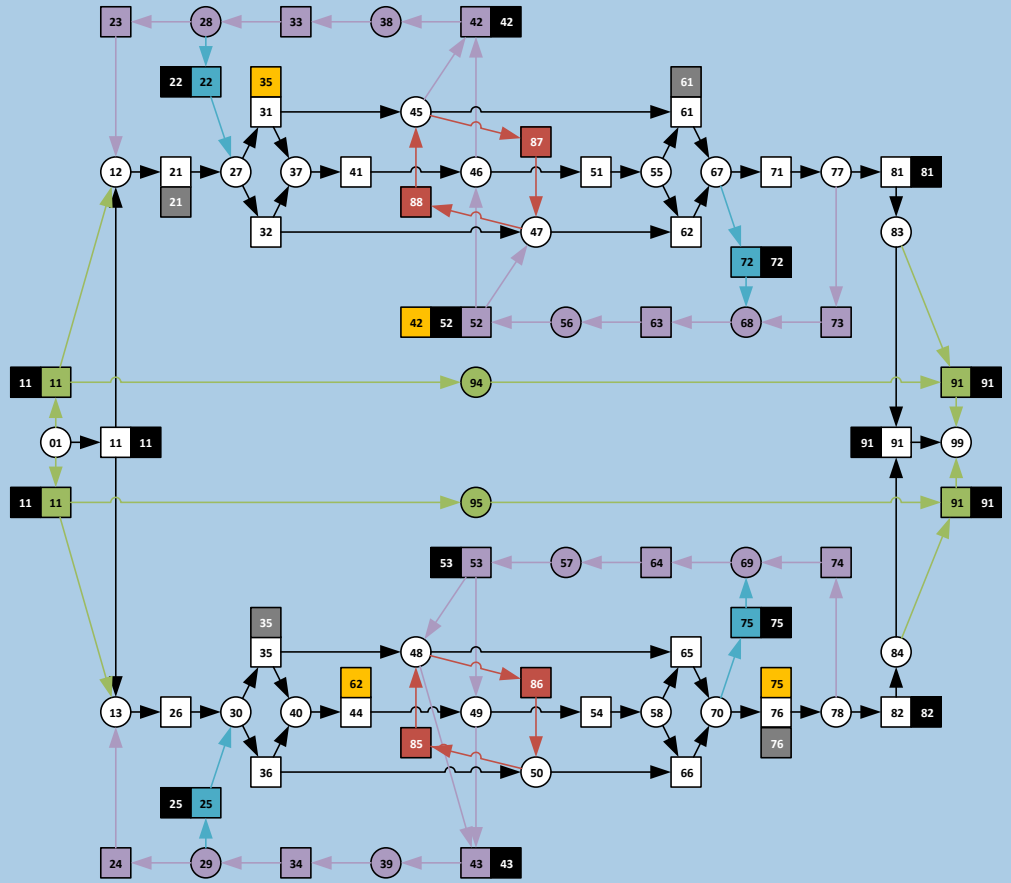
IEEE TASK FORCE ON PROCESS MINING

PDC 2020 Setup

Josep Carmona, Benoît Depaire, Eric Verbeek, PDC 2020 organizers

The master model

A. Dependent tasks (red)
B. Loops
   1. Simple (purple)
   2. Complex (blue)
C. OR constructs (green)
D. Routing constructs (black)
E. Optional tasks (grey)
F. Duplicate tasks (yellow)
G. Noise (added afterwards)



Data set

- 2x3x2x2x2x2 = 96 models
- For every model:
  - 2 training logs
    - With and without about 20% noise
  - 2 identical test logs
    - With about 50% noise
  - 2 identical ground truth logs
    - By classifying every test log using the model
- ✓ https://icpmconference.org/2020/process-discovery-contest/data-set/

Classification score

- Balanced mean:
  - True positive rate (TPR = TP/(TP+FN))
  - True negative rate (TNR = TN/(TN+FP))

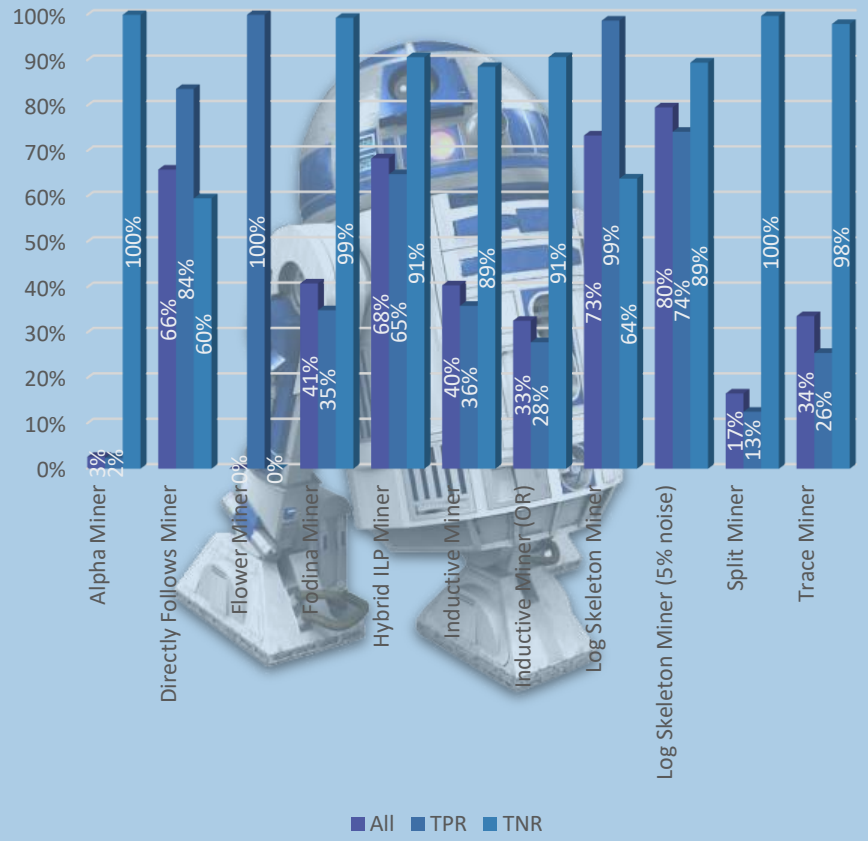
  - About 50% noise in test logs
  - If all positive (or negative), then 0.0
  - If 1.0, then all true

- Model independent
  - We only need to have a classifier whether a trace fits the model or not.

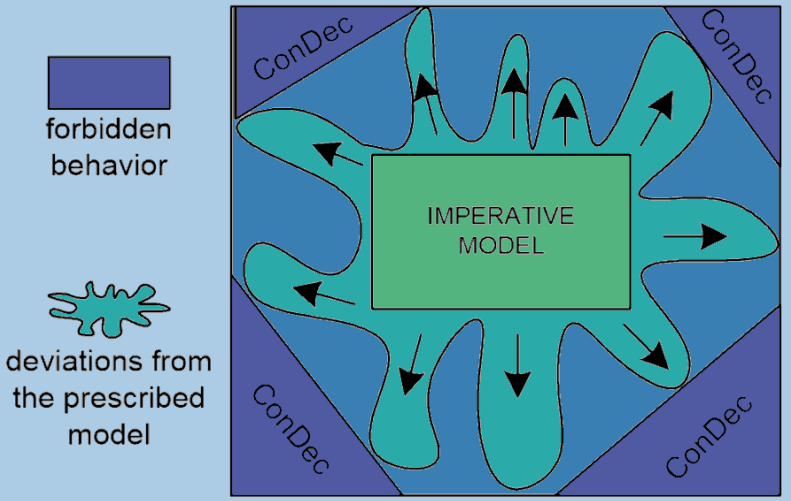
Automated contest

- Logs are not disclosed
- Working algorithms should be submitted
- Organizers (three times):
  1. run the algorithm on every training log,
  2. classify the corresponding test log using the discovered model, and
  3. Score the classified test log using the corresponding ground truth log.
- 11 example algorithms
  - ✓ https://icpmconference.org/2020/category/pdc-2020/example-discovery-algorithm/

Example algorithms



Automated contest

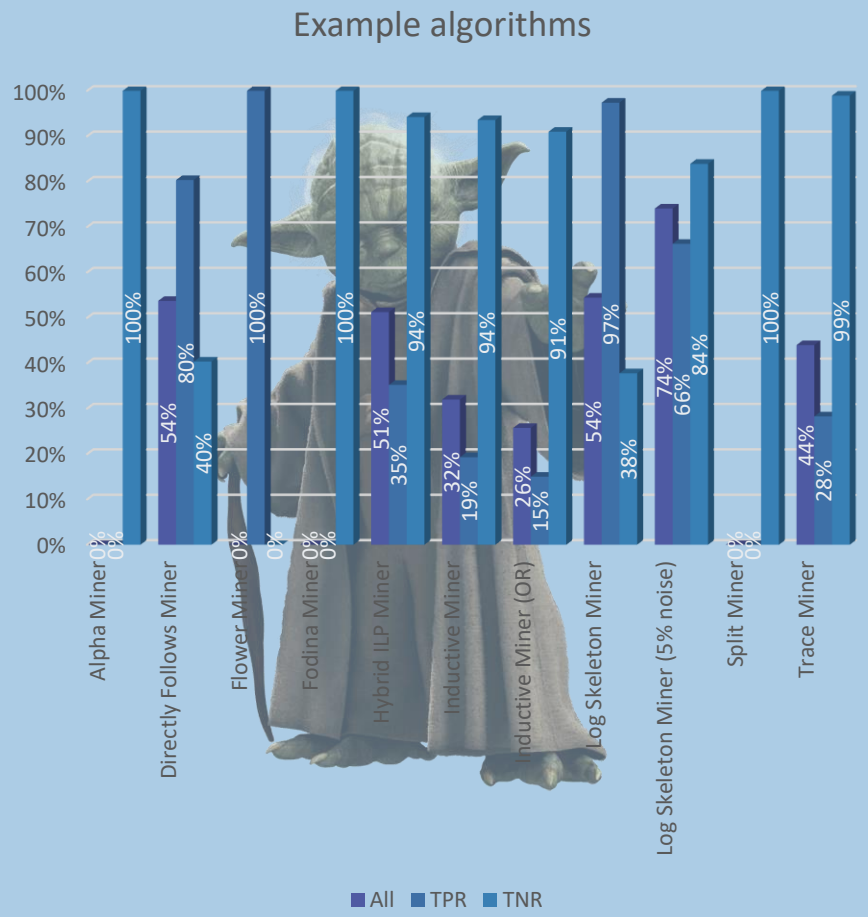


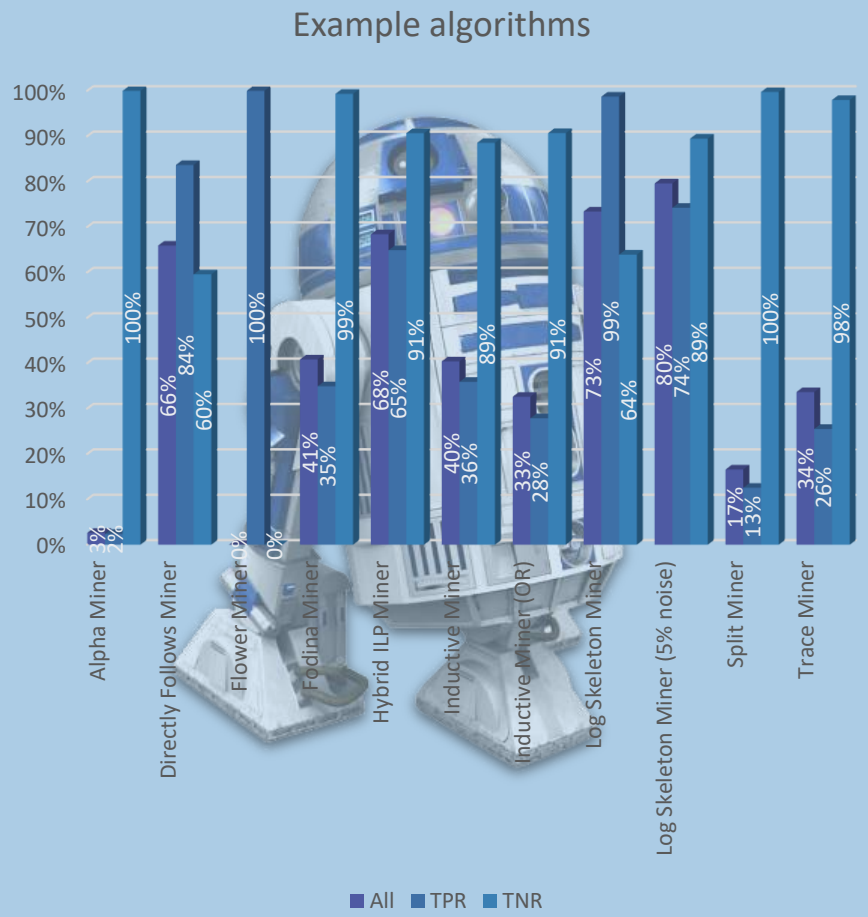
(a) Declarative ConDec vs. imperative languages

- DECLARE: Full Support for Loosely-Structured Processes – Pesic, Schonenberg, Van der Aalst, EDOC 2007

Manual contest

- Most complex training log and the corresponding test log are disclosed
- Classified test log should be submitted
- Organizers:
  1. Score the classified test log using the corresponding ground truth log.





Downloads

- Data set
  - 96 Models, 192 training logs, 192 test logs, 192 ground truth logs
- 11 implemented example algorithms
- All implemented submissions
- Classifiers
  - BPMN, DCR, LSK, PNML, XES
- Scorer

✓ https://icpmconference.org/2020/process-discovery-contest/downloads/

PDC 2020 Results

Josep Carmona, Benoît Depaire, Eric Verbeek, PDC 2020 organizers

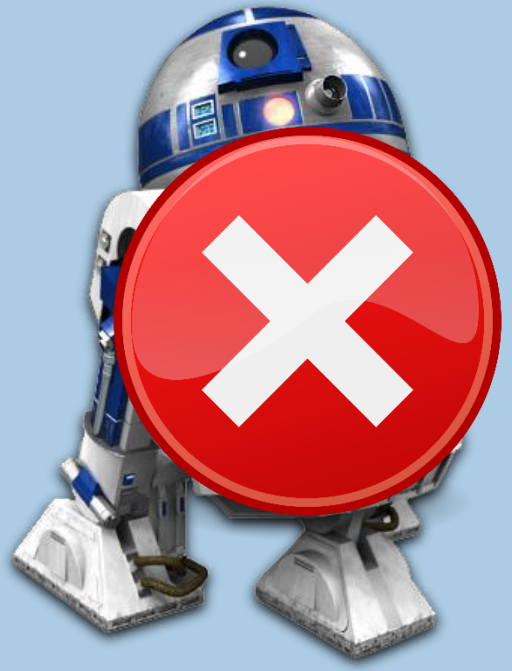
IEEE TASK FORCE ON PROCESS MINING

Manual contest

- No submissions
- No winner

Automated contest

- 10 submissions:
  - Tijs Slaats c.s.
    - DisCoveR [Light] (Open|Closed) World
  - Karl Heinrichmeyer and Robin Bergenthum
    - KokosMiner[2[T5]]
  - Sander Leemans
    - InductiveMinerIM(c|fa)
    - DirectlyFollowsModelMiner
- ✓ https://icpmconference.org/2020/process-discovery-contest/automated-contest/

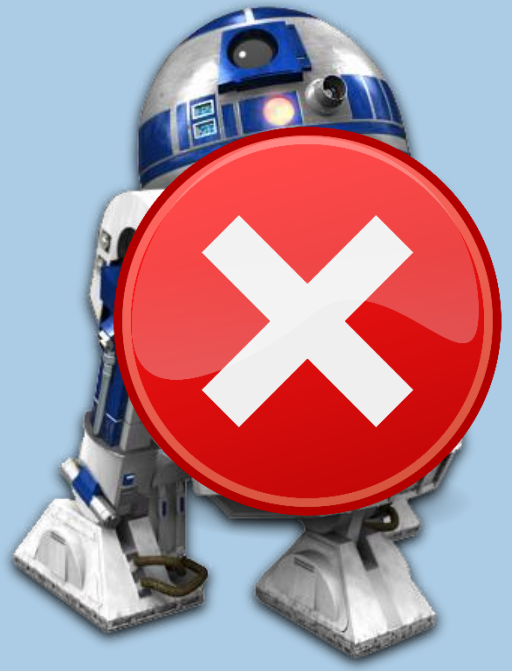


Automated contest

8. KokosMiner[2], InductiveMinerIMc

Insufficient results

- KokosMiner2
  - 100% if no noise, no duplicate tasks, no optional tasks

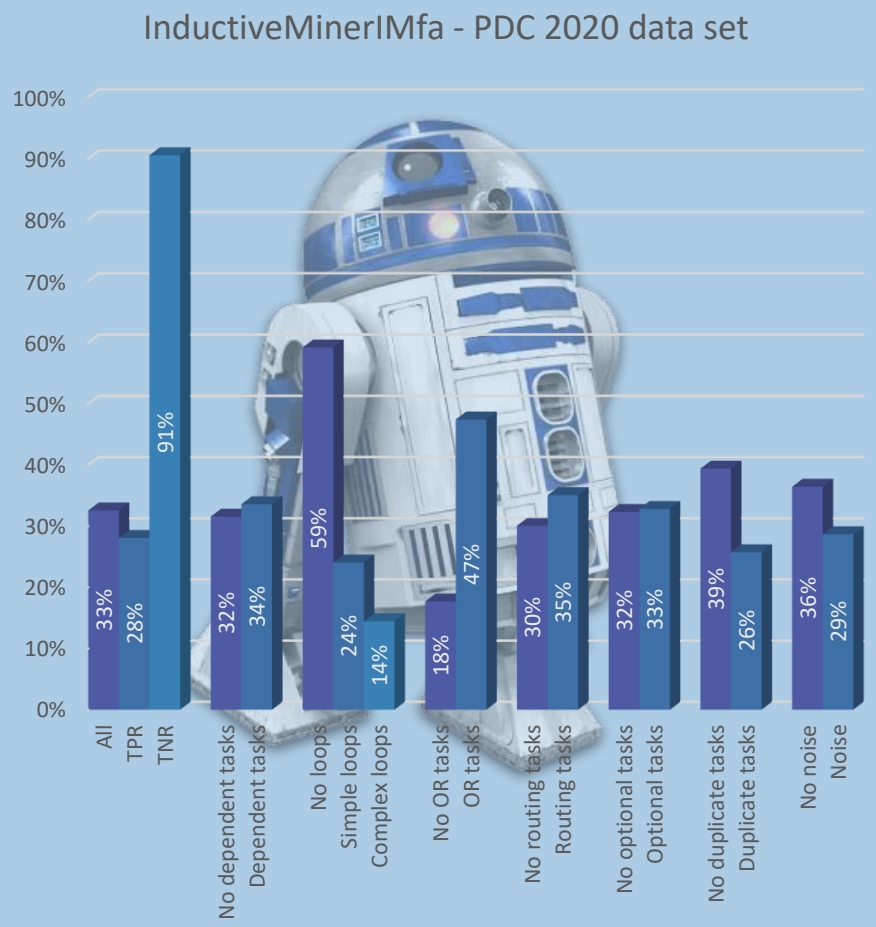


Automated contest

8. KokosMiner[2], InductiveMinerIMc

Insufficient results

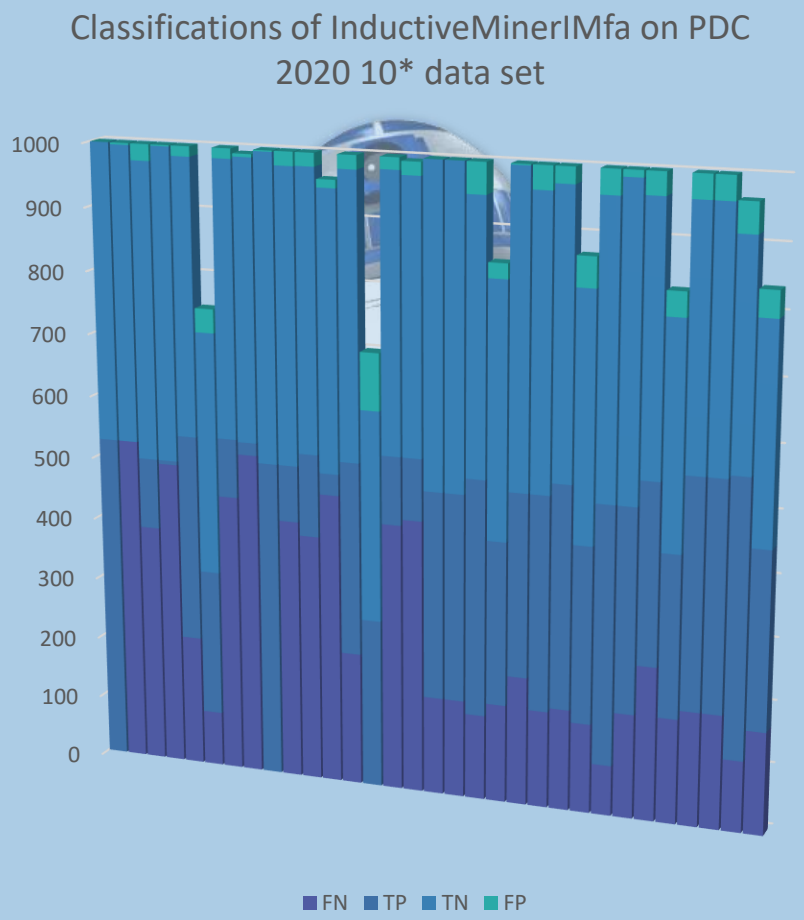
- InductiveMinerIMc
  - 100% if no noise, no duplicate tasks, no dependent tasks, and no complex loops (98% with dependant tasks)



Automated contest

7. 33% InductiveMinerIMfa
8. KokosMiner[2], InductiveMinerIMc

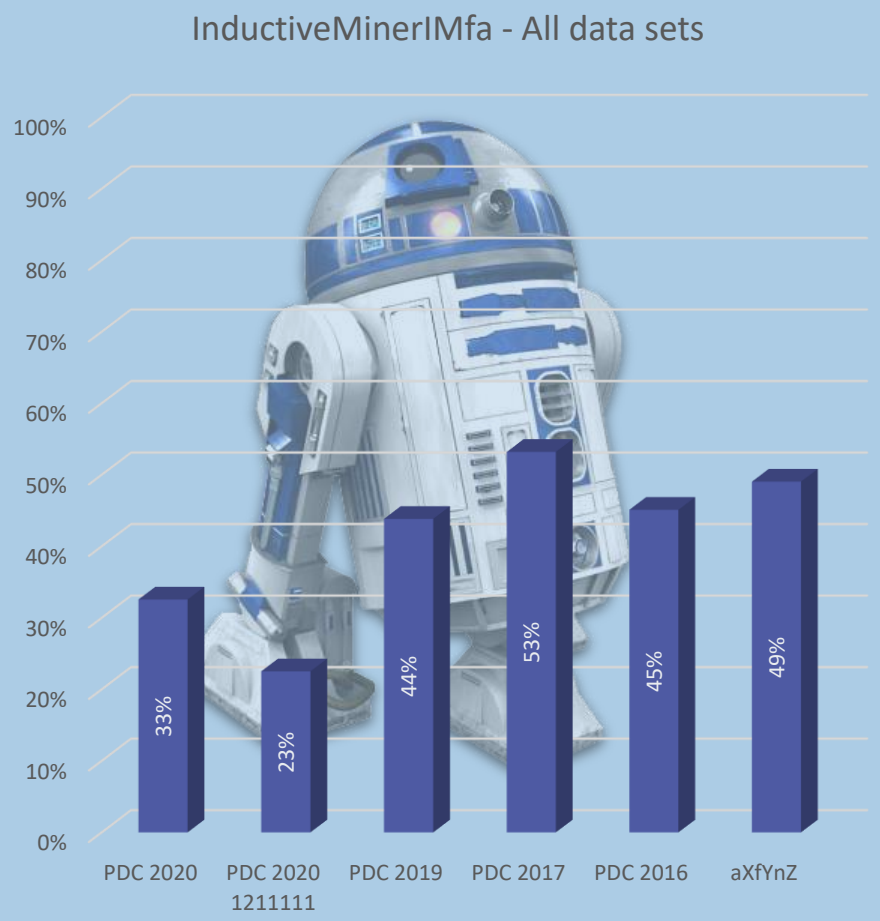
➢ 59% if no loops



Automated contest

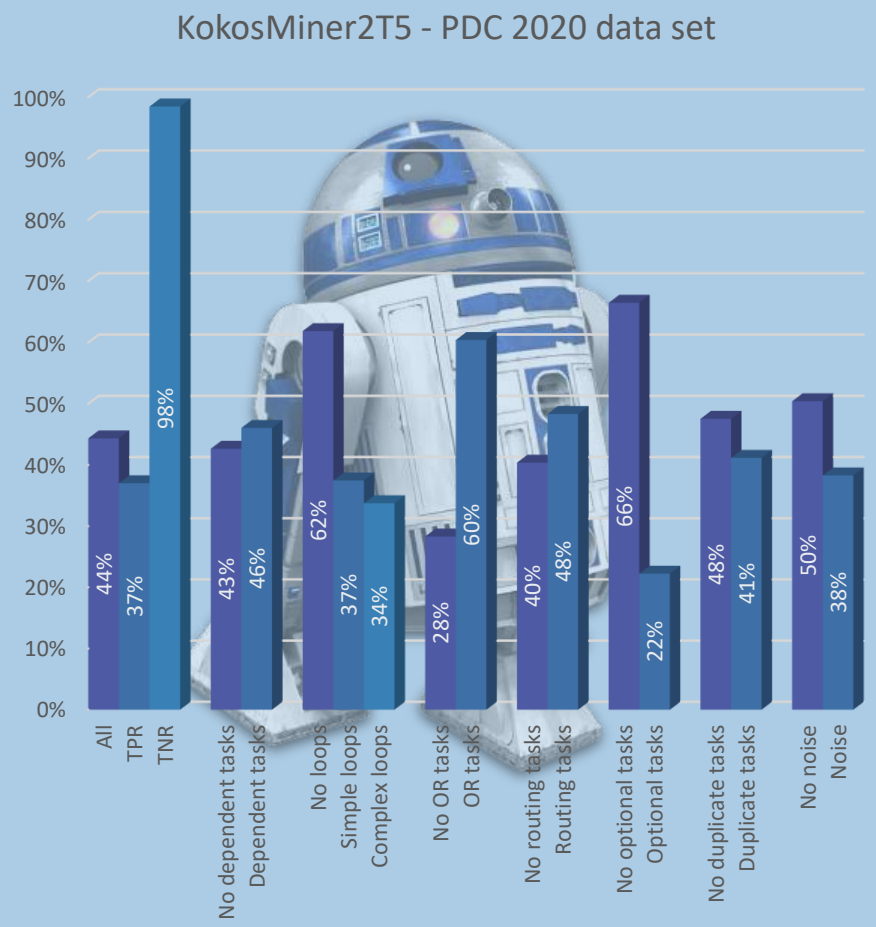
7. 33% InductiveMinerIMfa
8. KokosMiner[2], InductiveMinerIMc

➢ Non-deterministic!
- Minimal FN/TP/TN/FP over three runs



Automated contest

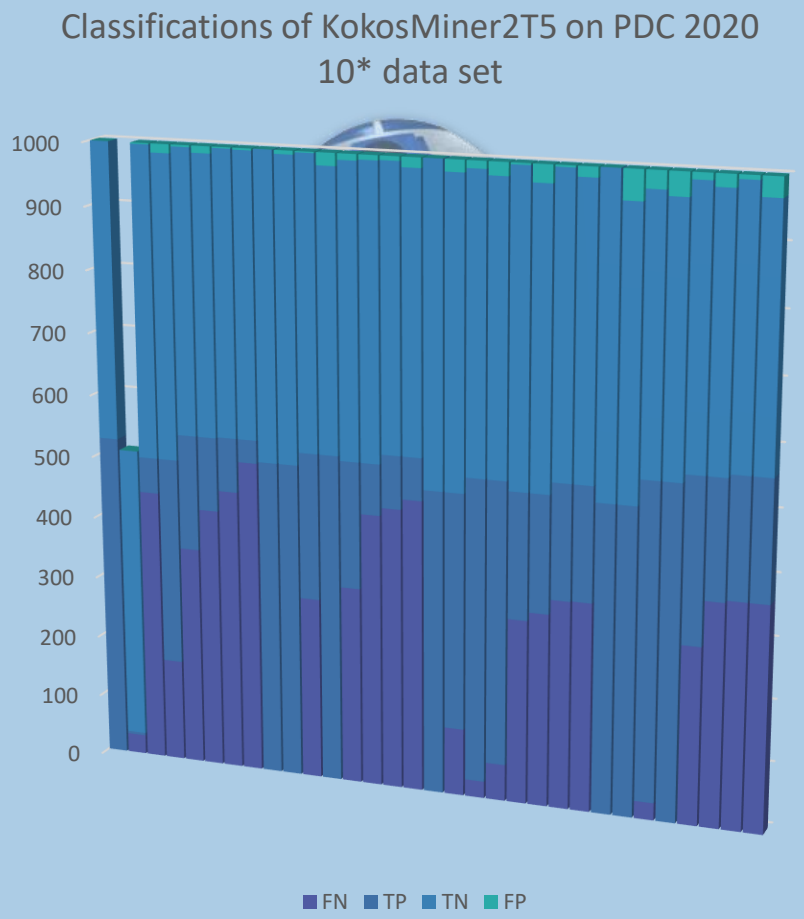
7. 33% InductiveMinerIMfa
8. KokosMiner[2], InductiveMinerIMc



Automated contest

6. 44% KokosMiner2T5
7. 33% InductiveMinerIMfa
8. KokosMiner[2], InductiveMinerIMc

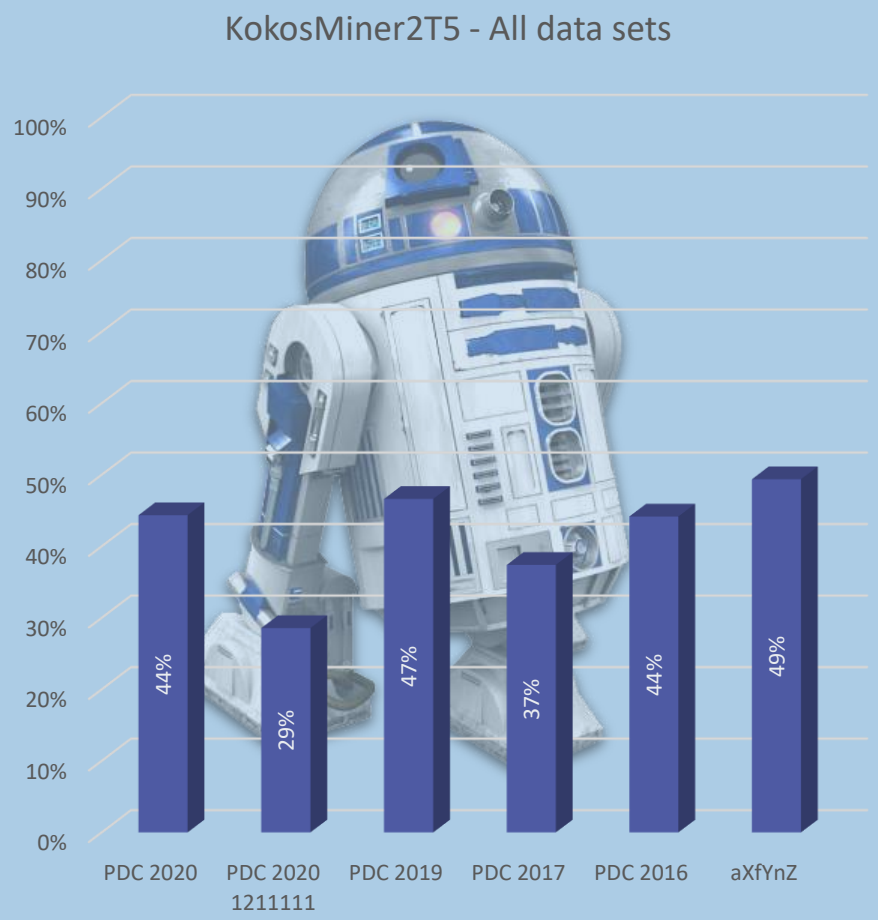
- 66% if no optional tasks
- KokosMiner2 scores 48% (single run)



Automated contest

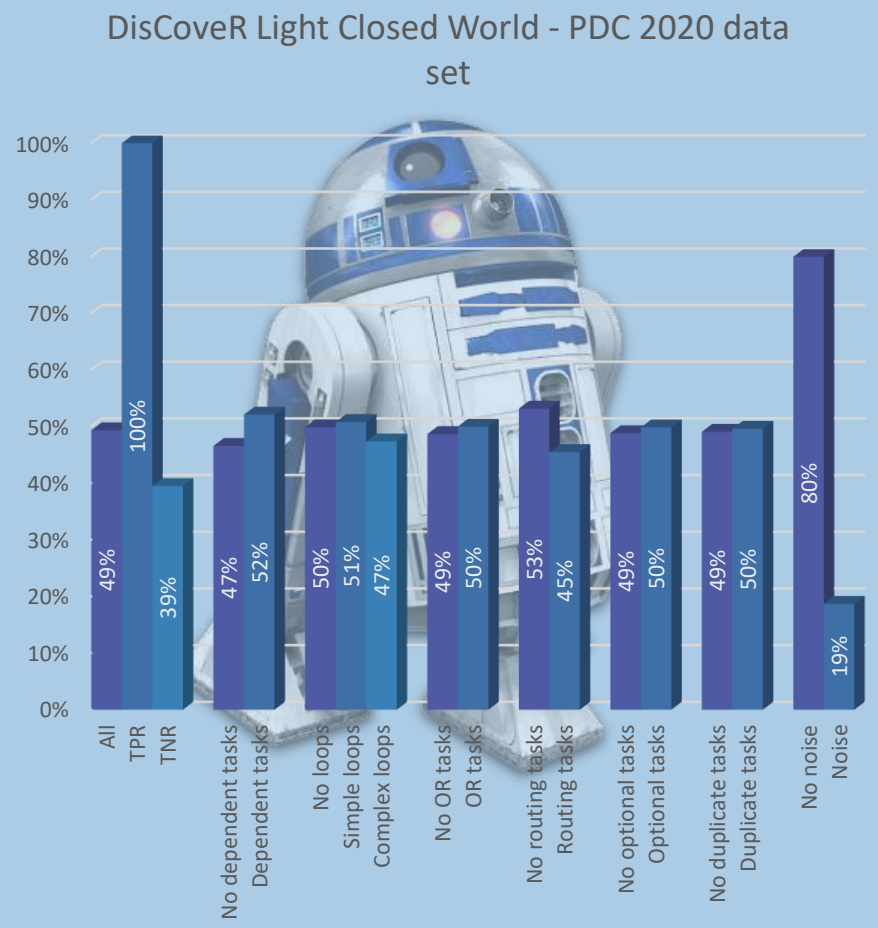
6. 44% KokosMiner2T5
7. 33% InductiveMinerIMfa
8. KokosMiner[2], InductiveMinerIMc

➢ Non-deterministic
- Due to timeout of 5 minutes (?)



Automated contest

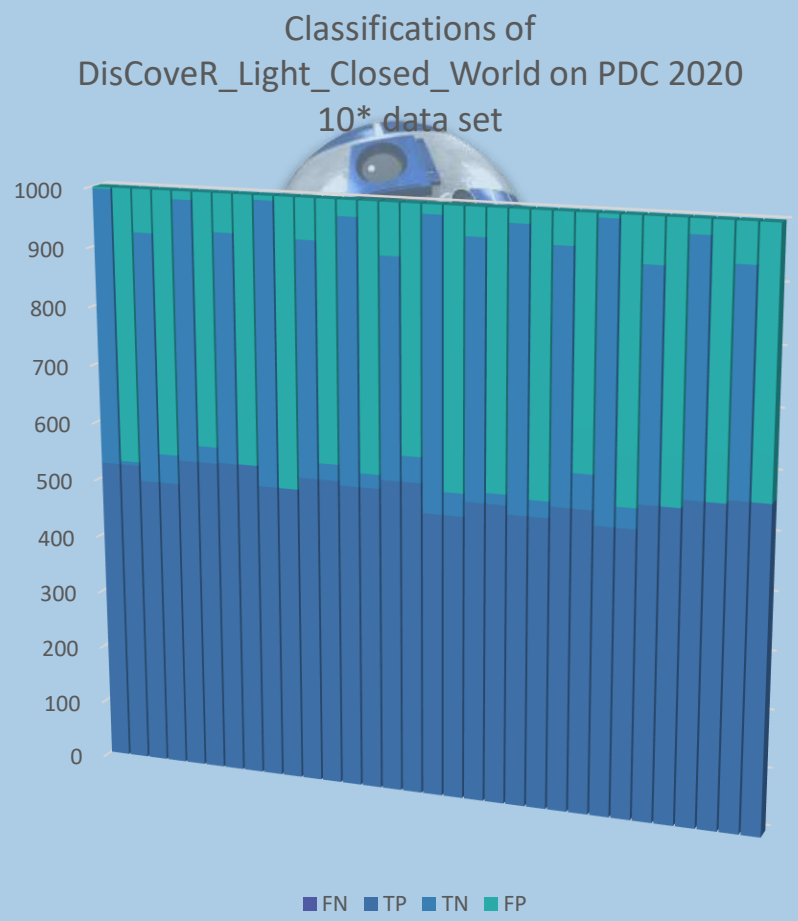
6. 44% KokosMiner2T5
7. 33% InductiveMinerIMfa
8. KokosMiner[2], InductiveMinerIMc



Automated contest

4. 49% DisCoveR Light (O|C) World
6. 44% KokosMiner2T5
7. 33% InductiveMinerIMfa
8. KokosMiner[2], InductiveMinerIMc

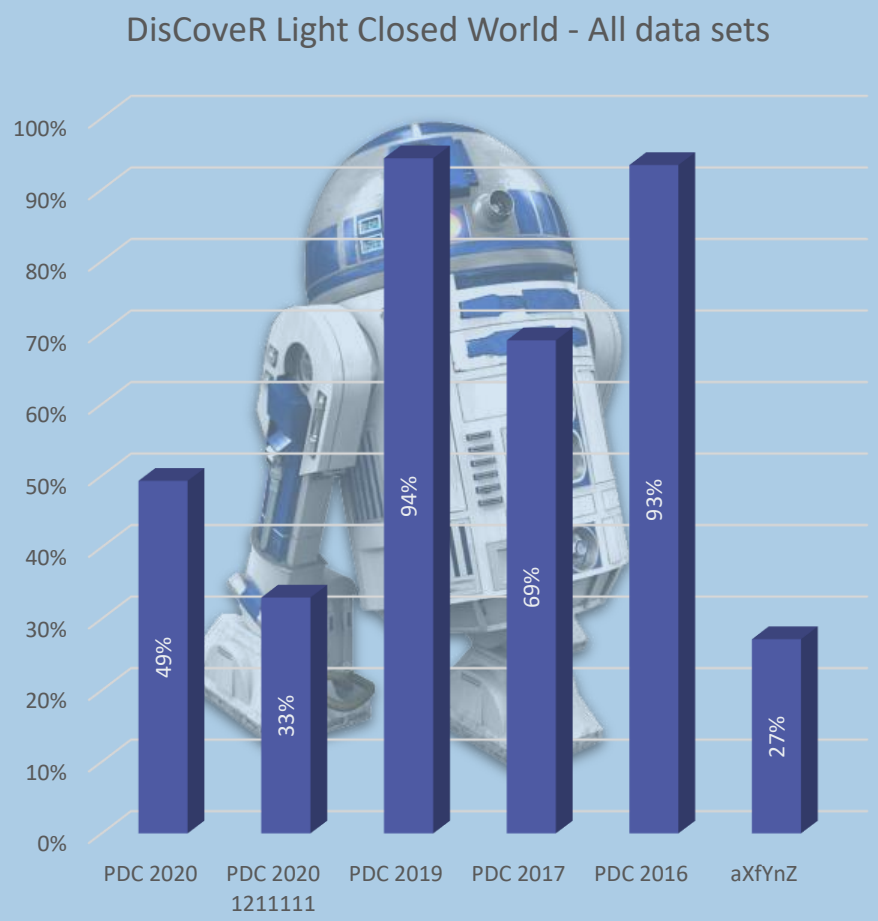
➢ 80% if no noise



Automated contest

4. 49% DisCoveR Light (O|C) World
6. 44% KokosMiner2T5
7. 33% InductiveMinerIMfa
8. KokosMiner[2], InductiveMinerIMc

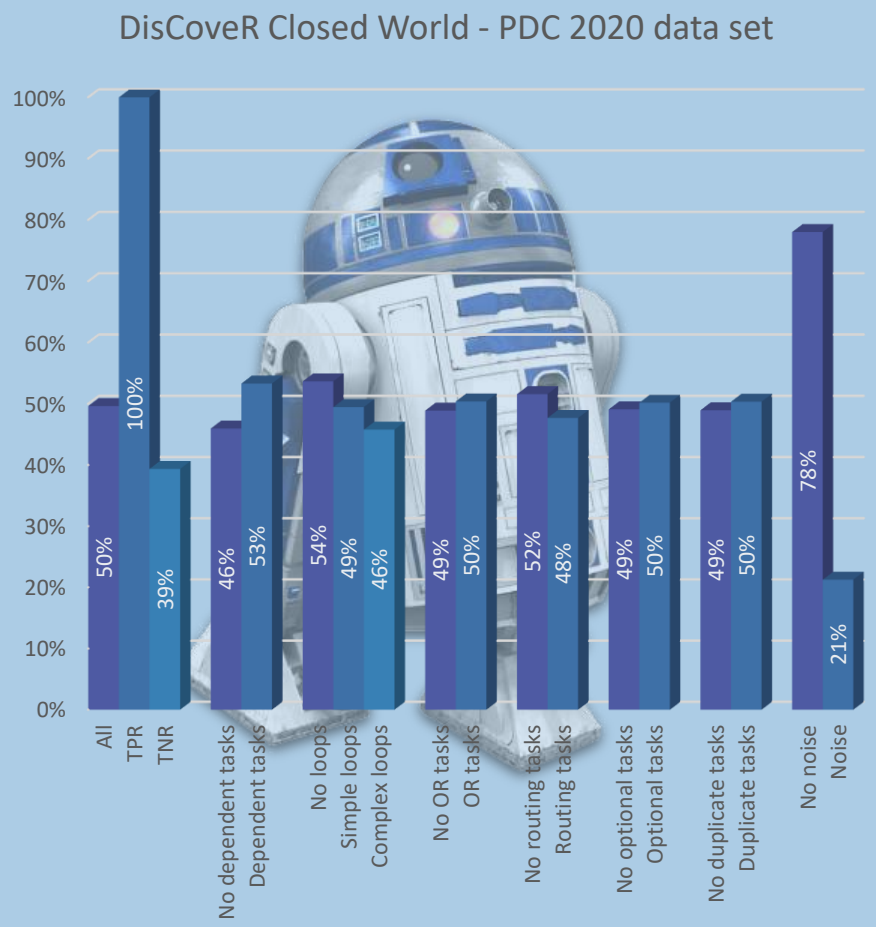
- Almost no FNs
- Many FPs if noise



Automated contest

4. 49% DisCoveR Light (O|C) World
6. 44% KokosMiner2T5
7. 33% InductiveMinerIMfa
8. KokosMiner[2], InductiveMinerIMc

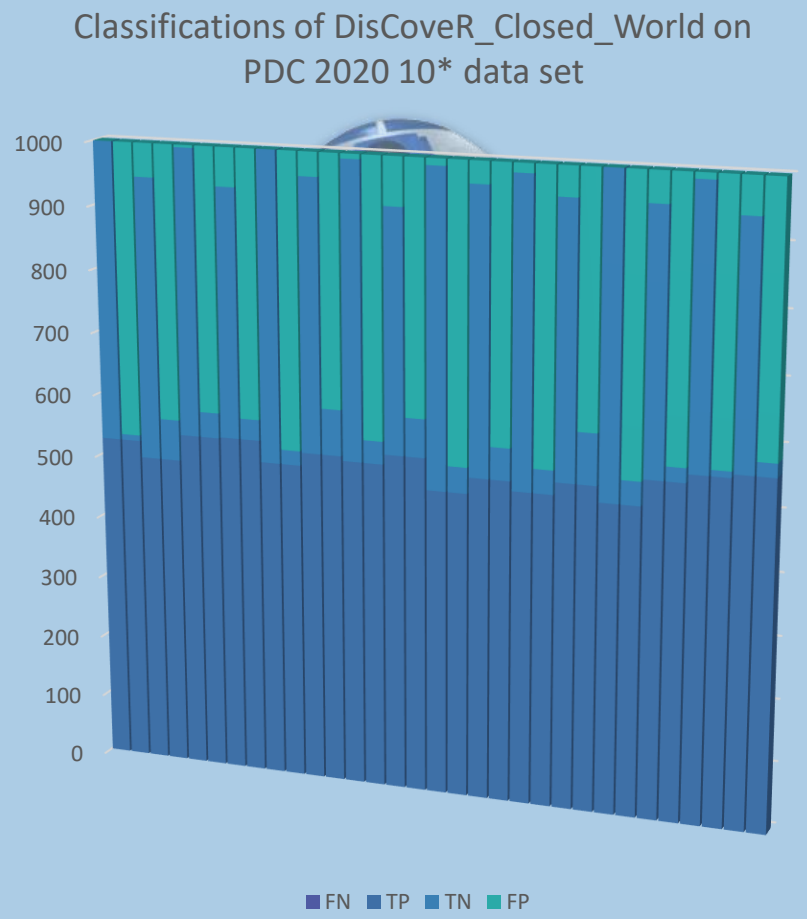
➢ Very good scores on PDC 2016 and PDC 2019 data sets



Automated contest

2. 50% DisCoveR (Open|Closed) World
4. 49% DisCoveR Light (O|C) World
6. 44% KokosMiner2T5
7. 33% InductiveMinerIMfa
8. KokosMiner[2], InductiveMinerIMc

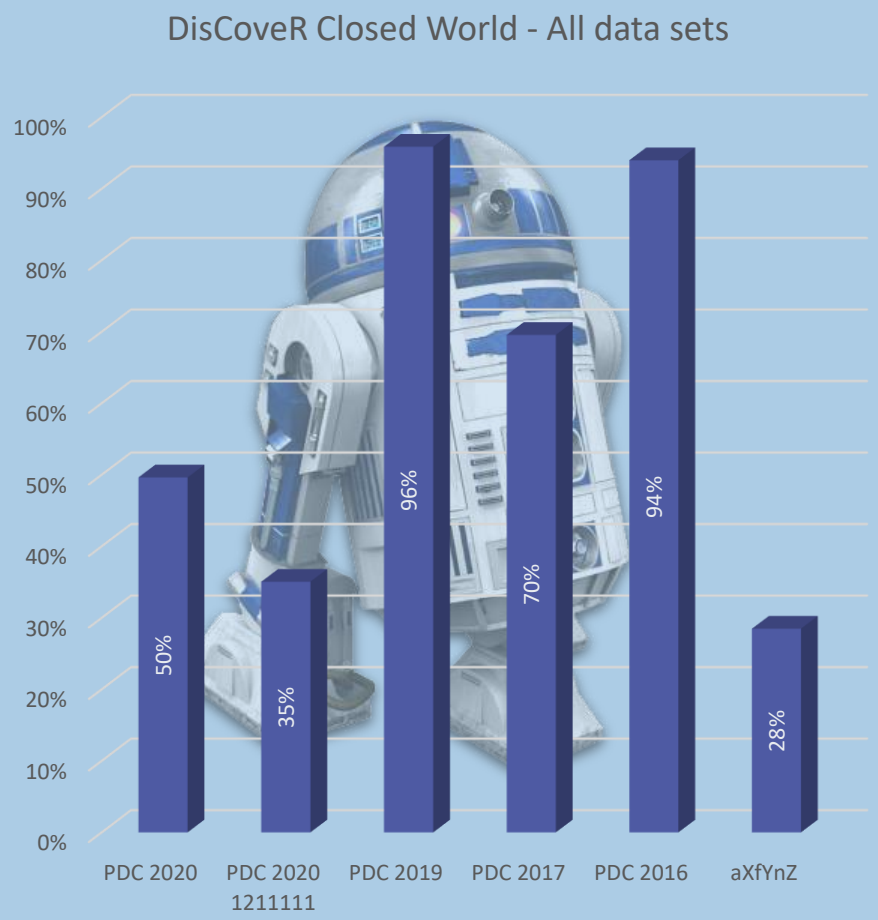
➢ 78% if no noise



Automated contest

2. 50% DisCoveR (Open|Closed) World
4. 49% DisCoveR Light (O|C) World
6. 44% KokosMiner2T5
7. 33% InductiveMinerIMfa
8. KokosMiner[2], InductiveMinerIMc

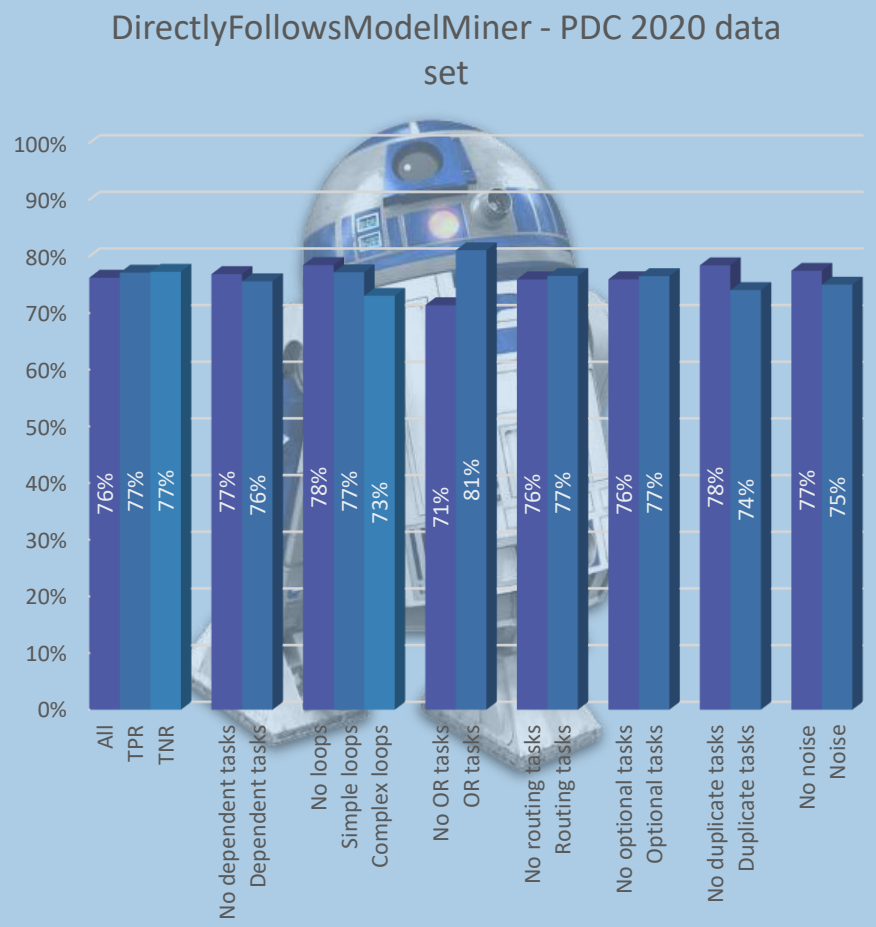
➢ Similar as for DisCoveR Light



Automated contest

2. 50% DisCoveR (Open|Closed) World
4. 49% DisCoveR Light (O|C) World
6. 44% KokosMiner2T5
7. 33% InductiveMinerIMfa
8. KokosMiner[2], InductiveMinerIMc

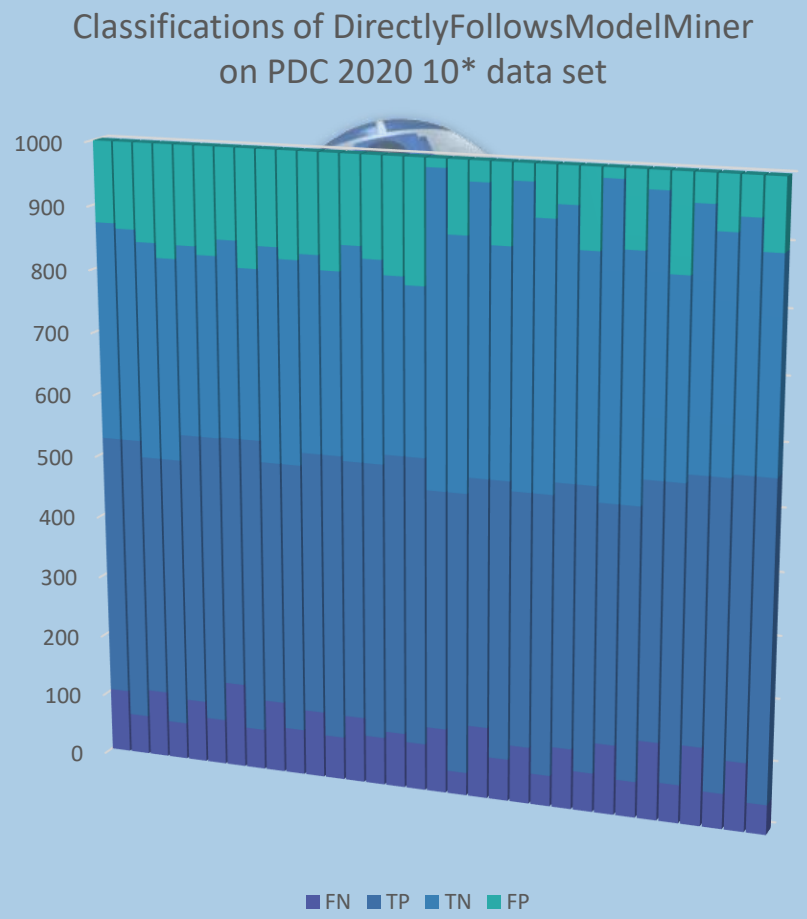
➢ Even better scores for PDC 2016 and PDC 2019 then DisCoveR Light



Automated contest

1. 76% DirectlyFollowsModelMiner
2. 50% DisCoveR (Open|Closed) World
4. 49% DisCoveR Light (O|C) World
6. 44% KokosMiner2T5
7. 33% InductiveMinerIMfa
8. KokosMiner[2], InductiveMinerIMc

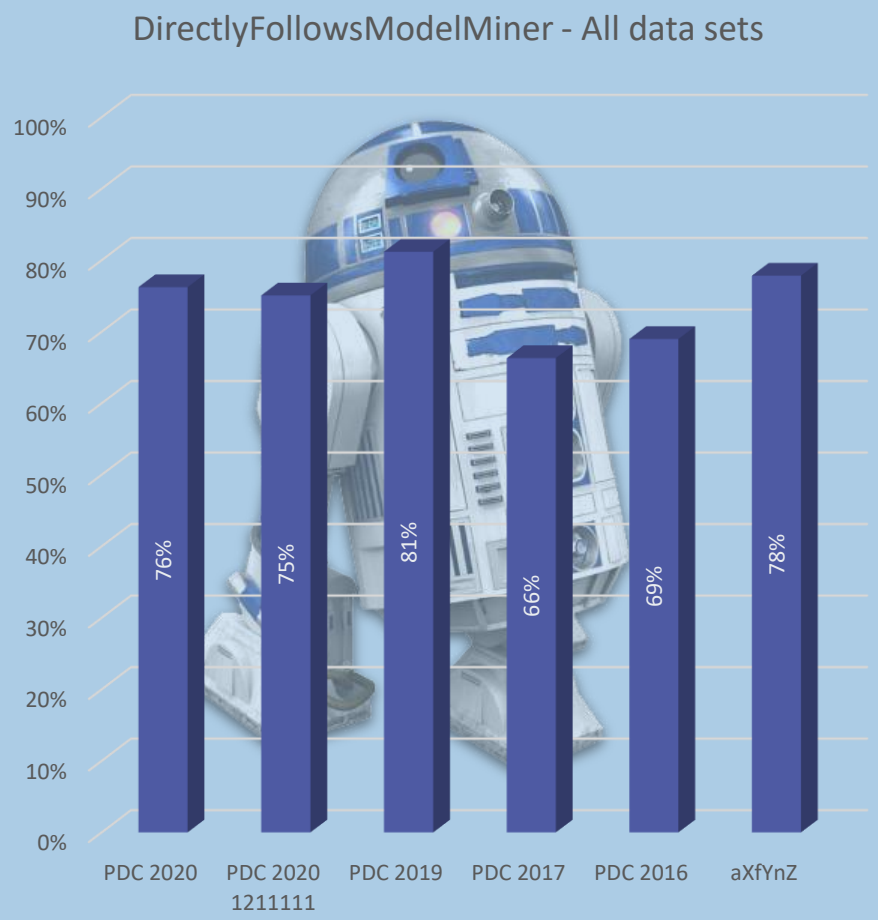
➢ Balanced results (71%-81%)



Automated contest

1. 76% DirectlyFollowsModelMiner
2. 50% DisCoveR (Open|Closed) World
4. 49% DisCoveR Light (O|C) World
6. 44% KokosMiner2T5
7. 33% InductiveMinerIMfa
8. KokosMiner[2], InductiveMinerIMc

➢ Some FNs, some FPs



Automated contest

1. 76% DirectlyFollowsModelMiner
2. 50% DisCoveR (Open|Closed) World
4. 49% DisCoveR Light (O|C) World
6. 44% KokosMiner2T5
7. 33% InductiveMinerIMfa
8. KokosMiner[2], InductiveMinerIMc

➢ Balanced results (66%-81%)

PDC 2020 Winner

SANDER LEEMANS:

DirectlyFollowsModelMiner

Josep Carmona, Benoît Depaire, Eric Verbeek, PDC 2020 organizers

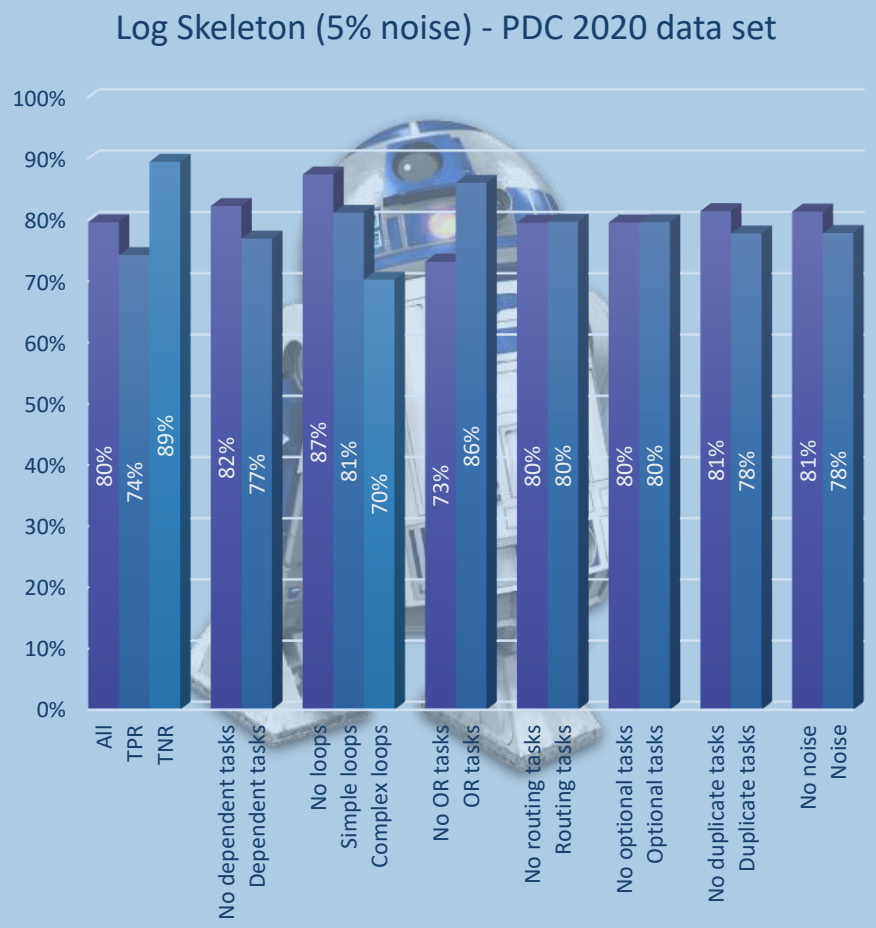
IEEE TASK FORCE ON PROCESS MINING

Backup slides

BEST EXAMPLE DISCOVERY ALGORITHM

Log Skeleton (5% noise)

Josep Carmona, Benoît Depaire, Eric Verbeek, PDC 2020 organizers

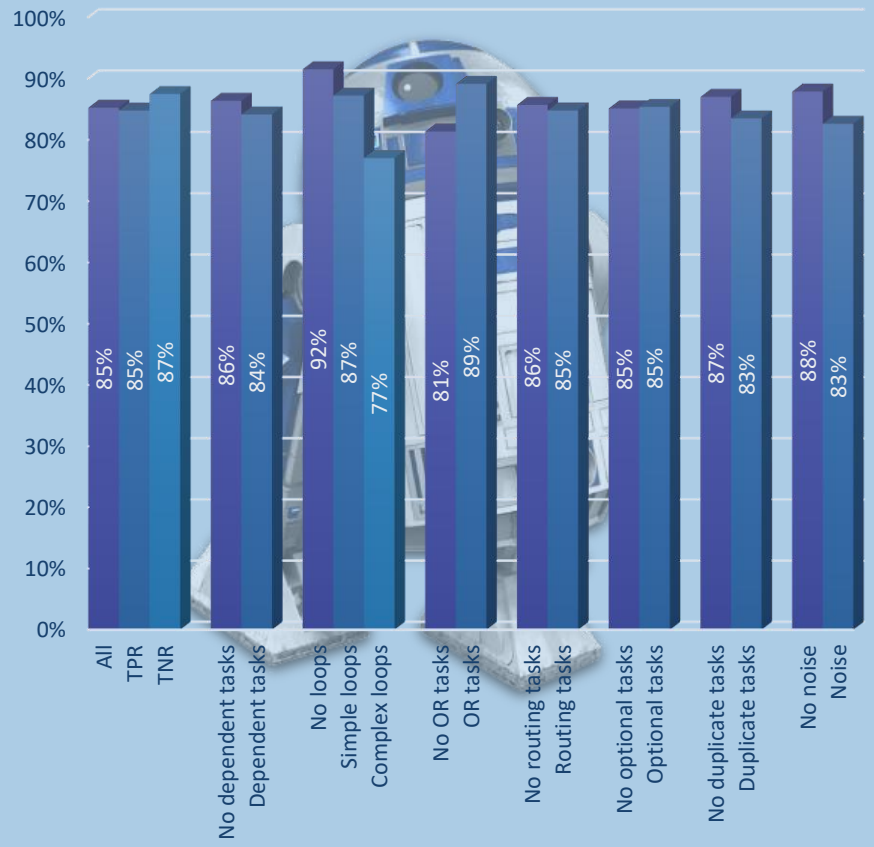


Automated contest

Best example algorithm

- 80% Log Skeleton (5% noise)

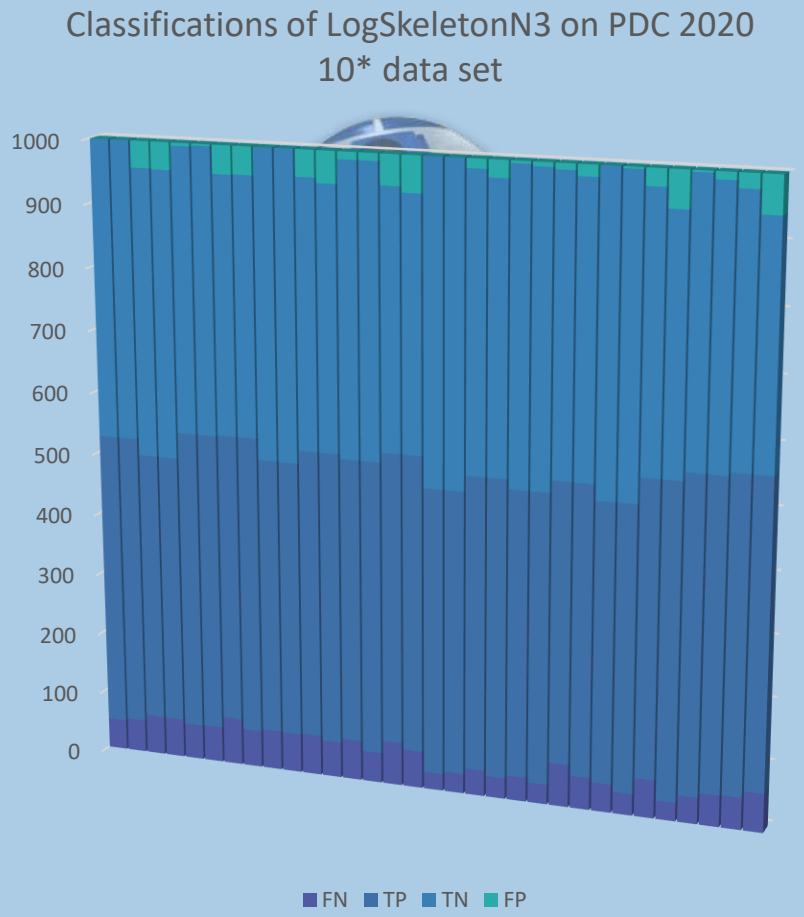
Log Skeleton (3% noise) - PDC 2020 data set



Automated contest

Better algorithm

- 85% Log Skeleton (3% noise)

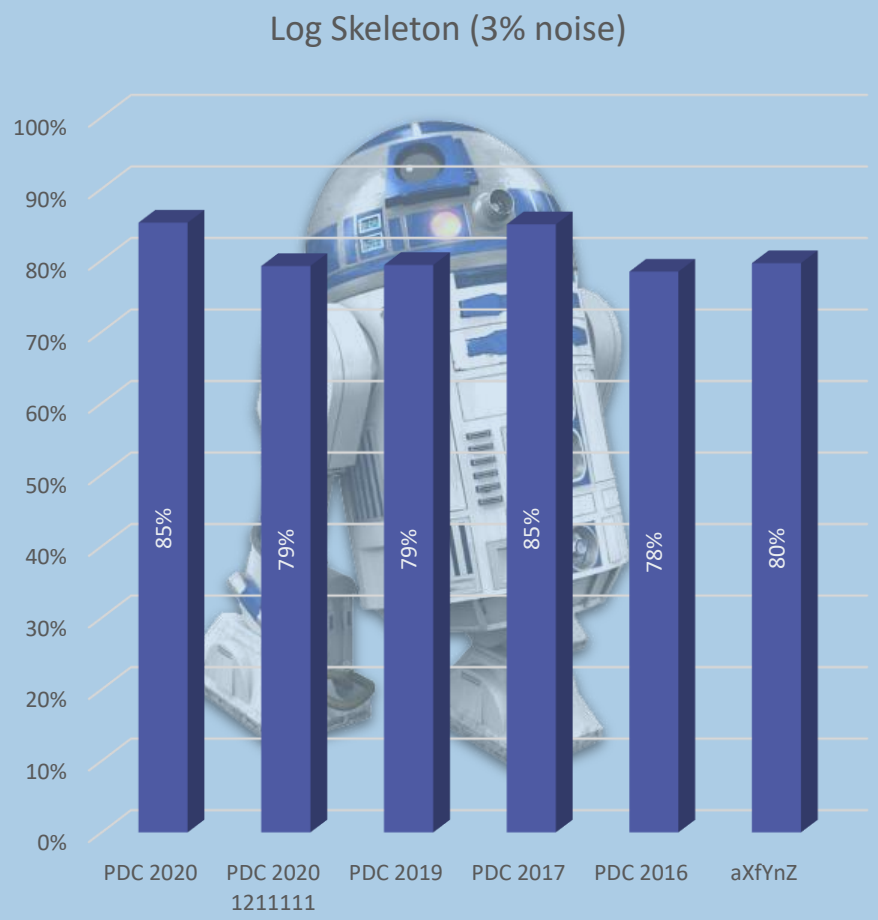


Automated contest

Better algorithm

- 85% Log Skeleton (3% noise)

- Few FNs, few FPs



Automated contest

Better algorithm

- 85% Log Skeleton (3% noise)
- Not better for all data sets