The Benefits of Classification: An Appointment Case Study

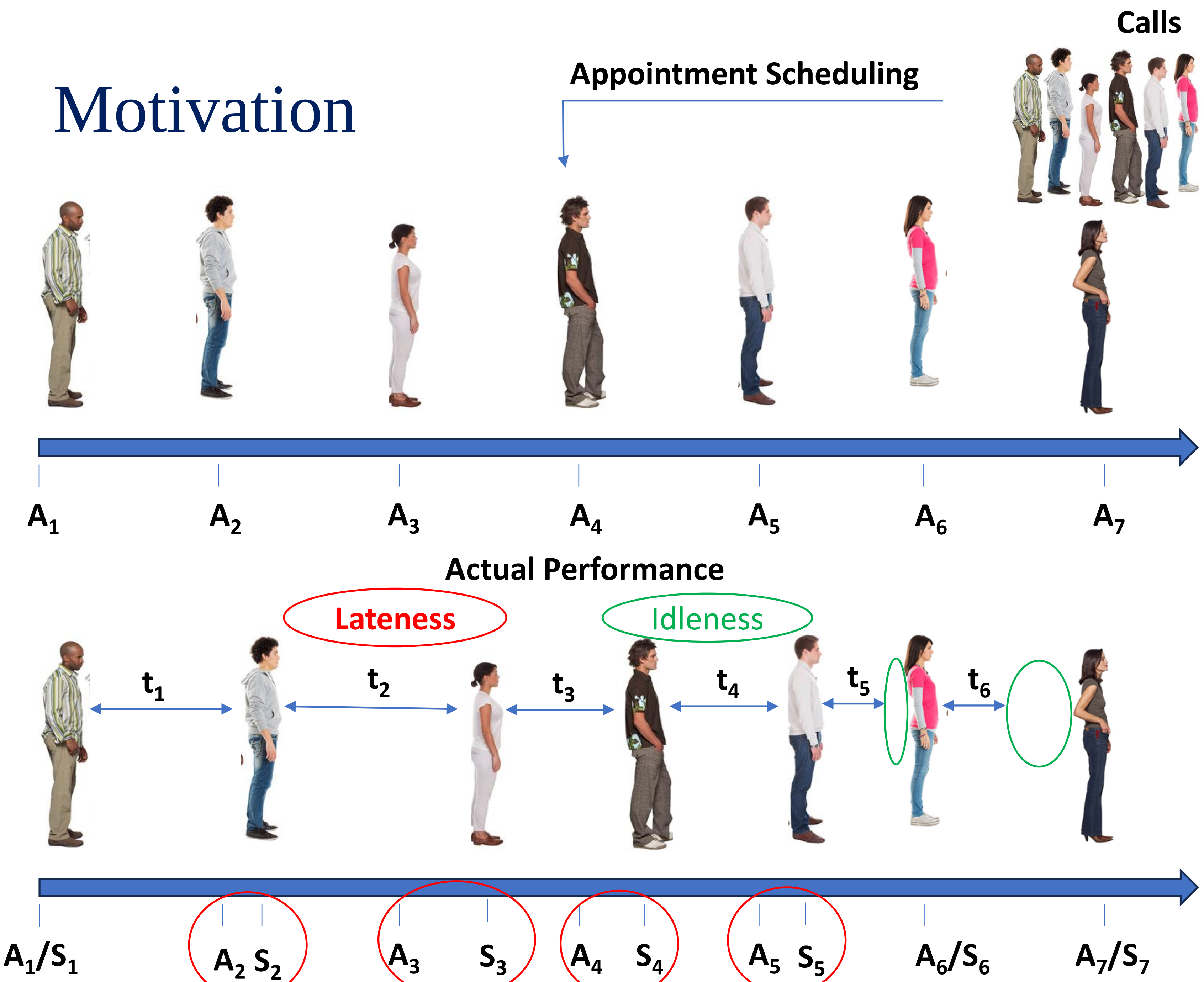
Positron Emission Tomography (PET)

Yariv N. Marmor (PhD)
Boris Shnits (PhD)
Illana Bendavid (PhD)

BRAUDE - College of Engineering,
Karmiel, ISRAEL

Valencia - ENBIS 2023
September, 13

Motivation



Calls
Appointment Scheduling
A_1 A_2 A_3 A_4 A_5 A_6 A_7
Actual Performance
Lateness
Idleness
t_1 t_2 t_3 t_4 t_5 t_6
A_1/S_1 A_2 S_2 A_3 S_3 A_4 S_4 A_5 S_5 A_6/S_6 A_7/S_7

Motivation

- Assumption:
  - Reduce treatment time variance will lead to better (expected) appointments (Kuiper et al., 2019).
  - Data mining (regression) will lead to variance reduction (better prediction) (Golmohammadi et al., 2023).
  - Better appointment will lead to better operational performance. (Christopher et al., 2022)
  - Appointment needs to be simple and easy to use (Savelsbergh & Smilowitz, 2016).
  - The more complex classification the more complex to operate in practice.

Methodology steps

- Classify patients by their service duration in a nested / hierarchal way:
  - “Best” classifier (using CART) – *F* options.
  - Folding tree in steps to get “Less Good” classifiers.
  - End with F classifications for each patient service duration.
- Use appointment algorithm (AA)
  - Start with the "best" classifier (of the previous step)
  - Repeat AA for different "days sequence" (random patients) until the operational outcomes meet an accuracy criterion (e.g., $HW/\mu<0.005$).
  - Repeat last point with the next "best" (not used) classifier
  - Repeat last point until the operational outcomes change (for the worse) statistically from the "best" classification.
- We identify the best practical classification as the one we used last (in the previous step) before we got statistically changes in performance.

Key Performance Indicator (KPI)

- EoD – End of Day

  The expected time the last customer/patient finishes service/treatment.

- Utilization

  The ratio between the sum of services duration done during regular shift hours and the shift length multiplied by number of servers.

- OT – Over Time

  The sum of services duration done after regular shift hours

- IT – Idle Time

  The sum of idle times between services

- OL – Offered Load

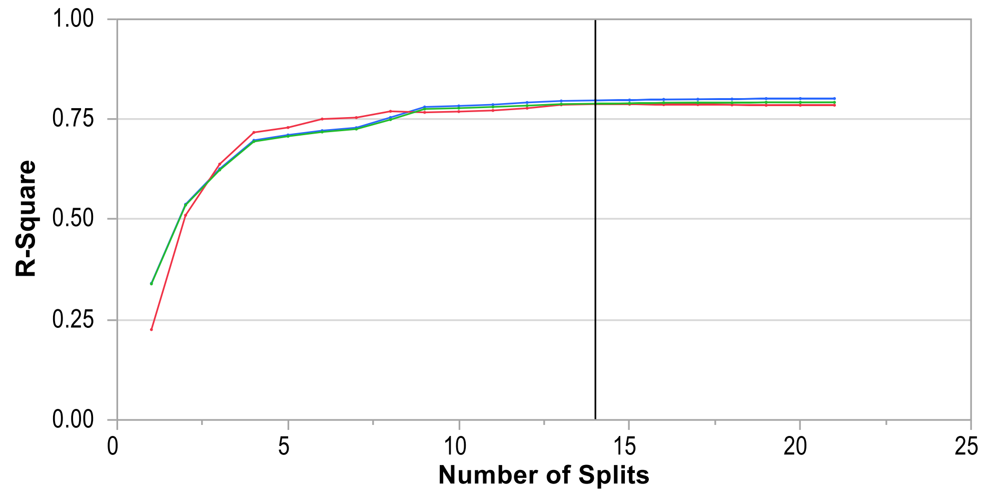
  The sum of expected service duration

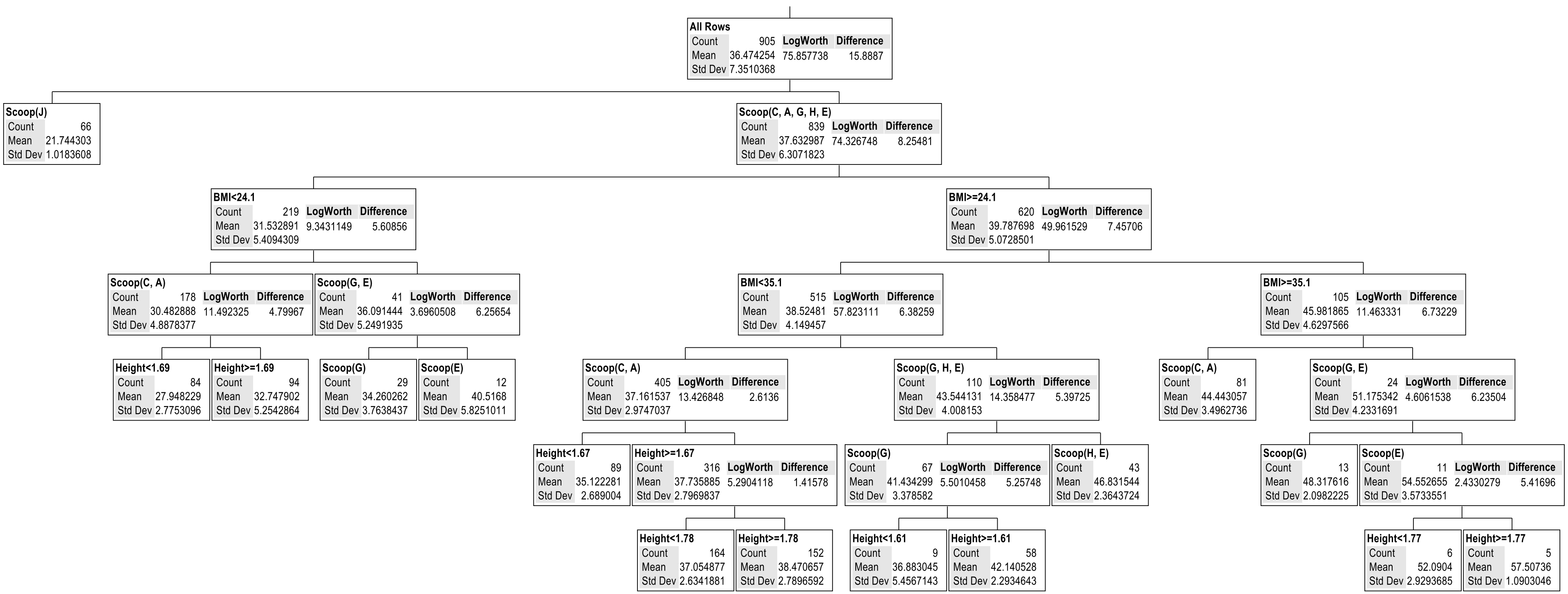
Case Study - Data

- Data – Mayo Clinic (October 2021):
  - 1,168 PET Scans; ~60 per day
  - Patient age (avg=65.7; sd=13.5)
  - Patient gender (36.6% females; 63.4% males)
  - Patient weight (avg=84.4kg; sd=20.9)
  - Patient height (avg=1.72m; sd=0.099
  - Scanner type (2 fast scanners which handle 54% of scans; 2 slow which handle 46% of scans)
  - Patient scan description (area to scan)
  - Scan duration
- Remark:
  - Since we have two type of scanners (faster and slower), we normalized the duration, so it fits the slower scanner times.
  - Appointment scheduling is based on QoS (Quality of Service), meaning the (expected) probability of appointment to start on time.

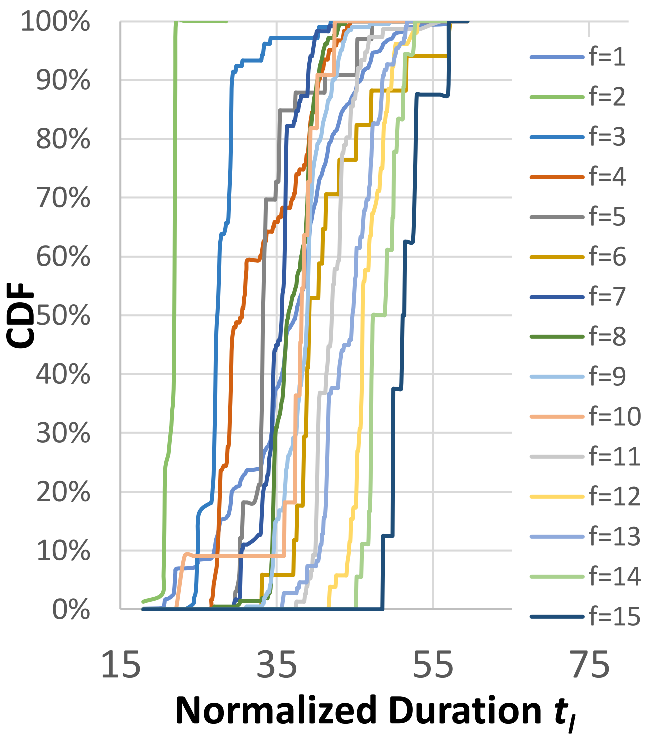
Case Study - CART

- 5-Cross Validation
- 20% for test



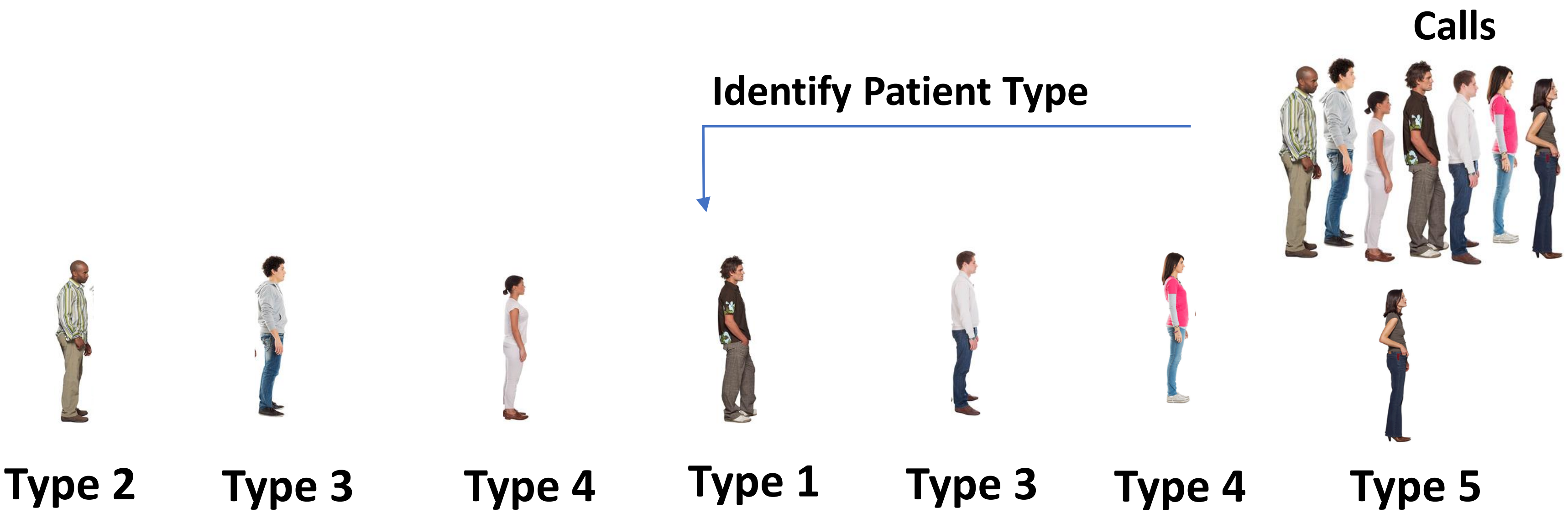


Case Study – Patient Classification by CART "Size"



t_l	Number of Classifications (in color- patient classification number)															l
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	
17.9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
28.5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	81
24.3	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	82
41.9	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	186
26.5	1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	187
44.4	1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	309
29.6	1	2	2	2	2	2	2	2	3	4	4	4	4	4	4	310
47.1	1	2	2	2	2	2	2	2	3	4	4	4	4	4	4	342
33.0	1	2	2	2	2	2	2	2	3	4	4	4	4	4	5	343
57.2	1	2	2	2	2	2	2	2	3	4	4	4	4	4	5	359
29.5	1	2	3	3	3	3	3	3	4	5	5	5	5	5	6	360
42.8	1	2	3	3	3	3	3	3	4	5	5	5	5	5	6	477
26.9	1	2	3	3	3	4	4	4	5	6	6	6	6	6	7	478
43.8	1	2	3	3	3	4	4	4	5	6	6	6	6	6	7	687
31.1	1	2	3	3	3	4	4	4	5	6	6	6	7	7	8	688
49.5	1	2	3	3	3	4	4	4	5	6	6	6	7	7	8	889
23.2	1	2	3	3	4	5	5	5	6	7	7	7	8	8	9	890
42.3	1	2	3	3	4	5	5	5	6	7	7	7	8	8	9	900
37.5	1	2	3	3	4	5	5	5	6	7	7	7	8	9	10	901
55.8	1	2	3	3	4	5	5	5	6	7	7	7	8	9	10	976
41.7	1	2	3	3	4	5	5	6	7	8	8	8	9	10	11	977
53.2	1	2	3	3	4	5	5	6	7	8	8	8	9	10	11	1028
35.7	1	2	3	4	5	6	6	7	8	9	9	9	10	11	12	1029
51.6	1	2	3	4	5	6	6	7	8	9	9	9	10	11	12	1137
45.1	1	2	3	4	5	6	7	8	9	10	10	10	11	12	13	1138
52.6	1	2	3	4	5	6	7	8	9	10	10	10	11	12	13	1155
48.6	1	2	3	4	5	6	7	8	9	10	11	11	12	13	14	1156
57.0	1	2	3	4	5	6	7	8	9	10	11	11	12	13	14	1163
56.8	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1164
59.4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1168

PMSN (Paced Multi Server Numerical-based) Algorithm



PMSN (Paced Multi Server Numerical-based) Algorithm

Calls

Identify Patient Type

Type 2 **Type 3** **Type 4** **Type 2**

47.1
26.5
57.2
24.3
44.4

Randomly get realization (assume 5 classes classifier)

t_l	Number of Classifications (in color- patient classification number)															l
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	
17.9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
28.5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	81
24.3	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	82
41.9	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	186
26.5	1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	187
44.4	1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	309
29.6	1	2	2	2	2	2	2	2	3	4	4	4	4	4	4	310
47.1	1	2	2	2	2	2	2	2	3	4	4	4	4	4	4	342
33.0	1	2	2	2	2	2	2	2	3	4	4	4	4	4	5	343
57.2	1	2	2	2	2	2	2	2	3	4	4	4	4	4	5	359
29.5	1	2	3	3	3	3	3	3	4	5	5	5	5	5	6	360
42.8	1	2	3	3	3	3	3	3	4	5	5	5	5	5	6	477
26.9	1	2	3	3	3	4	4	4	5	6	6	6	6	6	7	478
43.8	1	2	3	3	3	4	4	4	5	6	6	6	6	6	7	687
31.1	1	2	3	3	3	4	4	4	5	6	6	6	7	7	8	688
49.5	1	2	3	3	3	4	4	4	5	6	6	6	7	7	8	889
23.2	1	2	3	3	4	5	5	5	6	7	7	7	8	8	9	890
42.3	1	2	3	3	4	5	5	5	6	7	7	7	8	8	9	900
37.5	1	2	3	3	4	5	5	5	6	7	7	7	8	9	10	901
55.8	1	2	3	3	4	5	5	5	6	7	7	7	8	9	10	976
41.7	1	2	3	3	4	5	5	6	7	8	8	8	9	10	11	977
53.2	1	2	3	3	4	5	5	6	7	8	8	8	9	10	11	1028
35.7	1	2	3	4	5	6	6	7	8	9	9	9	10	11	12	1029
51.6	1	2	3	4	5	6	6	7	8	9	9	9	10	11	12	1137
45.1	1	2	3	4	5	6	7	8	9	10	10	10	11	12	13	1138
52.6	1	2	3	4	5	6	7	8	9	10	10	10	11	12	13	1155
48.6	1	2	3	4	5	6	7	8	9	10	11	11	12	13	14	1156
57.0	1	2	3	4	5	6	7	8	9	10	11	11	12	13	14	1163
56.8	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1164
59.4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1168

PMSN (Paced Multi Server Numerical-based) Algorithm

Considering fast pace (1.38)

Duration

Realiation	P1	P2	P3	P4	P5	P6	P7
	Type 2	Type 3	Type 4	Type 1	Type 3	Type 4	Type 5
1	47.1	36.0	47.2	20.7	35.9	52.4	37.6
2	26.5	34.4	46.1	21.6	38.8	45.6	46.8
3	57.2	34.6	45.8	21.1	40.8	41.3	43.6
4	24.3	40.0	46.0	21.9	34.7	45.8	47.0
5	44.4	37.4	43.1	22.1	39.6	38.8	47.1

Finish Time

P1	P2	P3	P4	First to Finish (macine)
m1(s)	m2(s)	m3(f)	m4(f)	
47.1	36.0	34.2	15.0	(m4) 15.0
26.5	34.4	33.4	15.7	(m4) 15.7
57.2	34.6	33.2	15.3	(m4) 15.3
24.3	40.0	33.3	15.9	(m4) 15.9
44.4	37.4	31.2	16.0	(m4) 16.0

=> A5=15.9 (80%)

15.9+35.9/1.38=41.9

Finish Time

m1(s)	m2(s)	m3(f)	m4(f)	First to Finish (macine)
47.1	36	34.2	41.9	(m3) 34.2
26.5	34.4	33.4	44.0	(m1) 26.5
57.2	34.6	33.2	45.5	(m3) 33.2
24.3	40	33.3	41.0	(m1) 24.3
44.4	37.4	31.2	44.6	(m3) 31.2

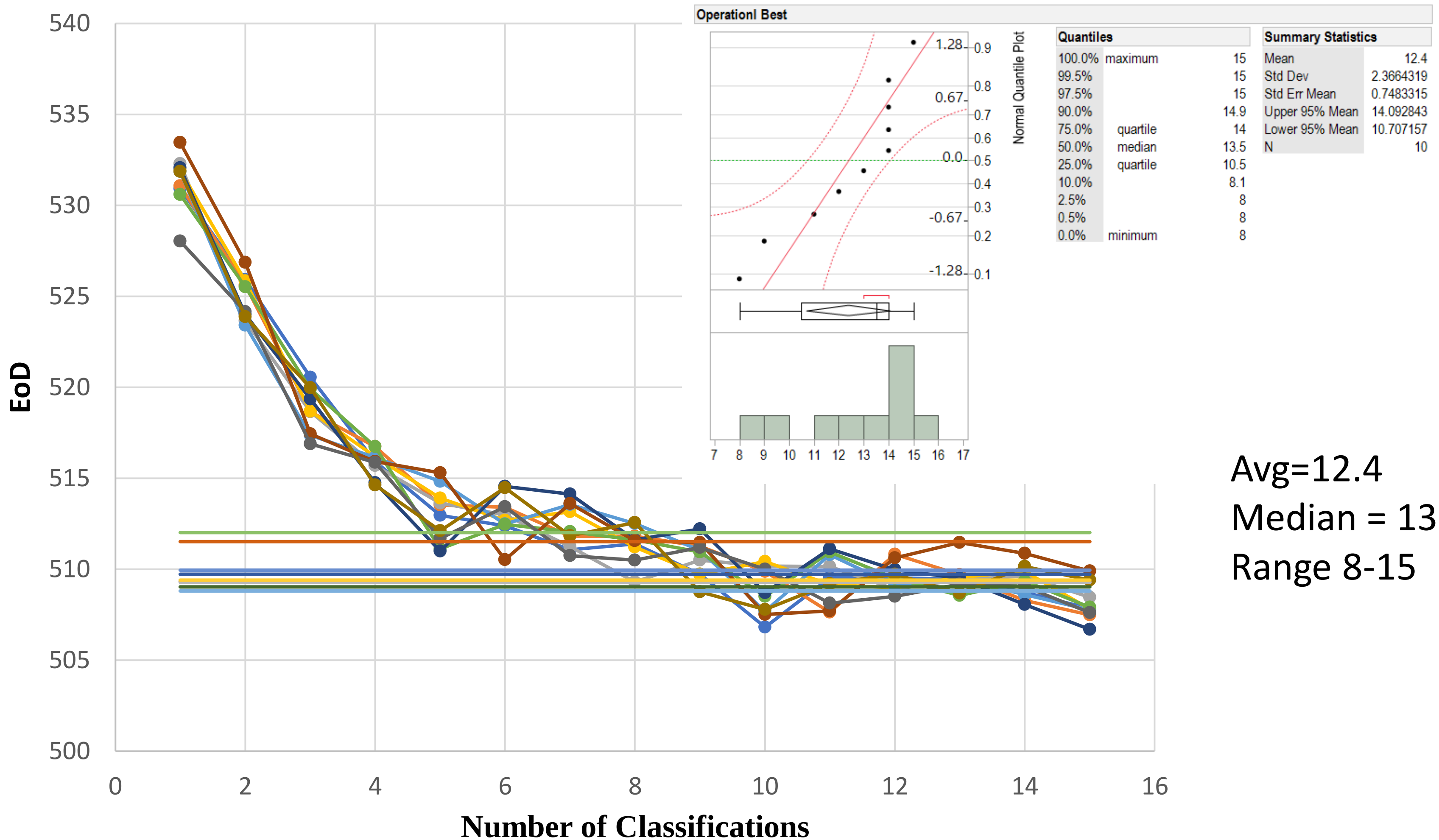
=> A6=33.2

Finish Time

m1(s)	m2(s)	m3(f)	m4(f)	First to Finish (macine)
47.1	36.0	72.2	41.9	(m2) 36.0
78.8	34.4	33.4	44.0	(m3) 33.4
57.2	34.6	63.1	45.5	(m2) 34.6
79.0	40.0	33.3	41.0	(m3) 33.3
44.4	37.4	61.3	44.6	(m2) 37.4

=> A1=A2=A3=A4=0; A5=15.9; A6=33.2; A7=36

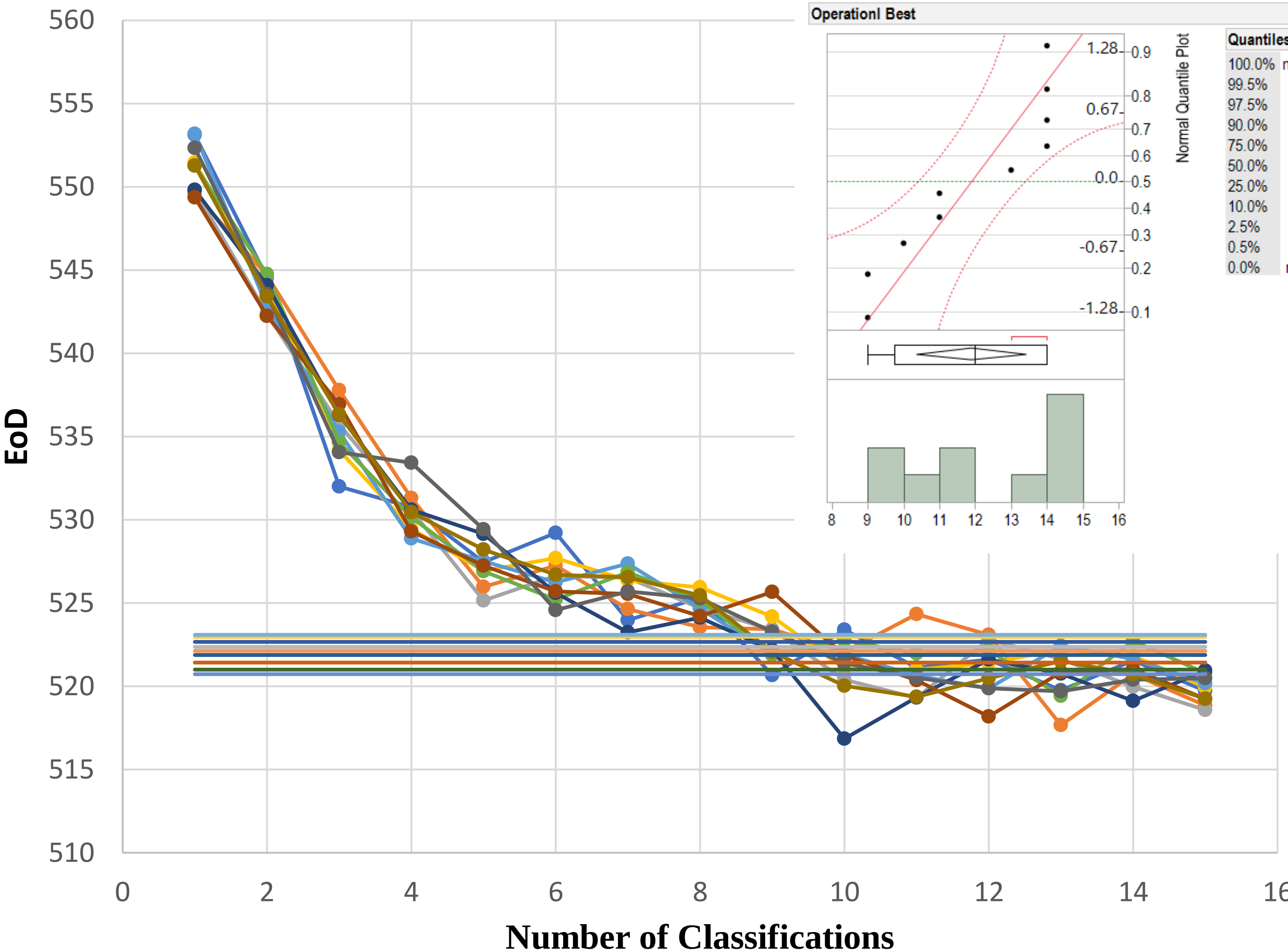
Case Study – Results (Operational Best)



Avg=12.4
Median = 13.5
Range 8-15

Line with markings– the expected EoD; Clear line – threshold to stop. QoS=0.7

Case Study – Results (Operational Best)



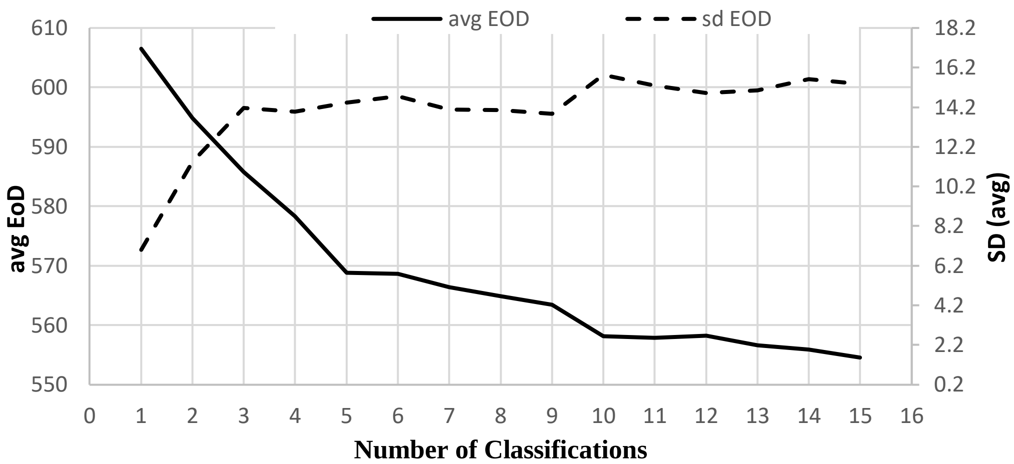
Quantiles		
100.0%	maximum	14
99.5%		14
97.5%		14
90.0%		14
75.0%	quartile	14
50.0%	median	12
25.0%	quartile	9.75
10.0%		9
2.5%		9
0.5%		9
0.0%	minimum	9

Summary Statistics	
Mean	11.9
Std Dev	2.1317703
Std Err Mean	0.6741249
Upper 95% Mean	13.424977
Lower 95% Mean	10.375023
N	10

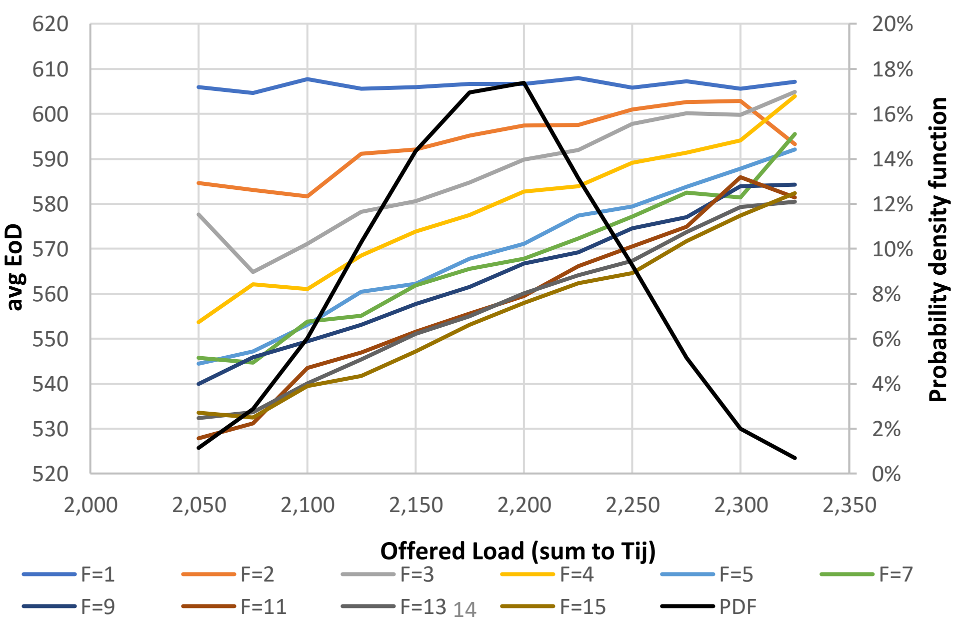
Avg=11.9
Median = 12
Range 9-14

Line with markings– the expected EoD; Clear line – threshold to stop. QoS=0.8

Case Study – Results (EoD) for QoS=0.95



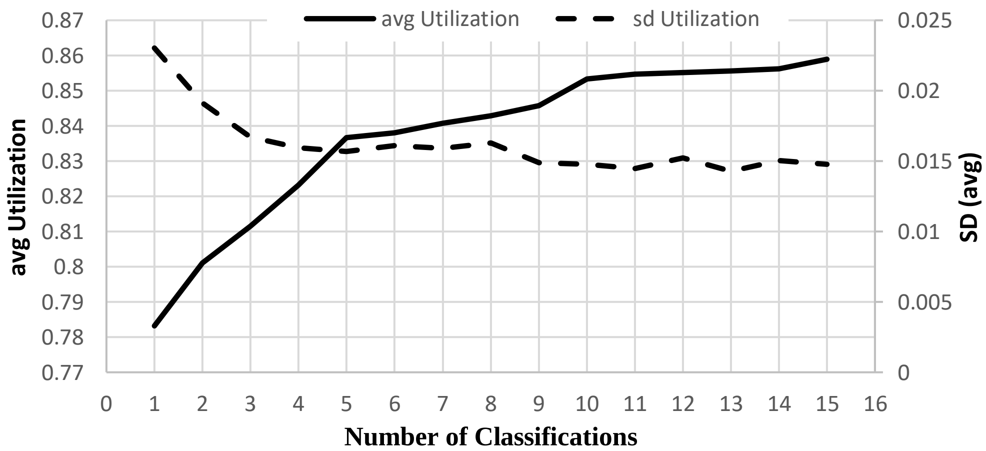
With the effect of Offered Load



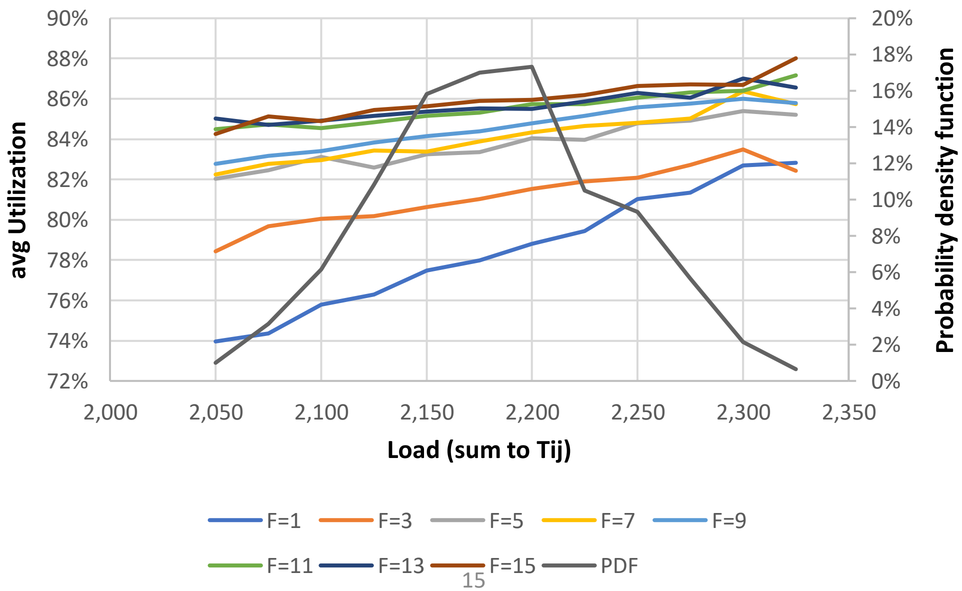
Findings:

- EoD **increases** with the distance to the best classifier.
- The EoD is **less robust** with Offered Load when closer to best classifier.

Case Study – Results (Utilization) for QoS=0.95



With the effect of Offered Load



Findings:

- Utilization **decreases** with the distance to the best classifier.
- The Utilization is **more robust** with Offered Load when closer to best classifier.

Conclusion and Future Research

- We introduce an algorithm for finding the best practical classifier that is easier to implement.
- Looking for the best classifier does not always translate into the best performance in practice.
- The complexity is needed in lower QoS and less needed in a high QoS system.
- For some KPIs, the best practice classifier is robust (i.e., Utilization), but sometimes it is the least robust option (i.e., EoD).
- Future research questions:
  - Can we get the same results with repeated CART algorithm?
  - Is there a connection between R-Square and the effect on our algorithm (big changes result in big effect)?

16

Questions / remarks?

E-mail: myariv@braude.ac.il