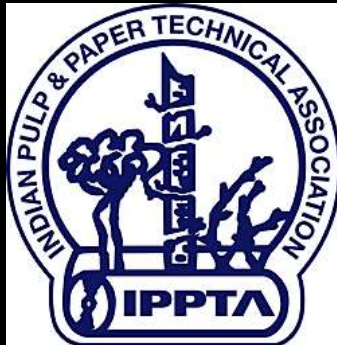


DATA MINING APPLICATIONS IN PULP & PAPER INDUSTRY



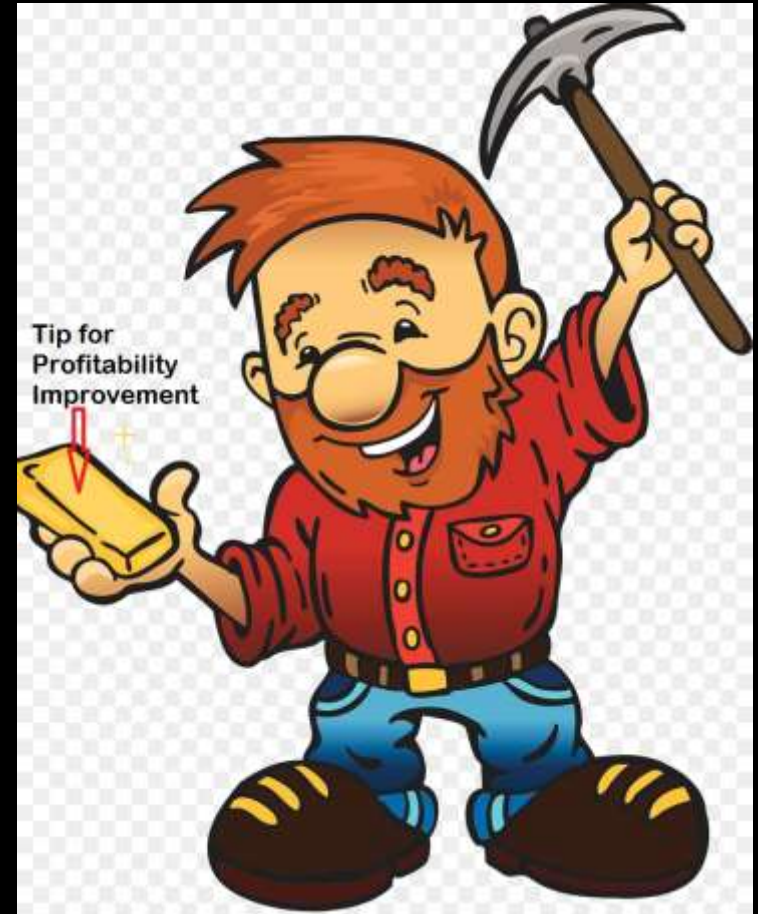
IPPTA Zonal Seminar,
Muzaffarnagar
August 17-18, 2018



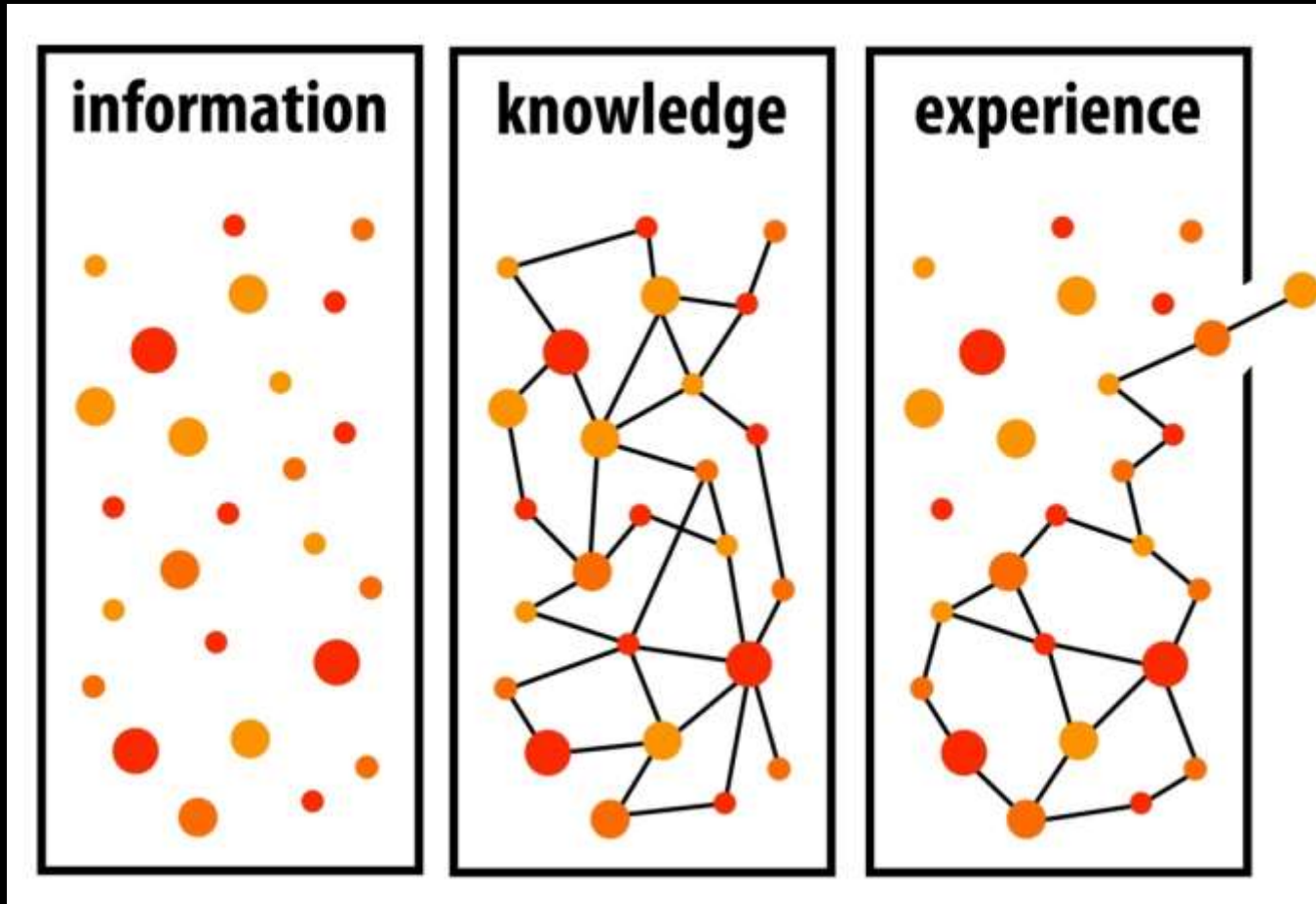
D K Singhal
Chandpur Enterprises Ltd.
Chandpur.

What is Data Mining

- In simple words, data mining is defined as a process used to extract usable data from a lot of any raw data. It implies analyzing data patterns in large batches of data manually or using any software.




Data Mining





First Pass Retention

- Measure Head Box Consistency
 - Measure Back Water Consistency
 - Calculate
-
- Variations due to measurement errors
 - Variations due to process parameters
 - Variations due to basis weight of paper
 - Variations due to quality of paper
- 

First Pass Retention

- The mill was getting FPR nearly 75%.
- The production team was quite confident that FPR is maintained same since long.
- No reported change in raw materials, operating change in parameters etc.
- Sometimes, FPR variations did appear.
- Every shift, FPR was being measured and recorded (first data only).

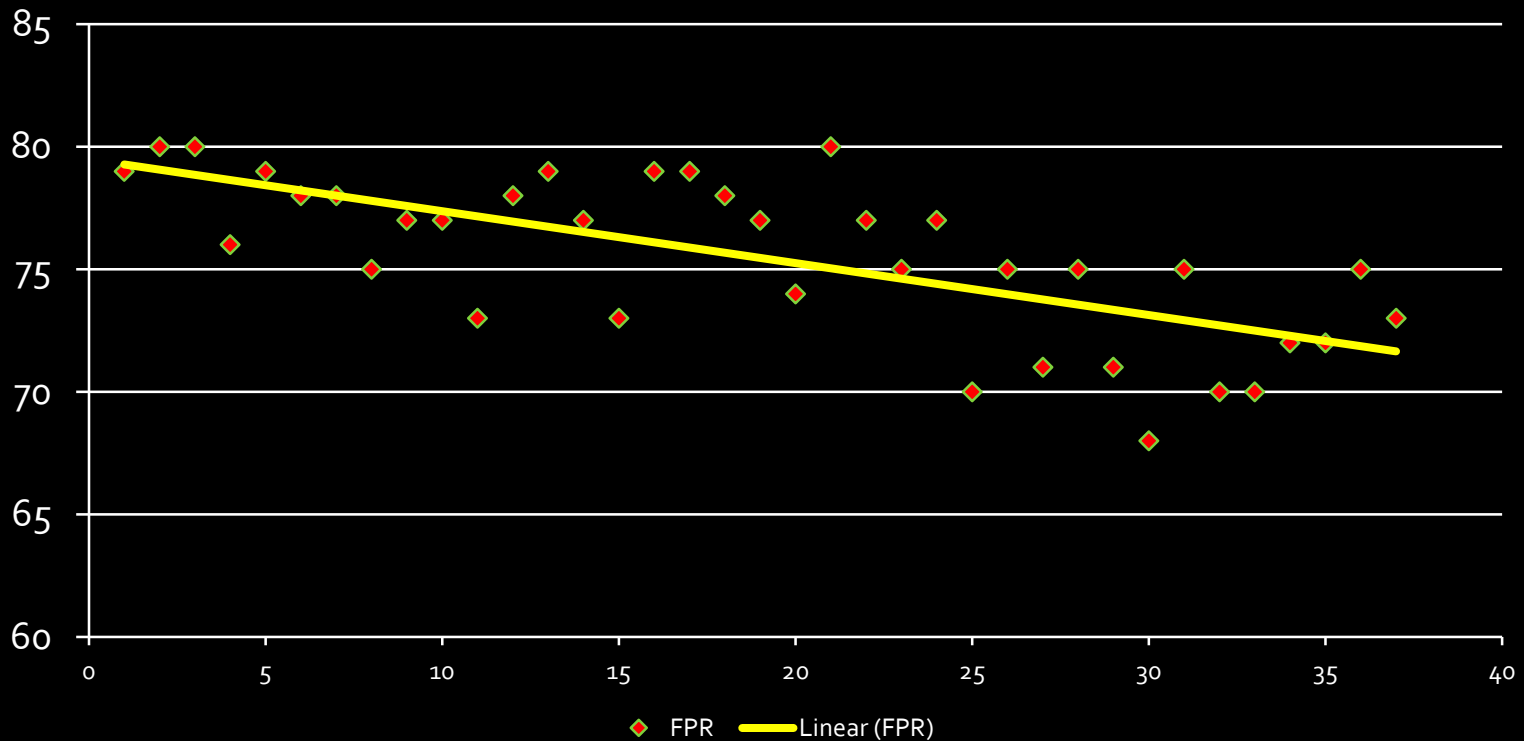


Data Mining for FPR

- Various qualities
 - Wide basis weight range
 - Other causes like retention aid dosage etc.
-
- Data filtering for quality and basis weight
 - Study of 6 months recent past

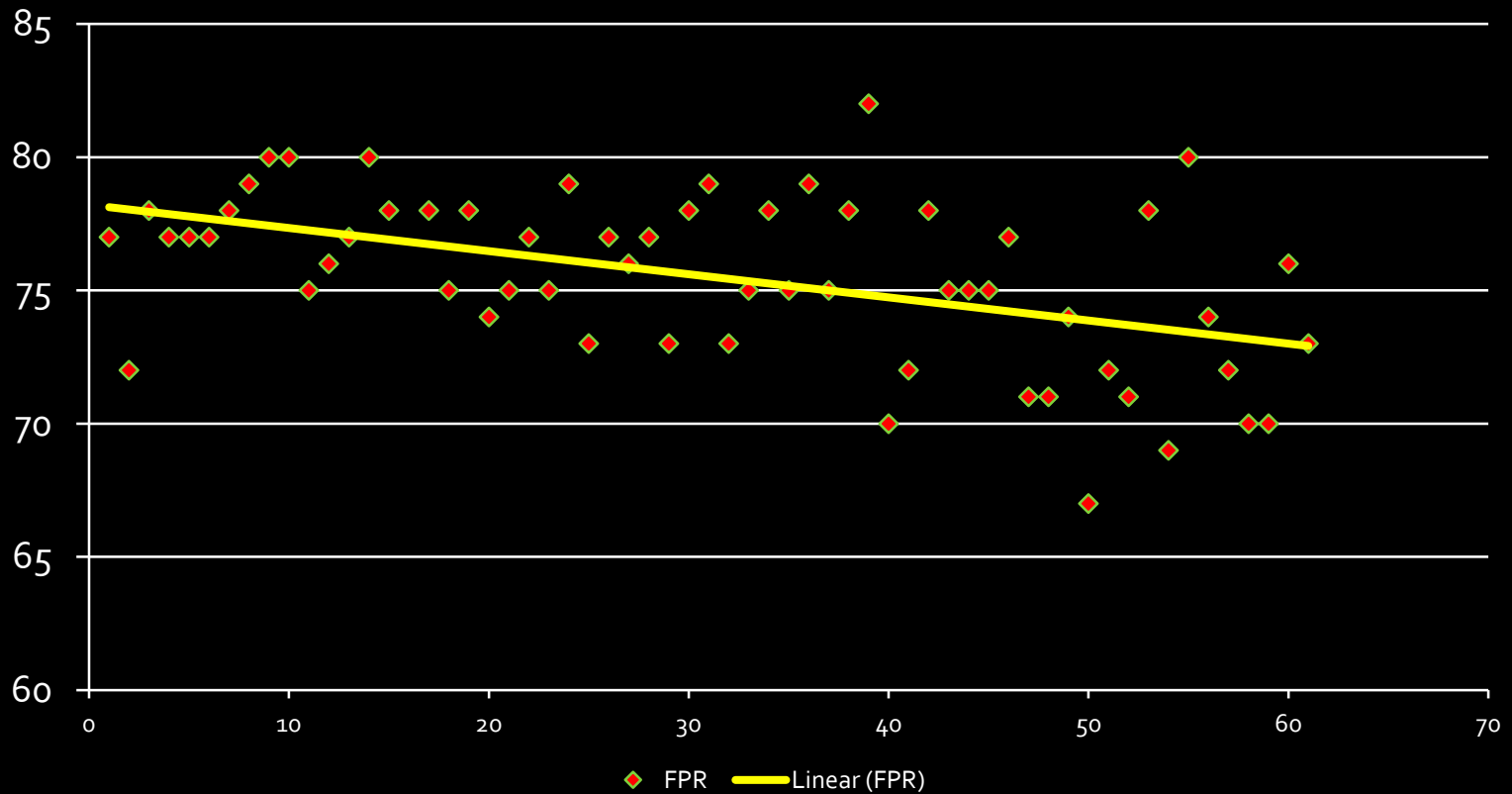
First Pass Retention

FPR for 45gsm
(Duration of Study : 6 Months)



First Pass Retention

FPR for 42gsm
(Duration of Study: 6 Months)





About Supercalender Rolls

- Supercalender cotton rolls are made up of cotton fibers mounted on a solid shaft.
- During operations, these rolls were getting damaged too early, and requiring recoating, causing significant downtime and cost.

Supercalender Roll



Supercalender Rolls Problem

- Mill's first supercalender.
- Inadequate supplier support to investigate the problem.
- Mill had experienced staff, but the cause of problem or possible solution was not visible.
- It was tried to mine the data (records) of such failures.

Preliminary Investigation

- Invoice Date
 - Mounting Position
 - Original Vs. New Shaft
-
- **Result: No Clue!**

Further Mining

- Were failures more frequent with-
 - Specific Day, say Saturday? (Astrology)
 - Particular Shift? (Time of day)
 - Specific Operator? (Negligence)
 - Specific Paper Grade or GSM? (Paper properties)

- **Result: Again No Clue.**

Clue Appears

- Most failures (around 90%) occurred within two hours of startup of plant when the plant was previously shut for more than two days.
- Was there something wrong with startup?
- **Result: Problem Solved!**

Basis Weight Profile

- Paper machine team of a mill indicated that they were not able to get GSM control beyond a particular accuracy.
- What was behind that 'particular accuracy'?
 - The machine operational team, or
 - The machine itself?



Management Questions

- Problem Nature: Temporary / Permanent?
- Probable cause of it!
- Any solutions?
- Who can solve the problem?



Constraints

- Wide basis weight range
- Changes in machine speed
- Small or big orders
- Different furnishes
- Several possible causes
 - Like head box, wire, press, dryer etc.



Action Line

- Compile the data by taking average of all rolls made every fortnight.
- Subtract average of each profile from individual value to get absolute profile.
- Colour coding to get quicker understanding.

Fortnightly Average Profile

	1	2	3	4	5	6	7	8	9	10	11	12	Var.	Avg.
JAN_1	45.3	45.5	45.5	45.5	45.2	45.2	45.4	45.3	45.2	45.5	45.4	45.3	0.3	45.4
JAN_2	43.0	43.3	43.0	43.2	43.2	43.2	43.2	43.2	43.2	43.3	43.3	43.0	0.3	43.2
FEB_1	42.0	43.0	43.0	43.1	42.8	43.1	43.2	42.1	42.8	42.9	42.8	42.7	1.2	42.8
FEB_2	41.2	41.9	41.8	41.8	42.0	41.6	42.0	41.8	41.8	41.5	41.8	41.7	0.8	41.7
MAR_1	49.0	50.0	50.1	50.2	50.0	50.1	49.9	50.2	50.0	49.9	50.1	49.8	1.2	49.9
MAR_2	45.8	46.6	46.3	46.5	46.4	46.5	46.7	46.6	46.7	46.4	46.5	46.3	0.9	46.4
APR_1	50.5	51.0	50.9	51.0	51.0	50.9	51.0	51.1	50.9	50.8	50.7	50.6	0.6	50.9
APR_2	49.8	50.4	50.7	50.5	50.3	50.2	50.3	50.7	50.5	50.5	50.2	50.1	0.9	50.4
MAY_1	49.8	50.4	50.7	50.5	50.3	50.2	50.3	50.7	50.5	50.5	50.2	50.1	0.9	50.4
MAY_2	43.7	44.9	44.5	44.5	44.3	44.5	44.5	44.5	44.5	44.2	44.4	44.1	1.2	44.4
JUN_1	46.4	46.6	46.6	46.3	46.5	46.7	46.3	46.3	46.4	46.3	46.3	46.3	0.4	46.4
JUN_2	38.5	38.2	38.4	38.2	38.1	38.4	38.4	38.4	38.2	38.4	38.2	38.3	0.4	38.3

Absolute Profile

	1	2	3	4	5	6	7	8	9	10	11	12
JAN-1-15	-0.1	0.1	0.1	0.1	-0.2	-0.2	0.0	-0.1	-0.2	0.1	0.0	-0.1
JAN-16-31	-0.2	0.1	-0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	-0.2
FEB-1-15	-0.8	0.2	0.2	0.3	0.0	0.3	0.4	-0.7	0.0	0.1	0.0	-0.1
FEB-16-28	-0.5	0.2	0.1	0.1	0.3	-0.1	0.3	0.1	0.1	-0.2	0.1	0.0
MAR-01-15	-0.9	0.1	0.2	0.3	0.1	0.2	0.0	0.3	0.1	0.0	0.2	-0.1
MAR-16-31	-0.6	0.2	-0.1	0.1	0.0	0.1	0.3	0.2	0.3	0.0	0.1	-0.1
APR-01-15	-0.4	0.1	0.0	0.1	0.1	0.0	0.1	0.2	0.0	-0.1	-0.2	-0.3
APR-16-30	-0.6	0.0	0.4	0.1	-0.1	-0.1	-0.1	0.4	0.1	0.1	-0.1	-0.3
MAY-01-15	-0.6	0.0	0.4	0.1	-0.1	-0.1	-0.1	0.4	0.1	0.1	-0.1	-0.3
MAY-16-31	-0.7	0.5	0.1	0.1	-0.1	0.1	0.1	0.1	0.1	-0.2	0.0	-0.3
JUN-01-15	0.0	0.2	0.2	-0.1	0.1	0.3	-0.1	-0.1	0.0	-0.1	-0.1	-0.1
JUN-16-30	0.2	-0.1	0.1	-0.1	-0.2	0.1	0.1	0.1	-0.1	0.1	-0.1	0.0



Absolute Profile

- The absolute profile indicated that the major problem lies with position 1, where basis weight is generally on the lower side.
- The problem was corrected; data shown in next slide.

Absolute Profile

	1	2	3	4	5	6	7	8	9	10	11	12
Jul_1	0.2	0.0	0.1	-0.1	-0.1	0.0	0.0	0.2	0.0	0.1	-0.2	0.2
Jul_2	0.0	0.0	-0.2	-0.2	-0.1	0.1	0.3	-0.2	0.0	-0.2	0.6	0.0
Aug_1	-0.1	0.0	0.2	0.0	-0.1	0.0	0.0	0.1	-0.2	0.0	0.0	-0.1
Aug_2	-0.1	-0.1	0.1	-0.1	0.0	0.2	0.0	0.0	-0.1	0.2	0.0	0.0
Sep_1	-0.2	-0.1	0.2	0.2	0.3	0.3	0.0	-0.6	0.0	0.0	-0.3	-0.2
Sep_2	-0.2	-0.1	0.0	0.0	0.0	0.1	-0.1	0.1	0.1	0.2	0.1	-0.2
Oct_1	-0.1	-0.2	0.1	0.1	-0.1	0.2	0.0	0.0	0.2	-0.1	0.5	-0.2
Oct_2	-0.1	0.3	-0.1	0.0	-0.1	0.2	0.0	0.0	0.0	-0.2	0.0	-0.1
Nov_1	-0.1	0.1	0.1	-0.1	0.1	0.0	-0.1	0.2	0.0	0.1	0.1	0.0
Nov_2	0.0	0.1	0.1	0.1	-0.2	0.2	0.0	0.1	0.1	0.0	0.1	-0.2
Dec_1	-0.2	0.0	-0.1	0.0	0.2	0.0	-0.2	0.1	0.0	0.0	0.0	-0.2
Dec_2	-0.2	0.1	-0.1	0.1	0.1	0.2	-0.1	0.1	0.3	0.0	0.0	-0.1

Online Calibration Validation

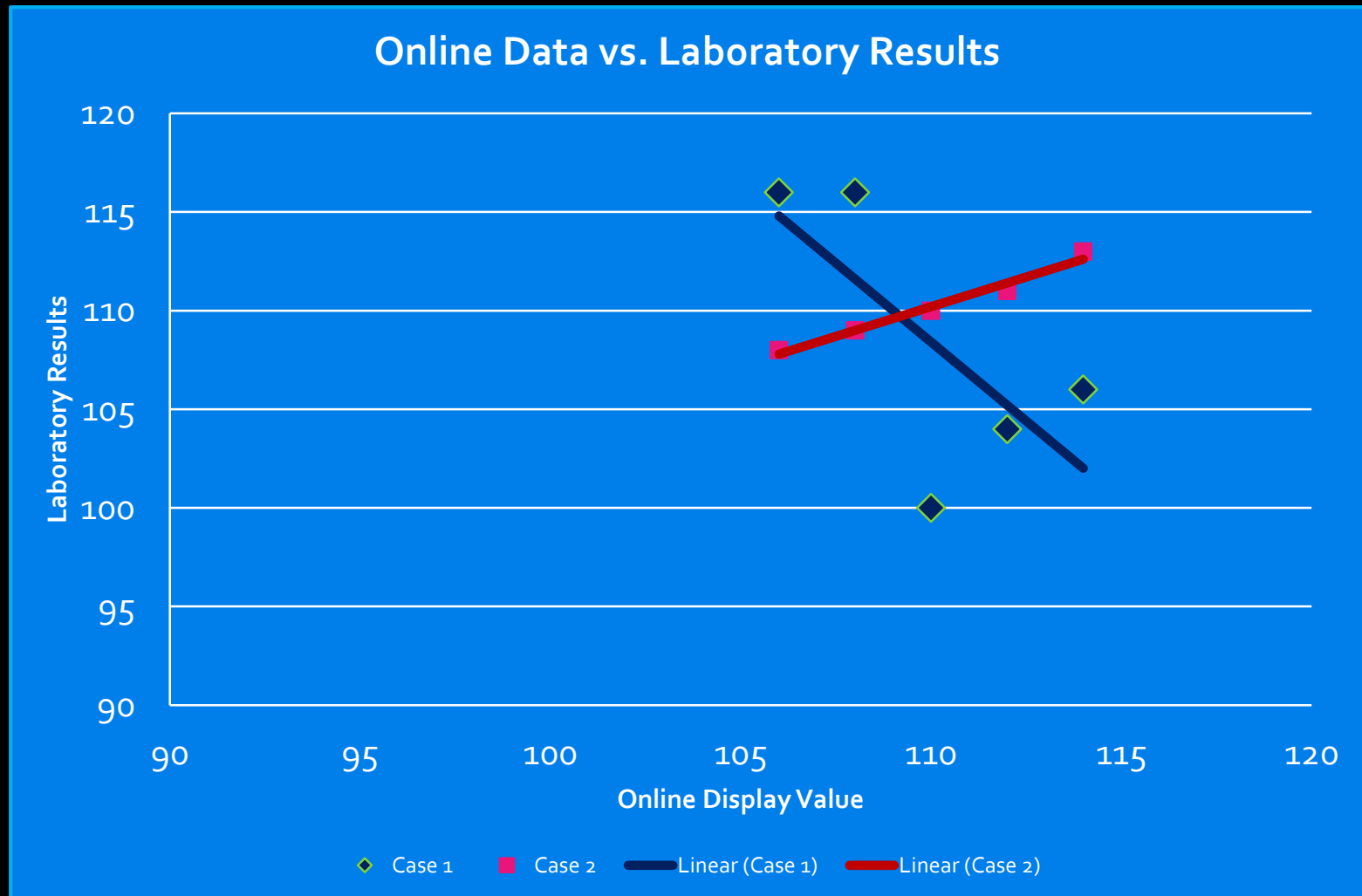
- The basic problem is that we need to calibrate an equipment by the methods which do give certain error.
- IS:3025-44 for BOD-
- 200 BOD sample with 200 ± 37 is OK.
- Furthermore, time required for testing BOD, COD is significantly high.

Calibration Validation

- Some of the laboratories take data for five consecutive days, and make a comparison.
- On the basis of that, they certify that the data displayed by the online matches with the laboratory data; hence the system calibration is OK. Let's have a look at some data-

COD OCEMS Display Value	106	108	110	112	114
Case 1	116	116	100	104	106
Case 2	108	109	110	111	113

Graphical Representation



Approach Used

- Collect data for several days.
- Use Microsoft Excel to get a linear trendline.
- Get equation of trendline in the form-

$$y = m x + c$$

- Where, **m** is the slope and **c** is the intercept on the x axis. Ideally, **m**=One; **c**=Zero.

Sample Data

Date	BOD				COD				TSS		
	Disp	Lab	% Err		Disp	Lab	%Err		Disp	Lab	%Err
1	11	10	9.1		94	92	2.1		15	10	33.3
2	10	9	10.0		85	80	5.9		14	10	28.6
3	10	11	10.0		86	92	7.0		14	10	28.6
4	11	12	9.1		98	116	18.4		16	20	25.0
5	12	13	8.3		107	120	12.1		18	20	11.1
6	14	12	14.3		118	104	11.9		19	30	57.9
7	13	13	0.0		110	116	5.5		18	20	11.1
8	13	10	23.1		109	92	15.6		18	20	11.1
9	12	10	16.7		107	108	0.9		18	30	66.7
10	10	10	0.0		90	80	11.1		15	10	33.3
11	11	13	18.2		96	116	20.8		16	10	37.5
12	10	12	20.0		99	120	21.2		22	20	9.1
13	10	11	10.0		98	116	18.4		16	20	25.0
14	9	9	0.0		87	96	10.3		15	10	33.3
15	12	13	8.3		118	116	1.7		19	20	5.3
16	13	14	7.7		122	132	8.2		15	20	33.3
17	10	9	10.0		94	80	14.9		13	10	23.1

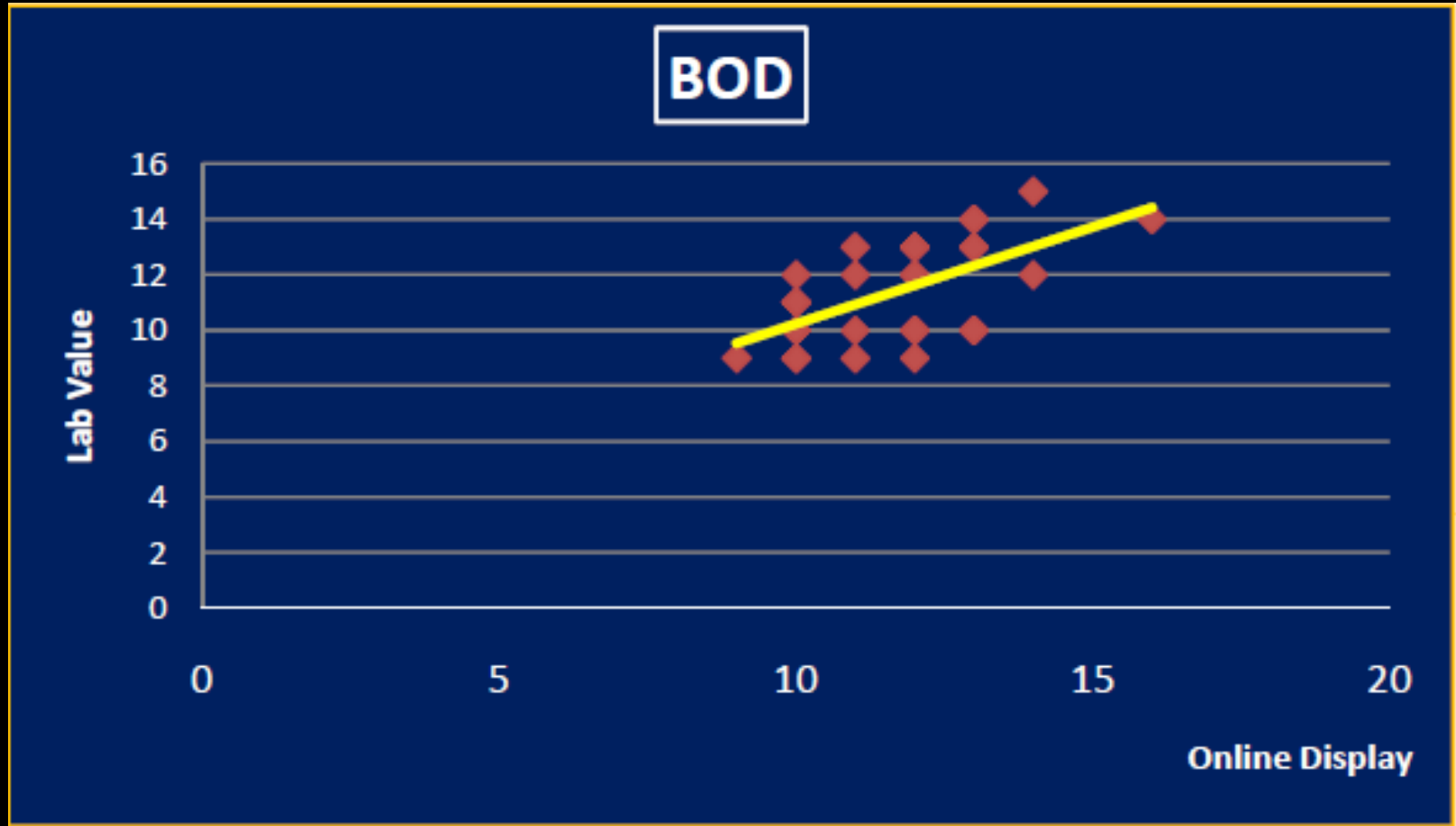


Data Comparison

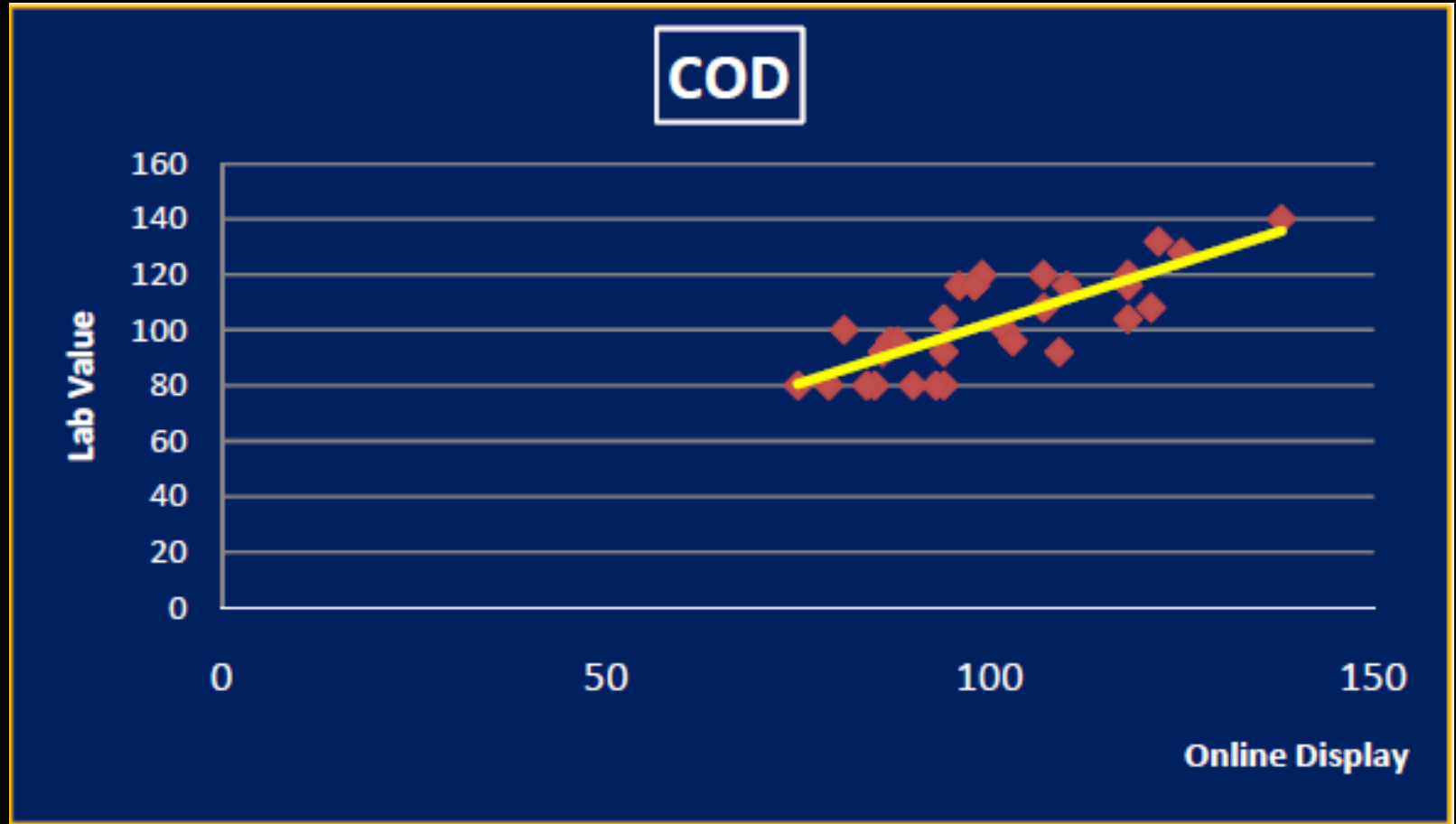
- Data for BOD, COD and TSS, for any single day varies significantly.
- However, average of both data sets match fairly.

- Let's see the trendlines.....

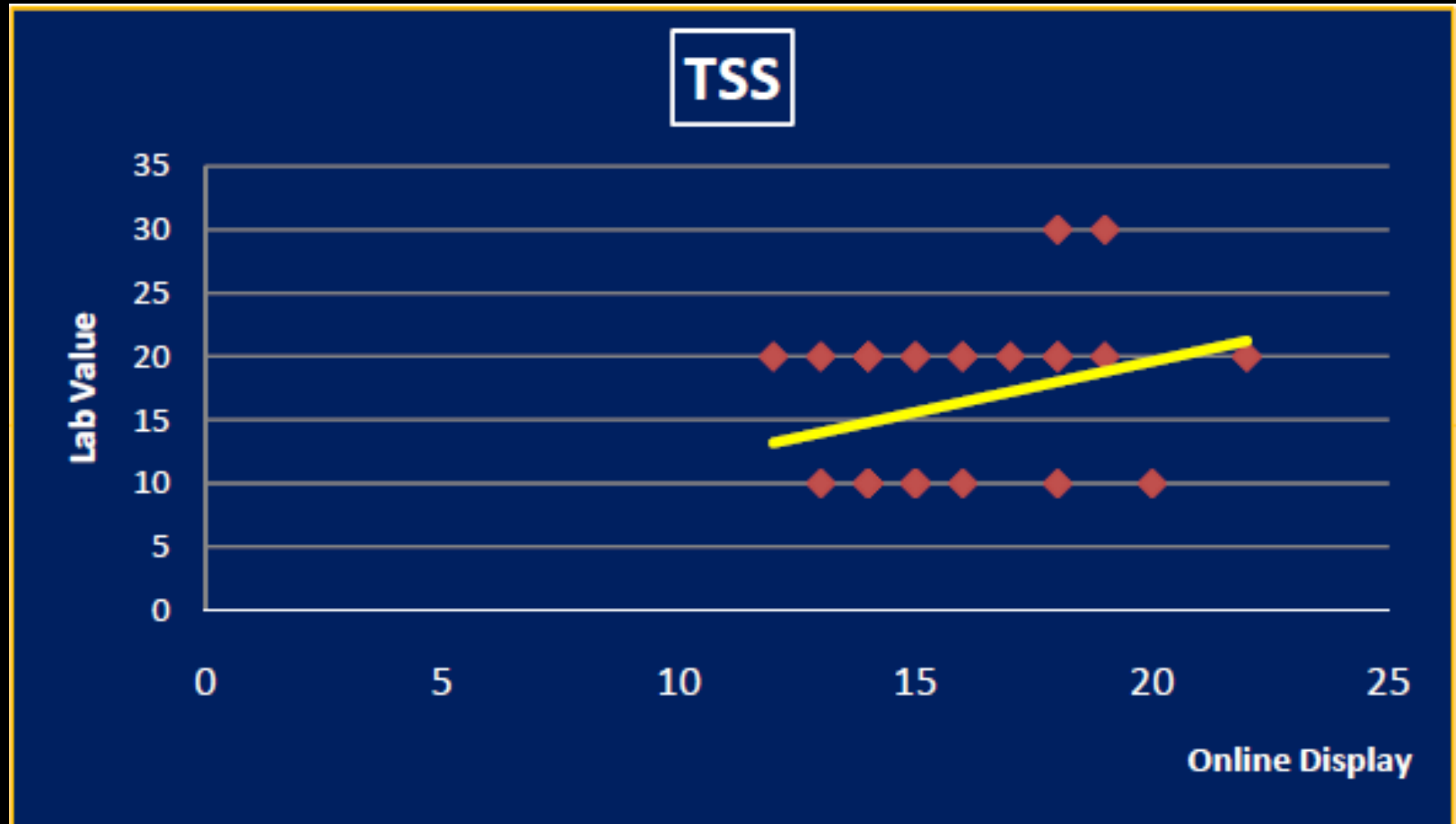
BOD ($m=0.7, c=3.3$)



COD ($m=0.88$, $c=14.8$)



TSS ($m=0.80$, $c=3.6$)





Issues in this case..

- Mill testing vs. NABL lab testing
- Internal testing procedure verification
- Mandate calibration vs. self responsibility



The Way Ahead...

- This presentation was made just to show the potential with data mining.
- For future, we need to-
 - focus on quality data generation.
 - develop data mining strategies.
 - automate data mining.



Thank You.