ISSN: 0379-5462
IPPTA: Quarterly Journal of Indian Pulp and Paper Technical Association
Vol. 32, E3, 2020, pp. 45-51

Technical Papers

Application of Data Mining in Pulp & Paper Industry

Abstract: *Lot of data collected and stored in record rooms or in the hard disc of computer can be extremely useful for the mills to improve quality, productivity, efficiency and economics. The present article is an effort to initiate a new thinking towards data mining. A few case studies have been discussed here to explain how the past data can be converted into useful experience and further into a tool for improvements.*

Key Words: *Data Mining, Paper, Management, Quality Management, Productivity, Quality Control*

D K Singhal
Director
Chandpur Enterprises Ltd.

Introduction

We have a lot of information available in the form of data which is gathered during the long period of operation, and process management. Every day, our operators as well as electronic data monitoring systems like SCADA and QCS collect a lot of information about our plant, and store the same in form of log books in hard copy as well as in soft copy versions.

All these records contain valuable information, which generally remains unused in absence of our efforts to extract useful information from it. Let us understand it with a simple example-

A purchase executive in a small mill noticed that the electric tube light was to the tune of 30 nos. every month. The information was shared with the department, and they decided to schedule purchase of 30 tubelight pack every month in a scheduled way. Another manager asked- “why do they need so much tubelights? Is the quality poor? Do we need a better brand of tubelights?” The electrical department informed- “There are around 278 tubelights installed in the plant, many of these are at places where we need to switch these ON for 24 hours. Thus, we are having an average life of more than 9 months. This seems satisfactory.” Everything was normal, but soon, an electrician indicated that there were 5 places where the tubelights have to be replaced almost every week. Well, these five fixtures were replaced, and the consumption reduced to typically 10-12 per month.

The question is, “How that electrician could get the information?” In fact, he just compiled the data available with him, and came to a useful conclusion. Now, we can understand, in several instances, a detailed analysis of data yielded significant savings as well as productivity improvement.

The present paper is aimed to share several such observations.

D.K.Singhal

First Pass Retention

This was a recent case, where in a meeting the mill technical team informed that all the paper machine operating parameters were running well within range, FPR included in the parameter list. The mill had a practice of checking machine parameters on daily basis, and a recording a summary of the same in a register. However, the mill was making several grades of paper in different basis weight range, on a single paper machine. Obviously, for each grade, the operating parameters had to be different.

To begin with, the FPR was selected as a trial parameter. A commonly used grade in two different basis weights was chosen for evaluation. For a particular grade and basis weight; date and retention were noted in a Microsoft Excel worksheet and a plot was made from the data. Next, a linear trendline was obtained for the given set of data. The data and trendline appeared as indicated in figure 1.

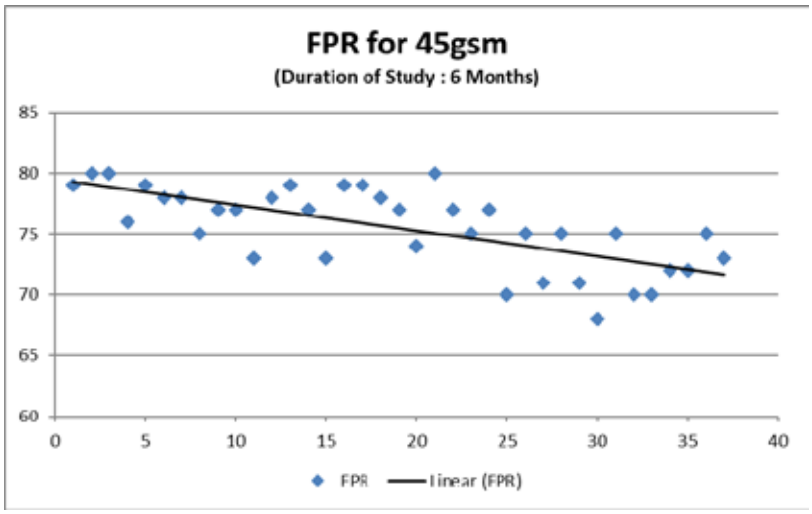

Figure 1: FPR for 45gsm paper

Interestingly, there was a decline in FPR with time. Not only this, the decline was nearly 1% point per month or so, so it remained unnoticed. To validate the findings, the same exercise was done for other grade also. The results are given in figure 2.

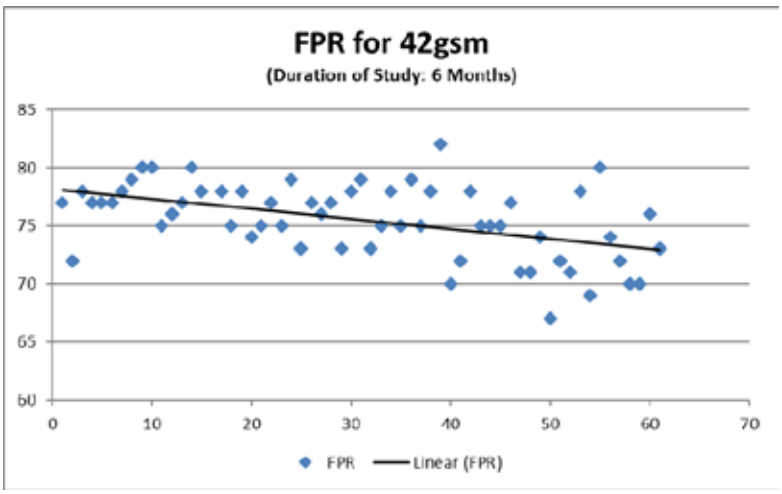

Figure 2: FPR for 42gsm paper

In fact, when the FPR was being discussed just a few days back in the plant internal meetings, everyone was of the opinion that the FPR is nearly constant for the past few months. But, as the data was plotted as indicated in figure 1 & figure 2, for a period of 6 months, a gradual decline became noticeable. After the plots were handed over the production team, they started looking back to process and took necessary actions to re-achieve the same FPR level, which was achieved within a couple of days.

Supercalender Roll Burnout

Cotton rolls are commonly used in supercalenders, though many mills have switched over to synthetic rolls. In a particular case the cotton rolls got burnt during operation. The mill was new, and hence, the coating manpower was arranged from an already running mill. Everyone in the mill was sure that the operators are competent enough, and there might be some other reason for premature failure of rolls. However, the mill kept on facing frequent roll burnout issue. After any such case, the roll had to be sent to supplier for recoating, which involved a lot of capital cost. The mill had adequate spare rolls, so production loss was not an issue, though recoating cost was getting a serious issue. Furthermore, the production suffered for lack of demand and there were frequent start-stop of production.

The matter was discussed with the supplier, who just tried to shrug off the issue. According to him, this was a routine and nothing much could be done to improve the life of the rolls, even though the mills felt that such a failure should be considered as premature failure. The issue remained existent for more than a year or so.

The mill decided to investigate into the problem using data mining. All the relevant data was put in Excel sheets, and parameter to parameter comparison was started. Following aspects were evaluated-

1. Invoice Date: Were the rolls from a particular invoice (lot) failed much earlier than others?
2. Were the rolls installed at a particular position failed more compared to others?
3. Some rolls had been supplied with the equipment, and others were prepared by the local supplier. Was there a quality issue either in locally supplied rolls or the original rolls?

All such evaluations failed to give any conclusive clue to the issue. Now, the direction of investigation was turned in a blind way, asking so called 'stupid' questions like-

1. Were the rolls failing more on any particular day, say Saturday? (Astrology)
2. Were the rolls failing more on any particular shift? (Time of day)
3. Were the rolls failing in the shift of some particular operator? (Operator negligence)
4. Were the rolls failing more while calendering any particular grade? (Paper properties)

However, detailed evaluation of data in such ways indicated a strange point- 'Most of the failures (around 90%) occurred within 2 hours of startup of plant when the plant was previously shut for more than 2 days.' Well, this was a great initial hint. The matter was discussed with the operational team, who themselves suggested that the startup should be in such a way that the roll temperature increases slowly to the desired operating temperature. A sudden startup was resulting in sudden temperature rise from within the rolls, and operators are interested in surface temperatures. As a result, the temperature inside the rolls increased beyond a critical value, and the rolls failed.

Now, this finding was really strange. The production team knew what should have been done, but they were not following the generally accepted practices, as they thought it was just a minor issue. Anyway, more than 8 years have passed and there has not been even a single instance of such failure since then.

MD & CD Profile Evaluation

The profile control is becoming from just an important parameter to a 'status-symbol' for many papermakers. With increase in automation, QCS and DCS systems, mills are able to manage profile control in much better way. However, to achieve maximum production, in some grades, it becomes necessary to disturb the profile though reluctantly. The typical causes for the same is non-uniform moisture profile due to poor clothing condition, problem in drying system, improper felt showering or hood related issues.

In several cases, particularly in Kraft paper with higher basis weights, increasing machine speeds is the main aim. No one obviously, would like to spend time and production to identify the instant cause of the problem. The problem becomes more complex when the point of problem shifts from one place to another in the cross direction. Under such circumstances, the management would like to know clear information about-

- Is the problem temporary or permanent?
- What can be probable cause of the problem?
- Do we have any solutions?
- Can we find someone who can solve the problem?

In a typical case, the mill was facing similar issues, and in need for a direction. The mills data was compiled to get something useful from it. The mills profile records were analyzed and average of the profile for all rolls made in batches of 15 days were computed. The average fortnightly profile for six months is shown in Table-1.

	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

Table 1: Fortnightly averaged profile

As you can see, it is practically difficult to conclude immediately from the data given in Table-1. To understand it better, from the individual basis weight figure, average of the profile was subtracted. This way, one can get the absolute profile which indicates at what position the basis weight is more and at which position it is less. Now, all the data was color coded for easy understanding. The same is indicated in Table-2.

	1	2	3	4	5	6	7	8	9	10	11	12
JAN_1	-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_2	-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	-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_2	-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_1	-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_2	-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_1	-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_2	-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_1	-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_2	-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_1	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_2	0.2	-0.1	0.1	-0.1	-0.2	0.1	0.1	0.1	-0.1	0.1	-0.1	0.0

Table 2: Fortnightly absolute profile for the above case

As clear now from the table itself, at position 1 (NDE), the basis weight is generally on the lower side, which indicates some problem from the machine side. The issue was identified, and corrective action was taken after which the problem reduced significantly.

	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

Table 3: Fortnightly absolute profile after the corrective action was taken

It is interesting to note that at a few points, the profile varies by 0.5-0.6, but the same gets corrected in the next fortnight. In fact, this happens because of some temporary problem, like dirty wire, headbox streaks, foreign material getting stuck in headbox lips etc. However, if such an issue persists for a longer duration, it gets noticed easily, and it is easier for the mills to take corrective action.

ETP Online Calibration validation

Effluent treatment plant is an important part of the whole mill today and almost every paper mill of India is proudly treating the effluent as per the regulatory norms. However, proper monitoring of ETP performance is a must to maintain the desired treated effluent properties. As per the guidelines, all of the paper mills in Ganga Basin have installed COEMS (Continuous Online Effluent Monitoring Systems), which give instant readings of desired effluent parameters.

As the measurement of BOD, COD and TSS by the COEMS is based on UV-Vis Spectrophotometry, while the testing in our laboratories is done using titration method, there might be some difference in both approaches. One cannot just say which method is more accurate or reliable. To give an example, a 1% Hydrogen Peroxide solution results in 220 COD, as the Hydrogen peroxide just interferes with the action of Potassium Dichromate during the test. Theoretically, there should be zero (or say negative) COD of this solution. However, using UV-Vis Spectroscopy, the COD measured is near zero, while the lab test result is highly positive.

Similarly, the testing of BOD, being a biological process, may give significant errors by the conventional methods using Winkler method to determine DO. As per the BIS standard (IS:3025-44), a standard solution of 200 BOD prepared by Glucose-Glutamic Acid (each added 150 mg to one liter of distilled water) solution giving a value of 200+37 should be considered acceptable. In case, the sample is of say 20 BOD, the percentage error might be more. Unfortunately no published document indicates the repeatability-reproducibility of BOD and COD testing.

Now, the Government mandate demands that the readings displayed by the COEMS should fall within +10% of laboratory value. This puts the mills in a complex situation, as one cannot know the laboratory value with accuracy due to the nature of tests involved. Mills have to get COEMS calibrated, and the common method is to get the effluent sample tested by any NABL approved laboratory and see if the laboratory result matches with the OCEMS measurement. Not only this, the bigger problem is that the test results are obtained after a week's time from the date

of sampling. In case it meets, a calibration validation certificate obtained from the NABL laboratory is submitted to pollution control board; and if it does not match, one has to ask the supplier to visit the mills, calibrate the COEMS accordingly, and repeat the whole exercise once again.

Some mills have tried to take daily samples for a few days, and get a report according to that. The basic problem is that if the readings do not vary significantly, it may give a misleading interpretation. Let us consider following two cases for COD as example-

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

We can see that in both cases, readings are within +/-10% range compared to display, but can you say that the calibration is acceptable in both the cases? Maybe you'd like to say instantly that the case 2 represents a better calibration.

For the same, it was decided to develop an easier, simpler and self test method for calibration validation. The laboratory chemists were given several samples and after we became convinced that they are testing properly, some standard test solutions were made. The BOD standard solution was made using GGA and COD standard was made using KHP. The results found by the chemists were in close proximity with the standard theoretically calculated values.

Now, the linear curve fitting was done for the whole month for the OCEMS value and laboratory results. As we know a straight line can be represented in the following form-

$$y = m x + c$$

Where, m is the slope and c is the intercept on the x axis. Ideally, m should be unity, and c should be zero, but a line having m and c near these should be considered to indicate that the calibration is proper.

To calculate m and c, Microsoft Excel functions were used.

m=SLOPE(Laboratory Data Range, COEMS Display Data Range)

c=INTERCEPT(Laboratory Data Range, COEMS Display Data Range)

This way, we get the data for a whole month, and for any parameter, we can compute m and c values. Just having a look at the values, it is now easier to evaluate the calibration status easily, using the following typical approach.

Approach:

1. Check the intercept and slope values for the previous month.
2. For BOD, the intercept should be between +/-3.0, slope between 0.6-1.2.
3. For COD, the intercept value should be between +/-25, slope between 0.6-1.2
4. For TSS, the intercept value should be between +/-10, slope between 0.5-1.5
5. In case the intercept and slope are found out of range, do rinse the sensor using standard solution properly and observe the data for next month.
6. In case, the intercept and slope values do not improve, ask the supplier to visit the site for calibration.

A sample data sheet is given in table 4.

If we look closely, we find that the BOD readings have a significant variation between the displayed on COEMS and those obtained in laboratory by actual testing. In several cases these variations are to the tune of 20-25%. But, as indicated earlier, we can use the overall data compiled to generate a trend line, or even its characteristics generally represented by m and c; which give better and more reliable information about the calibration status.

COEMS Calibration Monitoring Chart											
Sample Data for a Typical Month											
	BOD				COD				TSS		
Date	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
18	12	13	8.3		121	108	10.7		12	20	66.7
19	13	13	0.0		81	100	23.5		14	10	28.6
20	11	9	18.2		75	80	6.7		14	10	28.6
21	12	10	16.7		94	104	10.6		15	10	33.3
22	12	12	0.0		93	80	14.0		15	20	33.3
23	12	9	25.0		86	92	7.0		14	20	42.9
24	13	14	7.7		118	120	1.7		17	20	17.6
25	14	15	7.1		125	128	2.4		18	10	44.4
26	12	12	0.0		102	100	2.0		15	10	33.3
27	10	11	10.0		84	80	4.8		13	20	53.8
28	10	10	0.0		88	96	9.1		14	20	42.9
29	16	14	12.5		138	140	1.4		22	20	9.1
30	13	10	23.1		103	96	6.8		20	10	50.0
31	10	10	0.0		79	80	1.3		14	20	42.9
Avg.	11.6	11.4	9.8		100.2	102.6	9.3		16.1	16.5	32.3
Slope			**0.70**				**0.88**				**0.80**
Intercept			**3.3**				**14.8**				**3.6**

Table 4: Sample Data Sheet to Show Calibration Status

In Table 4, we can see that the slope is around 0.7-0.8 for all the three parameters, which can be considered sufficiently acceptable. We can also see that individual day readings of BOD might vary by as high as 20-25%, though the monthly average is just 9.8%. In fact, such detailed analysis helped the OCEMS supplier to calibrate the system easily, when needed.

Conclusion

Data mining is a relatively new word for paper industry. As clear, the data analysis and compilation can be used to analyze the process operating and control parameters for betterment of process. It is possible to use the data recorded in the past as a useful tool to analyze and improve the performance of our systems. Mills may explore the possibility of further improvement by analyzing their data themselves, and to further take advantage of this technique, may seek advice of experts in this area.