JPMTR 097 | 1706
DOI 10.14622/JPMTR-1706
UDC 004.6-022.56(676+655.1)

Review paper
Received: 2017-07-10
Accepted: 2017-07-30

Technologies for using Big Data in the paper and printing industry

Igor Karlovits

Pulp and Paper Institute, Ljubljana, Slovenia igor.karlovits@icp-lj.si

Abstract

The paper gives a review about the possibilities of using Big Data technologies in the print industry. Current situation and research are presented with a brief overview and description of the data mining process. Process parameters readouts and process modelling by using problem-solving methods open up new possibilities for production efficiency. Different authors have provided solutions for print and print quality related problems by data collection through sensor readouts and real-time monitoring of different production system variables. Major production techniques (offset and flexo printing) have been partly investigated and monitored through closed inline controls, or metadata communication with the use of Job Definition Format. The researchers have found possibilities in solving particular print or production related issues with the use of Big Data techniques or its subsets, but still, no integrated market ready solution exists. A theoretical framework for a corrugated production factory is shortly presented, outlining possible applications and connections for a fully integrated data mining system that could bring the Industry 4.0 in the printing sector.

Keywords: data mining, process control, printing, papermaking

1. Introduction

The use of Big Data for different applications from marketing, sales, production optimisation and maintenance is offering promises of so-called Industry 4.0, which relies on connected devices, large choice of versatile sensors and intelligent systems. Automation and data exchange of network devices controlled by artificial intelligence systems combined with data handling using cloud computing is expected to increase the productivity and value added part of the production by several percent. The EU has established a rather high goal of 20 % for the ratio of added value of production. The paper and graphic arts companies as a service-based business and in some applications as printed electronics producers are also prone to adapt to the changing business landscape which is emerging. On the basis of a PwC (2016) industry survey, the forest, paper and similar industries are currently automated around 38 % with predicted rise to 72 % in the next five years. The paper and printing industries are under pressure due to decreasing consumption of all paper grades, with packaging papers excluded, by a report made by CEPI (2017) which demands the change of technology to short run production, production optimisation or innovation into new products. As new technology and innovation are quite expensive for small and middle-sized printers, one of the viable solutions is to try to optimise print production processes. Besides Lean Manufacturing and other management tools available for decades, now it is possible to use “redressed” problem solving and analytical tools from the Big Data portfolio for all optimisation, decision-making and problem-solving challenges. Some of the steps in the data mining process are very similar to Root Cause Analysis (management tool for finding hidden or underlying causes for a problem by mapping causing issues), in combination of “5 Why” (repetitive interrogative questioning why has something went wrong, with the goal of finding cause effect relationship in a process) and Failure Mode, Effects, and Criticality Analysis used in classic business problem solving (George, et al., 2005), but contemporary data analytics provide more powerful resources due to increased speed of computing, sensors and networking solutions. The difference between classic business intelligence (reporting) and predictive business analytics is presented in Figure 1.

While the classic problem-solving methods already mentioned have some business value, their complexity is low and are mainly focused on solving one problem or a particular challenge in a total quality management system.

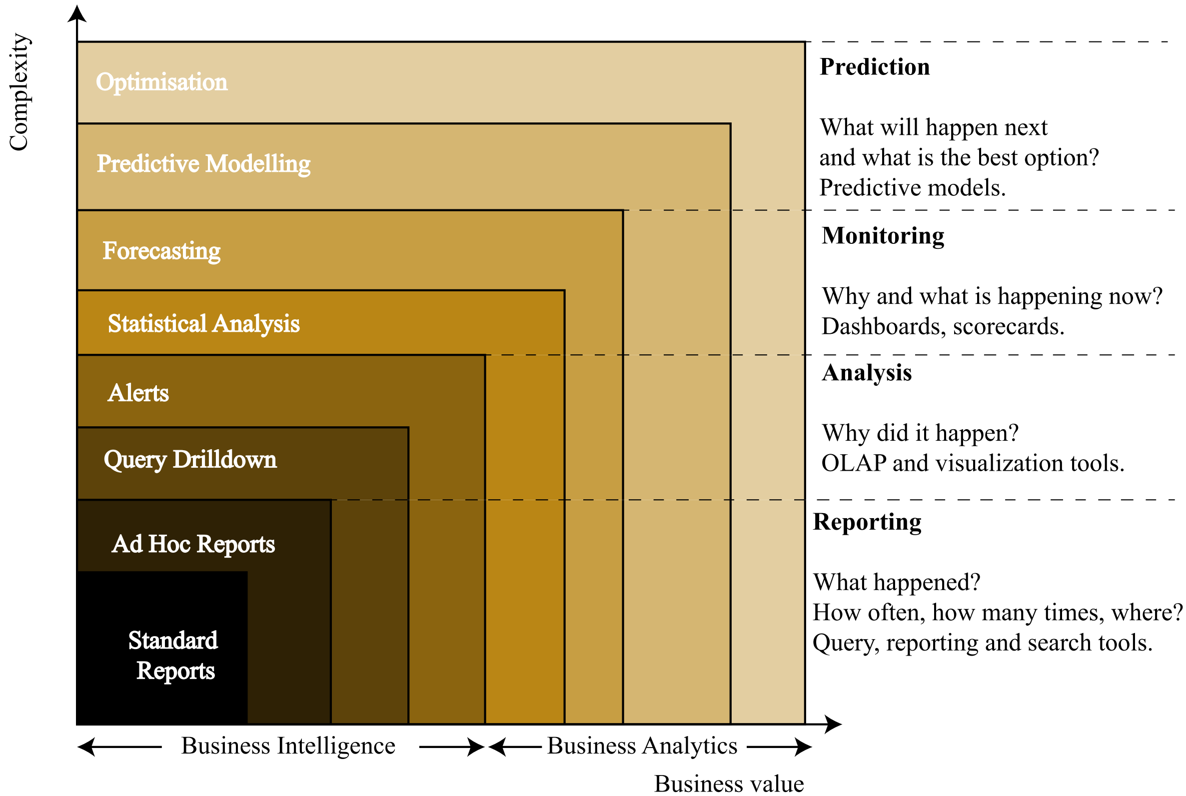

Figure 1: Difference between conventional and predictive analytics value and complexity, adapted from Practicanalytics (2011)

Some of these methods are reactive, while the predictive analytics are driven by more complex methods and software solutions with the aim of maximising business value for a whole company. On the other hand, business intelligence is more concerned with the business side of the company. Business intelligence looks for trends at the macro or aggregated levels of the business, and then drills up, down, or across the data to identify areas of under- and over-performance, while predictive analytics is about finding and quantifying hidden patterns in the data using complex mathematical models that can be used to predict future outcomes (Schmarzo, 2014).

Big Data is a blanket term for any collection of data sets so large or complex that it becomes difficult to process them using traditional data management techniques such as, for example, the relational database management systems (RDBMS). Data science and Big Data evolved from statistics and traditional data management but now are considered to be distinct disciplines (Cielen, Meyman and Ali, 2016). The relationship and the interconnection of different technologies can be observed from Figure 2.

One of the first attempts was the CRoss-Industry Standard Process for Data Mining (CRISP-DM) where the process consists of six steps or phases, as illustrated in Figure 3 (Bijlani and Bauer, 2016).

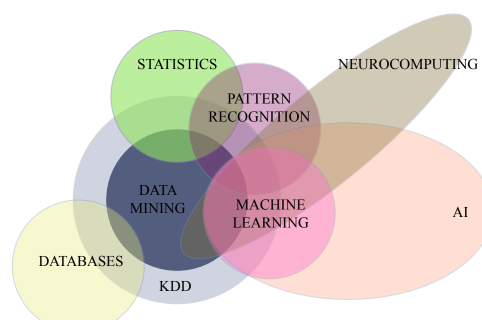

Figure 2: Universe of data science, adapted from Srivastava (2015)

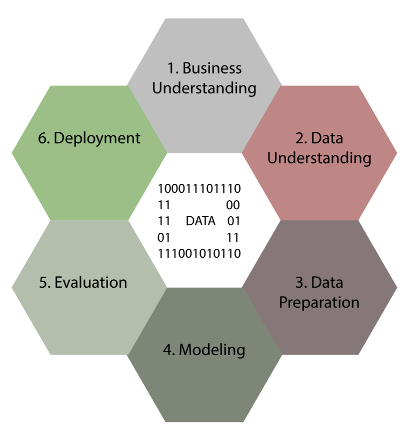

Figure 3: CRISP-DM Conceptual Model, adapted from Bijlani and Brauer (2016)

Data mining as a core of Big Data applies statistical and logical methods to large data sets. These methods can be used to categorise the data, or they can be used to create predictive models. Predictive models, however, transform these descriptions into expectations upon which we can base decisions. For example, the owner of a book-selling web site could project how frequently she may need to restock her supply of a given title, or the supply chain manager of the print house can order paper stock on the basis of previous data (from book seller). It is important to recognise that data mining cannot provide answers to every question, nor can we expect that predictive models will always yield results which will, in fact, turn out to be the reality. Data mining is limited to the data that has been collected (Bijlani and Bauer, 2016). An overview of potential problems and challenges as well as different applications can be found in Hassani and Silva (2015). Data and analytics can mainly be grouped into descriptive, predictive and prescriptive. A large share of the production and business analytics data are descriptive statistics which summarise what has happened during a process or an event. Predictive analytics, on the other hand, uses a vast array of data mining, machine learning and statistics and different modelling to analyse data and to enable some sort of predictions. These predictions are not deterministic as it will happen but probabilistic and will show what may happen. In most cases, the models predict missing numerical data on the basis of previously collected available data. For example, in a standardised printing production, if there was some minor failing out of data causing missing values, these models can predict what was the value or can predict what will be most probably the data, for example ink density or CIE $L^*a^*b^*$ value of the prints. One of the modifications of predictive analytics is the prescriptive analytics which adds another layer to the data handling the actionable data and the feedback which returns the results of the processed action. Its feature is a prediction of the possible consequences from different scenarios and these models can recommend the best option for a preselected results or expected outcome.

2. Literature overview

One of the early cases of using top down induction of decision tree modelling to reduce cylinder banding in printing was carried out at RR Donelley & Sons Company by Evans and Fisher (1994) in the early 90's where they managed to reduce bands in gravure printing from 538 in 1989 to 42 in 1993. This is a rare case of published data analytics regarding some new data presented and gathered by Hornbuckle (2016). The fact is that on average a commercial printer/press is only productive (generating revenue) 60 % of the time according to a study done by SpencerLab (2017). The typical work cycle of a printing press is presented in Figure 4.

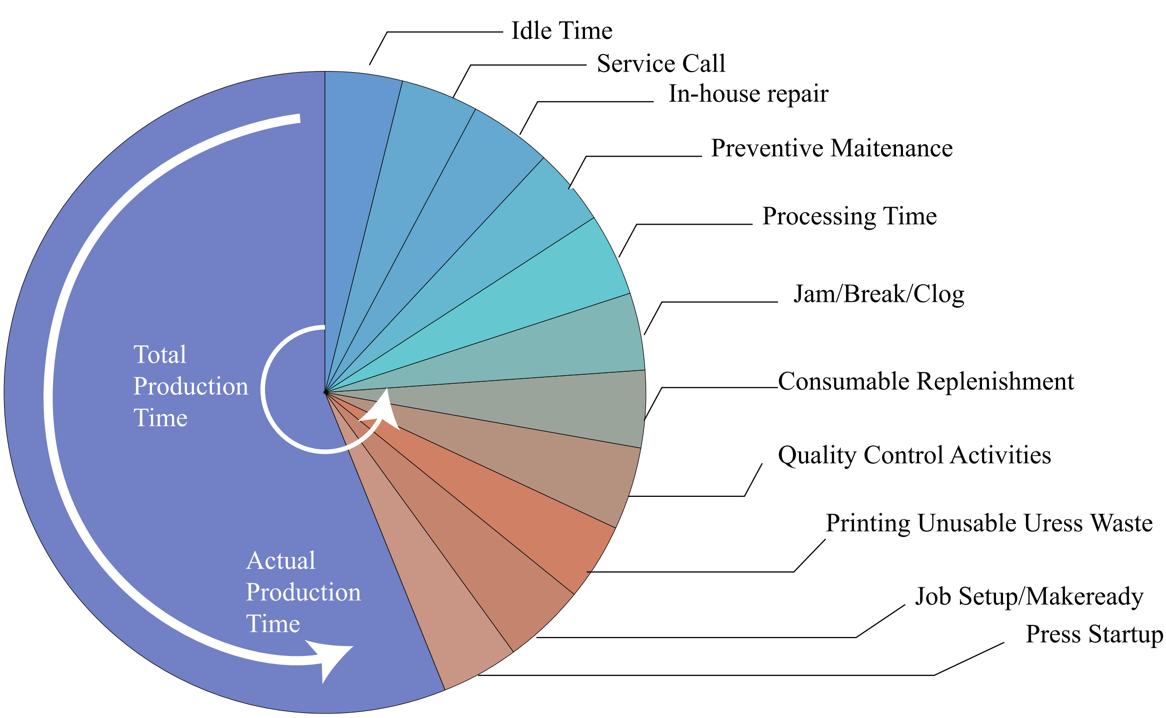

Figure 4: The work cycle of typical press, adapted from SpencerLab (2017)

In their study, they claim that based on their analysis of offset presses from manufacturers including, but not limited to, Komori, Heidelberg, Ryobi, Manroland, Goss and others, the improvement in revenue opportunity ranges from $ 500 000 to worth of $ 9 000 000 per device per year. The variation in additional revenue opportunity has to do with the type of press, its capacity and its duty cycle. From Figure 4, it can be noted that predictive analytics can help with maintenance scheduling and optimisation and also spare parts inventory or additional support materials which can lead to savings in ordering and warehousing. In the papermaking, some of the tools for production optimisation were published by Linnala and Hämäläinen (2001) who described the application of bi-level dynamic multi objective optimisation to the papermaking process, or more accurately, the paper web brake system. The results show that the approach was successful: capital costs were decreased while maintaining adequate process stability. However, the computing time requirements were relatively high; to reduce them, the operational optimisation on the lower level could be performed using a slightly simplified process model. A similar subject was examined in a paper by Hämäläinen, Madetoja and Ruotsalainen (2010). In paper industry one of the recent approaches was published by Fu and Hart (2016) where the authors collaborated with MWV mill which battled with significant quantities of internal rejects and production downtimes. The authors, due to the failure of the classic problem-solving methods to tackle the issues, deployed Big Data analysis to determine the root causes. They reported that 6 000 operating variables were selected with more than nine billion data points in the period of almost three years. The results indicated an action which changed process targets and technological change in the process of drying. In a paper by Jackson (2011), the author presents case studies involving Big Data techniques for value chain cost and return of investment (ROI) optimisation in the papermaking industry.

There were also studies trying to solve overall quality and printability issues like in a thesis work by Gerard Leloup (2002), where the author tries to integrate the quality factors and predicted values for flexography through the use of “Printability Coefficient”. The author used quality factors (mottling index, edge sharpness, dot gain, density, dot deformation) with different contribution importance to calculate the final p coefficient value.

The calculation was done for the total variation range of the different parameters and then reduced to a “united result” in percentage. With the two precedent results and the definition of units (c_i), it was also possible to calculate first the partial and then the global influence, in a percentage of “P units”, of each primary parameter onto the final “Printability Coefficient”. The result of the calculation is called unit influence coefficient (f_i). The calculation of the distance of the user’s values (x_i) to the references values (a_i) will then be the first step of the calculation procedure. The obtained values will be divided by the defined units (c_i) and then multiplied by (f_i). The sum on the index “i” is called p, which is presented in Equation 1:

$$p = \Sigma\{[(x_i \cdot a_i)/c_i] \cdot f_i\} \qquad [1]$$

Also, the same study claims that is a good visual correlation with the human observers assessment and the final conclusion is that there is a possibility that this method can be used in the printing industry. In another thesis by Lundström (2014) he used image analysis and machine learning methods to model offset printing. The author tried to solve problems of the data mining and modelling through the use of three levels of situation awareness. On the basis of chosen print quality parameters, the author used Random Forrest method of machine learning for training for a model so that the observed quality scores give a set of computed print quality attributes. He concluded that results obtained in the thesis strongly indicate that the computational intelligence-based approach can provide an objective print quality assessment. Whilst other researchers focus on a global approach where a large area of the print is scanned, the techniques developed in this thesis are able to compute useful print quality attributes from small test areas. It was demonstrated that the overall print quality scores obtained from human assessed print samples can be modelled with a good accuracy by using the print quality attributes acquired on-line from the printing press. In their study, Parola, et al. (2003) developed a software in MATLAB and used it in the web print measurement. They have tested their system on a newspaper press concentrating on paper web tension problems while investigating more than 2 000 customer reels. They have reported that the press components had major influences on the slackening of the web. Beside press parameters, there were correlations regarding paper properties which influenced press register accuracy. The end-user printer found the IQTension named software measurement module useful for troubleshooting and material evaluation. Similar problems regarding web printing were studied by Alzghoul, et al. (2009) where they used data mining techniques with two approaches to study web breaks. They used genetic search to analyse process results with a set of input variables providing the lowest average loss incurred in taking decisions. They have reported on average 93.7 % of test data set were classified correctly. Also, there were some statistical studies in flexo by Matulaitienė and Jurkonis (2013) in analysing a one year run of flexo press and mistakes using SPSS statistical software. One more study in flexo printing problem-solving data mining was carried out by Ejsmont, Krystosiak and Lipiak (2015). One large study regarding banding in gravure printing was presented by Gaudard, Ramsey and Stephens (2006), where they used JMP software for data analysis and data mining techniques on an available data set regarding banding problems. They found that banding has 11 categorical variables and 18 continuous variables in predictors. They have used both partition examples and classification trees as a subpart of Six Sigma project work.

3. Integrating the data collection

The paper production machines have a high degree of automation. Preventive maintenance of sensors/scanners and actuators in the machine level control system (MCS), process control system (PCS) and quality control system (QCS) are key enablers for achieving continuous availability of the sensors/scanners over the entire life cycle of the system. In addition, automatic controllers have to be constantly adapted to the altered conditions and goals of paper production and this closed production circle is enabled through proprietary software. Logs and sensors reading regarding different production parameters are implemented to wet end, dry end, and stock preparation process phases. In a paper by Gough, et al. (2007), the authors present a model-based predictive adaptive process controller on a number of challenging pulp and paper mill control loops including paper machine reel brightness control, lime kiln temperature profile control, slaker temperature control, and extraction stage pH control. The presented solution resulted in financial cost saving and enables paper makers to significantly improve their process control. Also, some mathematical modelling of pulp and paper production is presented in a work by Jansson, et al. (2004). One of the largest company in the paper sector, Voith, uses complex process and predictive analytics in the PaperMiner software for process optimisation where 6 000 readings are continuously taken and stored during the paper production. The PaperMiner and the process control software relies on self-organizing maps (SOM) (Bullinaria, 2004). The SOM method can be used for forecasts with good success. Once a SOM has been generated, the place of a modified machine setting can be determined on the two-dimensional map, and it can be used to determine the expected values for the desired target parameters, such as porosity, formation, etc. (Bamberger and Nicolas, 2005). Of course, the quality of such forecasts is greatly dependent on the number of data sets used to generate the map, and whether or not all major influences were included. Decision Trees are another method of analysis. In order to use Decision Trees, a target value to be analysed is first selected. The Decision Tree can then be used to find out which process adjustments must be made. Another product is OnV, which is predicting quality parameters based on readings from process data – in real time. As an example of sensor complexity in a sensoring system on a wet end of a paper making machine by Voith, necessary for good data array access for prediction purposes, is shown in Figure 5.

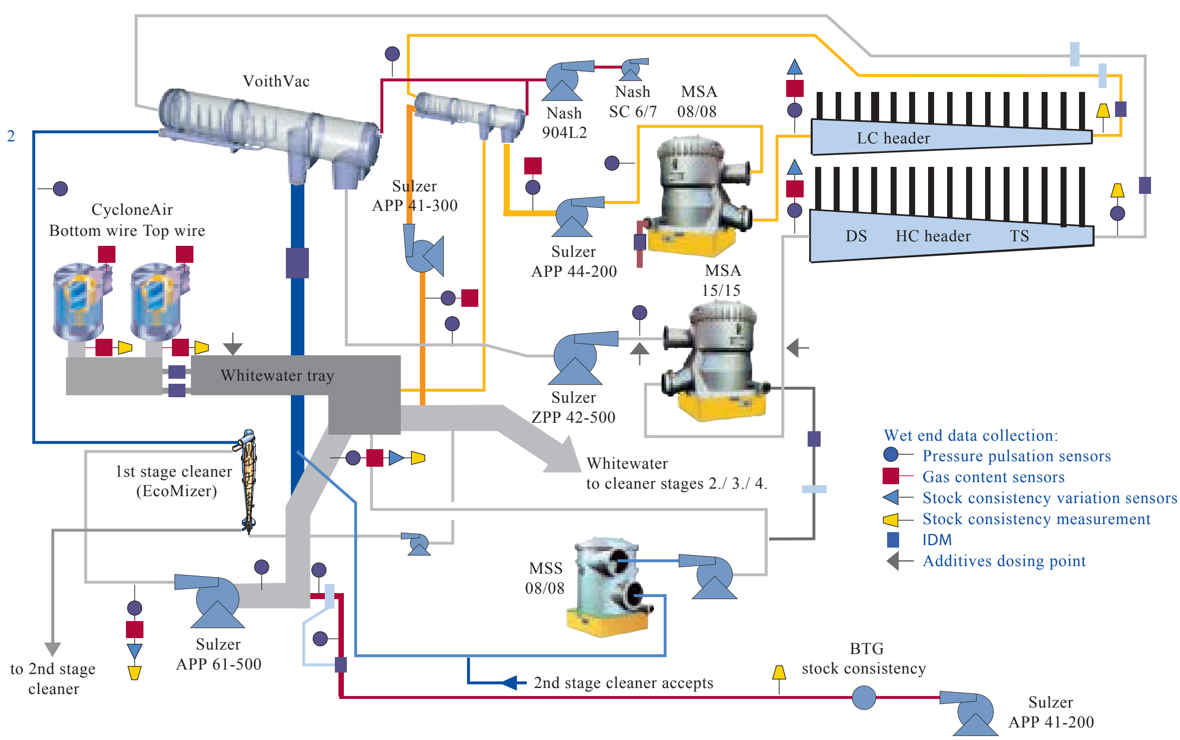

Figure 5: One of the industrial solution of sensor readings for process control by Voith, adapted from Stibl and Natterer (2005)

On the other hand, the printing and graphics arts industry is lagging in terms of integrated process automation and solutions for predictive analytics. There are several closed loop control systems for web printing with modelling and analytics (Kaestner and Nilsson, 2003; Verikas, Malmqvist and Bergman, 2000; Shankar, Ravi and Zhong, 2009) or Heidelberg's Prinect Image Control but authors report several problems of these systems (focusing just on the colour aspect of the printing reproduction):

- Uncertainty in both printing and papermaking industries about the main reasons causing too extensive variations of high-quality colour prints leading to insufficient overall print quality, customers' complains, and substantial economic losses.
- Lack of comprehensive knowledge of the interaction between paper, printing press, ink, and other constitutes of the printing process. This gap leads to difficulties in finding economically plausible means of adjusting technological parameters for optimising the papermaking and printing processes.
- Lack of robust tools for print quality predictions based on various paper, ink, printing press and process parameters.
- Lack of tools capable of online measuring of several print quality attributes, which aggregated into a print quality measurement provide print quality assessments well correlating with human evaluations.
- Lack of tools capable of explaining which technological process parameters are failing when the overall print quality is insufficient.

In a broader view, the printing industry has recently started to implement the Management Information Systems (MIS) which are large pieces of software covering from Customer Relationship Management (CRM), to scheduling and estimating of print, with more or less basic tools for process control and capabilities. One of the interface solutions, which enables the communication between different machines and computers, is the one based on Job Definition Format (JDF). This format allows systems from many different vendors to interoperate in automated and MIS centric workflows. Currently, this system is still developing and it is not a straightforward one (the papermakers have closed loop as the installations are mainly from one producer), because of a large number of equipment and software producers. There is a possibility to make a relational database through the JDF Storage Engine which can create some sort of base for the Data Mining and Analytics but there are no reports about a larger scale of industry implementation using this capability. The previously mentioned Prinect Image Control by Heidelberg is scanning the printed sheet and the reference print control strips and makes end loop quality control adjustment based on a preset data either by internal database or by JDF file. But it still lacks the sensoring of all process parameters (paper condition, ink viscosity and temperature, print pressure) and prediction of colour on the basis of these readings. As we can observe in Figure 6,

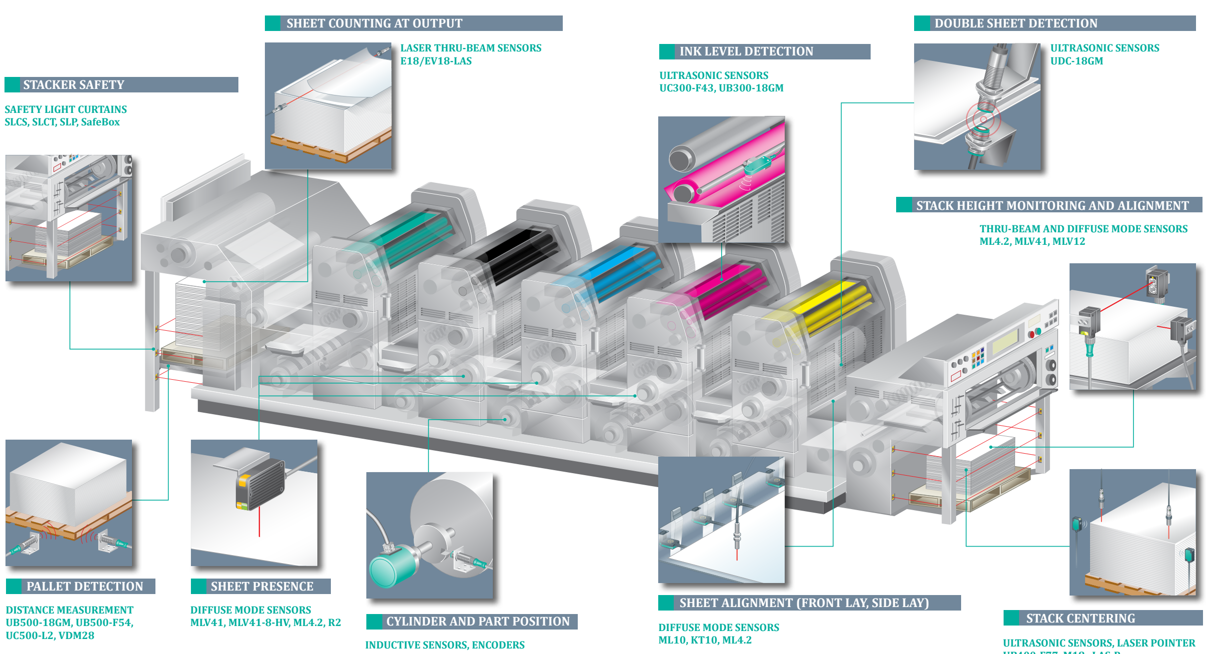

Figure 6: Sensor systems by Pepperl+Fuchs Inc. (2003) for sheet-fed offset printing press

a solution for an array of sensors for the sheet-fed offset printing press enables a lot of readings, but there is a lack of many sensors for process control and data gathering for process variables which can influence the final print quality.

The MIS software solutions or production control software enable business analytics about production speed, downtimes (where the causes are often manually put in by the operator) and can serve to improve efficiency regarding printing press operation and capacity handling. The joining of the two process control systems into one Big Data capable system is still a work in progress for the printing industry.

4. Theoretical blueprint for a corrugated factory

To show the possibilities of using Big Data and data analytics in the printing industry, the theoretical blueprint of a corrugated factory with data collection points is presented. One of the challenges is solving the "5 V" of Big Data (Volume, Velocity, Variety, Veracity and Value). In a book by Marr (2015) these concepts are referring to the challenges of a data mining system regarding the acquisition, usability and final added value of the collected data. In our theoretical framework, we shall try to integrate the production/process improvement and business analytics department. The basic corrugated factory, due to easier overview, will constitute from one corrugator, one flexo printing and die-cutting machine and one flexo folder gluer. The information stream will be separated into four levels and in four output users. The terminology and levels were adopted from Marr (2015). The four most important layers are:

- Data source layer (all sensor data and other machine information comes here, customer database, sales records, etc.)
- Data storage layer (distributed file system where all real-time and other data are stored)
- Data processing and Analysis where data is reduced with tools like MapReduce, and is prioritised to be analysed
- Data output layer where the end user is presented with statistics or Key Performance Indicators (KPI)

The data is streamlined and adapted for users/customers, financial users, internal processes/production and employees.

The data source layer would be collecting data from online continuous data measurement system regarding internal logistics (paper rolls) into and out from the warehouse to the corrugators. The system can be solved using RFID labels or other smart labels with geo positioning in the paper roll warehouse. The warehouse has to have a weather station which sends data regarding relative humidity and temperature, which affect the runnability properties of the used papers. The corrugators can be equipped with inline measurement system for adhesive preparation, inline paper roll temperature measurement during the preheating, and viscosity measurement of the starch adhesive. After glueing the separate layers, additional inline measurement devices would be installed for humidity, surface cleannliness and mechanical properties such as tensile strength, stiffness and warping. The machine's internal control system is a part of the Big Data system regarding speed and other machine's specific sensor data connected to the control screen. The data gathered from the machine can be used further for back-looping any warping or other corrugating problems (based on collected data) in the process and should be send to quality control and to the next machine in line, to adjust machine settings according to the already produced corrugated board. If the values are out of limits, the software automatically reschedules the job and cancel any bad material going further up in the process and making downtime on other machines. This data also goes to the financial users, employees concerned and supply chain management (regarding stock change).

The data feed can be used to predict any printing or converting problems based on the humidity and mechanical properties of the made corrugated board. For example, too low humidity will increase the possibility of cracking of die-cut and crease lines. The thickness or calliper of the board can also be used as a predictor for the flexo printing machine to adjust the gap between rolls and other transport elements which may alter the mechanical properties of the produced board.

On a flexo printing and die-cutting machine, the inline measurement of printing ink viscosity, colour values and die-cutting and creasing quality, respectively, can be monitored constantly to streamline the production process. The data from the creasing and die-cutting operation are transferred to the next working station, the flexo folder gluer, which has internal control (glue viscosity, glue gap detection and the side joining image processing), which is interconnected to the mechanical properties of the board and cutting and creasing quality.

All data (bad quality, rework, etc.) is stored for post analysis and data-based error prediction, while financial data of costs are automatically updated in the business part of the system. The production data is also stored per user for post analysis and prediction. The data mining and predictive analysis can be used to improve the productivity with sending data from the converting machines to the board producing machines for the current or next production.

5. Conclusion

The automation and the information technology supported analytics possibilities are widespread in different industry segments. They are widely used for marketing research and consumer behaviour analysis and modelling, while manufacturing examples are mainly used in closed-loop systems like the paper machines, which are fully automated. Printing industry has not yet adopted the full possibilities of the current technologies due to limiting factors, as not fully automated systems, incompatible data protocols and not fully developed possibilities in current solutions like JDF. There are gaps regarding sensor possibilities, connection and communication, and overall software implementations. They are mainly developed for the business side like downtimes, production speed and other parameters (scheduling, production cost), which can be combined into valuable key indicators or for example Overall Equipment Efficiency (OEE) ratings. On the other hand, production and process solving issues are still made in offline mode using traditional statistical software or not networked solutions. It seems that the printing industry will need a big leap forward in catching Industry 4.0 possibilities regarding data mining and analytic techniques for process improvement.

References

Alzghoul, A., Verikas, A., Hållander, M., Bacauskiene, M. and Gelzinis, A., 2009. Screening paper runnability in a web-offset pressroom by data mining. In: P. Perner, ed. *ICDM: Advances in Data Mining. Applications and Theoretical Aspects, 9th Industrial Conference*. Leipzig, Germany, 20–22 July 2009. Berlin, Heidelberg: Springer.

Bamberger, W. and Nicolas, A., 2005. Innovative trends in automation. *twogether: Paper Technology Journal*, 19, pp. 56–59.

Bullinaria, J.A., 2004. Self Organizing Maps: Fundamentals, *Introduction to Neural Networks*. [online] Available at: <http://www.cs.bham.ac.uk/~jxb/NN/l16.pdf> [Accessed 23 May 2017].

Bijlani, N. and Brauer, E., eds., 2016. *Data Mining for the Masses, Second Edition: with Implementations in RapidMiner and R*. North Charleston, USA: CreateSpace Independent Publishing Platform.

CEPI, 2017. *Key Statistics 2016*. [online] Available at: <http://digibook.digi-work.com/Digibooks.aspx/Get/cepi/1641/KeyStatistics2016_Finalpdf> [Accessed 17 May 2017].

Cielen, D., Meyman, A.D.B. and Ali, M., 2016. *Introducing Data Science: Big data, Machine Learning, and more, using Python tools*. Shelter Island, USA: Manning publications.

Ejsmont, K., Krystosiak, K. and Lipiak, J., 2015. Application of selected data mining technique in printing industry. In: *Conference: Polish Association for Production Management*. Zakopane, 2015. Opole, Poland: Publishing House of The Polish Society of Production Management, pp. 75–86.

Evans, B. and Fisher, D., 1994. Overcoming process delays with decision tree induction. *IEEE Expert*, 9(1), pp. 60–66.

Fu, J. and Hart, P.W., 2016. Leveraging mill-wide big data sets for process and quality improvement in paperboard production. *Tappi Journal*, 15(5), pp. 309–319.

Gaudard, M., Ramsey, P. and Stephens, M., 2006. *Interactive Data Mining and Design of Experiments: the JMP® Partition and Custom Design Platforms*. [pdf] Cary, NC, USA: JMP. Available at: <https://www.jmp.com/content/dam/jmp/documents/fr/events/explorers-resources/372455_interactive_datamining.pdf> [Accessed 9 May 2017].

George, M.L., Rowlads, D., Price, M. and Maxey, J., 2005. *The Lean Six Sigma Pocket Toolbook: A Quick Reference Guide to Nearly 100 Tools for Improving Process Quality, Speed, and Complexity*. New York: McGraw-Hill.

Gerard Leloup, L., 2002. *Measurement and Prediction Procedures for Printability in Flexography*. Doctoral Dissertation. Stockholm: Royal Institute of Technology.

Gough, B., Meade, D., England, G. and Kovac, S., 2007. Model-based predictive adaptive control of pulp and paper mill processes. In: *TAPPI Papermakers and PIMA International Leadership Conference*. Jacksonville, FL, USA, 11–15 March 2007. Peachtree Corners, GA, USA: TAPPI.

Hämäläinen, J., Madetoja, E. and Ruotsalainen, H., 2010. Simulation-based optimization and decision support for conflicting objectives in papermaking. *Nordic Pulp & Paper Research Journal*, 25(3), pp. 405–410.

Hassani, H. and Silva, E.S., 2015. Forecasting with Big Data: a review. *Annals of Data Science*, 2(1), pp. 5–19.

Hornbuckle, S., 2016. *Predictive Analytics: The New Productivity Weapon For Commercial Printers.* [online] Available at: <http://whattheythink.com/articles/81150-predictive-analytics-productivity-commercial-printing/> [Accessed 17 May 2017].

Jackson, C., 2011. Turning data into action. *Journal of Science and Technology for Forest Products and Processes*, 1(2), pp. 30–37.

Jansson, J., Lindberg, T., Dahlquist, E. and Persson, U., 2004. Process optimization and model based control in pulp industry. *TAPPSA Journal, African Pulp and Paper Week 2004*. [online]. Available at: <http://www.tappsa.co.za/archive2/APPW_2004/Title2004/Process_optimization/process_optimization.html> [Accessed 17 May 2017].

Kaestner, A.P. and Nilsson, C.M., 2003. Estimating the relative shrinkage profile of newsprint. *Optical Engineering: the Journal of SPIE*, 42(5), pp. 1467–1475.

Linnala, M. and Hämäläinen, J., 2001. Improvement of cost efficiency in papermaking with optimization tools. *Journal of Science & Technology for Forest Products and Processes*, 1(2), pp. 71–76.

Lundström, J., 2014. *Situation Awareness in Colour Printing and Beyond*. Doctoral Thesis. Halmstad University.

Marr, B., 2015. *Big Data: Using SMART Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance*. Chichester: John Wiley & Sons.

Matulaitienė, A. and Jurkonis, E., 2013. Statistical analysis of the flexographic printing quality. *Science – Future of Lithuania / Mokslas – Lietuvos Ateitis*, 5(6), pp. 598–600.

Parola, M., Kaljunen, T., Beletski, N. and Paukku, J., 2003. Analysing printing press runnability by data mining. In: *TAGA Proceedings 2003 Conference: 55th annual conference*. Montreal, QC, Canada: 6–9 April 2003, pp. 14–17.

Pepperl+Fuchs Inc., 2012. *Sensors for the Printing Industry*. Twinsburg, USA: Peperl+Fuchs Inc. [pdf] Available at: <http://dsassets.s3.amazonaws.com/Auburn-Armature/SensorsforthePrintingIndustryFlyer.pdf> [Accessed 21 May 2017].

Practicanalytics, 2011. [online] Available at: < https://practicanalytics.files.wordpress.com/2011/04/whatisbi.jpg> [Accessed 11 June 2017].

PwC, 2016. *Industry 4.0: Building the digital enterprise: Forest, paper and packaging key findings*. PwC. [pdf] Available at: <https://www.pwc.com/gx/en/forest-paper-packaging/pdf/forest-paper-packaging-key-findings.pdf> [Accessed 8 May 2017].

Shankar, N.G., Ravi, N. and Zhong, Z.W., 2009. A real-time print-defect detection system for web offset printing. *Measurement*, 42(5), pp. 645–652.

Schmarzo, B., 2014. *Business Analytics: Moving From Descriptive To Predictive Analytics*. Dell EMC Services. [online] Available at: <https://infocus.emc.com/william_schmarzo/business-analytics-moving-from-descriptive-to-predictive-analytics/> [Accessed 21 May 2017].

SpencerLab, 2017. *Productivity: Utilization & Availability*. Melville, NY, USA: SpencerLab. [online] Available at: <http://www.spencerlab.com/testing/productivity/> [Accessed 11 June 2017].

Srivastava, T., 2015. *Difference between Machine Learning & Statistical Modeling*. [online] Available at: <https://www.analyticsvidhya.com/blog/2015/07/difference-machine-learning-statistical-modeling/> [Accessed 11 June 2017].

Stibl, B. and Natterer, S., 2005. Voith process solutions – wet end process analysis as the foundation of an optimization of paper quality and process stability. *twogether: Paper Technology Journal*, 19, pp. 48–51.

Verikas, A., Malmqvist, K. and Bergman, L., 2000. Neural networks based colour measuring for process monitoring and control in multicoloured newspaper printing. *Neural Computing & Applications*, 9(3), pp. 227–242.