MODULAR AND OFFSITE CONSTRUCTION
MAY 21-24 2019 FAIRMONT BANFF SPRINGS HOTEL BANFF, AB, CANADA

Data Analytics Application for Non-Conformance Reports in a Cabinet Manufacturing Facility

Osama Mohsen[1*], Yasser Mohamed[2], and Mohamed Al-Hussein[3]

[1] *PhD candidate, Department of Civil and Environmental Engineering, University of Alberta*
[2] *Professor, Department of Civil and Environmental Engineering, University of Alberta*
[3] *Professor, Department of Civil and Environmental Engineering, University of Alberta*
[*] *Corresponding author's e-mail: omohsen@ualberta.ca*

ABSTRACT
Industrialization of building construction through offsite construction and modularization is an effective approach for improving performance of construction projects. In a modular construction approach, building components are produced in a well-controlled factory environment. The components are then delivered to site, in sequence, for installation by site crews. This process reduces construction waste, improves product quality, and minimizes onsite safety incidents. As the market conditions are rapidly changing, the demand for more customized and unique products is increasing. Customers increasingly demand customized dwellings to reflect their cultural tastes and personal preferences. Cabinets in the house, kitchen or otherwise, are building components that constitute a large portion of the visible customization that customers are interested in. This paper focuses on the analysis of records in Non-Conformance Reports (NCRs) at a cabinet manufacturing facility in Alberta, Canada. An NCR record represents a defect in any product that needs a repair or rework; it captures several attributes of the defective part, such as the job number, wood species, stain, the date and time when the record is created, etc. The systematic approach presented in this study employs data analytics to the collection, cleaning, and analysis of the NCR dataset. The dataset is first analyzed as per existing operations. Various data pre-processing techniques, including attribute and instance selection and transformation, are then applied to clean the dataset. The results show that most of the "Rework" results from administrative or product handling errors, while the majority of "Repairs" result from product finishing errors. The impact of repairing the defective parts is discussed, and recommendations to reduce the number of NCRs and thereby enhance the performance of operations are presented.

KEYWORDS
Modular Construction; Industry 4.0; Data Analysis; Knowledge-based Decisions; Cabinet Manufacturing; Non-conformance Report

INTRODUCTION
With the emerging trend of automation and data utilization in manufacturing, commonly known as Industry 4.0 or the fourth industrial revolution, data science and related fields are bringing forward promising concepts that have the potential to improve our ability to make smarter and more relevant decisions, and to cope with the future vision of production that requires processes

and products to be highly mechanized, automatized, and customized. Lasi et al. (2014) provides an outline of the two development directions by which the term "Industry 4.0" is defined: application-pull and technology-push. The term originated from a set of principles and recommendations outlined in a final report by the Working Group Industry 4.0 sponsored by the Federal Ministry of Education and Research in Germany. Lee et al. (2014b), Zhong et al. (2017), and Almada-Lobo (2015) address the recent trends of intelligent manufacturing, and the industrial transformations that result from such initiatives. They reviewed and outlined key related technologies, such as internet of things (IoT), cyber-physical systems (CPS), cloud computing, and big data analytics (BDA) that are used to enable intelligent manufacturing.

In today's data-oriented business environment, survival and achievements necessitate the ability to apply the concepts of data science to specific business problems in order to improve decision making. Provost et al. (2013) argued that there is strong evidence of substantially improved business performance via the application of data-driven decision making, big data technologies, and data-science techniques based on big data. Within an Industry 4.0 factory concept, Lee et al. (2014a) proposed a CPS framework that integrates sensor data collected from floor machines with algorithms that are capable of learning from such data. Their objective is to achieve a flexible fleet-wide system that helps machines to be self-aware of their own performance, actively warn about potential issues, and support decision-making for proactive maintenance.

Data collection and processing is a very pivotal step towards establishing the smart factory of Industry 4.0. The value of any data-driven project or study depends profoundly on the quality of the data collected. Regardless of the researchers' skill level or what journal the study is published in, poor collection and management of data result in a poor study (Matteson et al., 2017). Several techniques, methods, and tools are used for data collection in manufacturing. Altaf et al. (2015) and Altaf et al. (2017) developed an online simulation-based production control system in a wall panel prefabrication factory using RFID technology. Their system evaluates production performance based on real-time data acquired by the RFID system. Shen et al. (2014) proposed a hybrid system, using RFID and wireless sensor networks (WSNs), to interactively integrate a traditional RFID system and a WSN system for efficient data collection. The proposed system utilizes smart nodes that conjugate the functions of RFID tags, RFID readers, and wireless sensors. Chen et al. (2014) provided a comprehensive overview of big data and studied the four phases of the value chain of big data: data generation, data acquisition, data storage, and data analysis. For each phase, they discussed the general background, technical challenges and latest advances. Lee et al. (2015) proposed an architecture of a context intelligence platform for industrial big data analytics to handle location, sensor and unstructured data. The proposed platform consists of five layers including presentation layer, analytical layer, data layer, infrastructure layer and IoT layer. Lee and Mohsen (2008) provided an early research review of the application of context-aware computing to construction site work. They proposed a taxonomy of context that captures the dynamic nature of construction operations.

In a modular construction context, many work packages, or group of activities, within a project can be organized into production lines to produce specific outcomes. As a particular example, kitchen cabinets are normally designed as per customer needs and manufactured in factories. The industrial partner in this research is one of the largest fully-integrated cabinet manufacturers in Western Canada. They face several challenges typical of construction manufacturing facilities that handle a wide range of product customization. These include the need to (i) develop a robust workflow to account for diverse cabinet types and sizes; (ii) analyze current production processes and identifying bottlenecks; (iii) improve and streamline existing production lines to optimize the operations of different interrelated processes; (iv) integrate state-of-the-art solutions to create a safer work environment; and (v) enhance product design to minimize material and process waste while meeting customer needs.

METHODOLOGY

Research Overview

Within the context of adopting Industry 4.0 concepts, and in order to support decision-making, the extraction of meaningful information and patterns from available data is applied in this study. Current research focuses on the analysis of records contained in the Non-Conformance Reports (NCRs) at a cabinet manufacturing facility. When there is a defect in a cabinet part that needs a repair or rework, an NCR record is created that captures several attributes of the defective part such as the job number, wood species, stain, the date and time when the record is created, the responsible department, as well as other related attributes. The objective of this study is to develop a data-driven predictive model by analyzing past records of NCRs pertaining to cabinet manufacturing operations. Such a study will assist in identifying the probabilities that certain products will have deficiencies and would require more time and cost to reproduce. This will lead to being able to recognize problems before they occur, revise possible improvements in a timely manner, as well as make future enhancements aimed at improving the whole production process. In previous research, Mohsen et al. (2018) utilized discrete-event simulation to model the floor operations at the industry partner's cabinet manufacturing facility. Rework and repairs, it should be noted, are crucial activities that affect the overall throughput of the factory. Being able to analyze NCRs can give us a better understanding of the impact of rework and repairs on the floor operations and consequently benefit the overall enterprise. Specifically, rework and repairs, account for approximately 13% of the total floor operations. Any improvement to reduce and number of NCRs will highly enhance the factory operations and benefit the business.

Research Methodology

Factory floor operations are observed in order to gain deeper understanding of the practical challenges faced by work crews in the plant. A relevant dataset is obtained from the enterprise data warehouse. The dataset is then cleaned and analyzed, and conclusions are drawn. The cleaned dataset is then used to train the machine learning predictive models used to predict future error categories or groups of defective parts. This future prognostic ability will support decision making regarding resource utilization, staff assignments to different work stations, and scheduling of material delivery to sites.

DATA PREPARATION AND ANALYSIS

Dataset Summary and Preparation

The NCR records examined in this study cover a period of 24 months starting from January 2016 until the end of December 2017. In this section, the raw dataset is overviewed and prepared. In addition, the cleaned dataset is analyzed using RapidMiner Studio software (hereinafter RapidMiner). In the next section, conclusions of the study are deduced, and recommendations are then presented.

The NCR dataset contains 11,978 records. Each record has 18 attributes as shown in Figure 1 below. Most of the attributes are of nominal values. One attribute that will be fragmented and analyzed in more detail, the "PartDesc", is a mix of text and dimensions. The attributes used to categorize the type of defect are "NCRCategory", "NCRErrorGroup" and "NCRError".

1	2	3	4	5	6
ID	JobNumber	OldJobNumber	WoodSpecies	Stain	DoorCatalog
7	8	9	10	11	12
ProductName	PartDesc	CreatedBy	CreatedDate	CreatedByDept	Reason
13	14	15	16	17	18
DeptResponsible	NCRCategory	StartingDepartment	Type	NCRErrorGroup	NCRError

Figure 1. Attributes (column headers) of NCR dataset

Data preparation and pre-processing is a very important step in analyzing the raw data. With respect to the available data, the following is a selected list of steps taken to prepare the data prior to the analysis:

1. The "ID" attribute will be used as an identification attribute and will have no effect on the analysis of the data. Other identification attributes have been discarded as they contain redundant information or have no effect on the analysis.
2. The "PartDesc" attribute is a combined field that contains a description of the defective part and its dimensions. This attribute is split into three attributes: one that describes the defective part and two attributes containing the dimensions (length and width). Figure 2 below shows how this attribute is fragmented using MS Excel and custom VB code.

Description	Dimensions	Dimensions (cleaned)	Length (inches)	Width (inches)
013B00 WDoor	20 7/8 (530.225) 29 7/8 (758.825) 3/4	20 7/8 29 7/8 3/4	20.88	29.88
CM11 Crown Moulding	3 3/4 108 0	3 3/4 108 0	3.75	108
971 Door	13 3/8 (339.725) 25 7/8 (657.225) 3/4	13 3/8 25 7/8 3/4	13.38	25.88
871a Door	14 7/8 (377.825) 29 7/8 (758.825) 3/4	14 7/8 29 7/8 3/4	14.88	29.88
Fixed Shelf	20 11/32 28 3/4 0	20 11/32 28 3/4 0	20.34	28.75
000K60-082 BDoor	14 7/8 21 1/4 3/4	14 7/8 21 1/4 3/4	14.88	21.25
Gable, Left Lg	21 26 0	21 26 0	21	26
013B30 Drawer Front	32 7/8 (835.025) 5 3/4 (146.05) 3/4	32 7/8 5 3/4 3/4	32.88	5.75
017B00 WDoor	15 7/8 (403.225) 41 7/8 (1063.63) 3/4	15 7/8 41 7/8 3/4	15.88	41.88
017B00 WDoor	15 (381) 23 7/8 (606.425) 3/4	15 23 7/8 3/4	15	23.88
017B00 WDoor	15 (381) 23 7/8 (606.425) 3/4	15 23 7/8 3/4	15	23.88

Figure 2. Result example of splitting the "PartDesc" attribute

3. Some attributes have been omitted as a large number of values are missing.
4. In addition to the steps above, attributes are discarded if any of the following apply:
   a. Descriptive attributes having many values that cannot be used to identify a specific category, nor can it be converted into another useful type of data.
   b. Identifying attributes that do not, when used as-is, add relevant information.
   c. Redundant data that is captured within other attributes.

5. Moreover, records are deleted that contain missing values, as well as records where the length or width of the defective part is equal to zero. Missing values need to be removed or substituted to allow the development of predictive models while, at the same time, not adversely affecting the model performance.
6. Nominal values are then converted to numbers using "dummy coding" technique, commonly known as binarization. In dummy coding, for all values of the nominal attribute a new attribute is created. For each record, the new attribute which corresponds to the actual nominal value of that record gets the value 1 and all other new attributes get the value 0.
7. All numeric values are then normalized using range transformation [0.0 – 1.0].

Applying the above steps, the number of records in the dataset is reduced to approximately 11,470 records (about 95.5% of the raw dataset), and the number of attributes is reduced to be 8. In the next section, observations about the dataset and recommendations are presented.

Observations and Recommendations

Reducing the number of NCRs generated during the daily operations is one of the intended objectives, with an aim to streamline operations, decrease workflow disruption, and consequently increase the daily throughput of the production line. By doing so, delivery targets will be met, the profitability of the business will increase, and overall customer satisfaction will be achieved. Recommendations based on the analysis of the NCR dataset include:

a) Restructuring the type of collected data by focusing on collecting more relevant information for each defective part. For example, the finish time required to complete the repair/rework has not been specifically recorded.
b) The description field of each data point in NCR contains both a description of the defective part and its dimensions. It is strongly recommended to record these two data into separate fields. By doing so, the extraction of data and pre-processing of the whole dataset will be more convenient and less time-consuming. Overall, it is important to standardize all data fields in the dataset for easier data retrieval and processing.

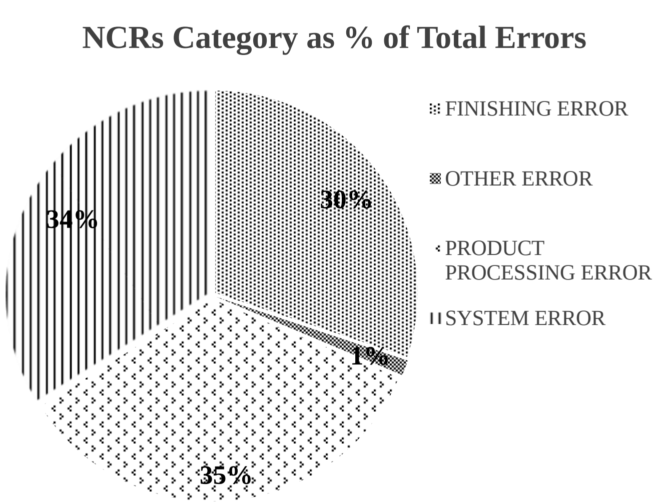

Figure 3. NCRs categories as % of total errors

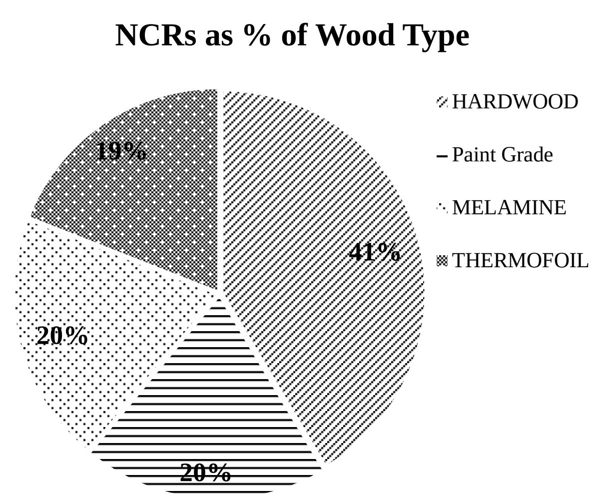

Figure 4. *NCRs as % of wood type*

c) The defective parts are found to be of two main types: rework and repairs. Most of the errors are the result of either (i) finishing error, (ii) product processing error, or (iii) system error. Most of the "Rework" results from "System" or "Product Processing" errors and the majority of "Repairs" result from "Finishing" errors. Figure 3 shows NCR categories as percentages of the total number of records.

d) The type of wood that accounts for the majority of defects is hardwood. Focusing on resolving rework/repair errors for this type of wood will significantly reduce the number of NCRs. Figure 4 shows the distribution of NCRs as a percentage of major wood types.

e) "Handling" mistakes caused most of the "Product Processing" errors, accounting for 43% of the total errors under this category. It is recommended to investigate and apply better handling practices of the material. This might require rearranging and/or regrouping work stations, revising the floorplan layout, and applying more efficient strategies to move the raw and finished material around the floor. "Administration" mistakes caused most of the "System" errors, accounting for 67% of the total errors under this category. It is recommended to investigate how the administration workflow and practices can be improved. This includes the flow of information from the office to the floor, procedures carried out within the office to convert the customer order into floor job, and conversion of data between different software platforms.

f) There is no apparent correlation between the number of NCRs and the time of the year, although there is a cyclic pattern that occurs over the time periods covered by the dataset. Figure 5 shows that the lowest number of NCRs occur during the month of June.

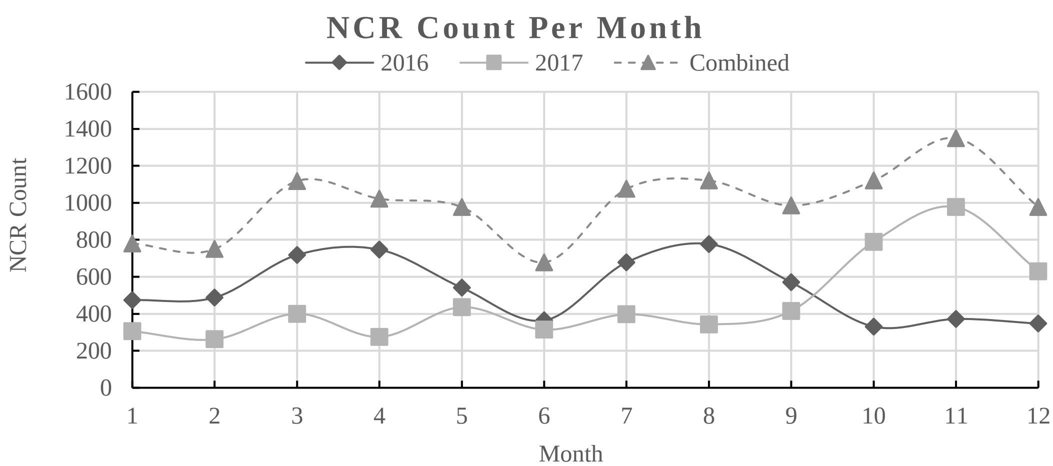

Figure 5. NCR count per month, 2016-2017

Data Analysis and Predictive Models

After the raw dataset has been preprocessed and transformed to the proper format, various machine learning models are trained and tested with the objective of creating a predictive tool that can be used to forecast the future categorization of defective parts. In the current study, a few predictive classification models are developed using RapidMiner, a commercial data mining software. From the variety of classification models available in literature, we opted to select the following simple, yet powerful, models: k-nearest neighbours (kNN), Naïve Bayes (NB), and Decision Tree (DT). Being able to forecast the category of an NCR record before a defect occurs will significantly increase our ability to plan for the rework or repair that would otherwise be less predictable. The impact of rework and repairs can significantly affect the overall operations, it should be noted, as discussed by Mohsen et al. (2018), that ignoring the rework in the simulation model can degrade the result by at least 15%. Including the rework cycle in the simulation model, meanwhile, increases the accuracy of the results, where the total number of days to complete an order is found to be approximately 14 days (which matches the actual duration). However, if the rework cycle were to be skipped in order to simplify the model, the number of days to complete the order is predicted by the model as 11.5 days, which is not representative of the actual duration. Therefore, forecasting the future probabilities of requiring a rework or repair can enhance the simulation model results.

CONCLUSION

The fourth industrial revolution, also known as Industry 4.0, has become more popular and more widely implemented due to recent advances in cyber-physical systems, big data analytics, cloud computing, and device and machine interconnectivity. In this paper, we have focused on utilizing a big data solution in a cabinet manufacturing facility, where data pertaining to NCR has been studied and analyzed. The objective of the analysis is to use data-driven predictive algorithms to forecast the category of a defective part of a cabinet. Predicting the category of a defective part can assist in making decisions about resource utilization, staff assignments to different work stations, and scheduling of finished product delivery to sites. Moreover, having an accurate estimate of the probabilities of occurrence of rework and repairs greatly enhances the accuracy of any simulation model used to imitate the floor operations at a cabinet manufacturing factory. Future research on this topic will focus on testing more complex classification models in order to predict not only defective parts, but also other aspects of construction manufacturing. In future research, the outputs of a BDA study will be used as inputs to various parts of the simulation model. The goal is to have real-time data-driven parameters which will enhance the accuracy and robustness of the model and consequently improve our ability to make better informed decisions about the overall business operations.

ACKNOWLEDGEMENTS

This research was funded through an Engage Grant (File No. EGP 514475-17) from the Natural Science and Engineering Research Council of Canada (NSERC). Special thanks are extended to the industry collaborator for their support throughout this study.

REFERENCES

Almada-Lobo, F. (2015). “The Industry 4.0 revolution and the future of Manufacturing Execution Systems (MES).” *Journal of Innovation Management*, 3(4), 16-21. <http://hdl.handle.net/10216/81805>

Altaf, M., Bouferguene, A., Liu, H., Al-Hussein, M., and Yu, H. (2017). “Integrated production planning and control system for a panelized home prefabrication facility using simulation and RFID." *Automation in Construction*, 85 (2018), 369-383. <https://doi.org/10.1016/j.autcon.2017.09.009>

Altaf, M., Liu, H., and Al-Hussein, M. (2015). "Online Simulation Modeling of Prefabricated Wall Panel Production Using RFID System." Proceedings of the 2015 Winter Simulation Conference, 3379-3390. December 6-9, Huntington Beach, CA, USA. <https://doi.org/10.1109/WSC.2015.7408499>

Chen, M., Mao, S., and Liu, Y. (2014). "Big data: A survey," Mobile Netw. Appl., 19(2), 171-209. <https://doi.org/10.1007/s11036-013-0489-0>

Lasi, H., Fettke, P., Kemper, HG., Fled, T., and Hoffmann, M. (2014). "Industry 4.0." Bus Inf Syst Eng, 6(4), 239-242. <https://doi.org/10.1007/s12599-014-0334-4>Lee, C.K.M., Yeung, C.L., and Cheng, M.N. (2015). "Research on IoT based cyber physical system for industrial big data analytics," Proceedings of the 2015 IEEE Int. Conf. Ind. Eng. Eng. Manage., Dec. 2015, 1855-1859. <https://doi.org/10.1109/IEEM.2015.7385969>

Lee, J., Kao, H., and Yang, S. (2014a). "Service innovation and smart analytics for Industry 4.0 and big data environment." Proceedings of the 6th CIRP Conference on Industrial Product-Service Systems, 3-8. <https://doi.org/10.1016/j.procir.2014.02.001>

Lee, J., Bagheri, B., and Kao, H. (2014b). "Recent Advances and Trends of Cyber-Physical Systems and Big Data Analytics in Industrial Informatics." Proceedings of Int. Conference on Industrial Informatics (INDIN), Porto Alegre, Brazil. <https://doi.org/10.13140/2.1.1464.1920>

Lee S., Mohsen O., (2008). "Application of Context-Aware Computing to Construction Field Work." Proceedings of the CSCE 2008 Annual Conference, 1(2), 483-490. June 10-13, Québec City, Québec, Canada.

Matteson, S., Sherrod, S., and Cetin, S. (2017). "Collection, storage, protection, and sharing issues with large-scale data sets." SAGE Research Methods Cases. <https://doi.org/10.4135/9781473999053>

Mohsen, O., Abdollahnejad, S., Sajadfar, N., Mohamed, Y., AbouRizk, S. (2018). "Modeling and Simulation of Cabinet Manufacturing Processes: evaluation and recommended controls." *Proceedings of the 17th International Conference on Modelling and Applied Simulation (MAS)*, September 17-19, Budapest, Hungary.

Provost, F., and Fawcett, T. (2013). "Data Science and its Relationship to Big Data and Data-Driven Decision Making." Big Data, 1(1), 51-59. <https://doi.org/10.1089/big.2013.1508>

Recommendations for implementing the strategic initiative INDUSTRY 4.0. Final report of the Industry 4.0 Working Group (2013). Accessed on January 02, 2019. <https://www.acatech.de/Publikation/recommendations-for-implementing-the-strategic-initiative-industrie-4-0-final-report-of-the-industrie-4-0-working-group/>

Shen, H., Li, Z., Yu L., and Qiu, C. (2014). "Efficient data collection for largescale mobile monitoring applications." IEEE Trans. Parallel Distrib. Syst., 25(6), 1424–1436, Jun. 2014. <https://doi.org/10.1109/TPDS.2013.122>

Wan, J., Tang, S., Li, D., Wang, S., Liu, C., Abbas, H., and Vasilakos, A.V. (2017). "A Manufacturing Big Data Solution for Active Preventive Maintenance." in *IEEE Transactions on Industrial Informatics*, 13(4), 2039-2047, Aug. 2017. <https://doi.org/10.1109/TII.2017.2670505>

Zhong, R.Y., Xu, X., Klotz, E. and Newman, S.T. (2017). "Intelligent Manufacturing In The Context Of Industry 4.0: A Review" Engineering, 3(5), 616-630. <https://doi.org/10.1016/J.ENG.2017.05.015>