ISSN: 2063-5346
Section A-Research paper

# Use of Machine Learning for Digital Manufacturing – Demonstration on an Industrial Use Case

**[1]Rishabh Agarwal, [2]Shrikant Bhat, [3]Rupesh Khare, [4]Supriya Singh**
ABB India
[1]rishabh.agarwal@in.abb.com, [2]shrikant.bhat@in.abb.com, [3]rupesh.khare@in.abb.com,
[4]supriya.singh@in.abb.com

***Abstract*— In Engineer to order (ETO) business, on-time project execution is a complex endeavour. This complexity is multiplied by customized engineering drawings, delay in customer approvals, heavy document and change management and mainly due to the participation of multiple stakeholders. This introduces high level of uncertainty during execution and impacts total throughput time (TTPT) and on time delivery (OTD) often resulting in lower inventory turnover ratio (ITR). The major factors contributing to uncertainty comprise type and rating of products, components used, order type, customer segments, resources involved, suppliers, production lines, contract terms and conditions, etc. Moreover, these factors have different impact across different phases of order management from marketing till dispatch. In this study, a Machine Learning based model is used considering more than 40 variables, both numeric and textual, spanning all the manufacturing phases. As a first step, the key driver analysis determined the likelihood of the delay with certain sales channel, component type, product variant, complexity of job, number of panels, and sales & project engineer levels. This is followed by developing a Supervised Learning Algorithms which identified the clusters based on order specifications and associated propensity of delay in servicing the orders. The accuracy of the models varied from 73 to 77%. These preliminary findings are not only promising to establish confirmation on some of the intuitive findings, but these also help in initiating operation excellence projects on many other important but non-intuitive findings.**

***Keywords*— machine learning; manufacturing analytics; MES; Engineered to order production**

## I. Introduction

One of the major challenges in applying machine learning (ML)/Artificial Intelligence (AI) algorithms for manufacturing analytics was unavailability and harmonization of data available across different phases as well as tiers of manufacturing. This problem is now circumvented to a large extent due to rapid spread of digitalization triggered by cheaper sensors, wider coverage of internet and increased management awareness about the impact of ML/AI towards decision making.

This publication captures one such use case wherein manufacturing analytics is applied to ABB's Engineered to Order business. The manufacturing workflow in such case covers the value chain data starting from marketing and sales channel getting the customer order to the logistics delivering the given order. Between these two functions the value chain covers important functions such as order handling, drawing approval, supply chain and manufacturing. Various specific factors such as customized engineering drawings, delay in customer approvals, heavy document and change management introduce complexity in managing such orders which in-turn impacts impacts total throughput time (TTPT) and on time delivery (OTD) often resulting in lower inventory turnover ratio (ITR).

This study focused on use of machine learning approaches to analyze the data across the manufacturing value chain and derive intelligence to support the management improve production planning and help in reducing TTPT while improving OTD and ITR.

## II. Problem Statement

### A. *Engineered to order manufacturing under consideration*

ETO manufacturing involves the creation of an entire product that is custom made to the desired specifications of a customer. The workflow starts with a customer order and spans through all the functions listed in Figure 1 to deliver the required product to the customer. In the scenario under evaluation, data is captured across the entire value chain that involves customer name, specific product type, components, leads from functions handling the customer order, region and sales channel, etc. For each order, the delay across various functions is captured.

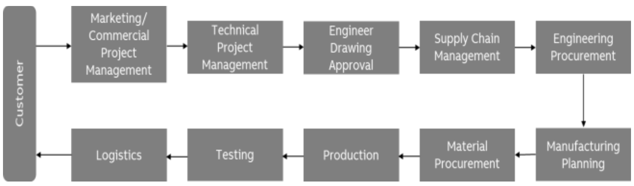


*Figure 1. Typical ETO manufacturing workflow*

### B. *Manufacturing analytics problem*

There are 10 different functions which are part of the sequential order handling process. At each step different variables have an impact on the on-time delivery of the

*ISSN: 2063-5346*
*Section A-Research paper*

order. The variables under consideration in the study (~40) are having attributes such as Type of equipment, variant, sales channel, complexity, payment terms, delivery terms, the resources involved, etc.

The challenge is to analyze at each stage of manufacturing the impact of variables that cause the most delay to on-time completion of order.

## III. Solution Approach

### A. Data Pre-processing and EDA

Historical data from 2018 was available for the study. A few variables contained missing values. After consulting with the domain expert since the number of missing values was small and there was no good method to fill the missing values, they were dropped from the analysis. After pre-processing 1605 data points remained in the study.

Each function has some key activities associated with it that are tracked during the manufacturing process. Each activity has a planned completion date which is generated after receiving the purchase order from the customer and it is based on the complexity and characteristics of the order. The actual date of completing the activity is also recorded. If the actual date of completion is on or before the planned date, then the order is on-time. If the actual date is later than planned date, then the order is delayed. This logic is used to make a categorical variable called “Delay Type” for each key activity of the functions.

Hence our problem can be defined as a classification problem with 40 variables as independent variable and the Delay Type binary variable as the dependent variable. This relationship between dependent variable and independent variable is investigated for each functional activity.

Exploratory Data Analysis is performed on the data to understand the main patterns and bivariate relationships between the variables.

### B. Proposed Model

As discussed in the previous section, our problem can be defined as a classification problem. Except two variables the other independent variables are categorical variables. Our dependent variable is a binary categorical variable identifying whether the order was delayed or on-time. We use decision tree for this problem statement. The objective of the decision tree is to end up with a subgroup of data points that's relatively high in the metric one is interested in. In the current study this metric is orders which are delayed or on-time. The decision tree takes each explanatory factor and tries to reason which factor gives it the best split. After the decision tree does a split, it takes the subgroup of data and determines the next best split for that data.

## IV. Results and Discussion

This section presents the results obtained through exploratory data analysis and using decision trees to identify the key drivers. The decision tree accuracy metric on test dataset is shown below.

*Table 1. Classifier Accuracy Metrics*

| *Classifier* | *Accuracy* | *Precision* | *Recall* |
|---|---|---|---|
| Decision Tree | 75% | 0.76 | 0.96 |

Based on exploratory data analysis we found some interesting patterns.

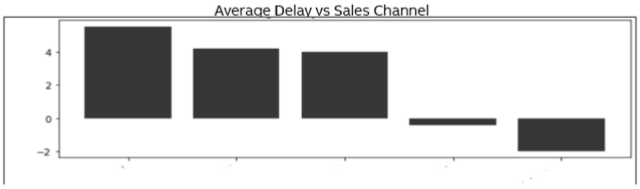


*Figure 2. Average delay of orders vs Sales Channel*

We see the average delay for an order varying significantly depending on the sales channel. Some channels have a negative average delay.

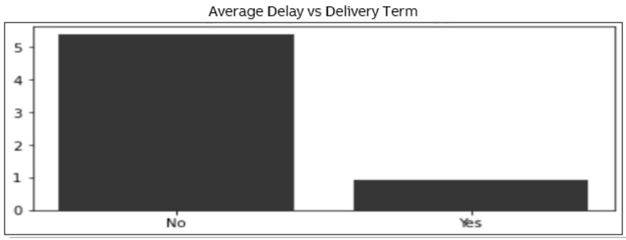


*Figure 3. Average delay of orders vs Delivery terms*

There was a clear difference in the average delay of orders depending on the Delivery Term.

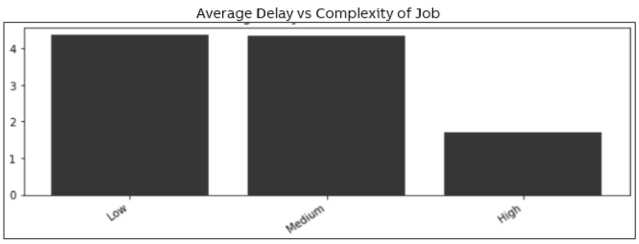


*Figure 4. Average delay of orders vs complexity of job*

During order handing over we observed that High complexity orders on average had smaller delays compared to low and medium complexity orders.

| Likelihood of getting Delayed | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Functional activities | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| Sales Channel | XYZ 1.46 times | XYZ 1.21 times | XYZ 1.11 times | XYZ 1.08 times | - | - | PQR 1.15 times | PQR 1.24 times | - | ABC 1.38 times | PQR 1.15 times |

*Figure 5. Likelihood of order delay across functions*

In a similar manner, we investigated for each function the key drivers of delay. It was observed that during the initial functional activities orders through sales channel XYZ are

*ISSN: 2063-5346*
*Section A-Research paper*

the most likely to be delayed. The likelihood of delay decreases as the order proceeds through the function. During the later stages in the order execution process, we see orders belonging to sales channel PQR and ABC tend to have a higher likelihood of delay. The likelihood values are calculated as a ratio of percentage of orders delayed through a channel to the overall order delayed percentage.

Decision trees also help us to identify sub-groups of order that tend to be delayed or on-time.

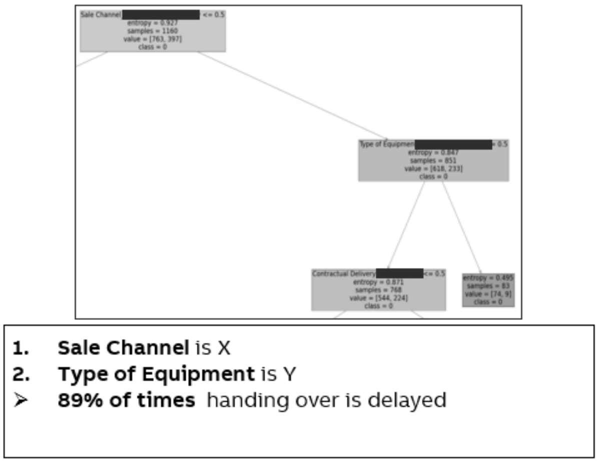


*Figure 6. Order segment that tends to be delayed*

One segment of orders that was identified by the decision tree was where Equipment Type Y was sold through Sale Channel X. 89% of orders were delayed in this order segment. This multivariate relationship provides key insights to the business.

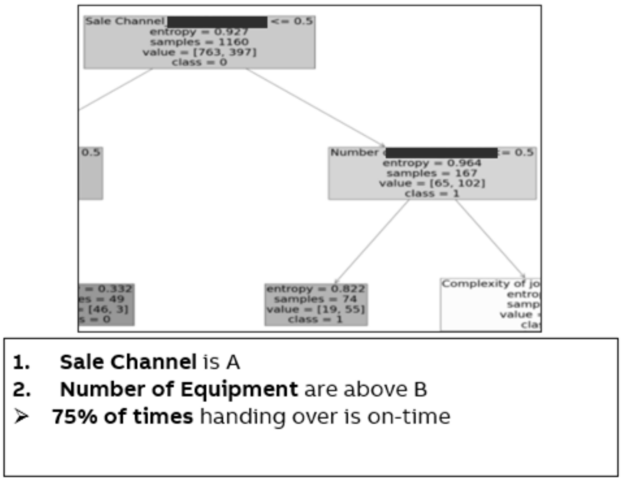


*Figure 7. Order segment that tends to be on-time*

Another sub-group of orders was identified where 75% of the orders were on-time. This group consisted of orders sold through Sales Channel A and Number of Equipment ordered were above B. This was counter intuitive to the existing belief since it was assumed that orders with high quantity of equipment should be more prone to delays.

## Acknowledgment

We would like to acknowledge the support from Mr. Ranjit Koppikar, Product Group Manager, ELDS, ABB, Nashik, India for providing this opportunity.

## References

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] Hashmia Hamsa, Simi Indiradevi, and J.Kizhakkethottam Jubilant, "Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm.," Procedia Technology, vol. 25, pp. 326-332, 2016.

[3] *Myles, Anthony J., et al. "An introduction to decision tree modeling." Journal of Chemometrics: A Journal of the Chemometrics Society 18.6 (2004): 275-285. (references)*