Modeling Manufacturing On-Time Delivery

by

Jennifer Collins

B.S. Aerospace Engineering,
University of Illinois at Urbana-Champaign, 2013

Submitted to the MIT Sloan School of Management and the Aeronautics and Astronautics Department in partial fulfillment of the requirements for the degrees of

Master of Business Administration and
Master of Science in Aeronautics and Astronautics

in conjunction with the Leaders for Global Operations Program

at the
Massachusetts Institute of Technology
June 2019

© 2019 Jennifer Collins. All rights reserved.

Signature redacted

MIT Sloan School of Management
Department of Aeronautics and Astronautics
May 10, 2019

Nikos Trichakis
Associate Professor of Operations Management, MIT Sloan School of Management
Thesis Supervisor

Paulo C. Lozano
Professor of Aeronautics and Astronautics
Thesis Supervisor

Sertac Karaman
Associate Professor of Aeronautics and Astronautics
Chair of the Graduate Program Committee

Maura Herson
Assistant Dean, MBA Program
MIT Sloan School of Management
Modeling Manufacturing On-Time Delivery

by

Jennifer Collins

Submitted to the Department of Aeronautics and Astronautics and the MIT Sloan School of Management on May 10th, 2019 in partial fulfillment of the requirements for the degrees of Master of Science in Aeronautics and Astronautics and Master of Business Administration

Abstract

Historically, Company X has undertaken multiple projects to improve On-time Delivery (OTD) goals but not all of these have resulted in significantly improved performance. As a government contractor, Company X is evaluated on OTD as a contractual performance metric. This project established the foundations for a real-time predictive and diagnostic machine learning model for OTD. A real-time machine learning model could correct negative impacts to the manufacturing system before they impact the customer. It would also provide an understanding of how corrective measures are expected to affect outcomes.

To begin, candidate variables correlating to OTD were considered and the relevant datasets were gathered. These spanned the enterprise, including Bill of Materials (BOM), material resource planning, supply chain, production, quality, and test datasets. Each dataset was evaluated to assess data quality, using a framework which considers accuracy, reliability, timeliness, completeness, comprehensiveness, accessibility, and availability. Datasets were also examined for missing data, with recommendations for new data collection that may improve future iterations of predictive models.

The datasets were integrated to reflect the time-phased dependencies of the hierarchical BOMs. This dataset structure best represents the manufacturing reality. The integrated dataset was aggregated on the basis of BOM indenture-levels, due to data sparsity.

To maximize understanding and interpretability of this proof-of-concept model, the machine learning methods considered were limited to decision trees. Given a total training and testing population of 207 deliveries, the model achieved an accuracy (F1 Score) of 86% and RMSE of 39%.

While preliminary modeling shows promise for future models, a number of issues need to be addressed: expansion of training and testing datasets, improvements in data quality, gathering missing data, and implementing IT systems better suited for accessing large datasets. Once improvements can be made in these areas, a true real-time predictive model for OTD may be a possible solution for Company X.

Thesis Supervisor: Nikos Trichakis
Title: Associate Professor of Operations Management
MIT Sloan School of Management

Thesis Supervisor: Paulo C. Lozano
Title: M. Aleman-Velasco Professor
Department of Aeronautics and Astronautics
The author wishes to acknowledge the Leaders for Global Operations Program for its support of this work.
Acknowledgments

I would like to express my sincere appreciation for all of the people involved in this work.

First, my project team at my host company helped me understand such a complex organization and were eager teachers. My project supervisor, Dave was never afraid to dive deep into the data and his operational knowledge rivals MIT professors. Kim, Jeff, and Toni took the time to help me pull data, code, debug, and model. Kristyn wrote me a hundred queries. Christy took the time to understand “machine learning” and lent her expertise to the project. Many other people at Company X answered endless questions with patience and kindness.

I would also like to thank my MIT advisors, Professors Nikos Trichakis and Paulo Lozano, who advised on both the technical aspects of the project, but also the realities of writing a thesis.

A big thank you to my LGO family, whom I care deeply for and feel supported by both in and out of the classroom.

Thank you to my parents, who have believed in me since day one, and never doubted that a girl from small town Illinois could become an MIT astronautical engineer.

The final thank you is for my friend, colleague, and editor- my husband. He has made many sacrifices to support my education and career and I would not have been able to do any of this without his love and support.
THIS PAGE INTENTIONALLY LEFT BLANK
## Contents

Acknowledgments ..................................................................................................................... 5
Nomenclature .............................................................................................................................. 8

1. Introduction ............................................................................................................................ 9
   1.1. Thesis Objective .................................................................................................................. 9
   1.2. Project Motivation ............................................................................................................. 9
   1.3. Industry Background ......................................................................................................... 11
   1.4. Company Background ..................................................................................................... 13
   1.5. Technical Approach .......................................................................................................... 14
   1.6. Model Performance ......................................................................................................... 15

2. Literature Review ................................................................................................................... 16
   2.1. Predictive Analytics in Operations ..................................................................................... 16
   2.2. Data Quality ..................................................................................................................... 17

3. Methodology .......................................................................................................................... 19
   3.1. Business Understanding .................................................................................................... 19
   3.2. Data Understanding ......................................................................................................... 21
   3.3. Data Preparation ............................................................................................................. 25
      1. Rationale for Inclusion or Exclusion ................................................................................... 25
      2. Dataset Integration Methodology ...................................................................................... 29
      3. Dataset Integration Example .......................................................................................... 32

4. Model Performance ................................................................................................................ 34
   4.1. Regression Decision Trees ............................................................................................... 34
   4.2. Classification Decision Trees ........................................................................................... 36

5. Recommendations & Conclusions ......................................................................................... 38
   5.1. Short Term Recommendations .......................................................................................... 38
   5.2. Long Term Recommendations .......................................................................................... 39
   5.3. Conclusions ...................................................................................................................... 40

6. Appendix .................................................................................................................................. 42
   6.1. List of Potential Independent Variables ............................................................................. 42
   6.2. Example Data Exploration Report ...................................................................................... 44

References .................................................................................................................................... 46
Nomenclature
CLIN: Contract Line Item Number
DP: Deliverable Product
EO/IR: Electro-Optical / Infrared
IT: Information Technology
MBOM: Manufacturing Bill of Materials
MRP: Material Resource Planning
MS: Make Span
OTD: On-time Delivery
PO: Product Order
PR: Product Requisition
PT: Production Time
QN: Quality Notification
RF: Radio Frequency
TPut: Throughput
WIP: Work in Progress
WO: Work Order
1. Introduction

1.1. Thesis Objective

This thesis serves as a proof-of-concept model for predicting on-time delivery (OTD) by using machine learning methods on operations data across a business unit. It serves as the foundation for the long-term goal of creating a real-time predictive OTD model at Company X, a government contractor. A predictive model could indicate necessary changes in order to meet OTD on a programmatic and / or factory basis. A real-time machine learning model could correct negative impacts to the manufacturing system before they impact the customer. It would also provide an understanding on how corrective measures are expected to affect outcomes.

The dataset structure takes into account the time-dependent nature of the build process, with special considerations for the complex systems with thousands of parts and lead times of up to 2 years. It also evaluates data quality, determines new data to be collected, and makes recommendations for improving data systems.

1.2. Project Motivation

While the trendiness of machine learning and big data analytics deserves to be acknowledged, this project was borne out of careful consideration for the final predictive outcome, complexity of production, and Company X's long-term goals. OTD is an essential metric both internally and externally for Company X. The business unit also has a unique set of operational characteristics that increase complexity of normal operations. Finally, applying machine learning to operational data across disparate datasets helps poise Company X for the next wave of transformational, data-driven production technologies.

Organizations which contract with the federal government are subject to Federal Acquisition Regulations (FAR), which require the collection of supplier performance information [1]. One of these performance metrics is the contractor's adherence to the required delivery schedule and is known at Company X as OTD. This key business unit metric is used to evaluate engineering, production, spares, and service deliverables. Externally, performance to OTD goals can impact future contract awards, as well as reputation with the customer. Internally, performance to OTD goals can impact management of working capital, operating profit, and production schedules.
OTD is calculated internally on a programmatic basis and aggregated to a business-unit level metric. The goal for OTD is typically in excess of 90% and high-volume programs with many contract lines can skew the aggregated metric. While it is meant to measure the operational execution on a contract, other factors may play a role in influencing outcome, especially contracts. For example, Company X may sign a contract which promises a delivery within the lead time of the final product. In order to complete a final delivery within lead time, production schedules need to be accelerated, which can be difficult when some of the longest lead time parts come from suppliers. OTD is a far-reaching metric that impacts both the customers, suppliers, and internal production, so it is important to manage on both a programmatic and business unit-wide basis.

Over the last decades, Company X has undertaken multiple projects to improve OTD goals but not all of these have resulted in significantly improved performance. Company X’s distinct set of operational characteristics has made this problem especially challenging, including high performance designs, unique capabilities, high mix / low volume production, and limited supply chains. Historically, achieving high performance was the number one priority for products; however, within the last few decades, operations and affordability have come under increased scrutiny. Furthermore, Company X is not in a commodity industry; it is one of only a handful of companies in the entire world that can make such high-precision and performance products. In certain areas, arguably no one else can make competitive products. Production also occurs in much lower volumes with higher mixes, using products that can have lead times of multiple years. Finally, the supply chain of Company X adds an additional layer of complexity, with a variety of sole source suppliers and often limitations to domestic companies for their components. All of these factors contribute to an extremely complex production environment, where “simply” changing one metric can be a significant challenge.

In the past, Company X has undertaken projects to better understand the driving forces for OTD. The company has analyzed regression models and correlation tables to see which most significantly impact OTD. However, limited conclusions have been drawn from this analysis. The amount of production data collected at Company X is impossible for a single person to analyze using current methods. Datasets are rarely combined across sources due to the significant effort required, and analysis can be stovepiped. For example, quality professionals look at the quality data, while test professionals look at the yield data. With historical analysis unable to provide sufficient holistic understanding of OTD behavior and dynamics, a new methodology was necessary.
In order to better understand OTD dynamics, as well as to drive innovation, a machine learning model was desired to predict OTD. Company X has undergone significant efforts in the past few years to clean and improve operations data. Company X wanted to explore the potential of combining data from disparate sources and systems. While financial benefits of creating this type of model are long term, this effort provides another step towards becoming a data-driven enterprise, where data can effectively be turned into insights. It also helps usher in the fourth industrial revolution at Company X, better known as Industry 4.0, which is expected to help gain greater efficiencies and transform relationships between humans and machines [2]. One of the key pillars of Industry 4.0 is Big Data and Analytics, including the real-time collection and comprehension of data from disparate sources.

In conclusion, this project was created to understand a potential application of predictive analytics on operations data at Company X on a metric that deeply affects production and sales. The sheer complexity of operations at this technology company makes for a great machine learning candidate, where ordinary models would not suffice for such a large scope.

1.3. Industry Background

Being a part of the defense industry instills a strong sense of mission to many companies and their workers, as well as a complex set of contractual requirements and regulations. Companies are understandably restricted to selling weapon systems to the US Government and its allies. While making products related to national security, there are also a set of threats presented from rogue hackers, as well as those sponsored by foreign governments. This puts additional emphasis on data security, which can act as an opposing force to data science priorities.

Occasionally, contracting requirements and vehicles can promote behavior which is antithetical to typical operating principles. For example, the US Government may often buy a company’s components inventory, so a company is not motivated to invest in safety stock. Purchasing inventory is generally done on a programmatic basis, with its own accompanying contract, so it can be very complex to pool inventory, even if components or subassemblies are common across programs. The premium placed on performance for weapon systems in the defense industry can also come at the cost of manufacturability. A weapon system with the highest performance in the world is often not designed with manufacturability at the forefront of design engineer’s mind. Yield issues that may arise in production are generally compensated for in the contract. In 2010, the Department of Defense (DOD) created Better Buying Power in order to
increase affordability and cost control, as well as other best practices to strengthen buying power [3]. This initiative demonstrated a shift in the DOD’s philosophy to more forcefully promote efficient and lean operations.

The defense industry must put a high importance on Information Technology (IT) security, especially considering their intellectual property on weapon systems. Data security has continued to be of interest to defense contractors considering. “The more than decade-long campaign by Chinese intelligence targeting U.S. defense companies, in an effort to help close the technology gap as China tries to become a U.S. military peer.” [4] Publicly known attacks include 2007 hack of Lockheed Martin’s F-35, “stealing data on the plane’s design and electronics systems” [5]. In 2013, a confidential report published by the Defense Science board released, “Designs for many of the nation’s most sensitive advanced weapons systems have been compromised by Chinese hackers” [6]. In 2018, Chinese government-sponsored hackers “stole hundreds of gigabytes of data related to sensitive Navy undersea warfare programs from a government contractor” [7]. While publicly released attacks have focused on designs, many defense companies have taken up increased security around operations data, as well. Their data postures have become increasingly defensive, which has made gaining access to multiple data systems across an enterprise challenging for many good-willed data scientists.

Data access and aggregation in the defense industry is also complicated by different levels of classification. These levels of classification including Top Secret, Secret, Confidential, and Unclassified. While classified information has much more stringent security requirements, there can still be complications with using unclassified data. For example, Classification by Compilation dictates that compiling information that is individually unclassified may in some circumstances, become classified [8]. Data scientists must work with care when aggregating and integrating datasets in the defense industry.

Big Data analytics has many applications in defense and security, including change identification in drone video surveillance and organized crime prevention by scanning / monitoring social media [9]. However, there is still much to be done in using these techniques internally on operations data. By learning to navigate IT environments with security requirements and understanding the unique operational aspects of the defense industry, operations data can be more thoroughly understood.
1.4. Company Background

Company X is a multiple business-unit company that sells a variety of defense products and services. As a technology company, they design, manufacture, and support weapon systems and defensive systems, with products deployed globally. Their business unit represents multi-billion dollars of sales.

As a heritage company, with a significant number of acquisitions, Company X has a complex IT environment. Considerable effort and resources have been put into IT systems to gain more commonality, not only across the business unit, but also the enterprise. As a large organization of over 10,000 workers, many people depend on the data that is generated every day. Changing a small aspect of a data system may have far-reaching consequences for the people responsible for finding and analyzing data. As a business unit, there is also an ongoing tug of war between corporate and business level requirements and needs for IT systems.

Company X has put significant effort in the past few years to create a culture of data, especially for operations professionals. Many of the assembly operations are manual, so deliberate effort has been put forth to get reliable, accurate data for both automated and manual processes. Business-unit level wide metrics are available on a business-unit-wide website, to compare and contrast various production metrics across programs. While the company has always been driven by performance data of its products, it is working to make the entire workforce aware of operations and important metrics like OTD.

Similar to wider defense industry trends, Company X has a focus on data security. Separate access permissions are required for disparate databases. Data sources are often separate due to the long history of the company and its evolution with IT technology, as well as for containing any unauthorized access. Strong communication skills are required to gain access to the necessary databases. There is also caution in sharing presentations which contain data or distributing certain datasets across programs.

In summary, Company X is a technology-focused company in the defense industry with special considerations around data and access. It has worked to become a more data driven company and has a complex IT environment, as a heritage company with manual production processes.
1.5. Technical Approach

In order to create a predictive model for OTD, a methodology from the Cross-Industry Standard Process for Data Mining (CRISP-DM) was adapted. This includes business understanding, data understanding, data preparation, modeling, and evaluation [10].

The first objective was to gain a firm understanding of the business case and assess the available data. Luckily, this project’s primary goal was to explore the art of the possible with Company X data and to set the foundation for a potential long-term return on investment. However, it was still important to understand the dependent variables that would provide the most value for Company X to predict. A list of desired independent variables was created, each that was hypothesized to potentially effect OTD. Each variable was ranked in priority and its data sources determined. Contingency plans were created in case of unavailable or missing data for both dependent and independent variables.

Using the list of desired variables and their corresponding data sources, an understanding of the data was developed. Raw datasets were downloaded for each desired variable. The goal was to capture the entire lifecycle of the first available delivery data, so the datasets reached back to the longest lead time of the first OTD data point for all datasets. For example, the OTD data was collected from January 2018 through October 2018 and the longest lead time is 2.3 years, so data was desired to be collected from September 2015 through October 2018. For each dataset collected, exploratory analysis was performed, including understanding each data field, plotting histograms, and looking for data trends over time. A data quality evaluation framework was created and applied across the variety of available datasets.

Once the datasets were collected, the data was prepared for modeling. Using RStudio, the data was first cleaned. Differing levels of data cleaning was necessary, since many of the datasets were pulled from disparate sources. Derived variables were formulated and calculated, as necessary. To downselect from the list of ~50 potential independent variables, a mix of correlation matrices and operational knowledge were used to select a final seven variables to input to the model. All of the selected independent variables were normalized, and missing values were assigned a value to indicate normal operations. The datasets were integrated using a Manufacturing Bill of Materials (MBOM) structure, which incorporated the time-phased nature of building the product. This integration technique is reviewed in detail in Section 3.3.2.

After the data preparation was complete, the dataset was modeled and evaluated using decision trees, for both regression and classification. It was also desired to choose a machine learning model that
maximized understandability by non-data scientists, who are the most likely end users of a predictive product. The model was evaluated on accuracy and error.

At the outset of the project, well-seasoned Company X data scientists warned that business understanding, data understanding, and data preparation would consume 80% of project time, while the actual modeling and evaluation would consume the remaining 20%. This mismatch in time was even more exaggerated in the execution of this project, with ~1% of project time spent on modeling and the rest on all of the preparation. This is likely partially attributed to the learning curve to navigating the data environment at Company X, which is discussed in further detail in Section 3.1. Section 5 includes recommendations to decrease the time spent on data preparation.

1.6. Model Performance

The model has been trained and tested on 207 unique deliveries ranging from January 2018 through October 2018. In order to test the model accuracy, data was partitioned into training (70%) and test (30%) sets, considering both continuous and discrete outcomes. For the OTD metric delivered to the government, the outcome is considered binary. So, a partial delivery is considered to be a 0% OTD (e.g., if three units were delivered out of a contractual obligation of four units, the delivery would be considered 0% OTD). However, for manufacturing and planning purposes, it is beneficial to know of a partial delivery, so a continuous outcome between zero and one was also considered (e.g., if three units were delivered out of a contractual obligation of four units, the delivery would be considered 75% OTD).

Decision trees were chosen in order to maximize understandability of the final model, for which end-users are not expected to be data scientists. Regression Trees yielded an out of sample F1 Score (accuracy) of 86% and an RMSE of 0.39. Classification Trees yielded an accuracy of 76% with 19% Type 1 Error and 5% Type 2 Error.

The model and dataset outcomes require further refining before implementation. One of the most glaring needs is for a larger dataset. Recommendations for improvement will be discussed in depth in Section 5. However, this model was a successful proof-of-concept, demonstrating the possibility to predict OTD using data from disparate sources across the business.
2. Literature Review

As predictive modeling and big data analytics becomes more prevalent throughout businesses, it is important to remember a few foundational principles. First, a model is merely a representation of reality, it is not reality [11]. All models require quality data to create quality outputs, and the time required to gather and prepare the data is often underestimated [11]. Despite the potentially significant commitment big data models require, the products from quantitative analysis provide insight on “how to best compete in the fast-paced, data-rich, high coupled business enterprise of the twenty-first century” [11].

There is evidence that using big data strategically improves performance. In a Harvard Business Review study across 330 public North American companies, those using data-driven decision making were, on average, 5% more productive and 6% more profitable than their competitors [12]. Operations research can be used to support big data analytics in operations and supply chain data [11]. This thesis follows a variety of previous work from predictive models in operations and data quality metrics.

2.1. Predictive Analytics in Operations

Previous research has been performed in predictive analytics for manufacturing and operations, including delivery in the trucking industry, identifying high performance factories, and addressing risk in supply chains.

In Alcoba’s and Ohlund’s research on prediction OTD in the trucking industry, they built a logistic regression model on a historical dataset given six explanatory variables [13]. The research took place at Coyote Logistics, a third party logistics provider. These variables were determined by interviewing supply chain professionals within the organization, categorizing these variables into six distinct groups, and creating a fishbone diagram of potential variables. The final dataset used stratified sampling to force the model to use a dataset with equal number of observations (50% on time and 50% delayed) [13]. Since the dataset had only 26,146 delayed data points, the dataset maintained these observations and randomly selected 26,146 on-time observations. Using a confusion matrix to evaluate the model, it was found to have a misclassification rate of 23.5%. Future work included collecting additional data and adding new, real-time independent variables.

In Chan’s research on identifying high performance factories, he found that the currently used factory metrics varied significantly in their ability to signal long-term performance [14]. The research took place
at Li and Fung, a garment manufacturing supply chain manager. The internally collected metrics considered were quality (first time pass rates), delivery (on-time delivery), and compliance (social and environmental) [14]. These metrics were the primary inputs to the model, but the model also incorporated externally collected variables including country, factory size, and factory year on entry. These variables were determined through a working group at the company. After brainstorming these variables, the final variables were simply selected on data availability. The Average Spend of Factories was modeled using Ordinary Least Squares, while the Firm Survival Estimate was modeled using a generalized linear model. Key takeaways focused on the statistical significance of the different independent variables. Recommendations at the end of the project included improvements in data integrity, and using new leading indicators of factory performance [14].

In Schmidt’s research on using predictive analytics to address risk in complex supply chains, she aimed to quantitatively understand the risk inherent at each factory within the supply chain [15]. This research was also conducted at Li and Fung. Data was gathered from supplier-recorded incident data, factory data, as well as country-level data from the World Bank and The Economist Group. The feature selection was completed by using variables which were proven to correlate to the probability of incident. The data was centered, scaled, and separated into training and validation / test. Using logistic regression, the model had an Area Under the Curve (AUC) of 91% [15]. She also considered the statistical significance of the different independent variables. Future recommendations centered on getting greater volumes of more accurate data.

From the available research, a better understanding of the applications of predictive analytics to operations can be seen. Each of these papers brainstormed independent variables and determined a methodology to down select to the appropriate final number of variables. Some struggled with data quality and availability.

2.2. Data Quality

The reliability of analytics results depends on the quality of the analyzed data for the specific mission [16]. Big data has relevant idiosyncrasies compared to traditional data sources, characterized by errors and missing values [16]. So, it is important to assess data quality when also considering predictive analytics.
*Data and Information Quality: Dimensions, Principles, and Techniques* defines eight different dimensions for data quality, which are reproduced below (and edited for conciseness) [17]:

1. **Accuracy**: correctness, validity, and precision on the adherence to a given reality of interest
2. **Completeness**: capability of representing all and only the relevant aspects of the reality of interest
3. **Redundancy**: capability of representing the aspects of the reality of interest with minimal use of informative resources
4. **Readability**: ease of understanding and fruition of information by users
5. **Accessibility**: ability of the user to access information
6. **Consistency**: capability of the information to comply without contradictions to all properties in the reality of interest
7. **Usefulness**: advantage the user gains from the reality of interest
8. **Trust**: how much information derives from an authoritative source, encompassing security considerations

In order to assess these data quality characteristics, there are a number of repeatable activities that require less resource specialization [18]. *Multi-Domain Master Data Management* suggests the following example elements of analysis:

- Completeness of key attributes
- Uniqueness
- Highest and lowest values
- Frequency distribution of key attributes
- Pattern analysis of attributes
- Data match analysis across sources

By quantifying key data attributes and characterizing their data quality, the dataset can be best understood for modeling projects and continued improvement of the available datasets.
3. Methodology

3.1. Business Understanding

While the potential financial benefits of accurately predicting OTD are long term, there are a number of business incentives:

- More predictable sales may help better manage working capital.
- Improved OTD may improve reputation with customers, potentially leading to higher bookings.
- The efficient deployment of resources for corrective actions may increase operating profits.

A predictive and diagnostic model for OTD could yield significant benefits to the business for working capital, bookings, and operating profits. However, in the short term, such a model could help Company X understand how it can deploy big data analytics on operations data and understand growth areas.

The goal of this project was to create a proof-of-concept model. In the beginning of the project it was made explicitly clear by the project team that a possible outcome could be that the currently available operations data would not be predictive of OTD. In this case, it was determined that the project would still be successful, if the outcome could include a data quality assessment, and a framework to assess what information / infrastructure was missing to complete a model in the future.

Before gathering data for the model, the IT landscape at Company X was analyzed. There are inherent complexities in gathering and integrating datasets in Company X that are unique to the defense industry. Ancillary government regulations, security concerns, and data spill avoidance work against a data scientist’s normal priorities. Additional complexity lies in navigating different software and architectures, as well as a combination of new and legacy systems. An illustrative IT system map is shown in Figure 1, which demonstrates the different locations and factors that impact data acquisition and integration for the project.
The security and firewalls throughout the IT environment require separate system access permissions. The access can be granted through an online portal or through a person that is in charge of the database. The time to receive access varies significantly. Some databases have different software used to query the database and data can be retrieved in different ways. Issues with the business-unit-owned databases are more easily resolved than the corporate owned, whose changes have even further ranging implications and can take significant time to address. Some standard queries exist, some must be customized, and others built from scratch. This all takes place on top of an extremely complex end product. While different layers of ownership, software, databases, and sources can inherently add its own security, it also makes the process of gathering and finding the right data require significant expertise and time.

Once a foundational knowledge of the IT environment was developed, a list of potential independent variables for OTD was created. By working with various subject matter experts, a list of ~50 variables was created, along with the definition, presumed database location, and a priority ranking. This list spanned business unit operations, including planning, supply chain, production, test, and design. An excerpt from the list is shown in Table 1, while a full list can be found in Appendix 6.1.
Table 1 Excerpt from List of Potential Independent Variables

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Definition</th>
<th>Data Category</th>
<th>Data Source</th>
<th>Prioritization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indenture</td>
<td>BOM Level</td>
<td>Design</td>
<td>DS #2</td>
<td>1</td>
</tr>
<tr>
<td>Length of Open QN</td>
<td>Length of time a QN is open</td>
<td>Quality</td>
<td>DS #2</td>
<td>1</td>
</tr>
<tr>
<td>Throughput</td>
<td>Cumulative completions over time</td>
<td>Production</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>Inventory Levels</td>
<td>Sum of inventory</td>
<td>Production</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>Lateness</td>
<td>Difference between expected and actual receipt date</td>
<td>Supply Chain</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>MRP Modifications</td>
<td>Number of times MRP was modified</td>
<td>Planning</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>Realization</td>
<td>Planned / Expected Test Time</td>
<td>Test</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>Dock to Stock</td>
<td>Time between receipt &amp; inventory placement</td>
<td>Production</td>
<td>DS #1</td>
<td>3</td>
</tr>
<tr>
<td>Contract Structure</td>
<td>Structure of awarded contract</td>
<td>Contracts</td>
<td>DS #3</td>
<td>4</td>
</tr>
<tr>
<td>Expedite Fees</td>
<td>Fees associated with expedited shipping</td>
<td>Supply Chain</td>
<td>DS #1</td>
<td>4</td>
</tr>
</tbody>
</table>

3.2. Data Understanding

Beginning with the top priority data variables, a strategy as determined to acquire the necessary datasets from the available data sources. In order to get the ~50 variables, twelve different datasets were downloaded. It was anticipated that the large number of disparate dataset may result in integration complexities. Their relationship to the different data sources are shown in Figure 2.
The goal was to capture the entire lifecycle of the first available delivery data, so the datasets reached back to the longest lead time of the first OTD data point for all datasets. For example, the OTD data was collected from January 2018 through October 2018 and the longest lead time is 2.3 years, so data was desired to be collected from September 2015 - October 2018.

The data structures across the datasets were relatively uniform, consisting of a material number, time stamp, and attribute grouping / values, with an illustrative data sample shown in Table 2. However, the MBOMs are structured differently, where the order of the dataset designates the hierarchical structure. For example, an end product is known as an Deliverable Product (DP, Indenture Level 0), and a DP's illustrative MBOM is shown below in Figure 3.

Table 2 Illustrative typical data structure

<table>
<thead>
<tr>
<th>Program</th>
<th>Factory</th>
<th>Material</th>
<th>Planned PT</th>
<th>Start Date</th>
<th>Posting Date</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>100</td>
<td>111-1</td>
<td>5</td>
<td>1/1/2018</td>
<td>1/4/2018</td>
<td>10</td>
</tr>
<tr>
<td>A</td>
<td>100</td>
<td>111-2</td>
<td>10</td>
<td>1/2/2018</td>
<td>1/12/2018</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>200</td>
<td>333-3</td>
<td>1</td>
<td>1/1/2018</td>
<td>1/1/2018</td>
<td>20</td>
</tr>
</tbody>
</table>
All data exploration was performed in RStudio. To perform this analysis, the data was first cleaned. The level of effort required varied significantly by dataset. Typical data cleaning activities included synchronizing dates from calendar date to manufacturing calendar date, removing units from numerical cells, and consolidating categorical attributes, when necessary. A typical data exploration analysis included evaluating uniqueness, minimum, maximum, attributes, and finding patterns. An example analysis is shown in Appendix 6.2. This analysis was used to better understand the data and the meaning behind various attributes.

After the data exploration was complete, the data quality could be evaluated. The system used for scoring the data quality was adapted from Blazent’s “Seven Characteristics that Define Quality Data” and *Data and Information Quality: Dimensions, Principles, and Techniques* [19] [17]. The system was also socialized within working groups at Company X. A three tiered system was used ranging from Good (green) to Needs Improvement (red) and is shown in Table 3.

<table>
<thead>
<tr>
<th>Scoring Variable</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td></td>
</tr>
<tr>
<td>No known errors in data</td>
<td>Known errors in data</td>
</tr>
<tr>
<td>Consistent source data without contradictions</td>
<td>Some inconsistencies between data sources</td>
</tr>
<tr>
<td>Data was posted to database on the same day</td>
<td>Data is posted to database within the month</td>
</tr>
<tr>
<td>Few missing materials</td>
<td>Some missing materials</td>
</tr>
<tr>
<td>Data includes the needed fields and materials</td>
<td>Data includes some needed fields and materials</td>
</tr>
<tr>
<td>Another person accessed data and/or was extracted in large chunks</td>
<td>Unable to gain access to database or extraction on weekly basis</td>
</tr>
<tr>
<td>Database was easy to access and extract data</td>
<td>All data is available through the last year</td>
</tr>
<tr>
<td>All data is available through the past 2.5 years</td>
<td></td>
</tr>
</tbody>
</table>
This data quality framework was then applied to the datasets downloaded for the project. The outcome of the evaluation is shown in Table 4.

Table 4 Data Quality Evaluation of Collected Datasets

While the specific suggestions and reasoning for the dataset quality evaluation cannot be disclosed, a more general understanding can be discussed through select examples. Through the process of downloading all of the MBOMs, it was discovered that there was an error in many of the MBOMs, causing some of the materials to be out of order, which resulted in a red, Needs Improvement for Accuracy. For Reliability, the cost data varied significantly for materials between sources, including Purchase Orders (POs), Purchase Requisitions (PRs), and the actual cost dataset. Timeliness was rated a red, Needs Improvement, for the cost dataset, since it was only updated on a quarterly basis. Many of the datasets suffered from Accessibility issues related to one of the current IT systems, which limits the download size of datasets. In certain cases, this required >50 separate downloads to get the required date range. Others too, suffered from Availability issues, with historical data abruptly stopping in the middle of the required date range. This means that the data was only archived up to a certain point. Company X has many good aspects to its data, but still has room to grow on its operations data quality journey.

Between evaluating data quality and searching for the desired potential independent variables, a number of missing datasets came to light. These datasets fit into three broad categories: lookup tables, historical data, and data connections. For the Lookup Tables, an example missing data set is the Major Program associated with each DP number. Since Company X has such a high mix of products, there are many different DPs. However, to date, there is not a comprehensive list of the DPs and their associated
program. In the future, having this list would make data aggregation and compilation go much more quickly. Considering historical data, Company X has not historically archived its MRP need dates, so it can be difficult to evaluate production performance of when a material was complete vs. when it was supposed to be complete (need date). This metric was anticipated to be highly predictive of OTD, but since the data does not exist historically, it could not be used for the predictive model. A final example for missing Data Connections is a DP end item number associated with each material number. Some parts are common across programs, but the materials are usually built with an end item in mind. This would help trace the material through the production process and avoid double-counting when integrating datasets into the MBOM structures. Company X has a number of areas where it can collect more data to make future analysis timely and potentially predictive.

3.3. Data Preparation

Once the datasets were collected, explored, and cleaned, the information was prepared for modeling. Due to the large number of potential independent variables, it was necessary to down select to a smaller list of variables. The final combined dataset needed to best represent manufacturing reality, including the time-phased hierarchical nature of building a DP. The following sections describe the rationale for inclusion or exclusion of independent variables, the dataset integration, and the procedures for missing data.

1. Rationale for Inclusion or Exclusion

While the project started out looking to gather approximately fifty different independent variables, the list was automatically withered down by the availability or missing data, as discussed in Section 3.2. The dataset was then only approximately thirty different potential variables. From there, the data exploration reports were used to determine sufficient uniqueness and completeness. Finally, a correlation matrix was used to help select the final independent variables that would be used for the modeling dataset. However, due to known imperfections in this methodology, these results were paired with operational knowledge to select the final independent variables.

The first step in down selecting the independent variables was to examine uniqueness and completeness. Uniqueness was defined as the percentage of the dataset that was equal to the mode. In Figure 4, the results of this analysis can be seen, where the frequency is the number of independent variables. As can be seen, there were eight variables with no variability, so these were discarded from the
modeling dataset. After examining many of these variables with no uniqueness, it was discussed with the dataset owner to determine why a number of the derived variables had no variability. Results were discussed, so dataset owners could explore the causes, as necessary.

![Histogram of Variable Uniqueness](image1.png)

*Figure 4 Histogram of Independent Variable Uniqueness*

The independent variables were also assessed for the amount of missing variables that occurred. Figure 5 shows a histogram of NAs that occurred within each individual dataset, including derived variables. Any independent variables which had more than fifty percent of NAs were discarded. Some of these variables overlapped with the discarded uniqueness variables.

![Histogram of Missing Values](image2.png)

*Figure 5 Histogram of Independent Variable Missing Values*
Once the list of independent variables had been culled down to approximately twenty variables, a correlation matrix was created. To create the correlation matrix, the data went through a preliminary integration, where the independent variables were joined based on material, manufacturing month, and manufacturing year. The correlation matrix used pairwise correlation due to missing variables throughout a single row. This was expected though, because of the way that the products are assembled. Company X has each material in an MBOM marked as make or buy. Not all of the independent variables are applicable to make parts and similarly for buy parts. For example, late delivery cannot be measured for a material that is made in house and throughput cannot be measured for a material that comes from a supplier. This known characteristic created some holes in the correlation matrix, which can be seen in Figure 6.

This method of integration was also imperfect, due to the way OTD is created and structured. The OTD dataset which was used, only consists of final deliverable items to the customer. For example, DPs, kits, spare parts, etc. have their own OTD data point. Lower level subassemblies that go into a DP, do not have an OTD value. So this correlation matrix would only show if a DP throughput is correlated to OTD, not if a lower-level material’s throughput correlates to OTD.

![Figure 6 Correlation Matrix of Preliminary Integrated Datasets](image-url)
Looking at the correlation matrix, some of the variable relationships can be visualized. In the squares with “?”, we can see where there are no pairs due to make / buy materials. For example, Material Yield (mat.yield in Figure 6) is only measured for make parts, so there is no correlation calculation for receiving supplied parts late (Late.Sch and Late.Stat in Figure 6). Some of the independent variables are slightly different variations on the same thing. For example, quality dispositions for internal quality problems (PE.Group in Figure 6) are the complement to receiving & inspection quality problems (RI.Group, in figure), which is why they have a strong negative correlation. Similarly, there are two different measures for receiving parts late (Late.Sch and Late.Stat), where one measures the time between delivery date to the handshake date vs. the contractually obligated date, respectively. These are very positively correlated.

The number and magnitude of changes to MRP (Avg.Change and Max.Qty.Change) are strongly negatively correlated with late deliveries. This is likely because of current MRP behavior, where the expected delivery date could be changed in order to avoid being considered late. While the correlation matrix was hoped to be used to down select independent variables, it’s clear that OTD has no strong correlations in this data structure, with correlation values between +/- 0.2.

Since the correlation matrix did not yield any strong positive or negative correlations of independent variables to OTD, one to two values from each representative dataset were chosen, using operational knowledge and experience. The final independent variables are shown in Table 5. Each of the variables was then scaled, though some were naturally scaled with the derived variable. Table 5 also shows how these variables were scaled. Further analysis could be performed exploring different independent variables for the modeling dataset.
Table 5 Final Independent Variable Selection & Normalization

<table>
<thead>
<tr>
<th>Final Dataset Variables</th>
<th>Variable Name</th>
<th>Definition</th>
<th>Normalized To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Throughput</td>
<td>TPut</td>
<td>Units produced per month</td>
<td>Monthly Avg</td>
</tr>
<tr>
<td>MakeSpan</td>
<td>MS</td>
<td>Days to produce a unit</td>
<td>Planned Production Time</td>
</tr>
<tr>
<td># of MRP Changes</td>
<td>MRPCh</td>
<td>Number of times MRP was changed</td>
<td>Monthly Avg</td>
</tr>
<tr>
<td>Test Time Realization</td>
<td>TR</td>
<td>Actual test time / expected test time</td>
<td>N/A (Actual / Expected)</td>
</tr>
<tr>
<td>Internal Process QNs</td>
<td>PE</td>
<td>% of QNs which are internally generated</td>
<td>N/A (% of Total QNs)</td>
</tr>
<tr>
<td>Yield</td>
<td>Yield</td>
<td>Passed Parts / Total Tested Parts</td>
<td>N/A (Pass / Total)</td>
</tr>
<tr>
<td>Price</td>
<td>Price</td>
<td>Material Price</td>
<td>Avg. Price</td>
</tr>
</tbody>
</table>

2. Dataset Integration Methodology

The final dataset used for the modeling needed to reflect two key manufacturing aspects: time dependency and MBOM structure. Time needed to be accounted for since a delay early on in the process may affect the final OTD. MBOM structure needed to be included since it reflects how the final product is actually built. By representing the time-phased nature of the MBOM in the dataset, the model would more closely represent manufacturing reality.

To construct the dataset with the down selected variables, a few key data points were needed, including the delivery date and lead time. While the final dataset uses multiple variables, a simplified version is discussed below, using throughput as an example. Figure 7 shows a simplified MBOM, where each node represents a different material. L0M0 (Level 0, Material 0) represents the final DP.

![Illustrative MBOM and Node Assignments](image-url)
The dependencies throughout the BOM are represented in Equation 1, where $TP = \text{Throughput}$, $D = \text{OTD}$ Delivery Date, $LT = \text{Lead Time}$, $t = \text{time}$ (manufacturing month). For simplicity only throughput is considered, and the operational effects of inventory and buffers are ignored.

**Equation 1 OTD as a Function of Throughput with Time Dependencies**

$$OTD \sim f(TP_{L0M0}(t = D), TP_{L1M1}(t = D - LT_{L0M0}), TP_{L1M2}(t = D - LT_{L0M0}), TP_{L2M3}(t = D - (LT_{L0M0} + LT_{L1M1}), TP_{L2M4}(t = D - (LT_{L0M0} + LT_{L1M1}), TP_{L2M5}(t = D - (LT_{L0M0} + LT_{L1M2}), TP_{L2M6}(t = D - (LT_{L0M0} + LT_{L1M2}))$$

Equation 1 states that the final OTD is a function of the throughput of the final DP at the time of the OTD delivery date, the $L1M1$ throughput at the delivery date minus the lead time of its parent, etc. These relationships can be applied across independent variables and MBOMs. However, the MBOMs vary significantly with up to 10+ levels, and 10,000+ unique materials. Operations data is not collected uniformly across materials; it is a strategic choice on where and when to collect data across the manufacturing process. Figure 8 shows an illustrative broken network with missing data points.

![Figure 8 Illustrative Broken MBOM Network](image)

In order to enforce data uniformity across MBOMs and better address missing data links, data was aggregated on a level by level basis. However, it was unclear how many levels the integrated dataset should represent. It was expected that data availability tends to decrease after a certain level. The preliminary integrated dataset used in the correlation matrix was then mapped to each of the MBOMs, removing time as a factor. So, for each material in the MBOM, the throughput, make span, etc. could be
seen within the MBOM structure. This dataset structure was used to analyze the data availability by level and is shown in Figure 9.

![Data Availability by Level](image)

*Figure 9 Data Availability by Level for Serialized Parts*

The results are not as expected, especially with the highest amount of NAs at Level 0. It would have been helpful to examine inventory numbers alongside manufacturing data, to see if some parts of the OTD BOMs are just being pulled out of inventory. Part of this could be explained by the OTD materials, in examining those which are kits or spares, especially those which do not have make parts and are simply taken in from a supplier, packaged, and sent out to the customer.

Other possibilities for the high number of NAs at each level could be the presence of Alternate IDs. Some MBOMs have a significant number of variants which can be built off of the same general part number. For example, a product may have two different sensors available, one for RF and another for EO/IR. Through the course of this project, it was realized that there was an upstream error in how the Alternate IDs were being pulled, so this information was not able to be extracted from the MBOMs. The MBOMs also have designated serialized parts. These parts are generally of higher monetary values, where unserialized parts tend to be commodities like O-rings, screws, and washers. Since significantly less data is kept on the unserialized parts, only serialized parts were examined. While the data availability by level did not come out as expected, it was decided to use the Levels 0, 1, 2, and 3+, where Level 3+ aggregated all levels higher than Level 2. This way, the model could still get analyze of how the levels affect OTD outcome. Further studies could be performed to optimize the number of levels to explore within the modeling dataset.
3. Dataset Integration Example

A data integration exercise is shown below, using three separate illustrative datasets, using one independent variable as an example. Table 6 shows an example from the OTD dataset, with the final delivered DP material number, the date it was to be delivered, and the OTD metric. Table 7 shows an example of the independent variable data set with each material number, its manufacturing date, and the throughput for the corresponding date and material. Table 8 shows an MBOM, including a material number, its level, and lead time. From these variables the cumulative lead time is calculated, which takes into consideration the lead time of each material’s parent. The Need by Date is then calculated as the delivery date minus the cumulative lead time.

Table 6 Illustrative Row from OTD Dataset

<table>
<thead>
<tr>
<th>DP</th>
<th>Delivery Date</th>
<th>OTD</th>
</tr>
</thead>
<tbody>
<tr>
<td>111-1</td>
<td>12/18</td>
<td>75%</td>
</tr>
</tbody>
</table>

Table 7 Illustrative Excerpt from Independent Variable Dataset (Throughput Only)

<table>
<thead>
<tr>
<th>Material</th>
<th>Date</th>
<th>Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>111-1</td>
<td>11/18</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>10/18</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>11/18</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>09/18</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>11/18</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>10/18</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 8 Illustrative MBOM with Lead Times

<table>
<thead>
<tr>
<th>Material</th>
<th>Level</th>
<th>Lead Time (LT)</th>
<th>Cumulative LT</th>
<th>Need by Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>111-1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>11/18</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>10/18</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>09/18</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>07/18</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>10/18</td>
</tr>
</tbody>
</table>

To integrate these three tables, the DP and OTD are taken from Table 6. Both average values and max values are recorded in order to account for variability. For Level 0, the throughput on the corresponding Need by Date is recorded, $TPut_{M=111-1}(t = 11/18) = 4$ units. Since there are no more Level 0 materials, then next up are Level 1 materials. There are two Level 1 materials and they have the same Need by date and two different throughputs $TPut_{M=1}(t = 10/18) = 4$ units and $TPut_{M=4}(t = 10/18) = 8$ units, respectively. For these two, the average and max are recorded. This continues into the
instance of Level 3, where no data is available for the correct month, so the data is given an NA. Table 9 shows the resulting row for the dataset integration for this particular OTD data point. Note that while the data has not been normalized in this example, it is normalized in the actual dataset.

Table 9 Illustrative integrated Dataset Row

<table>
<thead>
<tr>
<th>DP</th>
<th>OTD</th>
<th>L0 Avg TPut</th>
<th>L1 Avg TPut</th>
<th>L2 Avg TPut</th>
<th>L3+ Avg TPut</th>
<th>L0 Max TPut</th>
<th>L1 Max TPut</th>
<th>L2 Max TPut</th>
<th>L3+ Max TPut</th>
</tr>
</thead>
<tbody>
<tr>
<td>111-1</td>
<td>75%</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td>NA</td>
<td>4</td>
<td>8</td>
<td>10</td>
<td>NA</td>
</tr>
</tbody>
</table>

When integrating the final dataset, there were missing variables. These were handled by replacing the missing variables with numbers that represented normal operations, so that the final machine learning model would not see these numbers as indicators for the final OTD. For most cases, missing variables were replaced with a value of one. For example, normal operations would state that yield is 100%, so a missing yield variable would be replaced with a value of one.
4. Model Performance

The model was trained and tested on the integrated dataset, which represented 207 unique deliveries ranging from January 2018 through October 2018. The dataset was relatively balanced, with ~40% delinquencies and ~60% on-time deliveries. In order to test the model accuracy, data was partitioned into training (70%) and test (30%) sets, considering both continuous and discrete outcomes. Decision trees were the only machine learning modeling technique considered in order to maximize understandability of the final model, for which end-users are not expected to be data scientists. Both regression and classification models were tested.

For both modeling methods, RStudio’s “rpart” package was used to train and test the models. Models were trained on the training data and evaluated on test data. Trees were also pruned prior to evaluation.

4.1. Regression Decision Trees

For manufacturing and planning purposes, it is beneficial to know of the partial delivery, so a continuous outcome between zero and one was considered (e.g., if three units were delivered out of a contractual obligation of four units, the delivery would be considered 75% OTD). Figure 10 shows the ranking of variable importance, where the first part of the name represents the Level (e.g., L2 = Level 2) and the second part represents whether it’s the max or average value (Max or Avg), and finally is the independent variable (e.g., MRPCh, TR).

![Variable Importance](image)

*Figure 10 Regression Decision Trees Variable Importance*
From this figure, it can be observed that the first six variables within the variable importance are from Levels 2 & 3. This is interesting that the lower level assemblies drive OTD, instead of the higher level assemblies (Levels 0 & 1). The top six variables are a range, including the number of MRP Changes, the Test Realization, Throughput, and Price. The max value of Level 2 MRP changes is 30% more influential than the next variable. The number of MRP Changes appear in the top six variables three times. This raises questions less upon potential manufacturing changes, but more upon MRP behavior. For example, it is possible that MRP is being changed when it is known that a product will not be completed on time to move the goalposts, so then a product is considered on time. The top variables are also driven by maximum values, not average, which suggests that OTD is affected by variance.

The actual model outcome is shown in Figure 11. It can be seen that the model is largely driven by the number of MRP changes across levels and also by yield. This model implies that Level 3+ yield is the most important yield to consider, which makes sense intuitively. If the yields are low early in the process, it will likely negatively affect the subsequent level yields, as well.

![Figure 11 Regression Pruned Decision Trees Model](image)

Considering this behavior, it seems that changing MRP is only effective to a certain point. Once a certain threshold is reached, changing MRP has a negative effect on OTD. However, changing Level 0 MRP does increase the likelihood of meeting OTD. This model was evaluated by examining the root mean square error (RMSE) and the F1 score for accuracy. The RMSE was found to be 0.39 and the accuracy 0.86, and the variables considered are at least predictive.
4.2. Classification Decision Trees

When considering the OTD metric in the eyes of the government, the outcome is binary. So, a partial delivery is considered to be a 0% OTD (e.g., if 3 units were delivered out of a contractual obligation of 4 units, the delivery would be considered 0% OTD). So, in order to use classification trees, any delivery data point which had less than 100% OTD, was assigned an OTD of 0%. Figure 12 shows the rank of variable importance for the classification trees.

From this figure, it can be immediately noticed the outsized influence Level 2 max MRP Changes have over other variables, with 50% higher influence than the next leading variable. Here, the first three variables all relate to MRP changes, which is even more exaggerated than the Regression outcome. We can see the effect of the current MRP behavior. Following the MRP Changes, the next five variables in terms of importance are all in Level 3+. This demonstrates how the lower level assemblies drive the on-time deliveries, since these lower level parts constrain the higher level assemblies. The top non-MRP Change variables here are throughput, test realization and make span.

The model outcome is shown in Figure 13. It can be seen that after a certain threshold of MRP changes for Level 1 & 2, OTD will be negatively affected. However, for Level 0 MRP changes, the more often MRP is modified, the more likely OTD will be met. This is a prime example of moving the goalposts to meet the goal. The make span and test realization variables are counter intuitive, as it appears if more time is spent in production and testing than is currently planned, the product is more likely to meet OTD.
This model was evaluated using a confusion matrix, which can be seen in Table 10. The model was found to have ~20% Type 1 error and ~5% Type 2 error. The accuracy was found to be 76%. This version of the model was optimized for accuracy. However, future iterations of this model should optimize for Type 1 error, because it would negatively impact the customer to forecast an on-time delivery and then be incorrect. Whereas a Type 2 error would only internally impact Company X and may actually surprise the customer in a positive way: getting their delivery early.

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Actual</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>39</td>
</tr>
</tbody>
</table>
5. Recommendations & Conclusions

While this project was limited by time to six months of research, there are short term and long term recommendations that could improve future iterations of the model. Recommendations range from tactical (including plus or minus one month of lead time) to strategic (improving IT systems and evaluating MRP behavior).

5.1. Short Term Recommendations

There are a number of areas for improvement, including both data and modeling perspectives. This project was limited by time. However, now that the process of data collection and processing is known, there are a number of improvements that could be made in a relatively short amount of time.

On the data side, the most apparent problem is the small dataset size. This was partially due to the lack of data availability in OTD data with less complex MBOMs. However, in the future sub-assemblies and spare parts could be included in the model, as they frequently have their own OTD data. As time marches forward and the OTD database is backfilled, there will also be more OTD data points available in general, for both DPs and other hardware deliveries.

The process of dataset integration could be made more robust. One way to accomplish this is to take the average of independent variables value for plus or minus one month from the Need Date within the data integration process. This would better account for small schedule slips and correct for anomalous data. Another area to improve is the alignment of the OTD date to the independent variables. Currently, the dataset is integrated as if all manufacturing is done just-in-time. However, we know that this is not the reality for many programs. To make the dataset a better representation of reality, the DP manufacturing data could be aligned to the closest OTD date in the future. This would incorporate builds which may take place before the actual delivery date. Both of these methods could expand the dataset and make it more robust.

Within the currently available data, more data analysis could be performed to optimize model performance. A new methodology for independent variable selection could be tested. Or more simply, more iterations of the model with different groupings of the independent variables could be evaluated. Additional analysis could be performed to determine the optimal number of levels to examine within the MBOMs. More time could be put into determining the right mix of independent variables and levels.

While this proof-of-concept project focused on machine learning models with high interpretability, different methods could be applied. GBM, logistic regression, and other techniques could be used, to see
if the results are improved. This project also optimized for accuracy, but future iterations could minimize
false positives shown in Table 10, since false positives may jeopardize relationships with customers.

In the short term, this model can be improved by expanding available data, optimizing the
independent variables and levels, and experimenting with different machine learning methods.

5.2. Long Term Recommendations
While there are short term actions which will likely improve model performance, there are
fundamental issues that must be addressed to improve operations and data science at Company X. These
include examining the use and role of MRP, addressing data quality issues, and considering new IT
solutions. Many of these require significant additional time and resources.

Preliminary model results show that one of the largest drivers of OTD is frequently changing MRP. A
possible explanation for this phenomenon is that when a product is likely to be late, MRP will be modified
a number of times until it is no longer considered late. Obviously, MRP can legitimately be changed for a
number of reasons (e.g., contract modifications), but why are there so many? Some materials had their
MRP changed nearly one hundred times in a month. A factory manager could not possibly be held
responsible for making production to MRP when the quantity due may change three times per day.

MRP accuracy also needs to be continuously monitored for metrics like planned lead time and in
house production time. This past year, there were approximately 200 unique materials that were never
completed within their planned in house production time. Some of these materials had relatively high
production, producing >1,000 units, which were never produced on schedule. With thousands of unique
materials, Company X needs to determine how MRP can be more responsive to unrealistic targets in order
to make scheduling more accurate, while also holding factory performance accountable.

Data quality also needs to be continuously improved at Company X. Significant effort has been put
forth in the past to improve data quality. This must be viewed as a continuous journey as opposed to
various start and stop efforts. A data governance board has been put in place in order to identify,
prioritize, and resolve data issues. However, this type of group requires ongoing resources, time, and
support in order to be effective. Currently, data science is viewed as something that only affects the data
scientists, even though this area affects everyone who uses data. Data effects the entire business unit,
especially as Company X works to become a more data-driven organization. One back of the envelope
way to look at data quality investments is to look at the additional time and productivity of the data
scientists. If a data scientist has a salary of ~$130k/year [20], and the organization has ten data scientists,
and workers spend 10% of their time addressing data quality issues, investment could be rationalized as follows:

\[(130k/\text{year}) \times (10 \text{ data scientists}) \times 10\% = 130k/\text{year}\]

\[\left[\frac{(8 \text{ hours}}{\text{day}} \times \frac{261 \text{ working days}}{\text{year}} \times 10\%}{50 \frac{\text{working weeks}}{\text{year}}}\right] = 4 \frac{\text{hours}}{\text{week}} \text{ per data scientist}\]

Considering that a 10% efficiency from addressing data quality issues is extremely conservative, resources of $130k/\text{year}$ and of 4 hours/week per data scientist could be justifiably allocated. Looking at the time dedicated over the ten data scientists (40 hours), it easily supports the hire of one additional full time worker dedicated to solely addressing data quality issues. However, in addition to more time and money spent addressing data quality issues, agile procedures need to be put in place to resolve data issues while ensuring correctness, understanding impact, and considering security implications.

One of the last key long term recommendations is to consider implementing IT solutions which are more amenable to downloading large datasets in a secure, accurate, and fast environment. Due to current data download size limits, some of the individual data sets had to be downloaded >50 separate times in order to get the full timespan required. The current system represents a massive inefficiency in the time of the data scientists and introduce more error, since the subsequent datasets require additional processing and integration. In order to make a predictive OTD model and other predictive models feasible in the future, new IT solutions need to be implemented to allow quick and reliable access to large datasets.

Most of the long term recommendations point to developing and socializing a comprehensive business unit-wide data strategy. This includes making existing systems more responsive (e.g., MRP), understanding the path to high data quality across systems, and thinking through the next generation of IT solutions, which will satisfy Company X requirements. Company X has complex needs, and must consider the requirements of corporate, as well. The path forward to achieving Industry 4.0 requires wading through the complexity to determine an affordable, agile, and supported data strategy.

5.3. Conclusions

In conclusion, the proof-of-concept model to predict OTD using Company X data from disparate data sources across the business unit was a success. By understanding the IT systems, data quality, and operations driving the data, a model was able to be created to predict OTD. Short term, the model can be
improved by making the integration more robust and gathering additional data. Future work must be completed around improving data quality and IT solutions, but these can only be achieved with strategic investments in resources and time.
### 6. Appendix

#### 6.1. Notional List of Potential Independent Variables

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Definition</th>
<th>Data Category</th>
<th>Data Source</th>
<th>Prioritization</th>
</tr>
</thead>
<tbody>
<tr>
<td># Materials</td>
<td>Material count at each BOM node</td>
<td>Design</td>
<td>DS #2</td>
<td>1</td>
</tr>
<tr>
<td># Unique Materials</td>
<td>Unique material count at each BOM node</td>
<td>Design</td>
<td>DS #2</td>
<td>1</td>
</tr>
<tr>
<td>Parallelization</td>
<td># of production operations in parallel</td>
<td>Design</td>
<td>DS #2</td>
<td>1</td>
</tr>
<tr>
<td>Part Cost</td>
<td>Cost of producing or buying a part</td>
<td>Finance</td>
<td>DS #4</td>
<td>1</td>
</tr>
<tr>
<td>PR release vs. need to release date</td>
<td>Date of PR release vs. planned release date</td>
<td>Planning</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>WO finish vs. need date</td>
<td>Date of WO release vs. planned release date</td>
<td>Planning</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>PO release vs. need to release date</td>
<td>Date of PO release vs. planned release date</td>
<td>Planning</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>PR Version #</td>
<td>Max number of generated PRs</td>
<td>Planning</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>Final Release PR vs. Create</td>
<td>Difference between the PR release and creation dates</td>
<td>Planning</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>Purchase Order vs. Create</td>
<td>Difference between the PO release and creation dates</td>
<td>Planning</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>Disposition Quantity</td>
<td>Sum of total QNs</td>
<td>Quality</td>
<td>DS #2</td>
<td>1</td>
</tr>
<tr>
<td>Disposition Type</td>
<td>% of QNs by Rework, Void, Use-As-Is, Return, etc.</td>
<td>Quality</td>
<td>DS #2</td>
<td>1</td>
</tr>
<tr>
<td>Length of Open QN</td>
<td>Length of time a QN is open</td>
<td>Quality</td>
<td>DS #2</td>
<td>1</td>
</tr>
<tr>
<td># QNs Still Open</td>
<td>Count of open QNs</td>
<td>Quality</td>
<td>DS #2</td>
<td>1</td>
</tr>
<tr>
<td>Disposition Group</td>
<td>Internal or supplier quality issue origination</td>
<td>Quality</td>
<td>DS #2</td>
<td>1</td>
</tr>
<tr>
<td>QN Type Code</td>
<td>Internal or supplier part quality issue</td>
<td>Quality</td>
<td>DS #2</td>
<td>1</td>
</tr>
<tr>
<td>Throughput</td>
<td>Cumulative completions over time</td>
<td>Production</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>Throughput Range</td>
<td>Max(TPut) – Min(TPut)</td>
<td>Production</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>MakeSpan</td>
<td>Length of time to make a part</td>
<td>Production</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>Makespan Range</td>
<td>Max (MS) – Min( MS)</td>
<td>Production</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>Inventory Levels</td>
<td>Sum of inventory</td>
<td>Production</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>MRP Requirement</td>
<td>Number of materials planned to be complete</td>
<td>Production</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>Lateness</td>
<td>Difference between expected and actual receipt date</td>
<td>Supply Chain</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>PO Quantity</td>
<td>Number of parts ordered</td>
<td>Supply Chain</td>
<td>DS #1</td>
<td>1</td>
</tr>
<tr>
<td>% Materials from Supplier</td>
<td>Internal vs. supplier material count at each BOM node</td>
<td>Supply Chain</td>
<td>DS #2</td>
<td>1</td>
</tr>
<tr>
<td># Different Factories</td>
<td>Number of factories involved at each BOM node</td>
<td>Supply Chain</td>
<td>DS #2</td>
<td>1</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------------------------</td>
<td>--------------</td>
<td>-------</td>
<td>---</td>
</tr>
<tr>
<td>MRP Modifications</td>
<td>Number of times MRP was modified</td>
<td>Planning</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>MRP Change Quantity</td>
<td>Total change of quantity with MRP modification</td>
<td>Planning</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>MRP Change Time</td>
<td>Time between MRP modifications</td>
<td>Planning</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>First Pass Yield</td>
<td>Number of materials, which pass quality tests on first pass</td>
<td>Production</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>WIP Quantity</td>
<td>Sum of WIP</td>
<td>Production</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>WIP Range</td>
<td>Max(WIP) – Min(WIP)</td>
<td>Production</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>Rework Quantity</td>
<td>Material requiring additional work after failing inspection</td>
<td>Production</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>Rework Range</td>
<td>Max(Rework Qty) – Min(Rework Qty)</td>
<td>Production</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>Production Test</td>
<td>Production Order Type</td>
<td>Test</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>Rework Test</td>
<td>Rework Order Type</td>
<td>Test</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>Routing Sequence</td>
<td>Unplanned Test Operations</td>
<td>Test</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>Number of test operations</td>
<td>Number of test operations performed on a material</td>
<td>Test</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>Range of Test Run Time</td>
<td>Max(Test Run Time) - min(Test Run Time)</td>
<td>Test</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>Realization</td>
<td>Planned / Expected Test Time</td>
<td>Test</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>Range of Realization</td>
<td>Range of Planned / Expected Test Time</td>
<td>Test</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>Rework Tests</td>
<td>% of Test Time Spent on Rework</td>
<td>Test</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>Total Test Touch Time</td>
<td>Expected vs. planned test touch time</td>
<td>Test</td>
<td>DS #1</td>
<td>2</td>
</tr>
<tr>
<td>Supplier Award Type</td>
<td>Special designations for supplier types</td>
<td>POs</td>
<td>DS #1</td>
<td>3</td>
</tr>
<tr>
<td>Dock to Stock</td>
<td>Time between receiving supplier material &amp; going into inventory</td>
<td>Production</td>
<td>DS #1</td>
<td>3</td>
</tr>
<tr>
<td>Contract Structure</td>
<td>Structure of awarded contract</td>
<td>Contracts</td>
<td>DS #3</td>
<td>4</td>
</tr>
<tr>
<td>CLIN Complexity</td>
<td>Complexity assignment to contractual deliverable</td>
<td>Contracts</td>
<td>DS #3</td>
<td>4</td>
</tr>
<tr>
<td>Contract Modifications</td>
<td>Number of contract modifications</td>
<td>Contracts</td>
<td>DS #3</td>
<td>4</td>
</tr>
<tr>
<td>Expedite Fees</td>
<td>Fees associated with expedited shipping</td>
<td>Supply Chain</td>
<td>DS #1</td>
<td>4</td>
</tr>
<tr>
<td>Supplier Part Complexity</td>
<td>Measure of supplier part complexity</td>
<td>Design</td>
<td>DS #1</td>
<td>4</td>
</tr>
</tbody>
</table>
6.2. Example Data Exploration Report

Table 11 Uniqueness Characterization of Categorical Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Major Programs</th>
<th>Center</th>
<th>Factory</th>
<th>Process</th>
<th>Materials</th>
</tr>
</thead>
<tbody>
<tr>
<td># Unique Values</td>
<td>40</td>
<td>7</td>
<td>22</td>
<td>90</td>
<td>4,603</td>
</tr>
<tr>
<td>%NAs</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>% Unique Values</td>
<td>59.9%</td>
<td>74.2%</td>
<td>56.9%</td>
<td>16.0%</td>
<td>8.1%</td>
</tr>
</tbody>
</table>

Table 12 Uniqueness Characterization of Numerical Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Planned PT</th>
<th>Order Quantity</th>
<th>Actual Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td># Unique Values</td>
<td>40</td>
<td>7</td>
<td>22</td>
</tr>
<tr>
<td>%NAs</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>% Unique Values</td>
<td>41.7%</td>
<td>20.5%</td>
<td>91.1%</td>
</tr>
<tr>
<td>SD</td>
<td>12.8</td>
<td>72.9</td>
<td>27.7</td>
</tr>
</tbody>
</table>

Table 13 Data Exploration of Numerical Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>IHP</th>
<th>Order Quantity</th>
<th>Actual Quantity</th>
<th>Release Date</th>
<th>Start Date</th>
<th>Finish Date</th>
<th>Posting Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>240</td>
<td>1,300</td>
<td>1,300</td>
<td>7/13/2018</td>
<td>7/15/2018</td>
<td>7/15/2018</td>
<td>7/15/2018</td>
</tr>
<tr>
<td>Mean</td>
<td>6.6</td>
<td>66.3</td>
<td>5.5</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Median</td>
<td>2</td>
<td>50</td>
<td>1</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Mode</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>12/12/2017</td>
<td>05/30/2018</td>
<td>02/03/2018</td>
<td>01/10/2018</td>
</tr>
</tbody>
</table>
Figure 14 Materials Made within Planned PT by Center (Anonymized)

Figure 15 Throughput by DP and its Major Subassemblies
References


