TU Dublin, Tallaght Campus

AI & Supply Chain - Is it possible to predict when a supplier will miss a delivery target date using Machine Learning?

Jack Dillane

Department of Computing

Supervised by

Dr. Keith Quille

Department of Computing

25 April 2021

Declaration

I hereby certify that the material, which I now submit for assessment on the programmes of study leading to the award of Master of Science, is entirely my own work and has not been taken from the work of others except to the extent that such work has been cited and acknowledged within the text of my own work. No portion of the work contained in this thesis has been submitted in support of an application for another degree or qualification to this or any other institution.

Jack Dillane 25 April 2021

Acknowledgments

Firstly I would like to thank Dr. Keith Quille for his outstanding classes, tutorials & support throughout the past two semesters. I have been amazed by how much I have been able to learn in the past 8 months & it would not have been possible without his help & guidance.

I would like to thank my family for taking me in during this global pandemic and backing me as they have always done.

Finally, I would like to thank my girlfriend Erin for her never-ending love & support and being there for me through the good & the bad this past year. I wouldn't have made it this far without you.

Table of Contents

Abstract

This report is a project on determining if it is possible to predict when a supplier will miss a target delivery date to a manufacturing company using Machine Learning and an Artificial Neural Network. The report covers the generation of a dataset of delivery history from the supplier over a period of six years, the pre-processing and cleaning of the dataset and the analysis of the dataset for the most prominent features that impact late deliveries.

A number of Machine Learning algorithms were generated included Naïve Bayes, Support Vector Machines and a Decision Tree as well as an Artificial Neural Network. Each model was trained and tested against the dataset and examined for accuracy, precision, and recall.

The final Artificial Neural Network generated provided promising results showing that in the case of the data being analysed, it is possible to predict with an overall accuracy of 78 % whether a specific delivery from a supplier will be late or not.

Finally, the report discusses the meaning of the model results, future research and work that can be applied and some potential applications.

List of Figures

List of Tables

1 Introduction

Within Supply Chain Management & Procurement functions in any manufacturing facility, a key target is ensuring that their production lines are supplied with the right part, for the right price, at the right time. Measuring, understanding & managing supplier On Time Delivery (OTD) performance is therefore essential to ensuring that production components delivered are occurring at the "right time" as mentioned above.

The study of supplier OTD strategies and its impact on company performance have been extensively covered in the literature (Tan, et al., 1998; McLean, 2017) and across industries (Toyota (GB) PLC, 2013) for decades. From identifying factors of influence to implementing protective measures to minimise exposure to risk (Chopra, 2015) However, with the ongoing development of Machine Learning capabilities in recent years, the ability for Machine Learning to be used as a tool to help identify what scenarios and factors are at play when a company may be exposed to OTD risk and to also predict when it is likely that a delivery target will be missed is a relatively new field.

The focus of this project is to design a Machine Learning tool that attempts to predict when a company is at risk of a supplier missing a target delivery date. The study is based on a supplier and their delivery performance to a large manufacturing company based in Dublin, Ireland. It will attempt to achieve this by taking historical OTD data gathered over a five-year period in the company as well as pairing this with some external data features. A key goal of this project is to seek out complex relationships that influence a supplier's delivery performance. In particular, relationships that extend deeper & beyond those that have already been thoroughly explored in the literature and are common knowledge in today's industries.

On Time Delivery is a main performance metric in measuring an organisations Supply Chain efficiency. It is an indication of how well equipped a company is to support its customer's needs. The ability to understand when a company's supply chain may be at risk of missed deliveries potentially ahead of the event occurring would be of huge benefit.

2 Project Approach

2.1 *Objectives*

The objectives of this project are as follows:

Objective 1: Identify and gain access to attributes & data sets that may be of any influence on Supplier OTD. It is key to gain access to as much data as possible as when the project develops and we begin to uncover relationships, we can narrow the focus on factors of influence.

Objective 2: Clean the data provided. It is important to consider that not all data available and collected will be accurate. For example, delivery bookings of components are processed by operators manually. In the past bookings have been processed incorrectly & reversed at later dates. This may indeed have an impact on delivery performance overall, it is important to isolate internal vs. supplier impacts on performance.

Objective 3: Identify relationships. Using the data gathered this objective involves looking at attributes & understanding how they correlate to each other as well as the class factor of the project (i.e. will the parts be delivered on time).

Objective 4: Build a model for testing relationships. As the data available for testing has been gathered over a number of years it will potentially be possible to retrospectively predict OTD performance for deliveries already received. This will be of value in partly determining the accuracy of relationships & models identified.

Objective 5: Test the model on a set of unseen data, i.e. additional deliveries at the company over a set period of time.

2.2 *Data Requirements*

The largest collection of data for this project is gathered via the companies MRP database stored in SAP. This database contains Supplier OTD performance over a number of years for several thousands of part deliveries over hundreds of suppliers. The approach for this project is to take a single supplier for the test. Selecting a supplier that has been providing a high volume of parts over an extended period of time should give a large enough data set to conduct testing of value where trends may be identified. The data set includes attributes including part name, qty delivered, target delivery date, actual delivery date.

The supplier under evaluation in this project is an aluminium sheet metal supplier. Some additional external attributes including the price of aluminium & oil (relating to transport costs) will also be of input for the tests.

2.3 *Project Question*

The Project Question for this project is follows.

Is it possible to predict when a supplier will miss a delivery target date using Machine Learning?

2.4 *Approach*

This project will use a data driven approach to try and address the Project question using quantitative data supplied from the company's MRP databases & external online sources.

The analysis will involve a deductive approach of simulating on time delivery scenarios to be analysed & tested.

3 Literature Review

3.1 Introduction

The just-in-time (JIT) manufacturing methodology, also known as the Toyota Production System (TPS) involves aligning raw material deliveries from Suppliers directly with the production schedule at the point of assembly. Successfully implementing a JIT system like this allows companies to minimise the levels of raw material held on site, thus reducing capital tied up in inventory, and reducing levels of waste in events of excess stock being held.

In order to implement a system like JIT, companies will require rigorous internal forecast planning to ensure production will meet upcoming demands. Critically, it also requires that suppliers deliver goods at the target time to meet these production demands. As a result, supplier On Time Delivery (OTD) is a main metric of supply chain performance. Defined simply, Supplier OTD performance is a ratio of components that have been delivered at the required time to the total number of components shipped over an extended period of time (Chopra & Meindl, 2016).

In industry, recommended targets for OTD performance can range to as hight as 100% (Gehringer, 2020), highlighting the need for strict controls & balances within companies to reach these targets. Considering this, the tighter the parameters used for JIT increases the levels of supply chain risk, potentially inducing charges including expedite fees & increased shipping costs, particularly if the suppliers within the supply chain are not operating to the expected standard and missing delivery target dates.

Overall, it is clear that in order to operate as a lean manufacturer, understanding & successfully measuring how your suppliers are performing in terms of delivery performance is critical. The following sections will explore how Supplier OTD measurement has been researched in the literature as well as the future of predicting OTD performance, as is proposed in this project.

3.2 Measuring OTD

The successful measurement of delivery performance and also motivating supplier networks in the upkeep of OTD performance has been discussed for numerous years in the literature. This includes evaluating performance by using Formal quantitative rating systems, In-depth performance reviews and ongoing communication with suppliers (Giunipero, 1990).

3.2.1 Supplier Relationships

The focus of research has primarily revolved around active measurement of OTD & the engagement with stakeholders within the supply chain to promote success. For example focusing on risk management in order to deal with supply chain disruptions (Ambulkar, et al., 2015), collaboration with stakeholders in the events of scheduling conflicts and unexpected demands (Chen & Hall, 2007) or developing complex supply networks as a means ensuring continuous supply (Viswanadham & Gaonkar, 2003). These works are orientated at identifying risks in existing supply chains and addressing the relationship between supplier customer, reinforcing deliverables & expectations to minimise the levels of risk. However, while this approach may be driven by data in order to gain a current understanding of supply risk, they do not utilise the same data to look forward into when similar supply risks may occur. There appears to be some space in the literature for this. As noted by BSR (Business for Social Responsibility), a non-profit business network looking at the future of sustainable supply chain, a top priority for Procurement by 2025 is “improving Risk Prediction and Management” and also noting that the widespread adoption of technologies including Machine Learning as part of supply chain activities is the number one key force of change. (BSR, 2020)

3.3 OTD Performance & Predictions

Research into predicting OTD performance in supply chain networks is relatively low & is quickly becoming a key area of discussion, particularly with the expanding utilisation of Machine Learning (ML) and Artificial Intelligence (AI).

3.3.1 Artificial Intelligence & Supply Chain Management

The potential range of AI involvement supply chain & procurement is vast. At its core, "procurement as a corporate function is leveraging AI to streamline processes & improve decision making" (Soumendra Mohanty, 2018).

Focusing specifically on OTD performance prediction, Research has been conducted into predicting supplier disruptions by using a feature set of average orders, order book size, number of products per supplier & supplier agility leading to a successful prediction of late orders at 80% (Brintrup, et al., 2020).

Similar research has been conducted in using a Stepwise Regression approach to predicting supplier delivery times of aircraft engine parts based on closed purchase orders against planning forecasts (Banerjee, et al., Boston).

Research into predicting supply chain risk has also been conducted considering data including supplier tiers, products, orders and also deliveries. It also considered risk factors including quality metrics & capacity scores at suppliers, using decision tree learning to calculate risk to supply (Baryannis, et al., 2019).

There is a growing focus of research into accurately predicting when supply chain disruptions & missed deliveries will occur. This research has typically revolved around collecting & analysing supplier related data involving orderbook volumes, the number of products produced as well as factors of quality & capacity scores at the supplier.

The goal of this project is to build on some of the research already conducted, to test if it is possible to accurately predict when it is likely suppliers will miss a delivery. This will be done by utilising similar variables & models reviewed here and applying them to the company's available data det.

4 Methodology

4.1 *Data Selection & Preparation*

4.1.1 Overview

As discussed, the purpose of this project is to explore if it is possible to build a machine learning algorithm that is capable of predicting when a component supplier will miss a target delivery date to the manufacturing company. The company actively receives over four thousand unique components from one hundred and fifty suppliers. Many of these components vary in usage amounts across the manufacturing line, for example an M2 x 4mm bolt may be used four hundred times in each instrument assembly, versus an instrument cover, which is only used once. For this project, a single supplier was used for the analysis. Using a single supplier allows for a focused approach to understanding what variables impact their ability to deliver components on time. The supplier of choice was also a very active supplier with an orderbook of 481 individual components of varying demand & complexity. The chosen time period for this supplier's deliveries was six years (January 2015 – December 2020). This time period will give a significant volume of deliveries to be evaluated using Machine Learning Algorithms and also an Artificial Neural Network.

4.1.2 Attribute Selection

As this is an investigation into what will impact on supplier lateness there are a number of attributes that should be considered:

Component attributes: It should be expected that variations in the individual components being delivered will have an impact on delivery times. For example, it is intuitive to assume that a small, simple part that has a short lead time will be at a lower risk of being late compared to a large complex assembly with a long lead time. In the case of this project, the supplier being analysed supplies primarily aluminium sheet metal assemblies. These can range from simple brackets to large instrument panels & door assemblies that have additional componentry such as hinges, and gas springs included in the assembly. With the time constraints of the project, it is not possible to explore the specific complexity of each of the 481 parts being analysed, however, the **price** of the part being delivered will provide a reasonable estimate of the complexity of the part – low-cost parts are typically simple components and vice versa. The price data is readily available for each component being analysed.

It should also be important to consider some external factors that may be of influence. As mentioned, the supplier being reviewed is an aluminium sheet metal manufacturer, we should therefore consider material availability. An indicator of this can be the global **price of aluminium** at the time. Likewise, another potential indicator of market volatility may be the **price of oil**. The rate of change between these prices may also give an indication of market stability.

Order book volume: With varying demands for components throughout the year based on customer orders and forecasts, the orderbook for the supplier can change significantly. For example, some components are used in regular volumes and are often delivered in set batch quantities at regular weekly or monthly intervals. Other components, for example those used as spare parts, can have a much more volatile demand and there may be extended periods of time between deliveries as a result. The orderbook volume will have an impact on supplier's ability to deliver all parts on time. Each supplier has a finite capacity, and it should be expected that if they are overloaded with orders all due in the same time period, there is a risk of some components not being delivered on time. The analysis should include attributes of; the **quantity of parts delivered** for each delivery booking, the **time since each part was last delivered**, the **number of unique parts expected to be delivered that day** and the **cumulative number of unique parts and quantities delivered** over the time period being examined.

Supplier factors: There are numerous qualitative supplier attributes that may impact the supplier's ability to deliver components on time. This will include the supplier quality score (an internal metric based on the delivery of components conforming to the required specification) the percentage of business sales the company holds with the supplier, supplier location and size. In the case of this project, as a single supplier is being analysed, so much of this data will not be variable so was therefore excluded. However, with further research of analysing multiple suppliers this data would be useful to include.

4.1.3 Data Collection

There are preliminary datasets that were collected for the final dataset for the project, they are the; On-Time Delivery Report (OTD report), Material Price Record and the external Aluminium and Oil price datasets.

Internal Data Sources

The manufacturing company being analysed uses a Material Requirement Planning (MRP) platform to manage all its material purchasing, delivery booking and material consumption functions. With this, every individual component transaction is logged and stored within the MRP databases and is available for export. The primary dataset for this project is the OTD Report. This dataset consists of every delivery booking for all parts and suppliers, the date at which it was booked into the system, the amount that was booked in, and critically, when the expected delivery date was. This dataset can therefore give an indication of which specific deliveries were early, on-time & late. The data can be exported by querying the MRP system for the required date range and supplier. The dataset was exported for the time period of January 2015 to December 2020 in an excel format.

The secondary data source was the price of each purchased component, this is stored as an information record in the MRP database. It is possible to export all part prices by querying the data base with the associated parts the supplier provides. This Material Price Record dataset was exported to an excel spreadsheet.

External Data Sources

The external data of aluminium and oil price and their rate of change is freely available online (Index Mundi, 2021). This price dataset is available in monthly entries and can be exported from the website in an excel format.

Calculated Fields

Finally, some of the data within the final dataset will be calculated fields. These calculations were completed in Excel. For example, the “time since the last expected delivery” attribute is not indicated in the OTD report. This can easily be calculated in excel by sorting deliveries by part number and then expected delivery date (oldest to newest), then subtracting the expected delivery date from the previous expected date. In the case of the first delivery for the dataset, the entry was set to zero. Similar approaches were used for the other data fields using excel.

4.1.4 Data Preparation

The process of preparing the data involves cleaning and then merging the datasets. In the case of the OTD report, there are several fields included in the export that were not required, including the storage location, the operator who completed the booking and numerous identification tags. These were simply deleted from the dataset using excel.

Merging the datasets involved firstly identifying the ties between each dataset. In the case of the OTD report & price register, the common link between both was the part number. The common tie between the OTD report and the Aluminium & Oil price dataset was the month (in this case, the price for aluminium and oil was the same for all deliveries through that particular month). The dataset ties are shown in Figure 1. The merging of datasets was completed in excel.

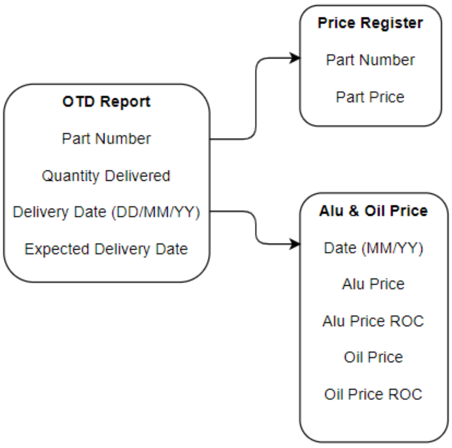

Figure 1: Dataset Structure

4.1.5 Class attribute

As the project goal is to identify when a supplier will miss a target delivery date, the class attribute in this case will be an indication of whether a delivery was late or not. Initially, the models built for the project used a regression approach to identify how late or early each delivery would be by a number of days (the delta between expected and actual dates of receipt). However, as will be discussed in the following sections of the report, the regression approach had significant performance issues. As a result, it was decided to change to a classification approach. In this case, any deliveries that were received after the expected delivery date were classified as Late (1) and any deliveries received on or before the expected delivery date were classified as Not Late (0).

The final dataset was then compiled into a CSV file, the list of all attributes and their source are shown in Table 1.

Attribute	Header Name (Jupyter Notebook)	Source
Expected Date of Receipt	expDOR	OTD Report
Actual Date of Receipt	DOR	OTD Report
Part Number	PN	OTD Report
Quantity Delivered	qtyDel	OTD Report
Price of Aluminium	alu	Alu & Oil Price
Rate of change of price of Aluminium	aluROC	Alu & Oil Price
Price of Oil	oil	Alu & Oil Price
Rate of change of price of Oil	oilROC	Alu & Oil Price
Cumulative delivered quantity	cummDel	Calculated field (OTD)
Time since last delivery	TSLD	Calculated field (OTD)
Delivery count (by part)	delCount	Calculated field (OTD)
Number of deliveries expected that day	noDelExp	Calculated field (OTD)
Number of deliveries received that day	noDelRec	Calculated field (OTD)
Part price	price	Price register
Late Delivery (**Class attribute**)	late	Calculated field (OTD)

Table 1: Attributes

4.2 *Data Pre-processing*

4.2.1 Data clean-up

As part of data pre-processing, the first step was to identify erroneous data within the entire dataset. This involved identifying which deliveries should be excluded from the analysis. The focus of the project is identifying which production component deliveries are at risk of being late, so any special, non-production orders from the supplier were removed from the dataset. These include orders where the manufacturing company supplied parts to the supplier for rework, or where there was no part number associated with the delivery (i.e. not a production part).

After the removal of these order lines, there were no other blank entries within the data set, so the initial clean-up operation was complete.

4.2.2 Outlier identification

True delivery date

There are a number of factors that can influence the integrity of the delivery data within the dataset. They will be discussed in the following sub sections.

Manual delivery booking

All deliveries from the supplier to the manufacturing company involve a manual process of a material handler booking each delivery into the companies MRP database on receipt. This booking represents the actual delivery date within the dataset being examined. There are cases over time where there is a backlog of deliveries received at the company and bookings may then be delayed by a number of days. As a result, some deliveries may be shown as late when the supplier is not actually at fault. Unfortunately, it is not currently possible to distinguish between delays as a result of the supplier and those from booking delays within the company. However, it is expected that the impact of this on the overall dataset will be minimal over the duration of the time period being examined.

MRP demand planning changes

The MRP platform used by the purchasing department within the company automatically creates component demand based on production plans and customer orders. The expected delivery date is often dictated by this software unless overruled by the material buyers. In some events, purchase orders may be received by the supplier with a given target delivery date. If there is a change to the production plans, e.g. the production team decide to no longer manufacture that assembly until a date in the future, the expected delivery date for the components may then be pushed out also. There are agreements with suppliers on a 'fixed period' of typically 4 weeks from the original target date where they will not push out the delivery in the event of a change like this. Similarly, in the opposite scenario where there is an unplanned order dropped-in from production, the expected delivery date to the supplier may appear even as soon as the same day, but realistically the lead time may be a number of weeks after that date. As a result of these changes, some deliveries will show both significantly early & late.

Additional external factors

Finally, there are additional factors that may influence the integrity of the delivery date information causing both early and late entries, most of which can be attributed to human error. This can include purchase order errors caused by incorrect data entry, delivery rescheduling with the supplier but not maintaining the MRP systems target delivery date, and also partial shipments, where some of the order may be fulfilled on the target delivery date to meet demand, but the remainder of the order is not shipped until a much later date causing a late delivery in the dataset.

Cleaning outliers

Overall, it is quite difficult to account for the errors caused by the factors previously mentioned. However, the data can be addressed wholly based on the distribution of delivery days. As can be seen in Figure 2 the mean value for delivery days is -10 (i.e., on average, deliveries arrived 10 days early). We can also see that the vast majority of the data is distributed within three standard deviations of the mean value. The data distribution figures are shown below in Table 2. As a base starting point, it was decided that any data entries that had delivery dates outside three standard deviations from the mean would be removed. While the total percentage of outliers account for only 1.61% of all data, it was chosen not to impute the mean value for these entries as there was sufficient additional data in the dataset.

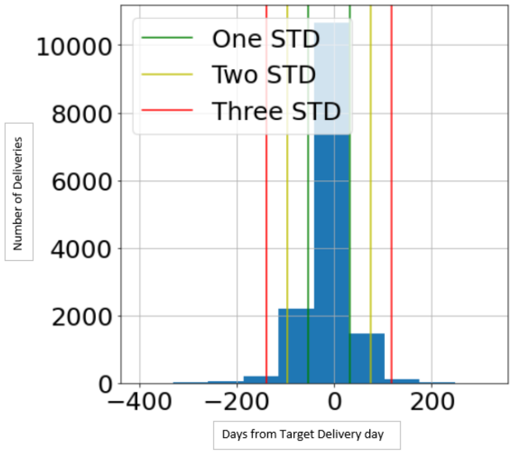

Figure 2: Delivery date distribution

Standard Dev	Data outside std	% Data outside	Max days late	Max days early
1	3222	21.9 %	32 days late	52 days early
2	786	5.34 %	74 days late	94 days early
3	237	1.61 %	116 days late	136 days early

Table 2: Delivery Date Distribution

Remaining attribute outliers

Regarding the remaining attributes in the dataset, there is no reason to suspect any incorrect data within the dataset, these values consist of either controlled data values from the company, historical pricing information or calculated fields. As a result, there was no additional pre-processing steps required on the feature entries.

4.2.3 Feature Selection

Following the gathering of the relevant attribute data, the next step is to begin understanding how these attributes have an impact on overall class result, in other words, how attributes positively or negatively impact delivery lateness. The following tests were conducted on the cleaned dataset using pandas.

Pearson's Correlation

The first test involved running a Pearson's correlation test of all attributes against the class variable. The results are displayed in Table 3. As can be seen from the results, there is no single attribute that appears to have any strong correlation with late deliveries, with the strongest relationships being found in the number of deliveries expected and Time since last expected delivery with a correlation of 0.2 positively & negatively respectively. With this, four attributes in total appear to have a stronger impact than the remaining. These attributes are discussed in the following sections.

Attribute	Correlation to Class attribute (Late)	Notes
expDOR	-0.04	
DOR	0.01	
PN	0.18	Noticeable positive correlation
qtyDel	-0.03	
alu	-0.04	
aluROC	0.01	
oil	0	
oilROC	-0.02	
cummDel	0	
TSLD	-0.02	
TSLED	-0.2	noticeable negative correlation
delCount	0.04	
noDelExp	0.2	Noticeable positive correlation
noDelRec	-0.08	
price	0.19	Noticeable positive correlation
Late	1	

Table 3: Attribute Correlation

Regression

Additionally, Regression was used to identify the most prominent features. The results of the Regression rest are found below in Figure 3. The features selected from Regression analysis were the same as those found in the Pearson's correlation analysis.

expDOR	18.55
DOR	0.52
PN	464.19
qtyDel	9.67
alu	23.06
aluROC	1.65
oil	0.0
oilROC	4.04
cummDel	0.1
TSLD	4.37
TSLED	600.66
delCount	24.89
noDelExp	612.92
noDelRec	99.71
price	540.44

Figure 3: Regression Feature Selection

Selected Features - Analysis

Attribute 1: price (Component Price)

The first attribute is price which represents the price of the item being delivered. The data has found that there is some small positive correlation between price and lateness. What this indicates is that parts with a higher price give some increased likelihood of being delivered late. This result is intuitive, in general the price of a part is dictated either by the complexity or by the size or materials used to manufacture. It is not unreasonable to assume that a component that takes more resources or steps to complete may be at greater risk of being late.

Attribute 2: noDelExp (Number of Deliveries Expected)

The second attribute is noDelExp which represents the number of unique component deliveries that were expected that day. Again, this attribute has some positive impact on lateness. This indicates that the more deliveries expected on a single day, the more likely it is that the delivery will not arrive when expected. Once again, this is an intuitive result, it is understandable that with the increased number of deliveries that are expected to arrive in a single day there is an increased risk that not everything will be delivered should there be a problem in manufacture.

Attribute 3: PN (Part Number)

The third and last positively correlated attribute is pn. The part number in the case of this data set is a numeric sequential entry, in other words, newer parts will have a higher part number than older. This correlation gives an indication that newer parts, which have a higher part number, are at greater risk of being late than older items. This may be understandable in the case that older parts may have been manufactured by the supplier for a longer period and as a result they may have more experience and ability in ensuring orders are completed in time.

Attribute 4: TSLED (Time since last expected delivery)

The last and main negatively correlated attribute is TSLED. This indicates the amount of time it has been since a specific part was last expected to be delivered to the company in days. What this means is if a part is delivered daily, the value is 1, if it is weekly it is 7, etc. This result is not necessarily intuitive, from the discussion on the previous attribute it could easily be assumed that the more frequently a part is delivered, it is likely the supplier will be spending more time and capacity manufacturing it. This would be expected to result better processes which should reduce the risk of deliveries being late. The data indicates that the opposite is occurring. However, there may be additional contributors to this result, including class imbalances, that will be discussed later.

4.3 Model Development

4.3.1 Initial Regression Model

As previously mentioned, the project was initially developed around using a regression model in order to attempt to predict delivery accuracy by a number of days before or after the target delivery date. This model struggled with very poor performance. Training & testing the entire dataset with Ten-fold cross validation on a linear regression model resulted in a root mean square error (RMSE) of 33.5 days, which is just under one standard deviation of the dataset distribution. The overall correlation coefficient of the data was also only 0.06. These findings were an indication that this approach would not be sufficient enough to provide any meaningful results as the range of average errors may find that deliveries predicted to be two weeks early will end up two weeks late. As a result of this it was opted to change the focus of the project to a classification model. Considering the key Research question of "Is it possible to predict when a supplier will miss a delivery target date using Machine Learning?" the most critical factor of consideration is if a delivery will be late or not, which can ultimately be classified as 'Late' or 'Not Late'. If it is possible to gain an accurate result of this, it would be more usable than an indication of how late a late delivery would be.

4.3.2 Machine Learning Models

The first models developed were Machine Learning Models built for binary classification with the aim that they can automatically learn and train themselves through experience based on the data provided. Machine Learning models will allow for development without explicit programming & should provide a good indication of the types of accuracies available from the data.

Model Training

Ten-Fold Cross-Validation (10FCV) was selected as the resampling procedure to evaluate the Machine Learning models performance on unseen data. 10FCV was selected as it generally results in a less biased estimates of model skills than other training methods including percentage split. With the large imbalance of the attribute values in the dataset (some individual components were delivered far more than others) the ability 10FCV to train and test the dataset on every instance provides the best possibility of a more generalised model and should also reduce the likelihood of over-fitting to the most prominent feature instances.

Model Selection

Naïve Bayes

The first model developed was a Naïve Bayes Algorithm. This classifier is based on Bayes theorem and an assumption that attributes are independent of one another. In other words, a Naïve Bayes classifier assumes that the presence of one feature is unrelated to the presence of any other features. Naïve Bayes is a fast & simple model to test and it generally works well in categorical problems compared to numerical variables. The Gaussian Naïve Bayes model was used for classification in the scikit-learn library

Support Vector Machines

The second model developed was a support vector machines (SVM) model. SVM is a supervised learning algorithm that can be used in both classification and regression problems. The algorithm works by plotting each observation in an n-dimensional space where n is the number of features in the dataset, in this case four. The resulting outcome is to then find the optimal hyperplane which bisects these points into the classifications of Late (1) and Not Late (0). A Linear model was tested giving a linear decision planes, all attribute arrays were normalised for fitting. These algorithms were imported from the scikit-learn library.

Decision Tree

The final Machine Learning algorithm tested was a decision tree. Decision trees are a supervised algorithm, that are based on a tree network of decisions. Beginning at the root, or first decision, of the tree, the algorithm breaks down a data set into smaller and smaller subsets while at the same time incrementally developing the decision tree itself. The final result is a tree with decision nodes and leaf nodes. Each decision node in this case has at least two branches. Each Leaf node represents a classification or decision. The key features of a decision tree are entropy & information gain, the algorithm partitions the data into subsets that contain instances of homogeneous values. If the sample is completely homogeneous the entropy is zero and if the sample is an equally divided it has entropy of one. The information gain is based on the decrease in entropy after a dataset is split on an attribute.

The goal output of a decision tree is about finding feature that returns the highest information gain. The algorithm was imported from the scikit-learn library.

4.3.3 Artificial Neural Network

In addition to the machine learning models an Artificial Neural Network was generated using Keras.

Network Topology

A sequential model was created, and various network topologies were tested until a suitable architecture was created. As the model developed is a binary classification problem with four features, four input dimensions are required. The model generated is fully connected, using the Dense class as defined in Keras.

A number of network topologies were trialled and tested including increasing the number of hidden layers and also the number of nodes in each layer using a trial & error approach until a suitable structure was found that was large enough to capture the structure of the problem, but also not being too large, resulting in overfitting, signified by the training data outperforming the testing data, or causing unnecessary amount of compute time for no additional performance gain. The final model resulted in four layers, with 30 nodes on the input layer, 20 nodes on each of the hidden layers and an output layer of 1, as required for binary classification.

The rectified linear unit activation function (ReLU) was used as the activation function on the first three layers and sigmoid was used on the final output layer. In general, better performance can be achieved using the ReLU activation function than tanh or sigmoid functions, however, using sigmoid on the outer layer ensures that the output of our network will always be between 0 & 1, as required by our classification model, and we can round to either class with a default threshold of 0.5.

Hyperparameter tuning

After defining the model, additional hyperparameters were tuned as part of compiling. The loss function to evaluate weights involves using cross entropy. This loss is defined as 'binary_crossentropy' in Keras and is used for a binary classification problems. The stochastic gradient descent algorithm 'adam' was selected as the optimiser for the model.

Adam was selected due to its efficiency and also as it automatically tunes itself and generally provides good results across a range of problems.

The model was then fit using the default keras batch size of 32 and trialled over a range of epochs until there was clear convergence in the plots of accuracy in the training & testing data, without excessive computing after convergence. A test & training split of 30/70% was also selected to ensure the model was tested on unseen data.

Final Network

After building, compiling & fitting the final model, the last step was to evaluate the model's performance and in particular, to allow it to be compared against the Machine Learning model results. To do this, the dataset was divided into defined training & testing datasets of 30% & 70%. Using scikit-learns 'classification_report' output it was possible to determine the precision, recall & f1 scores of the test set from the models predicted output vs. the actual result. The final model was run for 50 epochs. The graph of Training vs Test accuracy can be seen in Figure 4.

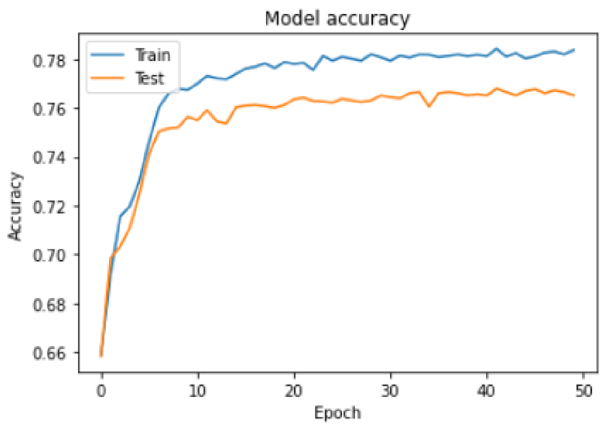

Figure 4: Final Network Training & Testing

Class imbalance

As will be discussed in the results section of the report, the high levels of class imbalance in the dataset appeared to have some influence on each model's performance. As a result, an additional test was carried out on the Artificial Neural Network with the same configuration described above but with the entire dataset under sampled to ensure equal distribution of both classes. This under sampling was completed using imbalanced-learn. The model was then retested.

5 Results & Discussion

5.1 Model Results

To begin, the results of each model discussed is shown below in the below tables.

	Class	Precision	Recall	F1
Naïve Bayes	Not Late	69 %	83 %	0.75
	Late	46 %	29 %	0.35
	Accuracy	64 %		

	Class	Precision	Recall	F1
Decision Tree	Not Late	37 %	24 %	0.29
	Late	12 %	21 %	0.15
	Accuracy	23 %		

	Class	Precision	Recall	F1
SVM	Not Late	67 %	97 %	0.79
	Late	49%	5 %	0.09
	Accuracy	66 %		

	Class	Precision	Recall	F1
ANN	Not Late	79 %	89 %	0.83
	Late	71 %	54 %	0.62
	Accuracy	77 %		

5.2 *Model Comparison*

5.2.1 Model Results

The results of each of the models vary. The following sections will discuss in some detail the results of each.

Decision Tree

Firstly, the Decision tree stands out due to its very poor performance overall. Total accuracy was found to be only 23%. Precision is the Positive prediction value, the ratio of correct positive predictions to the total predicted positives, given by the formula:

$$P = \frac{TP}{TP + FP}$$

For this model, precision for both Late & Not Late deliveries was found to be below 37% which is a very poor result particularly in the case of Late predictions indicating the model incorrectly predicted a Late delivery over four times more than its correct Late predictions. Similarly, the Recall (which in the case of a binary classification model, the recall of the positive class gives the Sensitivity & the negative class gives specificity) which is the true Positive Rate, the ratio of correct positive predictions to the total positives examples, given by:

$$R = \frac{TP}{TP + FN}$$

Both results for Late & Not Late predictions were found to be below 25%. This indicates that for both Late & Not Late predictions, with each correct prediction, there were over three incorrect predictions made in the same class.

Overall the Decision tree has shown very poor performance in testing, this may be due to overfitting of the data. Decision Trees can overfit as without some mechanism to the stop the splitting process in training, each leaf in the tree may be a sample in which case the model has fully learned the data set. Some steps that could help improve this would be dimensionality reduction or a random forest in where multiple trees are combined into an ensemble algorithm.

Support Vector Machines

The second model, SVM, initially looks to have some promising results, with an overall model accuracy of 66%. However, with examination of the report results it is clear that this model has also not performed well. The very high recall result for Not Late & opposite result for Late predictions indicate that the model has predicted the vast majority of results as being Not Late. When considering the overall dataset used in this project the class balance of Not Late to Late samples is ~66/34% as shown below in Figure 5. If the model simply predicts all results of being Not Late the overall accuracy will be 66%. This can easily be interpreted as a good performing model, but it is however a false flag. The F scores of each model also represent this. F score represents the harmonic mean of precision & recall. A score of 1 represents perfect precision & recall. It is given by:

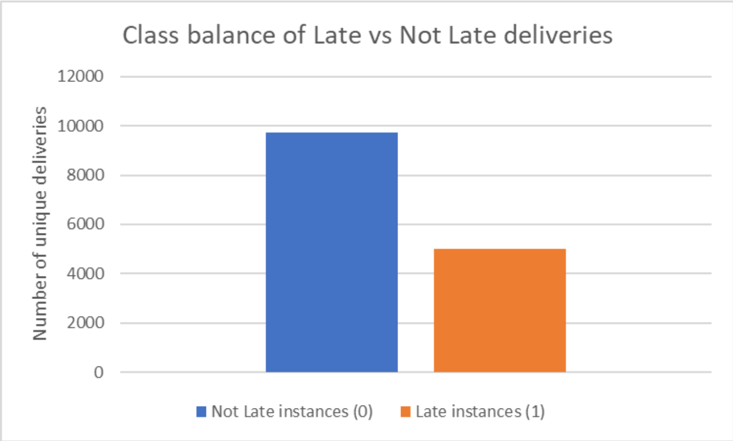

Figure 5: Class distribution

$$F = 2 * \frac{precision \ * \ recall}{precision + recall} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

The F score of 0.09 of the Late classification indicates a very low performance, the precision of the late score of 49% indicates that for the few Late predictions the model made, just below half were correct. Overall, this model has also performed poorly. Once again, this result may be caused by the class imbalance in the dataset.

Naïve Bayes

The Naïve Bayes model appears to have generated some positive results. While the overall accuracy is 64%, unlike SVM this model does have a distribution of predictions for both Late & Not Late results. The model has predicted with very recall the number of Not Late deliveries, the precision is also relatively high, but this is partly driven by the class imbalance in the dataset. As for the Late predictions the model's accuracy was 46% this indicates that it is only about as accurate as chance at predicting late deliveries. The low recall score of only 29% also indicates that the model did not predict enough deliveries as being late & is biased towards Not Late deliveries.

Overall, this model has a modest performance in that it is capable at predicting with some accuracy which deliveries will not be late. The implications & uses of this will be discussed in the following sections.

ANN

The final model, the Artificial Neural Network had the best performance of all. With a total accuracy of 77% and high scores in precision & recall across both classes shows that the model is capable of providing some valuable insight into predicting Late & Not Late deliveries from this supplier. Once again, the model performs better at predicting deliveries that are Not Late, however, this model is the only of all tested that can predict with greater than chance accuracy deliveries that will be Late by predicting 54% of all Late deliveries in test with a precision of 71%.

5.2.2 Class imbalance

The final test of the under sampled dataset gave the below results.

	Class	Precision	Recall	F1
ANN (class balanced)	Not Late	67 %	90 %	0.76
	Late	84 %	56 %	0.67
	Accuracy	77 %		

Overall, the impact of having balanced classes has had minimal effect on the performance of the ANN. This may suggest the model is generalised.

5.3 Discussion

5.3.1 Relevance & Usage

In order to understand the relevance of the results and any potential future usage of these models we should return to the Project Question: Is it possible to predict when a supplier will miss a delivery target date using Machine Learning?

With this some information is more valuable than others. For example, the risk of false negative prediction (i.e. prediction a delivery will not be late but the delivery was actually late) has a higher impact than a false positive (predicted delivery will be late and it was not late). With this, the recall of late predictions will be priority. As can be seen from the model results, only the ANN had a recall of greater than 50% meaning it correctly identified more than half of the late deliveries in the test data set. The ANN's precision was high in predicting late deliveries, which is also a benefit.

Secondly, another valuable insight these models may provide along with identifying which deliveries are at risk of being late is identifying which deliveries are at least risk. In other words, the ability to be able to accurately predict from a list of deliveries those are that are highly unlikely to miss the target delivery date would allow buyers are procurement professionals to prioritise their work with suppliers and focus on deliveries that are identified at risk and also those were the outcome is unclear.

This ability of the model to classify deliveries into high & low risk items has the potential to be key feature of the future for this project. Models developed can be used as supplementary tools to help in order planning, risk management and prioritisation with Supplier orderbooks.

5.3.2 Future Work

Overall, with time constraints this project is a starting block to a much larger project involving Supplier OTD and Machine Learning. In order to develop this work further.

Additional attributes and features

This project involved building Machine Learning models using quantitative data in the form of delivery history records and material pricing. In future, this should be supported with qualitative input in the form risk lists & quality scores for each supplier. These are human input features that should provide key insight into how the Supplier is performing at a given time. It would be required that this data is collected in a constructed form that would support the prediction model created.

Data cleaning

As discussed in the report due to the manual process of booking deliveries and the issues associated with volatile production planning & MRP it is very difficult to gain consistent data on early & late deliveries with erroneous data included throughout. A future process where the dataset is monitored as it is being developed (i.e. a process of cleaning data as it is being generated) may help improve accuracies and reduce the amount of outliers and noise within the dataset.

Supplier Expansion

The data should be expanded to include multiple suppliers. This project worked exclusively with a single supplier looking at their delivery history over a number of years. The supplier provided parts in a single commodity (sheet metal). This process allowed for a focused approach to answering the project question by removing additional variables including location, company size, additional commodities. If the work in this project can be further developed it would be beneficial to complete models on other suppliers and commodities and ultimately on the entire supply base for the company.

6 Conclusion

In conclusion, the project question has been answered, in that by examining the deliveries of a material supplier to the company over a period of six years it is possible to use machine learning to predict with some level of accuracy when a supplier will miss a delivery. Due to the limitation of data available and the ability to fully clean this existing data there are some limitations to the accuracy of the models created, however with future planning and additional features being included it will be possible to expand on the work already completed.

There is significant potential to build on the work of this project to create a model or interface to assist procurement & buying professionals and their suppliers as a supplementary tool. This can be in the form of prioritising their order tasks, evaluating risks by identifying deliveries expected to miss target delivery dates and also those that are at lowest risk, allowing them to focus on higher level tasks.

7 References

Ambulkar, S., Blackhurst, J. & Grawe, S., 2015. Firsm's resilience to supply chain disruptions: Scale development and empirical examination. *Journal of Operations Management,* Issue 33, pp. 111-122.

Armstrong, S., Sotala, K. & OhEigeartaigh, S. S., 2014. The errors, insights and lessons of famous AI predictions - and what they mean for the future. *Journal of Experimental & Theoretical Artificial Intelligence,* 3(26).

Banerjee, A. G. et al., Boston. *A Hybrid Statistical method for accurate prediction of Supplier Delivery times of aircraft engine parts.* 2015, Computers in Information Conference.

Baryannis, G., Danie, S. & A. G., 2019. Predicting supply chain risks using machine learning: The trade-off. *Future Generation Computer Systems,* Volume 101, pp. 993-1004.

Brintrup, A. et al., 2020. Supply chain data analytics for predicting supplier disruptions: a case study in complex asset manufacturing. *International Journal of Production Research,* 58(11), pp. 3330-3341.

BSR, 2020. *Future of Supply Chains 2025.* [Online]
Available at: https://www.bsr.org/en/our-insights/primers/future-of-supply-chains-2025
[Accessed 13 November 2020].

Chen, Z.-L. & Hall, N. G., 2007. Supply CHain Scheduling: Conflict and Cooperation in Assembly Systems. *Operations Research,* Issue 55, pp. 1072-1089.

Chopra, S. & Meindl, P., 2016. *Supply Chain Management Strategy, Planning, and Operation.* 6th ed. s.l.:Pearson.

Gehringer, A., 2020. *Rythm Systems.* [Online]
Available at: rythmsystems.com/blog/how-to-improve-on-time-delivery-with-the-right-kpis-culture-and-disciplines
[Accessed 12 November 2020].

Giunipero, L. C., 1990. Motivating and monitoring JIT supplier performance. *Journal of Purchasing & Materials Management,* 26(3).

McLean, T., 2017. *On Time, In Full.* 1st ed. s.l.:Taylor & Francis Inc.

Soumendra Mohanty, S. V., 2018. *How to Compete in the Age of Artificial Intelligence: Implementing a Collaborative Human-Machine Strategy for Your Business.* s.l.:Apress.

Tan, K.-C., Kannan, V. R. & Handfield, R. B., 1998. Performance, Supply Chain Management: Supplier Performance and Firm. *International Journal of Purchasing & Materials Management,* 34(3).

Toyota (GB) PLC, 2013. *Just-In-Time – Toyota Production System guide.* [Online]
Available at: https://blog.toyota.co.uk/just-in-time
[Accessed 15 11 2020].

Viswanadham, N. & Gaonkar, R. S., 2003. Partner Selection and Synchronized Planning in Dynamic Manufacturing Networks. *IEEE Transactions on Robotics and Automation,* 19(1).