MASTER

Predicting supplier material quality issues through procurement process deviations

Mulder, J.B.

Award date:
2021

Predicting supplier material quality issues through procurement process deviations

Master Thesis

Eindhoven University of Technology
Department of Industrial Engineering & Innovation Sciences

Author: J.B. Mulder
Student Number: 0870573

First Supervisor TU/e: dr. ir. R.M. Dijkman
Second Supervisor TU/e: dr. L. Genga
Third Supervisor TU/e: Z. Bukhsh
Company Supervisor ASML: R. Garenfeld

Date: 12-04-2021

I Preface

It feels a bit strange to be writing this part of my master thesis report, as I realize that my education at the Eindhoven University of Technology is actually coming to an end. I would like to thank Remco for guiding me through the final stages of the process and also Laura and Zaharah for taking the time to review my work. I would also like to thank the ASML team for giving me the opportunity to conduct my research and answer any of my many questions. Particularly, I would like to thank Robert, for taking his time for one on one meetings with me despite the time pressure and supporting me with getting my new job. Of course I also have to thank my parents for supporting me throughout my education (morally and financially) and helping me to pursue the highest level I could achieve. Without their support I would not have felt so free and secure to grasp onto every opportunity that presented itself to me throughout my time in Eindhoven and would not have been able to enjoy my time so much. I would also like to thank my girlfriend, Neeltje, for motivating me when I was less motivated and nudging me to continue to be and do my very best. Finally, I would like to thank the boys from STROPDAS for making the student years in Eindhoven (too) much fun and helping me to complete my education in 7,5 years. I wouldn't have it any other way and they really taught me a lot. Before you lies my master thesis report in which I have invested the knowledge I have gained over the last years. Through this report I hope to show my time in Eindhoven was well spent and I have deserved to claim the title Master of Science.

II Abstract

Using data to understand business performance is already common practice in many firms, however the full potential of this has yet to be achieved. Data concerning change logs of business processes are also becoming an increasingly more interesting source of information for businesses who are looking for new ways to implement and understand the knowledge they potentially hold. This research investigates if logged deviations in the purchasing process can be used to predict material defects which have been confirmed to be caused by the supplier. The purpose of the research is to identify a predictive and statistical relationship between purchasing process deviations and material defects. This is achieved by linking knowledge from business experts and literature and testing this using both statistical and predictive models. Change logs from a supplier were collected and linked to known material defects to create a dataset that was statistically analyzed and used to train and test machine learning prediction models. The results show that both the total number of deviations per order and the occurrence of certain types of deviations have a statistical relationship with the occurrence of material defects. Furthermore, multiple prediction algorithms proved to be able to successfully predict material defects at different levels of confidence, based on the balanced training and testing dataset. Testing the model on real and unbalanced data that had not been transformed, shows the gradient boosted trees model is still capable of predicting defects at an acceptable level. However, the model scores are much lower than those when tested on the balanced data. This shows the model is therefore applicable in practice and could help identify defective materials, but could also benefit from further improvements to ensure better prediction outcomes. In practice, this means that customers can evaluate suppliers by looking at the purchasing process and identifying if orders are at an increased risk or if an intervention is required. The model's output will prove to be more or less applicable to the customer, depending on the required accuracy and precision of material defect predictions. No previous research about the effect of process deviations on material quality was identified, but the research shows that there is a statistical and predictive relationship between the two. Further investigations into this relationship can be based on this research to better understand the impact of purchasing process deviations on supplier material quality.

Key Words: *purchasing process, process deviations, material defects, statistical relationships, predictive modelling*

III Executive Summary

Problem Formulation and Solution

ASML is a large firm that is dependent on multiple suppliers delivering and cooperatively designing components. Monitoring each supplier and ensuring they deliver according to the desired standards is a complex and difficult process that requires good cooperation between supplier and customer. One of these standards is the quality of the materials the suppliers produce. Defective materials can be costly and increase unexpected down time of machines. It is therefore desirable to reduce the number of material defects, understand what the cause of a defect was, and be able to predict defective materials before they cause costly issues or delays.

A tool that is capable of predicting material defects could help catch defects early in the products life cycle. By understanding and knowing what materials are at risk of being or becoming defective, materials could be inspected or replaced to prevent unexpected issues. It would also be desirable to identify the issue as early as possible, for example before the components are assembled in the factory. This was tackled by designing a prediction model that uses data that is available before the supplier's materials arrive at the factory. The prediction model uses data from the purchasing process to determine if a material is at risk of being or becoming defective. It does this based on change logs extracted from SAP linked to purchase orders to predict if a material notification caused by the supplier will occur. The output of the model is generated using gradient boosted trees in KNIME and shows what orders are predicted to be at risk, including a confidence indicator. The prediction is based on the occurrence of 18 different process deviations and the total number of deviations per order.

Model Practicality & Performance

The output of the model could be used to identify materials that are at risk and initiate precautionary actions. These actions could be physical inspections of the materials at risk or could also be used to build safety stock of risky materials. Furthermore, the model's output could be used to analyze the purchasing process and reduce risks caused by ASML through purchasing. However, practical application of the model depends on the accuracy and precision requirements for the supplier. The model is capable of predicting results with 95% accuracy, but it is biased towards producing false positives and produces far more hits than actual occurring defects. Also, the model does not predict every defect correctly, resulting in an output that misses potential defects and marks false positives. Depending on the desired follow-up actions, the model can be implemented into practice for high-level insights or should be improved upon to meet requirements.

Confusion Matrix

		Prediction		
		0	1	T
Actual	0	10456	625	11081
	1	12	135	147
	T	10468	761	11228

1/0: There was/was not a defect (MN)

Correctly Classified	10,591
Wrongly Classified	636
Accuracy	94.318%
Error	5.682%
Cohen's Kappa (κ)	0.282
True Positive Rate (Recall or Sensitivity)	0.918
True Negative Rate (Specificity)	0.944
Precision	0.177
F1	0.297

Table i: Performance output of the prediction model on real unbalanced data.

To better understand the performance of the prediction model on the real and unbalanced data, the results are depicted in *table i*. The confusion matrix can be used to inspect the predictions (columns) and the actual value (rows), where 0 indicates no defect, and 1 indicates there was a defect. The matrix shows 10,456 order items were correctly predicted to not have a defect, and 135 cases correctly predicted to have a defect. Twelve cases were false positives, no defect predicted when there was actually a defect, and 625 cases were false negatives, a defect was predicted when there was none. The Cohen's Kappa (κ) and F-score consider recall, specificity, and precision scores to provide a more accurate model rating. Cohen's Kappa shows how much better the model is than randomly guessing the outcome, where a score between 0.21 and 0.4 are considered an acceptable prediction score.

Implementation and Use

The current model could be used to implement an irregular manual prediction when this is considered relevant by quality managers. The current model, however, was only tested on one supplier and therefore it would be advisable to train a new model when applying the solution to other supplier with different materials. To do this, historic data of the supplier should be used to train the model and can be collected from three sources that must be linked to each other. Firstly, the purchasing process change data can be collected from SAP with help from the global operational procurement team. This team supplied the logged changes stored in multiple SAP data tables (EKKO, EKPO, CDHDR, and CDPOS) for the supplier which is then used to make the prediction on. The logged changes are then counted and for every order, the occurring changes and frequency is registered. Only orders related to new buys should be included. The purchase order numbers and purchase order items are then linked to the equipment numbers extracted from the CAD CAS dataset. Using the equipment number, the material notifications from the SQLD table can be linked to the purchasing process deviations. The SQLD data is stored by the data analytics team and used to monitor the MQP 3.0 score. Only the material notifications that are used to calculate the MQP score should be considered in this prediction. Following the data collection and linking, the data has to be prepared to train the prediction model. This can be achieved by balancing the data through random undersampling the majority class and oversampling the minority class using SMOTE. The data can then be bootstrapped to produce a bigger and less variable dataset. The dataset can then be used to train and test the gradient boosted trees model using 70% as training and 30% as testing data. Once the model is trained, the specific order items that must be inspected can be implemented into the model with the required data. The model will produce a prediction and a confidence indicator, this output can then be used as desired. Implementation of the model could be done by combining the tools Excel, Spotfire, and KNIME. However, if collected data is accessed directly by KNIME, the entire model could be executed in this program. In short, the model is based on specific change types and their frequencies to predict if the order will result in a material defect and with what confidentiality.

Potential Improvements

To ensure the output of the model becomes more reliable and useable, multiple improvements are suggested for future considerations. Firstly, finding the optimal parameters for producing the most accurate predictions should be considered. This research does not experiment with optimizing the prediction output and therefore reliability of the output could potentially be increased. Furthermore, as stated, the model was trained on only one supplier. By training the model on multiple suppliers, the model may become more generalizable and not require a separate model for each supplier. Quality experts and literature also mention material characteristics such as maturity and complexity to be of significance when predicting material quality. Implementation of material characteristics can also be done using the equipment number and should therefore be considered in future models. Finally, more complex models that take into account process patterns or more detailed information of the logged changes could also be used to make predictions, these models could potentially provide more reliable output because they are more detailed than simple occurrence counts of each deviation type.

IV Table of Contents

V List of Tables

VI List of Figures

1 Introduction

This research's objectives were to investigate the relationship and predictability of the supplier's material quality based on the occurrences and frequencies of process deviations in the purchasing process. The research provides insight into what factors have influence on supplier quality, the influence of process deviations and purchasing on material quality, the relationships between process changes and material defects, and the predictability of material defects based on purchasing process deviations. It does this by combining the knowledge of business experts and scientific literature, and testing this using statistical analyses and machine learning prediction algorithms. Literature and business experts provide a base for the research on which testing on actual data was executed. The research is based on an interpretation of the CRISP-DM methodology and was performed through incremental improvements of business understanding and modelling. The collected information, statistical analyses, and prediction modelling were tested on real data from a ASML, a large high-tech firm in the semiconductor industry. The data provided by ASML centralized around one of its suppliers, who was considered to be a good representation of the critical component suppliers. The data used in the analyses and tests on was one-sided and contained information about the purchasing process, the material characteristics, and the material defects of this supplier. The research can be used to better understand the influence of the procurement process on material quality and potentially help firms to better understand and predict quality of incoming materials based on knowledge that is available internally.

This chapter further explains the motivations behind this research and provides insight on the scoping limitations. Also, a brief explanation of the report structure is stated, to help the reader navigate through the report.

1.1 Motivation

Monitoring firm performance with key performance indicators that are based on large quantities of data is a common practice in many companies and often considered a key component to successfully managing a firm. It is also becoming increasingly more common to use data to predict future performance or identify upcoming issues before they have occurred. It is, however, still not completely clear what data can and should be used to make certain predictions, this is also the case when predicting supplier or material quality. Information on how to measure supplier quality, what information is relevant, and what factors influence supplier material quality are only sparsely addressed in prior research. This results in firms creating their own measurement techniques that are not based on scientific research, but rather on gut feeling or experience. Furthermore, because firms do not know how to correctly monitor supplier quality, it is difficult to know what they can do to improve quality and how to predict upcoming issues. Because of this, it is unknown if actions or processes performed by the firm have effect on the quality of materials delivered by the suppliers. And therefore, the actual source of a defect is uncertain, and it is unknown if it could have been prevented. In an environment where efficiency and quality are increasingly more important, it would be interesting to understand what factors influence this and how material defects could be predicted. By linking information on how to measure quality, factors influence quality, and looking at internal business processes, a firm can influence or recognize materials quality before materials arrive at the factory. This research attempts to expand on our knowledge of predicting supplier material quality through process deviations and material characteristics by linking scientific and practical knowledge and testing it on real data.

1.2 Scope

To ensure the research could be completed within the time restraints using the available information, the research was defined to fit within a manageable scope. Research concerning supplier quality prediction was limited and therefore valid sources are sparse. Considering this fact in combination with the scale of ASML and the quantities of data that were available, it was required to scope the research to achievable standards. To ensure the scope is clear, this section briefly states important scoping limitations and focusses on explaining why certain scoping decisions were made for this research. Firstly, the research was based on information and expertise held within ASML and focusses only on data that was available in their databases. This decision ensures the research could be conducted without cooperation from the suppliers. Supplier material quality was indicated using a form of material defect rate that is used internally, the material defects were internally stored as 'material notifications'. Process deviations were indicated using automatically logged value changes in the SAP database. The research only includes material defects, or material notifications (MN), recorded in 2019 and 2020, due to the limitations of data storage formats. Only the material notifications that were verified to be caused by the supplier were considered in this research. This is because the research attempts to predict supplier quality and cannot include defects caused in the ASML factory or by its customers. Furthermore, the research only takes in account a single supplier to keep the data manageable. The selected supplier was suggested by quality experts at ASML based on the availability of data, the firm's levels of standardizations, the importance of the supplier, and the overall generality of the supplier. The research was also limited to using only purchase orders (PO) for new buys (NB), implying all the materials were newly produced components and not refurbished or repaired parts. This option was selected to ensure that occurring material issues were certainly due to supplier production activities and not a result of poor repair work or other underlying material issues. Repair orders would also follow a different purchasing process, potentially making the data far too complex. Furthermore, only orders that were delivered to the local factory in Veldhoven were considered. This is because new products generally go to this location, the selected supplier only delivers to this location, and because the most accurate data is collected in the local factory. Finally, it is important to note that the research took into account the purchasing process of all purchase orders from 2018, 2019, and 2020. This was done to ensure all the relevant material notifications could be connected to a purchase order, as it can take some time before materials are delivered and even more time before defects occur. A summary of the scope is shown in the table below (*table 1*).

Scope Limitations	Value
Vendor	Single Supplier
Source of Cause	Verified Vendor
Factory	Plant in Veldhoven
Purchase Type	New Buys (NB)
MN date range	2019 & 2020
PO date range	2018, 2019, & 2020

Table 1: Overview of scoping limitations and their values.

1.3 Report Structure

The research is structure chronologically to help the reader understand the step-by-step approach to achieving the research outcomes. Following this chapter, the report first addresses the research goal and research questions that are focused on and answered by the research. Next, it explains what research method was applied and how the research was approached, based on nine iterative steps. This is followed by business understanding and a literature review of the relevant research fields and linked to each other to provide a better understanding of the context in which the research was conducted. Next, the data collection and preparation are explained to understand where the data comes form. Following this, the data is analyzed using both statistical analyses and machine learning algorithms. Finally, the output is expressed and insights produced by results are discussed and concluded. A more detailed description of each step can be found in the research methodology chapter.

2 Research Goal & Questions

This section explains the goal of this research and the research questions that are answered in the report to support the outcome.

2.1 Research Goal

The goal of this research is to analyze if the frequency of purchasing process deviations can be used as a predictor for supplier material quality.

The research attempts to identify a statistical and a predictive relationship between purchasing process deviations and the occurrence of supplier material defects as indicator for quality. Using process deviations as a predictor for quality was not yet been researched thoroughly and it was therefore unsure if it could be used as a reliable and accurate predictor. This research therefore had to identify if there is a predictive relationship between the occurrence of process deviations and material defects, if any specific deviations tend to result in defects, and how reliable these predictions are. Furthermore, the research should provide insight into applying process deviation defect predictions into a practical scenario and briefly touch on other data that could be added to the predictor to further improve defect predictability. Using the following research question as a guide to analyze the different aspects of supplier material quality prediction, a complete overview of key characteristics could be derived and were used to provide insights to support the research goal.

2.2 Research Questions

1. What are valid indicators of supplier material quality and what factors have influence on these indicators?
2. Do purchasing process deviations have influence on supplier material quality and what types of deviations have influence?
3. Is there a relationship between the occurrence of a process deviation or the total number of deviations and the occurrence of material quality defect indicators?
4. Is there a relationship between specific process deviation types and the occurrence of material quality defect indicators?
5. Can logged deviations from the standard procurement process be used to (reliably) predict the occurrence of material defects?

Successfully answering all research question should provide significant insights into the research goal and help clarify if purchasing process deviations could be used to predict upcoming material defects. Question 1 and question 2 are were addressed using scientific literature and business expert knowledge as explained in the business understanding and literature sections. Their function was to help understand what supplier material quality is and what is currently known about influencing and predicting it, especially through procurement process deviations. Questions 3, 4, and 5 were answered in the data analysis section and discussion, using both statistics and machine learning. Both question 3 and 4 provided an understanding of the potential relations between the procurement process and material defects. The purpose of question 5 was to test if process data can indeed be used for prediction theoretically and practically.

3 Research Methodology

The following section explains how the research was conducted. In this chapter details about what methodology the researched was based on and how it was executed in practice are explained. Also, an indication of the tools that were used and their general functions are provided.

3.1 Research Approach

The research approach was based on the Cross-Industry Standard Process for Data Mining (CRISP-DM) to support structuring and executing the research (Chapman, et al., 2000). This approach was selected because it is suited for research centralized around data mining and data analyses. The advantage of using CRISP-DM as a basis, is that it is an iterative process that can be used to improve a model based on recent findings. The deployment step was not included in this research because the final model was not directly applied into practice. A detailed description of the actual research steps is explained below and was conducted according to the structure in the following figure (f*igure 1*).

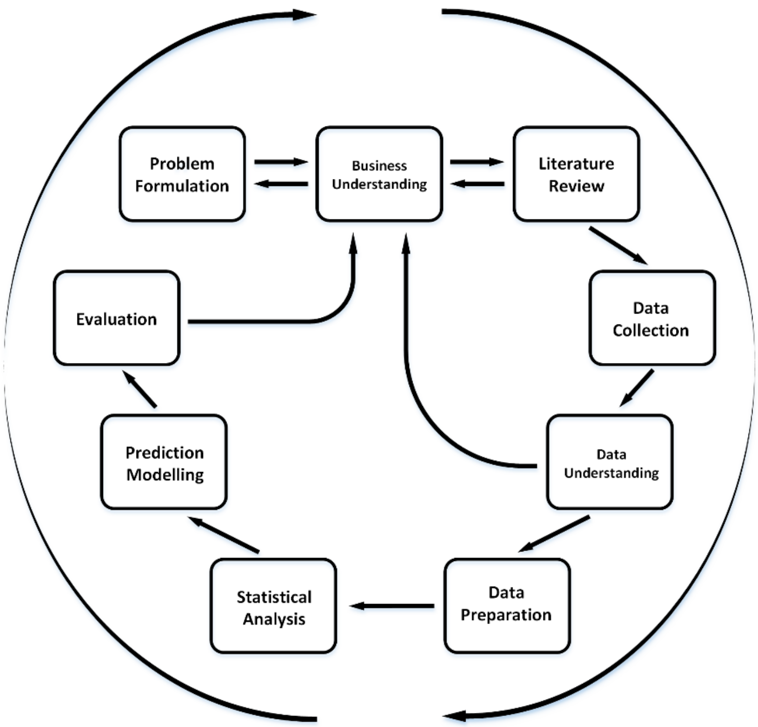

Figure 1: Research approach based on CRISP-DM cycle

3.1.1 Problem Formulation

Because the research was conducted in collaboration with ASML, the research started out with formulation and identification of a research problem. Multiple perspectives of a proposed general problem were taken into account to get an understanding of what research would be interesting to both ASML and scientific knowledge. With help from business experts, a high-level problem was formulated, this problem focused on supplier quality prediction. The proposed problem was then explored and modified using insights that emerged from both the business understanding and the literature review phases. Through iteration, a final problem was formulated that could be answered within the time restrictions and would provide potential new insights that could be valuable to ASML but perhaps also to scientific research in general.

3.1.2 Business Understanding

To understand what information was available and grasp what mechanisms or processes were of influence on the research problem, different aspects of the business were investigated using a range of accessible sources. The key outputs of the business understanding phase were insight into how quality was currently monitored, how it was measured, and what processes may be of effect on it. Firstly, business instructions and documents were inspected, to get a better understanding of the general procedures and information that was available. These documents also provided insight into how the business was supposed to work. More detailed information about the specifics of the business aspects were collected by speaking to business experts or product owners within the firm. The information collected through these conversations provided more detail about the actual processes, helped to identify relevant data sources and were used to understand the relationships between the different processes within the firm. Furthermore, the business understanding phase allowed the business experts to express their ideas on the research problem, providing unique insight into custom problems and perspectives. These expectations were later compared to scientific research from the literature review and linked back to better understand the business processes. Better understanding the internal mechanisms of the firm also helped to further scope the research and adjust the research goal. The information was however collected in an unstructured fashion and generally came from short meetings or messages.

3.1.3 Literature Review

The function of the literature review was to identify if current business indicators were valid according to scientific sources, if the expectations of the business experts were viable and realistic, and to understand how the findings from the statistical analysis and the prediction models could be explained. Investigation of supplier quality indicators and factors influencing them were conducted in a separate literature review of which the most important findings were used in this report. The review was conducted according to the principles stated by Randolph (2009) in A Guide to Writing the Dissertation Literature Review. The article states a literature review is performed by completing five consecutive steps, problem formulation, data collection, data evaluation, analysis and interpretation, public presentation (Cooper, 1984). The literature about the influence of process deviations and the influence of the procurement process specifically, were not covered in this literature review and were collected more arbitrarily following the initial search. This was done because the subjects were highly specific and little research about the topics could be found methodologically. The new sources were used to help bridge informational gaps that were not yet covered. Specific details about how the literature review was conducted are explained in the corresponding chapter.

3.1.4 Data Collection

Data was extracted from a variety of sources that were identified in the previous phases and considered to be important for the research. The collected data was required to provide insights into the purchasing process deviations and had to ensure the information could also be linked to a form of material quality. Because large quantities of data were available, it was important to only collect information that would be relevant for the research and fit within the scope, this was achieved by applying filters to the databases before extracting the required information. Information related to purchasing processes and purchase orders were extracted from multiple SAP tables by business experts and provided in Excel datasheets. The process data was collected and provided by business experts due to limited access to the databases and sensitivity of the data. All other data sources, such as data concerning material defects and equipment information could be accessed partially through annually archived datasets or through active datasheets that were used to monitor supplier performance and were update on regular intervals. These datasets or archives could be accessed using Spotfire, a designated data analysis tool commonly used internally. The goal of this step was to ensure all required data was localized and could be accessed. If information was found missing, this could occasionally be extracted from other archives by domain experts, else the information was considered ineffectual and removed.

3.1.5 Data Understanding

Following the collection of each dataset, the data was inspected and the information was interpreted. The goal of the data understanding step was to identify what information had been collected, how it could be used for statistics and prediction, and what data gaps were missing. The data understanding was conducted firstly by comparing the data to information from the business understanding phase. Any unclarities were resolved with the help of business experts or information found in the related business documents. Specific abbreviations and details to how values were calculated could be identified in documentation or, if required, explained in detail by business experts. In this step, the relationship between the different datasets were also investigated to ensure the collected datasets would be relevant to each other. This was achieved by comparing time stamp information, identifying if there was overlap between the datasets and occasionally requesting business expert to confirm or clarify uncertainties. In following steps is would also be relevant to connect the data to each other with the goal of producing one main source. To ensure this could be achieved, overlapping data columns were marked and checked. By iterating through the business understanding, literature, and data collection steps, all required data and the necessary information to link them was identified and collected. Data understanding was performed either using Excel or Spotfire, depending on the format it was stored in.

3.1.6 Data Preparation

The data preparation step deliverables were a master dataset with all identified data sources linked to each other and a training dataset for the prediction phase. The data preparation phase was performed in two stages to achieve this. First all collected data was merged into one master dataset by linking the different datasets to each other. This was done by adding information from one dataset to the rows in another dataset that could be linked based on unique identifiers. The incomplete data or data that could not be connected accordingly was eliminated because it could not be used for analysis. The master dataset would be the base set and was statistically analyzed in the corresponding section. Details about the content and how the mergers were executed can be found in the data collection and preparation chapter. Secondly, the data was transformed and prepared for predictive modelling. This step was conducted by removing unwanted or unreliable information that would not be used for prediction, removing outliers, balancing the dataset, and bootstrapping. Data balancing was performed by randomly undersampling the majority class and synthetically oversampling the minority class using SMOTE. The bootstrap was performed to produce a less variable and larger dataset. The resulting dataset is referred to as the balanced or predictive dataset, and an exact description of the transformations can be found in the data transformation subchapter. The first step of the data preparation phase was executed using Spotfire, the second step was executed in KNIME.

3.1.7 Statistical Analyses

The master dataset, prepared in the previous step, was analyzed using SPSS and KNIME to identify the relationships between material defects and process deviations. The goal of this step was to identify the relationships of the purchasing process aspects with material quality, and detail what relationships were significant. Depending on the data types, the chi-square test or the T-test was applied. The relationship between defects and the occurrence of deviations, the total number of deviations, and frequency of specific deviations were evaluated. A more detailed descriptions about the precise execution of each statistical analysis can be found in the data analysis chapter. Also in this section, the output of the analyses are presented and briefly evaluated, together with a description of how the output can be interpreted. To better understand the statistical output, the results were also compared to the expectations from literature and business experts. Details concerning this comparison can be found in the data analysis and the discussion chapters.

3.1.8 Predictive Modelling

The key deliverable of the prediction modelling step was to train and test different prediction models, identify what model worked best in this case, and investigate the models practicality. KNIME was used to generate six different prediction models, each using a unique algorithm. The algorithms were selected by identifying what options were available in KNIME, if these models were applicable for the research, how common they were, or how effective they are. The balanced dataset was used to train and test each model to ensure they all produced a prediction based on equal terms. The output of each model was analyzed using a scorer, which compared the prediction output to the actual value of the target data. This comparison produced a variety of scores and insights that could then be analyzed. The scores were used to compare the models to each other and identify the best prediction algorithm for this specific case. Details into how the models were set up and the output scores can be found in the corresponding chapters. Following the model generation and testing, unique insights provided by specific models were observed and evaluated. This was done by extracting exclusive information contained in each model. Finally, the most powerful prediction model and the input data were tested and validated. This was done by testing the relationship between the dependent and independent variables, testing the model using only statistically significant variables, and running the prediction model on the complete dataset with real data only. The output of this final test was recorded and analyzed using the same scoring indicators as the training sets.

3.1.9 Evaluation

The final step of the cycle was to evaluate the output of the previous steps and synthesize the findings of all steps. The outcome of this step can be found in the discussion and conclusion chapters of this paper. The outcome of the research was evaluated by linking information from literature research, expectations from business experts, statistical relationships, and the model predictions to each other. The goal was to explain what was expected, how the findings related to these expectations, and to understand why the output produced these results. The findings from the evaluation were used to better understand the business and make incremental adjustments or improvements to the research. The evaluation also helped to identify limitations of the research, which were collected and used to make suggestions on how to further improve the output or what other research should be performed to further validate the research findings. By connecting the information from each of the previous steps, a general explanation for the findings could be identified and used evaluate the validity of the outcome.

4 Business Understanding

The goal of this section is to provide insight into the context in which the research was performed and to get an understanding of the problem and expectations from a business perspective. The following section of the report provides a problem formulation, explains the perspective from business experts of the researched firm, and clarifies the mechanisms of the internal processes that are of interest for the research. The information stated in this chapter was collected from business documents, work instructions, and through business or domain experts. The information gathered from the experts was collected through various meetings, short interviews, or emails and were not recorded following a structured method. The business documentation and work instructions were used to clarify specific terminology and process standards or calculations.

Initial research showed that there was a desire to better understand if it is possible to predict material quality and to use that to improve company performance. Specifically, processes that are completed before the supplier materials are implemented in the factory, such as the procurement process, would help to minimize the impact of material defects. Business experts within ASML all consider deviations from the standardized processes as a potential influencing factor of the suppliers' quality. The actual impact of deviations in the procurement process on the quality of supplied materials is not considered to be strong but has also not been explored at any time. If prediction of material defects is possible, it could add value by reducing unexpected issues during the products usage.

4.1 Problem Formulation

The research was conducted within ASML's Sourcing & Supply Chain department. ASML is the leading manufacturer of chip-making equipment based on lithography. The S&SC department is responsible for creating and managing the supplier network of the firm. It strives to improve quality and reliability of its materials while reducing the supply chain costs. To keep track of supplier performance the firm focusses on the QLTCS (quality, logistics, technology, cost and sustainability) framework. Target scores for each of the QLTCS metrics are set annually and constantly monitored (ASML, 2021). Quality in general is tracked using multiple indicators, each based on different aspects of the supplier performance. Material quality of the suppliers is indicated using a material quality performance (MQP) indicator. The MQP indicates the number of defects as a percentage of the delivered materials over a time period and a lower percentage indicates better material quality performance of the supplier. Defective materials can be very costly because they result in, costly repairs or replacements, production delays or down time at the customers. It is therefore of high importance to ensure the materials ASML uses are of high quality. It would also help for the firm to be able to predict when materials or components are at risk of being or becoming defective to prevent or prepare for issues.

ASML does not produce all components in its factory, but rather assembles its machines by components produced in house and components delivered by suppliers. These components are often developed together with the suppliers, who are expected to deliver quality when the parts are ordered. ASML would benefit from understanding what they can do to help improve supplier quality or to predict if incoming materials from its supplier are at risk. To gain insights on this topic but ensure manageability, the research focusses on just one of ASML's many suppliers. The supplier, referred to as such for confidentiality reasons, was selected because it is a good representative for the scope of this research. There is sufficient data available on the supplier because it is one of ASML's core suppliers and delivers key components. Furthermore, the process is standardized but not yet optimal and with a sufficient number of defects to observe. The supplier is responsible for delivering high-tech mechatronic systems. More specifically, the company produces the base modules, called 'tubs', on which many of ASML's products are built. The relationship between ASML and the supplier is direct and collaboration is considered to be good. Significant amounts of information are shared and supplied equipment is tracked throughout its lifecycle. It is assumed that the supplier is a good representative for suppliers considered most important by ASML and therefore the results of the research can be generalized.

Because a broad range of components is sourced from suppliers, ASML cannot completely control the quality at which supplier deliver and does not always have all information the suppliers have. Supplier are subject to quality standards set by ASML, yet this does not imply they are always met. Also, being able to predict quality defects should result in the possibility to make improvements on them. Therefore, the data used to make the prediction should be observable before sourced components are implemented in the factory. It was observed that focusing on procurement process, that starts when a purchase requisition is received and ends when the order is delivered at the factory, could yield valuable and relevant information. More specifically, it was noted that it would be interesting to understand if the deviations in the purchasing process could be used to predict potential material defects.

4.2 Purchasing Process

To better understand how the purchasing process may influence the supplier's material quality, it is important to understand what the purchasing process looks like. As a whole, the procurement process is very expansive, but this section focusses only on the placement of purchases at standardized suppliers, starting from a purchase requisition and ending after the confirmation of delivery. The information used to analyze the process can be extracted from data concerning purchase orders (PO). A purchase order is triggered when the ERP system or the factory delivers a purchase requisition (PR). The purchase order is an electronic ticket that refers to a quotation from the supplier and indicates the product or service, quantity, 'need by date', and price for the product or service (ASML, 2021). The quotation used to make the PO is already agreed upon by ASML and the supplier during the new product introduction phase.

In an ideal situation, when the sourcing department receives a purchase requisition, a purchase order is automatically created in the ERP system. The PO must then be approved by the procurement department before it is sent to the supplier. Once the supplier receives the PO it must first confirm to ASML that the purchase can be executed successfully. After the confirmation the supplier can start production immediately. When the materials are received at the ASML factory the PO receipt is triggered and the purchase order is closed successfully. A high-level overview of the standard process is depicted below, in *figure 2*.

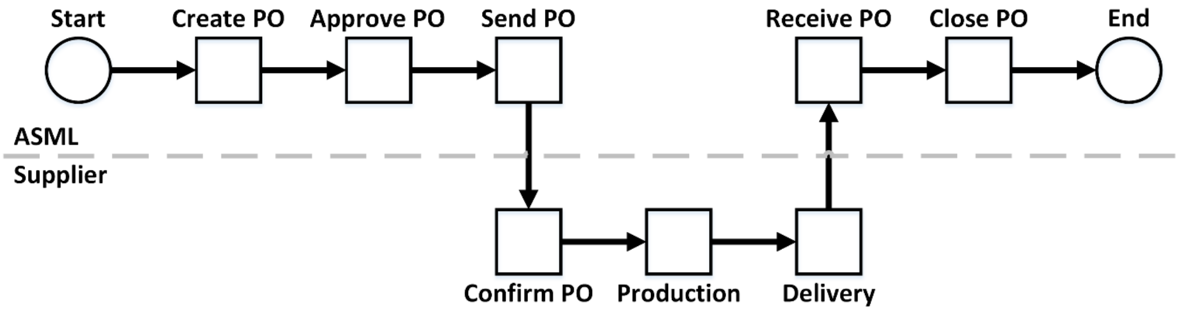

Figure 2: High-level overview of purchase order (PO) process.

Due to multiple factors, deviations from the desired purchasing process may occur. These deviations are logged in the related data tables in the ERP system and can be linked to the applicable purchases orders. A few simple process deviations are, for example, price changes, delivery date changes, or quantity changes. These deviations can occur throughout the purchasing process and can be caused by both the supplier and buyer. Besides being logged in the ERP system, the process is currently also monitored using process mining tools to better understand what deviations occur most frequently and in what sequence. Data concerning the frequency of process deviations was extracted from the ERP system by data experts and relevant data was identified collaboratively. Four different data tables were inquired to get a complete overview. Deviations are detected, or recognized, as a change to a data field in the ERP system. When a data field is changed, the time of occurrence, the old and new value of the data field, who made the change, or any other relevant information is logged by the ERP system and stored. The collection and preparation of the data is explained in more detail in the data collection chapter.

According to the domain experts it is to be expected that there is some relation between deviating from the standard procurement process and the quality of materials delivered by suppliers. Both the logistic and quality supplier managers (LSM & QSM) suggested that changes in the delivery quantity and delivery date may result in pressure on the supplier and subsequently result in a higher risk of quality failures. Both were however skeptical on how big the impact would be and if it could be used for prediction. The suppliers QSM also suggested that the maturity level of the components would be of bigger impact on quality and did not expect price changes would be relevant. The LSM added that volume and production capacities at the supplier were also more likely to be of influence on the quality. The operational procurement process manager also expected that a higher total number of deviations occurring could potentially have a negative influence on the quality. In short, it was expected that a higher frequency of deviations, order quantity changes, and delivery date or time changes could be of influence on or related to material defects.

4.3 Material Quality

Supplier performance is measured according to the QLTCS (quality, logistics, technology, cost and sustainability) framework at ASML. Each aspect of the framework provides a score for the supplier and can be used to monitor the correlating performance fields. This research focusses on the material quality aspect of the suppliers, which is best observed by focusing on the quality aspect of the QLTCS framework. Material quality can be indicated using a range of metrics as explained in the literature, however in this research the focus lies on a form of the defect rate of materials. At ASML, material issues and defects are recorded as a material notification (MN). These MN’s can cover a broad range of problems. A material notification is generally created to record and track suspected non-conforming products at ASML and are crucial for ensuring quality improvement. There are three types of notifications, the ‘customer quality notification’ (CQN), the ‘dead on arrival’ (DoA), and regular MN’s. Respectively, they imply being detected by the customer and resulting in a complaint, a component not working when arriving at the customer, and a non-conformity detected during the manufacturing process (ASML, 2021). In this research, the latter is of particular interest.

The definition of a material notification also implies that it is not only used to indicate defective materials however, but rather a collection place for complaints or issues related to materials. Only certain MN’s are actually related to material defects and relevant for this research, this is explained in the next paragraph. The MN’s related to the material defects are used to calculate the ‘material quality performance’ (MQP) performance indicator. The MQP is the leading quality indicator and indicates the percentage of materials that were delivered with defects from the supplier. These defects can be detected in the local factory or at the supplier and also accounts for materials that worked initially but broke while in use. Generally, the MQP is measured as a three-month moving average (MQP 3.0) and indicated as such. The MQP score is calculated by dividing the number of MN’s within the MQP scope by the total number of deliveries the supplier made within the predefined timeframe. In short, a lower score indicates better performance and the target is commonly set to just 0.15%. An issue with the MQP and its targets was that it was not based on scientific sources, but rather on a general assumption that it would suffice as a valid quality indicator

The material notifications that are relevant for the calculating the MQP are only those related to material issues that are caused by the suppliers. This means all service-related MN's, issues that occur internally, or caused by the customer are not used to calculate the supplier's quality score. The MN's specifically used in the MQP have all followed a predetermined route once the defect was detected, this route is visualized in the figure below (*figure 3*). This is done to ensure they were indeed caused by the product vendor. Once a defect is detected, it is recorded in the ERP system and the material is reviewed by a quality inspector. If the inspector suspects the issue was caused by the supplier, he sets the suspected cause field to vendor. This triggers a request for a root cause analysis (RCA) at the supplier, who then performs the requested analysis. The supplier then returns the results of the RCA, either suggesting what went wrong or declining the issue was caused by the vendor. The RCA is then judged by the supplier quality manager at ASML and approved or declined. In the data used in this research, the source of cause has always been identified as being the vendor. Therefore, the outcome of the RCA should always be that the issue was caused by the supplier. The RCA cycle time is considered a secondary quality indicator of the supplier, even though it does not precisely fit as a material quality indicator. It is calculated as a percentage of RCA requests completed within the predetermined timeframe. Because all defects caused by the supplier must follow this route, this information can be used to scope and filter out unrelated material defects.

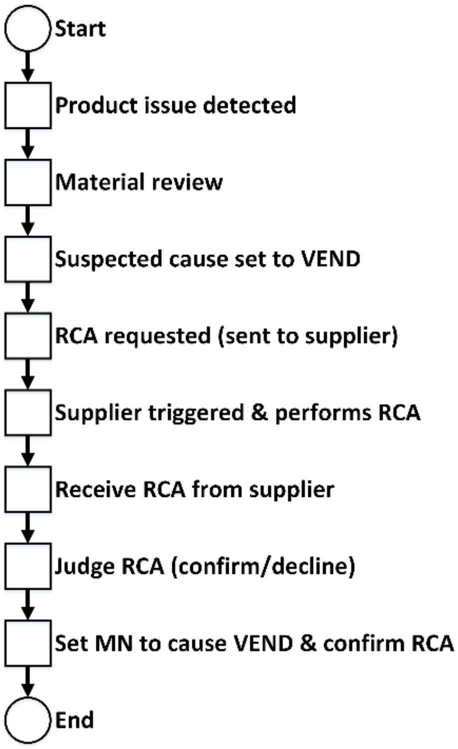

Figure 3: High-level overview of taken route for Material Notification (MN) with source of cause (SoC) vendor.

Together with its suppliers, ASML strives to achieve a minimal MQP score. This could be achieved in two ways, by producing more materials or by reducing the number of defects. Because many materials that are produced by the suppliers are highly complex, expensive and of low volume, only reducing the number of defects is a practical solution according to MQP manager. Both the QSM and MQP manager agreed that predicting upcoming MN's could be useful to prevent issues throughout the production process and help to take preventive actions. For example, if an incoming material is at risk of being defective, preventive checks could be performed to remove defective materials early in the process. Also, understanding what actions can be taken to reduce the risk of defects could help improve supplier material quality incrementally.

5 Literature

This chapter will focus on providing a scientific base for the research by explaining what performance indicators of supplier quality there are, what factors may have influence on these quality indicators, and the effect process deviations and the procurement process have according to literature. The literature will also be used to validate the ideas suggested by business experts and help understand the statistical and predictive outcomes of the research. The literature can also be used to evaluate the current processes used to manage quality and help understand in what type of environment application of prediction tools is valid. The literature review was partially conducted according to Cooper's five steps (1984) used in the method of Randolph (2009) and by using an ad-hoc approach to finding data to fill the gaps, details will also be explained in this chapter.

In the literature, supports was found for using a defect rate as a quality indicator and that it would be a suitable measurement of supplier material quality. Therefore, predicting material notification occurrences, used to calculate the defect rate, could be considered for predicting supplier material quality. Reliability was also identified as a valid material quality indicator but was mentioned less frequently. The literature also suggested that the delivery time window, schedule flexibility, and demand uncertainty influence supplier quality. This is in line with the business experts' observations that delivery date changes could influence quality. It was also suggested that price deviations could be of influence, contrary to what the experts thought. Material complexity and maturity was also suggested to influence material quality. Furthermore, the literature suggested that sharing information and good communication between supplier and customer has influence on quality. Other factors such as decentralized manufacturing, monitoring suppliers with IT support, and performing root cause analyses are processes that are already employed at ASML and suggested by literature to influence supplier quality. Predicting defects would also allow for more targeted inspections of materials, which was also suggested to influence material quality. Literature also supported that a higher number of process deviations could have a negative influence on performance and gaining insight into deviations, for example through predictions, could be used to reduce the negative effects. Previous research on service quality also suggests that good communication between supplier and customer helped to improve procurement process efficiency and standardization, which in turn increased performance.

5.1 Measuring Supplier Quality

According to Noshad & Awasthi (2015), supplier quality can commonly be evaluated using four categories, product, service, process, and organizational quality. Product quality evaluates criteria that the materials or products delivered by the supplier must adhere too, for example, the number of products that are delivered with a defect. An example of service quality is how quickly a supplier can address issues raised by its clients. It is simply a measure for the level of service the supplier provides to its customers. Process quality scopes towards flow of products or information within the supplier and between its customers. Examples here would be delivery of the product to the client, but also the communication forms between the two parties. Organizational quality evaluates the qualifications and systems an organization makes concerning its supply chain. For example, quality certification or the use of quality monitoring systems. The four categories may have some level of overlap, which would allow certain criteria to fit under multiple categories. The literature review centralized around supplier quality produced the table on the following page (*table 2*) and provides insight into the different supplier quality indicators used and how common these are based on the frequencies they were mentioned.

Articles	Product Quality		Service Quality				Process Quality			Organizational Quality					
	Defect Rate	Reliability	Corrective Actions	Quality Integration	Non-Conformance Complaints	Warranty	Deliveries	Management Practices	Production Disruptions	Supplier Training	TQM Practices or Six Sigma	Reward & Penalty System	Inspections & Serviceability	Certification	Quality Monitoring Systems
(Ang, Davies, & Finlay, 2001)	X			X	X			X							
(Baird, Hu, & Reeve, 2011)	X			X	X	X	X						X		X
(Chen & Yang, 2003)	X		X		X	X		X	X		X		X		
(Chen, Chen, & Li, 2005)													X		X
(Chen, Yeh, & Yang, 2004)	X		X		X	X			X						X
(Chin, Yeung, & Pun, 2006)	X		X					X		X		X	X	X	X
(Chuah, Wong, Ramayah, & Jantan, 2010)	X			X		X	X	X					X		
(Clemons, Baddam, & Henry, 2020)				X											X
(González-Benito & Dale, 2001)	X	X		X						X		X	X	X	X
(Yoo & Cheong, 2018)								X			X	X			
(Huo, Zhao, & Lai, 2014)	X	X	X	X	X		X							X	
(Kaynak & Hartley, 2008)	X			X				X		X					
(Lee, Yen, & Tsai, 2008)	X	X	X			X	X		X			X			X
(Lin, Kuei, & Chai, 2013)				X				X		X					X
(Maestrini, Luzzini, Caniato, Maccarrone, & Ronchi, 2018)	X	X		X		X									X
(Mellat-Parast, 2015)	X				X		X								
(Noshad & Awasthi, 2015)	X	X	X	X			X	X		X		X	X	X	X
(Noshad & Awasthi, 2018)	X	X	X	X	X		X			X				X	X
(Park & Hartley, 2002)	X	X			X		X								
(Quigley, Walls, Demirel, MacCarthy, & Parsa, 2018)	X				X							X			
(Salimian, Rashidirad, & Soltani, 2020)	X			X	X		X						X		
(Shin, Benton, & Jun, 2009)					X				X				X		X
(Wang & Fang, 2019)	X			X							X				
(White, Iammartino, & Fossaceca, 2019)	X				X			X		X	X				
(Xu, Peng, Pavur, & Prybutok, 2020)	X	X	X					X		X	X		X		X
(Yu & Huo, 2019)				X						X				X	
(Zhang, Guo, Huo, Zhao, & Huang, 2019)			X	X	X										
(Zsidisin, Petkova, Saunders, & Bisseling, 2016)			X												
(Zu & Kaynak, 2012)	X			X					X						X

Table 2: Supplier quality performance indicators stated in each article.

This research is focused specifically on the quality of supplier materials and predicting product defects, therefore only product quality will be used to indicate supplier quality. While no research was conducted to identify the best set of key performance indicators for supplier material quality, the defect rate was identified to be the most commonly used indicator according to the evaluated literature sources. The most basic defect rate is considered to be the number of defects divided by the total number of components received from the supplier. This is also the product quality information that can be easily recorded and managed, however the issue with this method is that it does not provide much insight. Because the base approach does not record any in-depth information (Xu, Peng, Pavur, & Prybutok, 2020), it does not provide a good overview or even a realistic indication of the situation. A more precise alternative could be, for example, the defective value (DV) mentioned by Chen, Yeh, & Yang (2004), in which the defect rate is an average of three factors closely related to defects. They suggested using the number of rejected lots, the number of rejected pieces, and the number of rejects from inspections. This method provides more detail concerning defective products, batches, and where the defects were identified, and can therefore be used to generate more precise insights. Secondly, product quality can be described in the form of reliability. Reliability provides insight into the consistency of the product quality and therefore provides a different perspective and other insights when compared to the defect rate. Reliability can be used to indicate how regularly supplier products are defective and how consistent the supplier's product quality is (González-Benito & Dale, 2001). Calculation of product reliability can be achieved by recording the number of good products between each defective product and transforming this information into a probability score. This information can then be used to calculate the consistency of defects and qualitative good products. However, when comparing the reliability score and the defect rate to each other, it was identified that monitoring the defect rate would be a more compatible quality indicator to use. This was based on two key factors. Firstly, the goal was to predict defective materials in each order, which was more compatible with the defect rate. Secondly, a form of defect rate was already being monitored and maintained at the researched firm, allowing for easier access to historic data.

Based on the information extracted from the literature above, it was considered safe to state that indicating supplier material quality based on the MQP score is a valid method. While the MQP score indicates only the most basic form of the defect rate, it is also the most common indicator, ensuring generalizability of the research. Furthermore, because the MQP was based on the occurrence of material notifications as defects, it was concluded that predicting material notifications was also a valid method to predict supplier quality. To clarify, predicting material notifications was considered the same as prediction material defects, which in turn could be used to calculate the MQP score, which was considered to be the same as calculating the defect rate.

5.2 Influencing Supplier Quality

Supplier quality is influenced by a broad range of factors that can be categorized into three groups for a better overview. These categories are people, process, and product factors. Process factors are related to processes, procedures, and standardizations concerning supplier quality. Examples are quality tools or sets of rules and regulations related to quality management. Product factors are related to the specific components or products that the suppliers produce and deliver. This factor can also encapsulate the design or development of the product. Factors related to the people or firm's performance, such as communication and cooperation, are classified as people factors. Examples of people factors are quality leadership, supplier relationships and good communication between supplier and customer (Noshad & Awasthi (2018) & Lin, Kuei, & Chai (2013)). While these factors may be relevant, they are not easily evaluated through process deviations and not in line with the research scope. Because of this, the following section will focus only on process and product factors that were identified in the literature review. The following tables (*table 3 & 4*) present the factors that were identified to have influence on supplier quality according to literature. Identification of the influencing factors can be used to understand why process deviations have influence on material quality.

Influencing Factor	Sources
Integration or use of quality management IT systems	Huo, Zhao, & Lai (2014); González-Benito & Dale (2001); Xu, Peng, Pavur, & Prybotuk (2020); Chin, Yeung, & Pun (2006); Lin, Kuei, & Chai (2013); Chen, Chen, Li (2005); Maestrini, Luzzini, Caniato, Maccarrone, & Ronchi (2018); Baird, Hu, & Reeve (2011); Noshad & Awasthi (2018)
Customer integration and information sharing	Huo, Zhao, & Lai (2014)
Supplier integration and information sharing	Huo, Zhao, & Lai (2014); Kaynak & Hartley (2008); González-Benito & Dale (2001); Zhang, Guo, Huo, Zhao, & Huang (2019); Salmian, Rashidrad, & Soltani (2020); Zu & Kaynak (2012); Yu & Huo (2019); Baird, Hu, & Reeve (2011); Noshad & Awasthi (2018)
Collaborative development and design of products	Wang & Fang (2019); Kaynak & Hartley (2008); Xu, Peng, Pavur, & Prybotuk (2020); Lin, Kuei, & Chai (2013); Chuah, Wong, Ramayah, & Jantan (2010); Zu & Kaynak (2012); Zsidisin, Petkova, Saunders, & Bisseling (2016); Baird, Hu, & Reeve (2011); Noshad & Awasthi (2018); Noshad & Awasthi (2015)
Quality information and performance records management	Wang & Fang (2019); González-Benito & Dale (2001); Zu & Kaynak (2012); Baird, Hu, & Reeve (2011); Noshad & Awasthi (2018); Noshad & Awasthi (2015)
Decentralized manufacturing	Wang & Fang (2019); Park & Hartley (2002); Zu & Kaynak (2012)
Unified quality standard	Wang & Fang (2019); Kaynak & Hartley (2008); Ang, Davies, & Finlay (2001)
Quality certification and standardization	González-Benito & Dale (2001); Zu & Kaynak (2012); Yu & Huo (2019); Noshad & Awasthi (2018); Noshad & Awasthi (2015)
Supplier selection and number of suppliers	González-Benito & Dale (2001); Salmian, Rashidrad, & Soltani (2020); Zu & Kaynak (2012); Zsidisin, Petkova, Saunders, & Bisseling (2016); Chen & Yang (2003); Baird, Hu, & Reeve (2011); Noshad & Awasthi (2015)
Just-in-time (JIT) logistics	González-Benito & Dale (2001)
Delivery frequency	González-Benito & Dale (2001)
Delivery time window	González-Benito & Dale (2001)
Delivery container standardization	González-Benito & Dale (2001)
Supplier quality penalization	González-Benito & Dale (2001); Quigley, Walls, Demirel, MacCarthy, & Parsa (2018); Yoo & Cheong (2018); Lee, Yen, & Tsai (2008); Chen & Yang (2003)
Physical or on-sight inspections	González-Benito & Dale (2001); Chuah, Wong, Ramayah, & Jantan (2010); Noshad & Awasthi (2018); Wang & Fang (2019); White, Iammartino, & Fossaceca (2019)
Quality development and continuous improvement	Xu, Peng, Pavur, & Prybotuk (2020); Zhang, Guo, Huo, Zhao, & Huang (2019); Chin, Yeung, & Pun (2006); Clemons, Baddam, & Henry (2020); Yu & Huo (2019)
Quality data analysis	Xu, Peng, Pavur, & Prybotuk (2020); Lin, Kuei, & Chai (2013); Noshad & Awasthi (2018)
Supplier quality incentivization	Yoo & Cheong (2018); Chin, Yeung, & Pun (2006); Noshad & Awasthi (2018)
Production schedule flexibility	Chuah, Wong, Ramayah, & Jantan (2010)
Supplier evaluation framework	White, Iammartino, & Fossaceca (2019); Noshad & Awasthi (2018); Noshad & Awasthi (2015)
Problem solving and root cause analyses	Chen & Yang (2003); Noshad & Awasthi (2018); Noshad & Awasthi (2015)

Table 3: Process factors influencing supplier quality performance. (21 instances)

Influencing Factor	Sources
Product mix flexibility	Huo, Zhao, & Lai (2014); Chuah, Wong, Ramayah, & Jantan (2010); Salmian, Rashidrad, & Soltani (2020)
Product volume flexibility	Huo, Zhao, & Lai (2014)
Product customization or modularity	Zhang, Guo, Huo, Zhao, & Huang (2019)
Product component conformance or variance stack-up	Shin, Benton, & Jun (2009); Lee, Yen, & Tsai (2008)
Product purchase price	Lee, Yen, & Tsai (2008)
Demand uncertainty or product churn	Zu & Kaynak (2012)
Knowledge of product requirements	Zsidisin, Petkova, Saunders, & Bisseling (2016)
Production competence and abilities	Zsidisin, Petkova, Saunders, & Bisseling (2016); Baird, Hu, & Reeve (2011)
Product or component complexity	Zsidisin, Petkova, Saunders, & Bisseling (2016)
Product design changes	Zsidisin, Petkova, Saunders, & Bisseling (2016)

Table 4: Product factors influencing supplier quality performance. (10 instances)

Focusing on process factors (*table 3*), it was found that, according to Zu & Kaynak (2012) decentralizing manufacturing of materials influences the level of quality and also the ability to monitor it. This statement helps support the relevance of being able to predict incoming material quality from external suppliers. Furthermore, physically inspecting materials is still a core task when ensuring materials are up to standard (González-Benito & Dale, 2001), therefore knowing more precisely what materials to inspect could improve efficiency of the quality assurance process. Tracking and monitoring quality information with help from IT systems and using this to share information and provide feedback to suppliers was also identified as a core influencer of supplier quality in numerous researches. This suggests that not only monitoring, but actively applying quality predictions can be used to significantly influence material quality. Specific key performance indicators that should be used for monitoring quality are scarcely recorded in literature, however multiple sources do state that maintaining a unified quality standard between customer and supplier can help improve quality (Kaynak & Hartley, 2008). Furthermore, using these standards to penalize supplier for underperformance (e.g. warranty claims) or incentivization for performing to standard could have effect on quality levels (Yoo & Cheong, 2018). To ensure quality defect are indeed a result of supplier failure, Chen & Yang (2003) state that it is important to investigate the root cause of defects. These analyses can then be used to prevent future reoccurrences and help improve supplier quality. Prior research also suggests that this level of collaboration between supplier and customer could improve product quality, especially through quality development and continuous improvements (Noshad & Awasthi, 2018). Also, González-Benito & Dale (2001) state that the delivery time window could influence supplier quality, suggesting that changing time windows may directly impact supplier material quality. Closely related to this is the flexibility of suppliers' production schedules. More flexible schedules are better at coping with changes in demand and have influence on the overall level of quality a supplier can deliver (Chuah, Wong, Ramayah, & Jantan, 2010).

Product specific factors (*table 4*) are generally related to materials specifically and could therefore also be used to understand predictability of material quality. Prior research states that product complexity is an important factor influencing quality of products (Zsidisin, Petkova, Saunders, & Bisseling, 2016). Generally, more complex components have a higher risk of defects. Zsidisin et al., (2016) also states that design changes that are made throughout a products lifecycle are of influence on the quality. While design changes may lead to improvements in the long term, the maturity of these changes may be related to quality risks. Zu & Kaynak (2012) also states that demand uncertainty could have effect on quality. Furthermore, Huo, Zhao, & Lai (2014) state that the suppliers flexibility of switching between products and production volume are of relevant influence on the quality of suppliers. This suggests that volume changes or orders of different materials could influence the quality a supplier delivers. Finally, the price of a product could have influence on the quality according to Lee, Yen, & Tsai (2008). This suggests that changes in price may also have effect on the quality of a product.

The literature suggested that the business experts were correct to observe that changing the delivery date of an order could influence supplier quality. It could even be remarked that product factors were considered to have more impact on supplier material quality than process factors. The delivery time window, schedule flexibility, and demand uncertainty were stated as possible influencers of supplier quality. Contrary to what the experts think, from literature it was deducted that price does have influence on quality. This could suggest that changes of product price may have influence on the quality of delivered materials. While not quite within the research scope, literature also indicated that product complexity and maturity are of influence on material quality. This is also what the quality managers considered to be the most important influencing factor. Furthermore, the literature suggested that sharing information and communicating about the materials with the supplier has influence on quality. This was considered to be somewhat related to the frequency of order reconfirmations. Therefore, frequent changes in a procurement process could be considered poor communication or a result of this. Other factors such as decentralized manufacturing, monitoring suppliers with IT support, and performing root cause analyses are all processes that are already employed at ASML and therefore may also be of influence on the quality of materials but could not be taken into account explicitly. Finally, because the developed prediction tool could help identify materials at risk of defects, a more targeted form of physical inspections could be applied in practice. This in turn was also considered a valid method to improve supplier quality.

5.3 Influence of Process Deviations

It was found difficult to identify literature that focused on the impact of process deviations on operational performance of a firm or its business processes. Dijkman, Turetken, IJzendoorn, & de Vries (2019) however did find that exceptions in business processes, which were interpreted as process deviations, do in fact lead to worse operational performance. Specifically, they found that business process exceptions lead to longer throughput times and that un-expectancy resulted in an increase of throughput time in comparison to expected exceptions. These findings were be used to suggest that process deviations eventually result in the potential occurrence of quality defects, because changes in the lead time or delivery time window could have this effect (Chuah, Wong, Ramayah, & Jantan, 2010). Furthermore, Munir, Jajja, Chatha, & Farooq (2020) stated that managing risks, or un-expectancies and changes, will have influence on operational performance of a firm.

Because there is little research about the influence of process deviations on operational performance, the key source was interpreted broadly to help support the research. The outcome of this research would help add knowledge to this field and further support these claims. The previous research suggested that a higher frequency of unexpected process deviations could have influence on decreasing performance, this was also considered to be plausible according to the operational procurement process manager. In other words, process deviations were considered a valid indicator of performance and therefore could also have influence on quality.

5.4 Influence of Procurement Process

The influence of the purchasing or procurement process on supplier performance and quality has also not been researched extensively (Hawkins, Gravier, Berkowitz, & Muir, 2015). However, according to Hawkins et al., (2015) there is a relationship between service quality and the performance of the procurement process. According to Noshad & Awasthi (2015), there is absolutely some degree of overlap between product and service quality, which was considered to be a valid indicator that the procurement process may also have influence on product quality. Hawkins et al., (2015) also found that customer commitment to suppliers helps define lead time requirements. In turn, they found that effective lead time requirement management and good communication directly leads to better service quality from the suppliers. Therefore, allocating more time to sourcing teams and improving the procurement process should be considered a realistic method to improve supplier quality. Immonen, Hallikas, & Pynnonen (2016) also support this and include that this leads to better long-term performance and improvement of complex systems. Furthermore, according to Glas (2018), better service from customer to supplier and better time management on the customer side improves supplier satisfaction. This in turn leads to better overall performance of the supplier, suggesting that effectively managing the procurement process from customer side helps to improve supplier quality performance. Finally, literature also stated that involving the purchasing department earlier in development of new materials or improvement materials could help improve the supplier-customer relationship, resulting in better service quality (Picaud-Bello, Johnsen, Calvi, & Giannakis, 2019).

The literature concerning the influence of procurement on service quality should only be used carefully and not weighted to much because of significant differences between service and product quality. However, the research on this topic does suggested that good communication between ASML and the supplier is important and that performing the procurement process in a structured and well-defined manner does influence firm performance. These findings helped to further support that the influence of procurement process is relevant to firm performance, and perhaps also to product quality based on similarities to service quality. This was also in line with the expectations of the business experts, who considered good collaboration in all processes was important.

5.5 Executing the Literature Review

The literature review was partially conducted according to the principles stated by Randolph (2009) in *A Guide to Writing the Dissertation Literature Review*. The article states a literature review is conducted by completing five consecutive steps (Cooper, 1984). These steps are problem formulation, data collection, data evaluation, analysis and interpretation, public presentation. These steps were applied to collect the data from the measuring supplier quality and influencing supplier quality sub-chapters. Because very little research about the influence of process deviations or the procurement process were identified, a more ad-hoc approach with snowballing was taken.

5.5.1 Literature Review

Data was collected from a total of five scientific sources. In this case the data bases Science Direct, Scopus, Web of Science, IEEE, and ACM Digital Library were used for collection. A predefined cluster of keywords were used in each search. The search engine of each data base looked for these terms in the title, abstract and key words of the stored article and returned the relevant articles. Details on the search terms and precise input can be found in the appendix (*appendix A*).

Following the initial search, papers were filtered out to ensure relevancy. Details on the number of articles found and deleted can also be found in *appendix A*. First, the papers titles, topics, and keywords were scanned for further specification of a subset of relevant articles. Next, the abstract was analyzed to confirm the paper would be relevant. If the paper was found to be relevant it was stored for more in-depth research. Any duplicates, generally papers found in multiple databases, were also eliminated in this phase. Finally, the discussion and conclusion of each article was analyzed and articles that did not fit the research criteria were excluded. Exclusion criteria were based on the research goals and questions. If the discussion and conclusion did not seem to add relevant information or answers to the goal or questions, the papers were excluded.

After producing the final list of articles using the previous criteria, each article was analyzed in detail. Key information was highlighted and different performance indicators and influencing factors were noted. The extracted information was then used to form and fill *tables 2, 3, and 4* from the previous sections.

5.5.2 Ad-hoc Approach

The literature on the influence of process deviations and the influence of procurement processes were not within the scope of the initial literature review and therefore researched in a later phase to help fill research gaps. Because literature about both topics was found to be limited, literature collected according to this approach was collected by identifying specific research outcomes or details. These outcomes or details were used to bind literature about process deviations, purchasing processes, and material quality. When a relevant paper was identified and deemed valuable according to the title, abstract, and key words, the related literature and cited literature was also inspected for relevancy, resulting in a snowball effect. However, due to the limited research performed on both these topics, little extra research could be added using this technique. The relevant information that could be of value to the research was then extracted and recorded in the same style as performed in previously mentioned information extraction. Because this process was not based on a validated methodology, it is important to emphasize that these sections of the literature could be less generalizable than the literature collected through Coopers approach.

6 Data Collection & Preparation

The following section explains what data was used, how it was collected, and how it has been prepared for statistical analysis and machine learning predictions. Data about the process deviations, or changes, in the procurement process were collected by extracting the logged changes of data tables and fields. This data would be used to provide insight into what changes occur, how they relate to the occurrence of material notifications, and if they can be used to predict upcoming material defects. The purchase order data was extracted from multiple SAP tables with help of data experts. This data provided a base on which defects could be linked to purchasing processes. Data concerning the material defects was also collected by extracting the relevant material notifications from the SQLD frozen databases. This data was used to identify materials or equipment had been validated to be defective and whether the defect was caused by the supplier. Following this, equipment data was extracted from the CAD CAS database to ensure the material notifications could be linked to the corresponding purchase orders and items. Using the equipment data was critical to ensure every defect could be correctly linked to a purchase order and the process deviations that occurred on the order. The connected dataset contained 11,538 orders and 147 material notifications and was considered to be the 'master' dataset. Critical for the research, the relationship between the material notifications and process deviations was confirmed to be causal. The master dataset was used to run the statistical analyses. The dataset was also transformed into the 'prediction' dataset for training and testing prediction models. This transformed dataset was required to balance the data and remove out of scope information that could cause noise. Data balancing was performed by randomly undersampling the majority class and performing a SMOTE on the minority class. The data was also bootstrapped to reduce variance of the predictions. All data was collected in and transformed using Excel, Spotfire and KNIME. Details on the collection and preparation processes are explained in the sections below.

6.1 Procurement Process Data

The procurement process data was critical for identifying what process deviations occur, to quantify how often these deviations are registered, and what deviations are relevant for the scope. This information could then be used as the attributes on which predictions could be based and to test relationships to the material defects. The datasets key values were the number of instances a change occurred for each of the 18 selected deviation types, the total number of deviations, and a change occurrence indicator. Furthermore, the purchase order number, purchase order item number, material number, and equipment number were included to later link the data to material defects.

Before explaining how the data was collected, it is important to understand the deviations, or change logs, that were included in the scope, and what these mean. The selection of relevant changes was based on the suggestions of business experts and the literature study. In the following table (*table 5*) the changes are listed and described. The table and field name come from SAP and explain what information was changed (field) and in which file (table). EKES changes are confirmations from suppliers, EKET changes are related to changes in the schedule, EKKO changes are related to management, and EKPO changes are specific to an order. The data name and short description columns indicates how the individual changes are referenced to throughout the research. Note that certain changes are very similar to each other however are in fact different, this is because the deviations occur in different tables and are therefore not always directly related.

Table	Field	Data Name	Description	Short Description
EKES	EINDT	EKESEINDT	Supplier confirmation of the delivery date (reconfirmation)	Supplier Reconfirm Date
EKES	MENGE	EKESMENGE	Supplier confirmation of the delivery quantity (reconfirmation)	Supplier Reconfirm Quantity
EKES	UZEIT	EKESUZEIT	Supplier confirmation of the delivery time window (reconfirmation)	Supplier Reconfirm Time Window
EKES	ERDAT	EKESERDAT	Creation date of supplier confirmation (reconfirmation)	Supplier Reconfirmation Creation
EKET	EINDT	EKETEINDT	Scheduled line item delivery date (rescheduling)	Reschedule Date
EKET	MENGE	EKETMENGE	Scheduled delivery quantity (rescheduling)	Reschedule Quantity
EKKO	FRGKE	EKKOFRGKE	Release approved purchasing document (reapproval)	Reapprove Order
EKKO	EKGRP	EKKOEKGRP	Purchasing group or manager change (reapproval)	Change Manager
EKPO	AEDAT	EKPOAEDAT	Last change indicator (resetting)	Reset Last Change
EKPO	ZCOMM_ ZONE_RPRNT	EKPOZCOMM_ ZONE_RPRNT	Send the purchase order reprint (custom change)	Resend Order
EKPO	NETPR	EKPONETPR	Set line item net price (resetting)	Reset Price
EKPO	BSTAE	EKPOBSTAE	Supplier confirmation of received order (resetting)	Reconfirm Order
EKPO	LOEKZ	EKPOLOEKZ	Deletion indicator of purchasing document (resetting)	Delete Order
EKPO	SCHPR	EKPOSCHPR	Order placement with ‘to be determined’ price (resetting)	Price Not Set
EKPO	MENGE	EKPOMENGE	Set purchase order quantity (resetting)	Reset Order Quantity
EKPO	PLIFZ	EKPOPLIFZ	Set planned delivery time supplier requires (resetting)	Reset Order Date
EKPO	EREKZ	EKPOEREKZ	Close order before invoice is completed (resetting)	Reset Status Closed
EKPO	ELIKZ	EKPOELIKZ	Delivery completed or closed (resetting)	Reclose Order

Table 5: Change indicators and their descriptions

Data concerning the procurement process was collected from the SAP database with help from the procurement department. The extraction was executed by business experts due to confidentiality reasons and access restraints. The data was extracted from the corresponding table names and filtered to only show orders within the research scope. Using PO numbers, line item numbers, and document numbers the data could be linked to each other. Currently the procurement department monitors the procurement process through process mining tools using the data stored in SAP databases. SAP logs the occurrences and time or date changes of purchase order statuses. The tool then provides clear insight into each step that occurred in a purchase and can be used to find the 'as is' process and deviations from the desired route. This same data source was used to identify changes that occur for each purchase order and purchase order item.

Purchase orders are generally made when tools and materials are acquired, but also when repair products are acquired. The data therefore only contains 'new buy' material purchase orders for scoping reasons. The initial data consisted of six data tables described below (*table 6*). The data contained all purchase orders from 2018, 2019, and 2020 placed at the supplier. As mentioned before, this range was selected because the earliest purchase order date connected to a material notification was from early 2018. Therefore, all identified material notifications could be linked to the corresponding purchase order.

Sheet	Description	Rows	Granularity
PO header	List of all placed purchases orders and corresponding general information	3006	Header
PO line items	List of all placed purchase order line items (individual materials) and corresponding details	11538	Line item
Change header	Changes of PO data logged in SAP	65538	Header
Field changes	Changes of PO line item data logged in SAP	466837	Line item
Field overview	List of changes that occur in data per source and their corresponding abbreviations and definitions	NA	SAP table fields
Pivot	Pivot table with overview of number of times each change occurred in the dataset	NA	Counter

Table 6: Procurement process data table descriptions.

To ensure the data could be used for prediction and linked to the material notification data, it had to be transformed to contain a table in which the number of occurrences of each relevant change is monitored. To achieve this, the focus was first placed on identifying what field changes logged in the data could be considered relevant. In certain situations, a field change could trigger a chain reaction of changes in the data. For example, a price change of one material would also lead to a price change of the PO net value. To prevent this, either the trigger change or final change of the corresponding change fields was selected and other related logs were excluded. Furthermore, certain change fields were marked as irrelevant by business experts and were also excluded from the data. These irrelevant fields were primarily considered such, because either the same value was logged in multiple fields of one or more tables, or because they were general change indicators related to document administration. The business experts who advised the process selection were the operational procurement process manager, the LSM, and QSM of the corresponding supplier. Also, the changes that were relevant according to literature were kept in the dataset even if business experts did not agree they could have influence. An overview of the remaining change logs and their definitions were stated in *table 5* on the previous page. A complete overview of all change fields, their SAP definition, and if they were considered within the scope can be found in *appendix B*. To get a better understanding of how many changes the data contained, an overview of the number of changes, maximum value, minimum value, and average number of occurrences is expressed in the following table (*table 7*).

Table & Field Labels	Change Occurrences	Max	Min	Average
EKES	**21,315**			
EINDT	21,162	43	0	1.834
MENGE	151	10	0	0.013
ERDAT	1	1	0	0.000
UZEIT	1	1	0	0.000
EKET	**17,945**			
EINDT	17,792	26	0	1.542
MENGE	153	5	0	0.013
EKKO	**5,362**			
FRGKE	5,323	22	0	0.461
EKGRP	39	1	0	0.003
EKPO	**363,960**			
AEDAT	187,373	92	0	16.238
ZCOMM_ZONE_RPRNT	172,730	93	0	14.969
NETPR	1,808	15	0	0.157
BSTAE	926	2	0	0.080
LOEKZ	923	3	0	0.08
SCHPR	122	2	0	0.011
MENGE	67	2	0	0.089
PLIFZ	7	1	0	0.001
EREKZ	3	1	0	0.000
ELIKZ	1	1	0	0.000
Total Deviations	**408,582**	**205**	**0**	**35.412**

Table 7: Table and field change occurrence of relevant process deviations.

Following the identification of relevant changes, a table was created with the PO number, PO line item, equipment number, and material number as row identifiers and the range of changes as column headers. Each change was then counted and assigned to the correct row using a pivot table. Two specific changes, from the EKKO table, were not assigned a line item specifically because they apply to the entire PO and not just one of the line items on the order. These changes were processed separately but otherwise managed exactly the same. The change EKKOFRGKE, purchase approval, is a mandatory manual step and therefore only instances where the change occurs more than once were counted. Finally, the purchase order changes were merged with all PO line item data in Spotfire. This resulted in the final overview containing information about the purchase itself and the linked process deviation. To improve readability, all empty columns and columns with identical values were removed from the dataset.

6.2 Material Notification Data

Data about material notifications was needed to indicate material defects. This information was used to help identify what purchase orders and order items were related to material defects and was used as the target output to train the prediction model. Key information and data points that were extracted, were the equipment number and the material number. With this information the data could be linked to the procurement data to indicate what rows actually resulted in the occurrence of a defect. However, to select the correct material notifications, it was important to only include the material notifications that were used to calculate or influence the MQP score. This was achieved by applying filters related to the scope and ensuring the predetermined criteria were met. Only material notifications related to a defective material and verified to be caused by the supplier through a root cause analysis were included.

Data concerning the material notifications was collected from the Supplier Quality Live Dashboard (SQLD) database and archived versions of the dashboard with help from the dashboard manager. The material notifications were used to indicate defects and in turn used to calculate supplier quality. The number of MN's divided by the total number of delivered materials is considered to be the most basic form of the defect rate explained in the literature. The SQLD database was used as a central location to store and display all material notification data extracted from the SAP environment since the start of 2019. The purpose of the database was to provide an overview of the occurrences of material notifications and allow for easy access to information concerning each notification. The data was extracted from SAP and appended to the SQLD set on a weekly basis. Annually, at the start of each year, all data was frozen and archived into a year overview. The data used in this research concerns all material notifications from 2019 and 2020 and was merged into one larger dataset of 279 columns and 1,390,342 rows, before filtration.

The data was first filtered to only contain information that fits within the scope of the research. Four core filters were applied concerning the desired vendor number, source of cause, plant, and equipment number. Details concerning the filters are displayed in *table 8* below. The equipment number was not directly related to the scope, however, it was crucial for connecting the material notification data to purchase orders and was therefore later included in this step.

Filter	Column Name	Value
Vendor Number	[Vendor Nr]	[7382801]
Source of Cause	[part_of_object_SoC]	[VEND, SVEN]
Plant	[Plant]	[NL01]
Equipment Number	[equipment]	[NOT BLANK]

Table 8: Initial data filtration for research scoping.

Following this, multiple column filters were applied to make the data more readable and eliminate unnecessary or unreliable information. Firstly, empty columns were deleted and columns where all data was identical were deleted because they add no further value. This was followed by deletion of duplicate columns or columns which contained identical information. These duplicate columns were a result of the SQLD dataset also being a built on the combination of multiple data sources. Finally, all columns that were considered to be irrelevant for the scope of the research were deleted. Post filtration 55 data columns were remaining with potentially valuable information or mandatory data to keep track of the notifications.

Finally, the duplicate rows had to be eliminated to ensure no material notification and material is considered twice in the final data set. This was achieved by deleting one or more rows if both the notification number and equipment number were equal to each other. Deleting rows based only on notification numbers could mean notifications that involved multiple materials could be deleted. Also, deleting only duplicate equipment numbers could mean new notifications were also lost. The filtered dataset contained a total of just 219 material notifications that fit the scope precisely and could potentially be linked with purchase order data.

6.3 Equipment Data

The function of the equipment data was to make linking the procurement data and the material notification data possible. The equipment data contained information about the original purchase order and item number and about the corresponding equipment number. This made it possible to link the material notifications to the correct purchase orders and items based on the equipment number. Therefore, the only information that was strictly necessary and used were the columns containing the purchase order number, the item number, and the equipment number.

To ensure the data concerning the purchase orders and the material notifications could be linked accordingly, the Configure as Designed (CAD) and Configures as Supplied (CAS) dataset was used. This dataset contained 3,256,161 rows of information on all the different components, or equipment, concerning the high-level design principles and supply details. In short, this data could be used to monitor and track materials and any relevant information about each material. The equipment number is a kind of serial number identification linked to every larger or more complex material delivered by certain suppliers. A limitation of the equipment number is that not all materials are assigned one, therefore linking defects and orders this way is only an option for more complex components.

Filter	Column Name	Value
Vendor Number	[Vendor_number]	[7382801]
Equipment Number	[Equipment]	[NOT BLANK]
Material Number	[Material_number]	[NOT BLANK]
Purchase Order	[Purchase_order]	[NOT BLANK]
Purchase Order Item	[Purchase_order_item]	[NOT BLANK]
Purchase Requisition Date	[Purchase_requisition_release_date]	[NOT BLANK]

Table 9: Data filtration of CAD CAS dataset.

As the dataset was only required to link two other datasets, the data was filtered to contain only materials from within the scope and a limited number of columns is maintained. The vendor number, equipment number, material number, purchase order, purchase order item, and purchase requisition date. These columns were needed to verify the data and link it correctly. The table above (*table 9*) details the filters applied to each column. The first three data field filters, vendor number, equipment number, and material number were required to link the data to the correct material notifications. The purchase order and order item were required to link the purchasing process data to the correct notifications, and the purchase requisition date was required to identify how far back the purchasing data would need to go to ensure all material notifications would be included. Next, the duplicates in the dataset were removed by deleting all rows in which the equipment number, purchase order, and purchase order item were equal. This was done to prevent one material notification from being linked to multiple purchase orders and new duplicates occurring after the data connection step.

It should also be mentioned that numerous equipment numbers could not be identified in the CAD CAS dataset. This was because the dataset did not cover all equipment prior to late 2018 as the data was not yet systematically extracted from SAP at that time. To further complete the dataset and get information about materials from 2018, missing equipment data was manually looked up directly in the SAP database. With help from a domain expert on data storage in SAP an extra 61 items could be identified and added to the equipment data. A drawback of this data was that it could only be verified using the material number and therefore less reliable.

6.4 Data Connection

The following section explains how the material notifications were linked to the purchase orders and the corresponding procurement process deviations using the equipment dataset. The two datasets were joined to allow a statistical analysis between changes and the occurrence of material defects and to form a dataset that could be used to train and test the prediction models. Firstly, the equipment data was joined with the material notification data by matching the equipment number and material number of each row. The material number was used as a precaution to verify the correct rows were joined. This join resulted in 147 unique rows of material notifications and their correlating purchase orders (PO) and purchase order item numbers. This step reduced the total number of MN's from 219 to just 147, simply because the data of the remaining 72 notifications could not be linked to a purchase orders that were within scope. This issue arose because data prior to 2019 was not stored as meticulously and therefore information was lost or never archived.

Following this, the material notifications data with the equipment data is linked to all purchase orders from the supplier. This was done by linking the equipment number and the material number in the purchase order data to the same values in the combined notification and equipment dataset. By doing this it became possible to link every material defect to the original purchase order number and the purchase order item numbers. The item number was relevant because it showed which material of the order was affected if multiple materials were purchased in one order. Because the material defects are now linked to the purchase orders, it also became possible to see what changes occurred in the purchasing process and allows the relations between the defects and changes to be observed. The resulting combined dataset consisted of 11,538 rows. These rows contain all the new material purchase orders placed at the supplier since 2018 and all verified vendor caused material notifications registered in 2019 and 2020.

The dataset that results from the data connection step was considered to be the master dataset of this research. This dataset was used to perform the statistical analyses in the following chapter but still required some processing before being usable for the predictive modeling. To provide some insight into the final dataset, the following table (*table 10*) provides a summary of the data and information.

Number of PO's (unique)	3,006
Number of PO Line Items (unique)	11,538
Number of Materials (unique)	1,126
Total Number of Deviations	408,582
Average Number of Deviations (per line)	35.4
Max/Min Number of Deviations	205/0
Number of Orders with a Deviation	11,310
Total Number of MN's (on PO Line Items)	147

Table 10: Summary of final dataset

6.5 Validation Unrelated Datasets

To ensure the PO change logs could be used as a predictor of material notification occurrences it was critical that both datasets were independent of each other. More specifically, it was important to verify that an occurrence of a material notification did not result in a change or deviation of a purchase order. This section briefly explains how the datasets were validated to have no causal relationship to each other. To verify this, the 'last changed on' column of the purchase orders was compared to the 'notification date' column of the material notifications. It was observed that for 30 out of the 147 cases the purchasing data was last changed after the MN was declared. These 30 cases were referred to as 'late changes' and were each checked to see if any of the logged changes connected to it occurred after the material notification was recorded. If no change occurred to the relevant data tables and fields after the MN was declared, then there could be no causal relationship. Just three cases were actually changed after a notification, however none of these were related to the relevant fields. Considering this, it was opted that even if a change was logged after the material notification, it was not a result of the material notification. Next this was verified with business experts to test if they suggest the same. Business experts from the purchasing team, who make the changes to the PO items, agreed that this should be the case and the data is unrelated. It was stated explicitly that they never log a change as a result of a material notification. An extra validation was implemented following the initial data analysis. The prediction algorithm was also trained and tested on a dataset without the 30 purchase orders with late changes. No significant difference was found between the output with or without the extra 30 rows. Through these three checks it was assumed that the data was indeed unrelated.

6.6 Data Transformation

To ensure the data could be used to train machine learning algorithms and provide reliable predictions, the data was transformed to meet specific requirements needed to do this. This dataset was considered the 'prediction' data and was used as training and testing set for each of the prediction algorithms. This step was performed specifically to remove data that could cause noise in predictions and to create a more balanced dataset to improve the predicted output. The transformation of the master dataset into the prediction dataset allowed the prediction output to be directly related to process deviations of the procurement process and removed data that is not within the scope of the research. The drawback of the following data transformation was that it reduced case specific insights and transparency that the original data had. This happened because data was both eliminated and synthetically created in the balancing process.

Firstly, a binary target column was generated. This column, called MN Occurrence, contained a 1 for every row with a material notification or a 0 if no notification was linked to it. It was created by identifying which rows contain a unique material notification number and appending the data. This column could be used by the prediction algorithms as the target column. It also allowed the model output to be compared to the test data predictions. Furthermore, any unnecessary columns containing data that should not be used to make a prediction were eliminated. The remaining columns consisted of the change logs counters, the total number of deviations and the occurrence of deviations. Following this, the numeric outliers were eliminated using the interquartile range (IQR) to flag outliers of the total number of deviations per row. The objective was to eliminate the most extreme outliers only, therefore the IQR multiplier (k) was set to the common standard of 1.5. Outlier removal was also tested on all change log columns but was finally not executed. This was because too many data rows would be eliminated as a result. More specifically, over 70% of the PO line items with a material notification would be eliminated and therefore too few target datapoints would remain for prediction. Any rows missing data would have been removed from the dataset, however no missing data was identified because it had been filtered out during previous steps.

Due to the imbalance of the target data, a large portion of the majority class was undersampled by randomly discarding rows. Undersampling was necessary due to the very high imbalance between the minority and majority class. Just 1.3% of the data belonged to the minority class, the class with a material notification. The risk of undersampling was that it could result in eliminating critical data points and increasing variance. To reduce the number of rows that would be randomly eliminated and maintain a larger dataset, the minority data was sequentially oversampled by applying the synthetic minority oversampling technique (SMOTE). SMOTE was used to double the number of rows with a material notification, by creating synthetic observations using the 5 nearest neighbors of other minority observation. The undersampled and oversampled data was then concatenated into a more balanced prediction dataset. For every minority class, two majority class rows were maintained. This was done to maintain a slight imbalance towards the majority class and to further reduce the number of rows that would have to be eliminated at random, ensuring the data would be slightly more realistic. The following table (*table 11*) provides an overview of the output of both balancing techniques and concatenated data.

Balancing	**Group**	**Rows Removed**	**Rows Added**	**Rows Output**
Random Undersampling	Majority	10,494	-	588
SMOTE	Minority	-	147	294
Concatenated				882

Table 11: Data balancing output using random undersampling and SMOTE.

Following the balancing of the data, it was important to reduce the potentially increased variance created by random undersampling. A potential solution to this was to apply bootstrap sampling. Another advantage of this was that the test and training data sample could be increased to a larger prediction set. Also, because a SMOTE was applied to the minority class, the data could not be used to inspect real data, reducing the effect of bootstrap drawbacks. The bootstrap was used to increase the number of observations by 200% to a total of 1,764 observations to help increase the size of each data fold used for training and testing. Holdout data from the bootstrap was discarded. The final prediction dataset consisted of 1,764 cases of which 1,158 did not hold a material notification and 606 cases that did result in a material notification.

7 Data Analysis

In this chapter, an explanation of the data analysis process and the initial results are explained. The data was first inspected using statistical analyses. The relationship between material defects and the occurrence of process deviations, the total number of deviations, and the specific deviations was tested using statistics. The information gained from this, together with literature and knowledge from business experts was then used to test if procurement process deviations could be used to predict material defects. The predictions were based on the previously mentioned prediction dataset. Six different prediction models were tested on the same data using mostly standard parameters. The models were compared to each other and insights from the models were extracted. Certain models provided better insights than other. Finally, the best scoring model was tested and validated on real data and the results are briefly explained.

7.1 Statistics

To help clarify if supplier quality could be predicted from process deviations and how the predations could be formed, a statistical analysis of the relations between material notifications and process deviations was performed. The statistical tests were performed on the complete (and unbalanced) dataset, but the numeric outliers of the total deviations were removed as explained in the data transformation set. First, a statistical analysis between the occurrence of a material notification and the occurrence of any deviations was tested using the chi-square test. The result of this test suggested there is a one-way correlation from the occurrence of deviations on material notifications. Next, an analysis testing if a relationship between the total number of deviations and the occurrence of a MN was performed using the T-test. The test results showed there was a relationship between the number of deviations and the occurrence of defects. To verify the outcome of the T-test, an extra analysis was run on the balanced dataset. Finally, the relation of each type of deviation, or logged change, to the occurrence of a material notification was checked using the T-test. Out of the 18 different changes, half of these proved to be statistically related to the occurrence of a material notification. The tests were run using SPSS and in the following section each of the statistical tests and its output is explained.

7.1.1 Material Notification Occurrence Dependency of Process Deviations

The first objective of the statistical analysis was to inspect if there was a relationship between the occurrence of any process deviation and the occurrence of a material defect. The results suggested that there was a one-sided relationship between the two variables. More specifically, that the occurrence of a material notification was related to the occurrence of a process deviation. This suggested that process deviations were, to some degree, a cause of material defects. Inspecting if there was a causal relationship helped to understand the basic concept of the research. The outcome was in line with expectations from business experts and literature, and could also be observed physically observed in the data quite easily. This was because all purchases that could be linked to a material notification had at least one logged change.

The statistical test was executed using the chi-square test. This method was selected because it could be used to compare two categorical factors to each other. In this case, a two by two comparison was made between material notifications and process deviations. Both factors could be either one, implying yes, or zero, implying no. In both cases, one indicated the occurrence of a MN or the occurrence of at least one change. The hypothesis tested by the statistical analysis is that there is a relationship between process deviations and MN occurrences. The null hypothesis suggests there is no relationship. The complete output of the test can be inspected in the appendix (*appendix C*), below the crosstabulation (*table 12*) and chi-square tests (*table 13*), containing the most valuable output, are shown.

Crosstabulation			MN Occurrence	
			0 (no)	*1 (yes)*
Process Deviations	*No*	*Count*	229	0
		Expected Count	226	3
		% of Total	2.0%	0.0%
	Yes	*Count*	10853	147
		Expected Count	10856	144
		% of Total	96.7%	1.3%

Table 12: SPSS output of crosstabulation from chi-square test for MN and process deviation occurrences

Chi-Square Tests	*Value*	*df*	*Asymptotic Sig. (2-sided)*	*Exact Sig. (2-sided)*	*Exact Sig. (1-sided)*
Pearson Chi-Square	3.101[a]	1	0.078		
Continuity Correction [b]	2.153	1	0.142		
Likelihood Ratio	6.098	1	0.014		
Fisher's Exact Test				0.078	0.047
Linear-by-Linear Association	3.101	1	0.078		
N of Valid Cases	11229				
a. 1 cell (25.0%) has expected count less than 5. The minimum expected count is 3.00 b. Computed only for a 2x2 table					

Table 13: SPSS output chi-square tests for MN and process deviation occurrences

The case processing summary identified 11,229 valid data rows that were used in the analysis. This was 100% of the input data and was a result of the data filtering performed in the data collection phase. Note this slightly lower total cases was a result of the numeric outlier removal. To ensure the chi-square test was valid, the crosstabulation was inspected first. To use the Persons Chi-Square test, the expected count value must be higher than 10 for each cell, this assumption was not met by the cell reflecting no process deviations and a MN occurrence. Because the basic assumption for the test were not met, the Fisher's Exact Test was used for validation. The 2-sided exact significance was 0.078 and was therefore not significant. This meant the null hypothesis was accepted, concluding there was no two-sided relationship. The one-sided exact significance was however 0.047 and therefore the null hypothesis was rejected, because it lies below 0.05. This implied that there was a one-sided association between the two factors. In other words, the state of one column could influence the other, but not the other way around.

From the chi-square test table, it became clear that the assumption that no more than 20% of the expected count is less than five was not met (*table 13, a.)*. In this case, the likelihood ratio could be used to better understand the association between the data classes. The ratio had a two-sided asymptotic significance of 0.014, which implied the null hypothesis was rejected. This meant that the occurrence of a material notifications was dependent of the occurrence of a process deviation. This was also in line with expectations, because all purchases that could be linked to a material notification had at least one logged change.

7.1.2 Relation Between Material Notifications and Total Number of Deviations

The next objective was to inspect if the total number of deviations that occurred on an order was related to the occurrence of material defects. It was found that there was a significant relation between the total number of deviations and the material notifications. This suggested that the sum of all changes that occur on a purchase order, could give an indication that there was or was not a material notification connected to the order. This was also in line with expectations from literature and experts. Validation of the test on a more balanced dataset also resulted in the same output. It was however important be critical of the outcome because of the distribution of the number of deviations was not per definition normally distributed.

The relationship was tested using the independent groups T-test. The T-test could be used to compare the mean of a dependent variable to two independent groups. These two dependent groups were the purchase order line items with a material notification and those without a notification. They were compared to an independent variable, in this case the total number of deviations that occur per order item. The hypothesis was that there is a relationship between the total number of deviations and the occurrence of a material notification. The null hypothesis would state there was no relationship. The test output can be viewed in the figure (*figure 4*) of three combined tables below.

Group Statistics					
	Group	*N*	*Mean*	*Standard Deviation*	*Standard Error Mean*
Total Deviations	No	11082	32.731	24.072	0.229
Total Deviations	Yes	147	39.871	18.305	1.510

Levene's Test				
	F	*df 1*	*df 2*	*p-value*
Total Deviations	14.309	1	11227	0.000

Independent Groups Statistics
Confidence Interval (CI) Probability 95.0%
Differences are reported of the groups: No – Yes

	Variance Assumption	*t*	*df*	*p-value (2-tailed)*	*Mean Difference*	*Std. Error Difference*	*CI (lower)*	*CI (upper)*
Total Deviations	Equal	-3.582	11.227	0.000	-7.140	1.993	-11.047	-3.233
Total Deviations	Not Equal	-4.675	152.773	$6.4e^{-6}$	-7.140	1.527	-10.157	-4.123

Figure 4: Independent samples T-test (complete data)

The group statistics provided a quick overview of the input data. Again, the dependent variables were the occurrences of MN's against the total number of deviations. It was also noticeable from the N-value that the two categories were highly unbalanced. The mean and standard deviation of both groups, however, were not too far apart. Using Levene's test in the following table, the significance (p-value) was calculated at 0.000. This implied that equal variances were not assumed for both groups. This could also be expected from the group statistics, because there was a visible difference between the standard deviations. Following this, the second row of the following table could be used. The independent samples test stated that the two-tailed significance was 0.000, which implies the null hypothesis was rejected. This means that there was a relation between the total number of deviations occurring on an item order and the occurrence of a material notification.

It was however very important to be critical of this outcome. Firstly, as stated previously, the two independent groups were highly unbalanced. The instances with a material notification were just 1.3% of the instances without a deviation. This imbalance could skew the results and not provide a proper image of the situation. Furthermore, the T-test requires the dependent variable to be approximately normally distributed. In the following histogram (*figure 5*) it is clear that there are some similarities to a normal distribution or log-normal distributions, but it is not a perfect fit and other distributions may better fit the image.

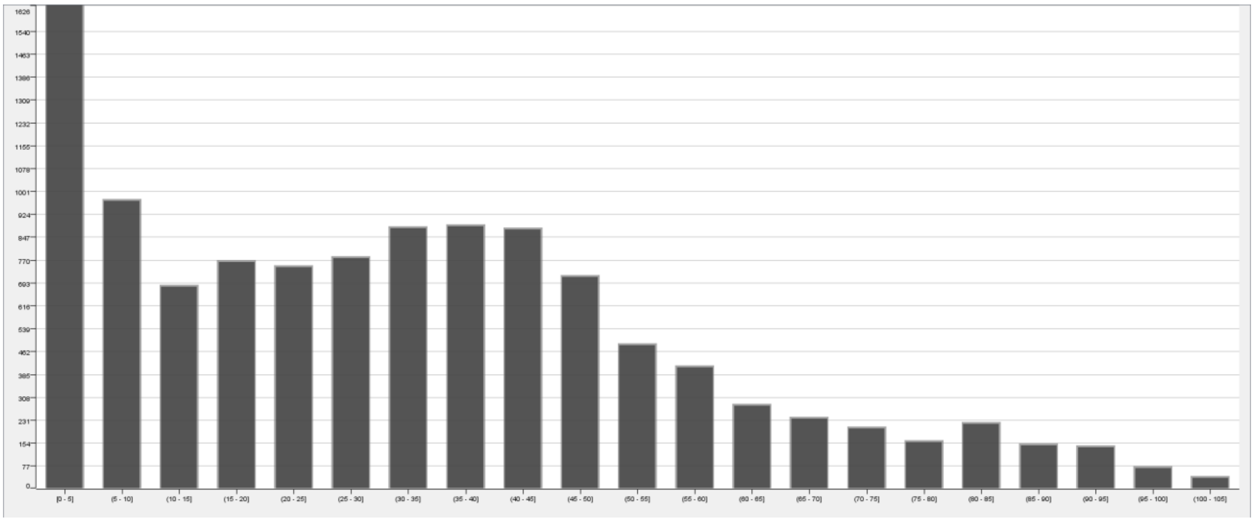

Figure 5: Histogram of total number of deviations (bin size 5)

To test if the unbalanced data was of influence on the statistical outcome, a training dataset was subjected to the same independent groups T-test. The dataset that was constructed for the machine learning prediction algorithm was used in this situation. The output of these tests can be seen in the following figure containing three tables (*figure 6*). It is important to state that the sample sizes were much smaller, yet still have a similar mean and standard deviation. Again, equal variances were not assumed, and the two-tailed significance was also 0.000. This outcome again implies that there was a relation between total number of deviations and material notification occurrences. This helps to support the idea that the relationship was not dependent cause by the imbalance between the target classes.

Group Statistics					
	Group	*N*	*Mean*	*Standard Deviation*	*Standard Error Mean*
Total Deviations	No	1158	32.221	23.182	0.681
Total Deviations	Yes	606	40.077	18.009	0.732

Levene's Test				
	F	*df 1*	*df 2*	*p-value*
Total Deviations	38.507	1	1762	$6.78e^{-10}$

Independent Groups Statistics
Confidence Interval (CI) Probability 95.0%
Differences are reported of the groups: No – Yes

	Variance Assumption	*t*	*df*	*p-value (2-tailed)*	*Mean Difference*	*Std. Error Difference*	*CI (lower)*	*CI (upper)*
Total Deviations	Equal	-7.272	1.762	$5.30e^{-13}$	-7.8557	1.080	-9.974	-5.737
Total Deviations	Not Equal	-7.859	1513.904	$7.31e^{-15}$	-7.8557	0.999	-9.817	-5.895

Figure 6: Independent samples T-test (balanced prediction data)

7.1.3 Relation Between Material Notifications and Individual Process Deviations

The final goal of the statistical analysis was to inspect if there is a relationship between the number of instances a specific deviation occurs and the material notifications. Out of the 18 different deviation types, half of the deviations were identified to have a significant relationship to the occurrence of material notifications. This would suggest that specific changes, and not just the sum of all changes, have influence on the occurrence of material defects. The types of deviations and their descriptions that had a relation to the occurrence of material notifications can be found in the table below (*table 14*). Notable is that specifically quantity and date changes seemed to be related to material notifications, as was suggested by experts and literature. Pricing changes were also related as literature suggested but was not expected by the experts. Confirmations were also related to material notifications, however, this was not supported by literature or business experts. Again, it was important to be critical of the outcome because of the distributions of the deviation types.

Data Name	Description
EKESEINDT	Supplier confirmation of the delivery date (reconfirmation)
EKETEINDT	Scheduled line item delivery date (rescheduling)
EKETMENGE	Scheduled delivery quantity (rescheduling)
EKKOFRGKE	Release approved purchasing document (reapproval)
EKPONETPR	Set line item net price (resetting)
EKPOBSTAE	Supplier confirmation of received order (resetting)
EKPOLOEKZ	Deletion indicator of purchasing document (resetting)
EKPOSCHPR	Order placement with 'to be determined' price (resetting)
EKPOMENGE	Set purchase order quantity (resetting)

Table 14: Process deviation and description with significant relation to material notifications

Because there was a relationship between the total number of process deviations and the occurrence of material notification, it seemed like a logical next step to test if there was a relationship between any specific changes. Again, the independent groups T-test was applied to test the relationships. The table below (*table 15*) provides a high-level overview of the output, details of the tests can be found in the appendix (*appendix C*). The hypothesis was that there is a relationship between the deviation type and the occurrence of a material notification. The null hypothesis suggested there was no relationship.

Deviation Type	Short Description	Levene Test p-value	Equal Variance Assumed	Independent T-test p-value	Null Hypothesis
EKKOEKGRP	Change Manager	0.148	Equal	0.471	Accepted
EKKOFRGKE	Reapprove Order	0.000	Not Equal	0.000	Rejected
EKESEINDT	Supplier Reconfirm Date	0.000	Not Equal	0.000	Rejected
EKESERDAT	Supplier Reconfirmation Creation	0.818	Equal	0.908	Accepted
EKESMENGE	Supplier Reconfirm Quantity	0.165	Equal	0.488	Accepted
EKESUZEIT	Supplier Reconfirm Time Window	0.818	Equal	0.908	Accepted
EKETEINDT	Reschedule Date	0.001	Not Equal	0.000	Rejected
EKETMENGE	Reschedule Quantity	0.000	Not Equal	0.002	Rejected
EKPOAEDAT	Reset Last Change	0.000	Not Equal	0.058	Accepted
EKPOBSTAE	Reconfirm Order	0.000	Not Equal	0.000	Rejected
EKPOELIKZ	Reclose Order	0.818	Equal	0.908	Accepted
EKPOEREKZ	Reset Status Closed	0.690	Equal	0.842	Accepted
EKPOLOEKZ	Delete Order	0.000	Not Equal	0.000	Rejected
EKPOMENGE	Reset Order Quantity	0.000	Not Equal	0.001	Rejected
EKPONETPR	Reset Price	0.000	Not Equal	0.000	Rejected
EKPOPLIFZ	Reset Order Date	0.542	Equal	0.761	Accepted
EKPOSCHPR	Price Not Set	0.015	Not Equal	0.000	Rejected
EKPOZCOMM_ _ZONE_RPRNT	Resend Order	0.000	Not Equal	0.382	Accepted

Table 15: High-level overview of the relation between deviation types and defects using independent groups T-test

The test was run by applying the same method used in the previous section, where the relation between material notifications and total number of deviations was inspected. Because the results were now produced for 18 different variables and not just one, only a high-level overview of the output is projected. Details of the process are not explained here, but the process from the previous section was repeated for all 18 tests. In short, check Levene's test, select equal or not equal variance, inspect the independent T-test p-value, and accept or reject the null hypothesis. The statistical analysis of the deviation types did provide better and interesting insights into what specific factors may have had influence on the occurrence of material notifications. Simply explained, if the null hypothesis was accepted as illustrated in *table 15*, there was no statistically significant relationship between the frequency of occurrences of that specific deviation type and the occurrence of a material notification. The rejected cases, however, imply that there is a statistically significant relationship between the two. Again, it is however important to remain critical of these results because it is assumed that the change occurrence adhere to a log-normal distribution based on their related histograms.

7.2 Machine Learning Predictions

To test the predictive capabilities of procurement process, change logs on the occurrence of material defects, six prediction models were tested. These models were a decision tree, logistic regression, probabilistic neural network (PNN), gradient boosted trees, random forest, and multi layered perceptron (MLP). The goal was to identify if any of the algorithms could provide accurate predictions of defective materials using only data about the frequency of procurement process deviations. Each algorithm was trained and tested using the balanced prediction dataset resulting from the data preparation phase. The results showed that the gradient boosted trees model produced the best score based on accuracy (97.7%), Cohen's Kappa (κ) (0.949), and the F-score (0.966). This is considered to be very good, suggesting the procurement process deviations were valid indicators of material defects and can absolutely be used to predict the quality of new materials from the suppliers. When the model was tested on real and unbalanced data, the results were far less reliable however. Accuracy, Cohen's Kappa (κ), and F-score were just 94.3%, 0.282, and 0.297 respectively. This was primarily because the model resulted in more false positives, often predicting a notification would arise when this had not actually been the case. The score was however still considered to be in the acceptable range, suggesting it was better than randomly guessing and could be of practical value. Using the decision tree, which scored good based on Cohen's Kappa value, it was possible to observe what types of deviations were used to make the predictions. The indicators EKETEINDT (Reschedule Date), EKESEINDT (Supplier Reconfirm Date), EKPOZCOMM_ZONE_RPRNT (Resend Order), and EKKOFRGKE (Reapprove Order) were used to do this. Interestingly this suggests that the statistically related changes were also better suited for basing predictions on. This was validated by removing all statistically unrelated deviations in an extra test, the results turned out to be of similar levels. Combining the results of the validation tests suggested that the models were somewhat overfitting but could provide relevant practical insights depending on the required accuracy and precision standards. They also clearly showed that certain procurement process deviations were of influence on material quality from the supplier, and that purchasing process deviations can be used as a predictive indicator.

7.2.1 Algorithm Scores

The following tables (*table 16 & 17*) show how the six tested models score using a variety of indicators. The scores were calculated by comparing the prediction to the actual value and entered into a confusion matrix. This information was then used to calculate the different indicators of which the most relevant were considered accuracy, Cohen's Kappa (κ), and the F-score. The accuracy is a simple indicator, while the F-score and Cohen's Kappa provide more insight into the general performance by relying on more aspects of the output.

<table>
<tr><th>Algorithm</th><th>Decision Tree</th><th>Logistic Regression</th><th>Probabilistic Neural Network (PNN)</th></tr>
<tr><td>Confusion Matrix</td>
<td><table><tr><td></td><td>0</td><td>1</td><td></td></tr><tr><td>0</td><td>992</td><td>116</td><td>1158</td></tr><tr><td>1</td><td>135</td><td>471</td><td>606</td></tr><tr><td></td><td>1127</td><td>587</td><td>1764</td></tr></table></td>
<td><table><tr><td></td><td>0</td><td>1</td><td></td></tr><tr><td>0</td><td>1056</td><td>102</td><td>1158</td></tr><tr><td>1</td><td>255</td><td>351</td><td>606</td></tr><tr><td></td><td>1311</td><td>453</td><td>1764</td></tr></table></td>
<td><table><tr><td></td><td>0</td><td>1</td><td></td></tr><tr><td>0</td><td>1092</td><td>66</td><td>1158</td></tr><tr><td>1</td><td>131</td><td>475</td><td>606</td></tr><tr><td></td><td>1223</td><td>541</td><td>1764</td></tr></table></td></tr>
<tr><td>Correctly Classified</td><td>1,463</td><td>1,407</td><td>1,567</td></tr>
<tr><td>Wrongly Classified</td><td>301</td><td>357</td><td>197</td></tr>
<tr><td>Accuracy</td><td>82.937%</td><td>79.762%</td><td>88.832%</td></tr>
<tr><td>Error</td><td>17.063%</td><td>20.238%</td><td>11.168%</td></tr>
<tr><td>Cohen’s Kappa (κ)</td><td>0.626</td><td>0.523</td><td>0.746</td></tr>
<tr><td>True Positive Rate (Recall or Sensitivity)</td><td>0.777</td><td>0.579</td><td>0.784</td></tr>
<tr><td>True Negative Rate (Specificity)</td><td>0.857</td><td>0.912</td><td>0.943</td></tr>
<tr><td>Precision</td><td>0.802</td><td>0.774</td><td>0.878</td></tr>
<tr><td>F1</td><td>0.789</td><td>0.691</td><td>0.828</td></tr>
</table>

Table 16: Algorithm scores for the decision tree, logistic regression, and probabilistic neural network (PNN)

<table>
<tr><th>Algorithm</th><th>Gradient Boosted Trees</th><th>Random Forest</th><th>Multi Layered Perceptron (MLP)</th></tr>
<tr><td>Confusion Matrix</td>
<td><table><tr><td></td><td>0</td><td>1</td><td></td></tr><tr><td>0</td><td>1145</td><td>13</td><td>1158</td></tr><tr><td>1</td><td>27</td><td>579</td><td>606</td></tr><tr><td></td><td>1172</td><td>592</td><td>1764</td></tr></table></td>
<td><table><tr><td></td><td>0</td><td>1</td><td></td></tr><tr><td>0</td><td>1119</td><td>39</td><td>1158</td></tr><tr><td>1</td><td>92</td><td>514</td><td>606</td></tr><tr><td></td><td>1211</td><td>553</td><td>1764</td></tr></table></td>
<td><table><tr><td></td><td>0</td><td>1</td><td></td></tr><tr><td>0</td><td>1064</td><td>94</td><td>1158</td></tr><tr><td>1</td><td>165</td><td>441</td><td>606</td></tr><tr><td></td><td>1229</td><td>535</td><td>1764</td></tr></table></td></tr>
<tr><td>Correctly Classified</td><td>1,724</td><td>1,633</td><td>1,505</td></tr>
<tr><td>Wrongly Classified</td><td>40</td><td>131</td><td>259</td></tr>
<tr><td>Accuracy</td><td>97.732%</td><td>92.574%</td><td>85.317%</td></tr>
<tr><td>Error</td><td>2.268%</td><td>7.426%</td><td>14.683%</td></tr>
<tr><td>Cohen’s Kappa (κ)</td><td>0.949</td><td>0.832</td><td>0.665</td></tr>
<tr><td>True Positive Rate (Recall or Sensitivity)</td><td>0.955</td><td>0.848</td><td>0.728</td></tr>
<tr><td>True Negative Rate (Specificity)</td><td>0.989</td><td>0.966</td><td>0.919</td></tr>
<tr><td>Precision</td><td>0.978</td><td>0.929</td><td>0.824</td></tr>
<tr><td>F1</td><td>0.966</td><td>0.887</td><td>0.773</td></tr>
</table>

Table 17: Algorithm scores for the gradient boosted trees, random forest, and multi layered perceptron (MLP)

To better understand the output of each model, the table content is explained below. Starting with the confusion matrix, the column headers indicate the predicted value and the row headers show the actual value, were 0 indicates no material notification and 1 indicates there was a notification, or defect. Ideally the maximum values would be in both top left and bottom right for a perfect prediction. If the output is closer to the ideal confusion matrix output, the number of correctly classified data is higher and wrongly classified is lower. The accuracy score and error depict the percentage of, respectively, correctly or wrongly predicted output. The Cohen's Kappa (κ) shows the inter-rate reliability, which takes into account the possibility of chance. It particularly provides insight into how much better the model scores than randomly guessing the predictions (κ=0). The true positive score, or recall, provides insight into how reliable the model is in detecting the occurrence of material notifications. The true negative, or specificity, provides insight into the probability no MN occurs when the model does not expect one. The precision helps to describe the random errors in the data. Finally, by combining the recall and precision score, the output can be used to calculate the F-score (F1) and get a better understanding of the models test accuracy. It is also important to mention that recall provides insight into the how well a class is detected, while precision helps understand how trustworthy the prediction is. For the Cohen's Kappa, recall, specificity, precision, and F-score a value closer to one would be considered better than a value close to 0. The F-score provides a better measure of the model accuracy than simply taking the percentage of correctly classified values. However, it does not take into account the true negative. To also take into account this group, Cohen's Kappa is a better indicator when predicting a binary classifier (Hand & Christen, 2017). According to Landis & Koch (1977) the Cohen's Kappa scores can be ranked to indicate the model performance. A good score would be between 0.61 – 0.80 and very good would be even higher. A model would be considered moderate between 0.41 – 0.60 and acceptable between 0.21 – 0.40. A lower score would be considered poor and suggests the model cannot be used to make a prediction.

As stated previously, the results show that the gradient boosted trees produced the best score based on accuracy (97.7%), Cohen's Kappa (κ) (0.949), and the F-score (0.966). This was to be expected when comparing it to the decision tree and the random forest models, because it scores slightly higher in general comparisons between the models. The worst results came from logistic regression with an accuracy, Cohen's Kappa, and F-score of 79.8%, 0.523, and 0.774 respectively. Based on the Cohen's Kappa value, the gradient boosted trees and random forest models would be considered very good, the PNN, MLP, and the decision tree models produced good scores, and a moderate score was achieved by the logistic regression model. This indicated every model could be used for a least a decent predicting of material defects based on the transformed dataset.

7.2.2 Modelling Actions & Parameters

All algorithms were run in KNIME to allow for quick testing and setup, and to reduce time required to individually program different models. Furthermore, using KNIME would ensure each algorithm was handled similarly and the input data was exactly the same. The algorithms were generally set up using default values because the goal was to see if prediction was possible and not to optimize the output. In this section the specifics of the parameters are explained for each model to provide some insight into how the prediction outcomes were generated.

The algorithms were all set up to make a prediction for the occurrence of a material notification. This was done by setting the target columns to make a prediction of the MN_occurrence data field. If required, the selected features were set to the 18 different change columns, the total number of changes, and the change indicator. Every model was cross validated 7-fold using a stratified dataset for both training and testing to improve the reliability of the results. This was achieved by using the X-Partitioner and X-Aggregator nodes that allow for running a model multiple iterations and collecting and combining the outcomes. The datasets were stratified to ensure the balance of defects is the same in both training and test data. For each iteration, the nodes ensure that 70% of the data was used to train the model, and 30% was used to test it. This division was chosen because it is a common amount to use for training and testing. The results of each test set were then collected and aggregated into one datasheet. This output data was then observed by the Scorer node that produced the confusion matrix, correctly and wrongly classified values, accuracy, error, and Cohen's Kappa score. The recall, specificity, precision, and F-score were calculated afterwards using the information from the confusion matrix. Following these general settings and steps taken for all models, the specific model parameters are explained below.

7.2.2.1 Decision Tree

The Decision Tree Learner node and the Decision Tree Predictor node from KNIME were used to produce a prediction using a decision tree. The node was set to use the Gini Index to calculate the split, because it is a simple and common method to split categorical target variables. To reduce the tree size and overfitting the minimal description length (MDL) pruning method was selected. MDL was selected because it is the only pruning method available using the learner node but also because it helps to increase generalization performance and therefore a higher prediction quality. The minimal number of records per node was set to 50 following brief experimentation. The value was selected because the corresponding Cohen's Kappa score was still significantly high enough to be considered good. When the value was set to one it would still only produce a 'good' Cohen Kappa score, but would be at risk of overfitting. Also, the decision tree remains readable when the minimal number of records is not too small. All other parameters were set to the default value and the details can be inspected in the appendix (*appendix D*).

7.2.2.2 Logistic Regression

The Logistic Regression Learner and Logistic Regression Predictor nodes from KNIME were used to predict material defects using logistic regression. Because logistic regression works best on normalized data, the input data was normalized using Z-score normalization, using a dedicated normalization node. The node was set to use the stochastic average gradient (SAG) solver because this approach converges quickly and provided better output than the alternative iteratively reweighted leas squares approach. The model was allowed a maximum of 200 epochs with an epsilon of $1.5e^{-5}$ to determine if the model was converged, these settings were default values of the node. The solver uses a fixed learning strategy with the default, but also most common learning rate of 0.1. Regularization was left in the default setting, uniform. Other parameters can be found in the appendix (*appendix D*).

7.2.2.3 Probabilistic Neural Network

The PNN Learner node and the PNN Predictor node from KNIME were used to produce a prediction based on a neural network. The nodes default values were maintained after brief experimentation with the theta values proved these settings provided the best results. The model was trained using the dynamic decay adjustment (DDA) method, on which the PNN Learner node is based, and generates rules using numeric data. The theta minus, set to 0.2, and theta plus, set to 0.4, were used as adjustment thresholds for the high-dimensional Gaussian functions that defined the rules. The Gaussian functions were defined by a center vector and a standard deviation. The function only covered non-conflicting instances during training (KNIME, 2021). All parameter settings are depicted in the appendix (*appendix D*).

7.2.2.4 Gradient Boosted Trees

The Gradient Boosted Trees Learner and Gradient Boosted Trees Predictor nodes from KNIME were used to predict defects using gradient boosted decision trees model. Firstly, the tree depth was limited to 10. The selected tree depth resulted in the best output when comparing default settings ranging from 4 to 10. The risk of using a higher value to limit the tree depth is that it could lead to overfitting. The model was set to learn 100 decision trees with a default learning rate of 0.1. The number of models was limited to 100, because too high values tend to result in overfitting the model. However, for smaller datasets, higher values do produce better results and therefore cannot be too small. Bagging, or row sampling was not performed, because the dataset was already bootstrapped in the data transformation phase and without bagging the learner would be trained on the complete training set. Attribute sampling was applied using a square root sample of the total number of attributes and was fixed for the entire tree. This option was selected because it is the default setting and produced good results. Details of the parameters can be found in the appendix (*appendix D*).

7.2.2.5 Random Forest

The Random Forest Learner node and the Random Forest Predictor node from KNIME were used to produce a prediction with a random forest model. As was the case for the decision tree, the split criterion was set to the Gini Index. This option is a common choice and also set as the default split for the learner node. The tree depth is by default not limited and it was not required to reduce computing time. The minimum node size was set to one as is standard. Identical rows could potentially cause issues with this value, but this did not occur in tests. The node was set to produce 100 models for prediction. Generally, a value between 100 and 500 is selected through tuning, however a value higher than 100 did not prove to be significantly better through experimental testing. The other parameters and details can be inspected in the appendix (*appendix D*).

7.2.2.6 Multi Layered Perceptron

The RProp MLP Learner and Multi-Layer Perceptron Predictor nodes from KNIME were used to predict material defects using multilayered feedforward networks. The node uses the RPROP algorithm to perform an adapted version of the weighted updates according to the behavior of the error function (KNIME, 2021). The model was allowed a maximum of 100 iterations because higher values did not seem to improve the model significantly. The number of hidden layers was set to 3 with 5 neurons per layer. These values were identified through quick experimentation and proved to result in the best Cohen's Kappa score. Other parameter settings can be viewed in the appendix (*appendix D*).

7.3 Decision Tree Insights

While the decision tree model did not provide the best prediction output compared to all other models, a unique characteristic of the decision tree is that it can quite easily provide insight into what deviations were used to split the data into the target groups. This insight could be used to better understand what information was being used to make the prediction. The following figure shows the decision tree produced by one of the seven iterations as a result of the k-folds method (*figure 7*). This particular tree scored an error rate of 17%, which was somewhat comparable to the overall performance of the model. Notable are the changes EKESEINDT (Supplier Reconfirm Date), EKETEINDT (Reschedule Date), and EKKOFRGKE (Reapprove Order) which respectively represent the reoccurrence of supplier confirmation of the delivery date, scheduling line item delivery date, and the releasing the approved purchasing document. These changes were all marked as statistically significant to the occurrence of a material notification in the statistical analysis. Two of these changes are related to the delivery date as was to be expected from experts and literature. The reapproval of documents cannot be directly linked to literature of insight from business experts, but could be related to communication. The final change was a custom log, EKPOZCOMM_ZONE_RPRNT (Resend Order), which indicates the purchase order was reprinted and sent to the supplier. Interestingly, this change type was not statistically relevant according to statistical analysis and was therefore somewhat unexpected. However, this change could be somewhat related to literature. The document reprint generally suggests information was lost or was changed, which in turn could be related to the levels of communication. Because the decision tree model did not provide a perfect score and was not the best model, it should not be immediately considered that these changes are the core indicators of material notifications. However, the decision tree does provide some insight into the way the model operates and shows that statistically relevant deviation types are of influence on the prediction.

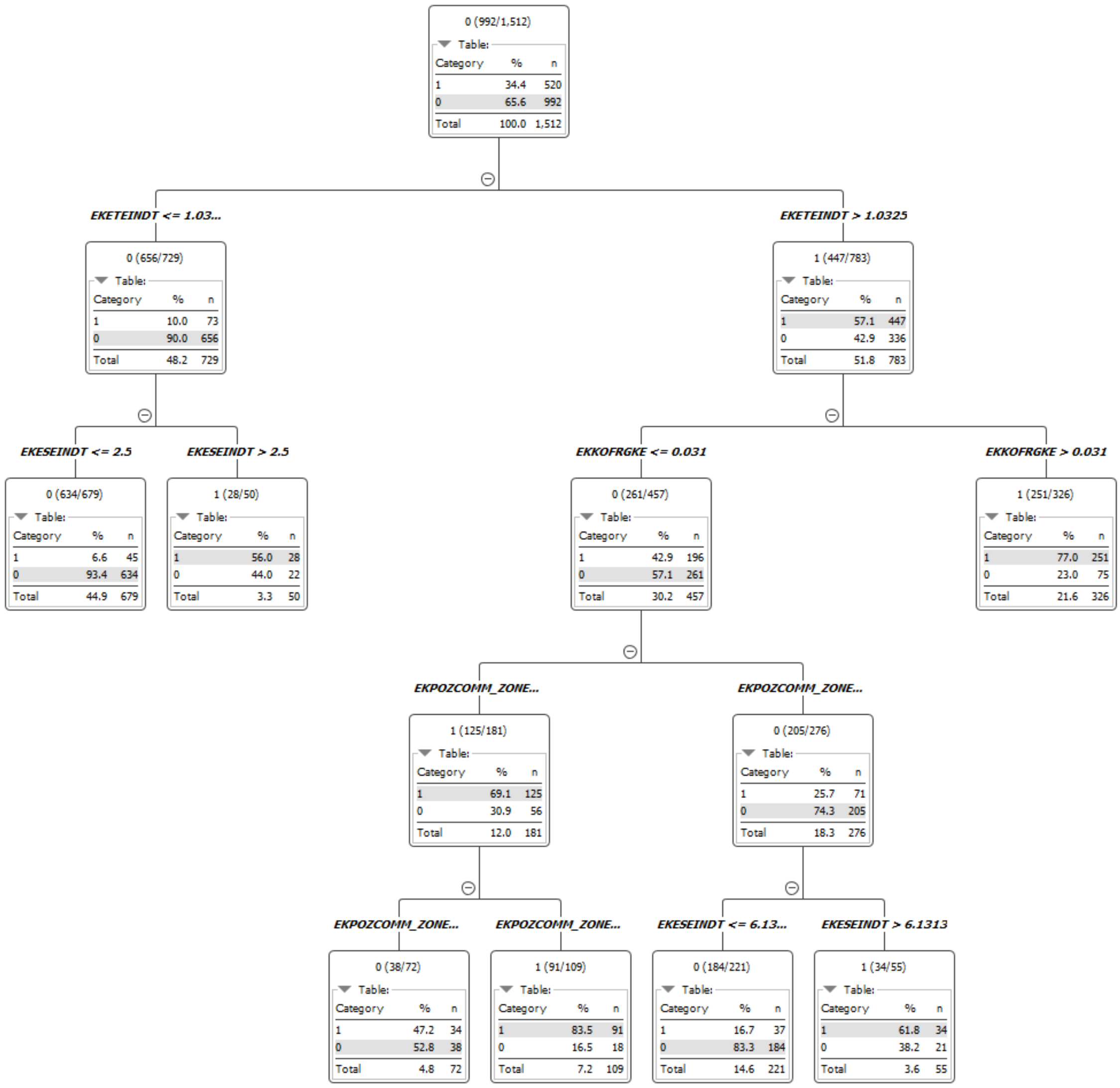

Figure 7: Decision tree generated in model training (error rate is 17%)

7.4 Validation

Following the initial output scores of the models and insights gained from the decision tree, validating the output was required to help understand the findings and clarify how well material defects can be predicted using process deviation data. In the following section the relation between process deviations and material notifications was checked, the influence of only statistically related deviations was tested, and the best scoring model was tested on the actual data from the research supplier.

7.4.1 Confirming Process Deviation and Material Notification Data is Unrelated

In the data preparation phase, it was observed that the material notification data and process deviation data was required to be unrelated for the predictions to be of any value. It was also stated that this was tested to ensure no causal relation was in place, this test is briefly explained below. 30 notifications were identified to have a logged change after the notification was made, raising concerns that there could be a causal relation. Domain experts stated changes were never made to purchase orders as a result of material notifications, however, to verify this, a prediction was also made without the 30 'late change' cases. The gradient boosted trees model was trained and tested on data that excluded the 'late change' cases. The gradient boosted tress model was used because this model proved to provide the highest score using all notifications. Training the model without the 30 cases, the model produced an accuracy of 98.035%, and Cohen's Kappa of 0.957. This result was slightly better than the original result but only marginally. The output suggested that the notification and deviation data was indeed unrelated. This was concluded because it was expected that the scores of the model without the 30 cases would be significantly poorer than with the cases, if they did indeed have a causal relationship with material notifications.

7.4.2 Prediction Based on Only Statistically Relevant Attributes

Insights from the statistical analysis provided information into what types of deviations were significantly related to the occurrence of material notifications. This information could also be used to limit the attributes used by the prediction models and to produce a prediction based only on are statistically relevant attributes. To test if this would be applicable and produce valid scores, the gradient boosted trees model was also trained and tested using only the statistically relevant process deviations as listed in *table 14* in a previous section, the total number of deviations, and the deviation occurrence indicators. The output of this test produced an accuracy of 96.259% and a Cohen's Kappa score of 0.916. This outcome was just slightly lower compared to when all logged deviations were used as input and can also be considered a very good prediction. The output does suggest it is better to use all changes and not just the statistically significant attributes, but shows the difference is minimal. The output could suggest that perhaps the occurrence of specific deviation types is also relevant in relation to the occurrence of other deviation types, even if they are statistically insignificant. This could show that patterns are perhaps of influence, but not very significant in this case. The output also proves that the model can make a decent prediction using less attributes, which is also an interesting insight and could be valuable if less data is directly available. These findings are also somewhat in line with the expectations of business experts and literature, considering multiple deviation types and frequencies could be of influence. Interestingly, when using only statistically insignificant deviations and counts, the accuracy drops to 79.535% and Cohen's Kappa is just 0.506. This is significantly lower than the previous output and shows that the statistical relevance of the deviation types is of influence on the quality of the prediction output. The prediction output could however still be considered good and indicates that deviations in general are a reasonable predictor of material defects.

7.4.3 Model Test on Real and Complete Data

To investigate if the prediction models could actually be applied for practical real-world scenarios, the model was finally tested on the complete and untransformed dataset used for the statistical analysis. This dataset consisted of real purchase order and order item data and material notifications that had not been synthetically produced or bootstrapped. The dataset was therefore much more variable and also highly unbalanced compared to the training and test data previously used. The prediction was made using the gradient boosted trees model that had been trained on the balanced and transformed prediction data because this model previously produced the best results prediction output. The output scores of the model are shown in the table on the following page (*table 18*).

Quite notable was the significantly larger number of purchase orders that were inspected in this test and the imbalance between the two classes, this also has an effect on how the output should be read. Looking at accuracy and error the model performed pretty well. This could also be said for the recall and specificity of the model. However, when inspecting the precision, Cohen's Kappa, and F-score it becomes clear that the model scores significantly lower than expectation based on the transformed data output. However, the according to Landis & Koch (1977) method of scoring, the output could still be considered an acceptable prediction. This indicates the model is still significantly better than randomly guessing the output.

<table>
<tr><td>Confusion Matrix</td><td><table><tr><td></td><td>0</td><td>1</td><td></td></tr><tr><td>0</td><td>10456</td><td>625</td><td>11081</td></tr><tr><td>1</td><td>12</td><td>135</td><td>147</td></tr><tr><td></td><td>10468</td><td>761</td><td>11228</td></tr></table></td></tr>
<tr><td>Correctly Classified</td><td>10,591</td></tr>
<tr><td>Wrongly Classified</td><td>636</td></tr>
<tr><td>Accuracy</td><td>94.318%</td></tr>
<tr><td>Error</td><td>5.682%</td></tr>
<tr><td>Cohen’s Kappa (κ)</td><td>0.282</td></tr>
<tr><td>True Positive Rate (Recall or Sensitivity)</td><td>0.918</td></tr>
<tr><td>True Negative Rate (Specificity)</td><td>0.944</td></tr>
<tr><td>Precision</td><td>0.177</td></tr>
<tr><td>F1</td><td>0.297</td></tr>
</table>

Table 18: Gradient boosted trees model output score using real and complete data

The reason the model scores lower is because the number of false positives has significantly increased compared to previous results. This in turn resulted in a much lower precision, which directly influences the F-score and the Cohen’s Kappa. The percentage of false positives is about 5.6% of the total number of cases without a material notification. This was somewhat concerning when considering the number of false negatives was around 8.1% of cases with a material notification. Which indicates many defects would not be detected. Multiple reasons could be mentioned as a cause for the increased number of false positives. Firstly, the model could be overfitted to the prediction data causing predictions to be less reliable. Secondly, because the data was balanced to train the prediction model, a large amount of orders without material notification were randomly excluded. This bias towards material notification occurrences could result in the model being more prone to mark an order as a risk case. Finally, it could be possible that the defect has not yet occurred but has yet to be detected. While this last possibility is not very likely to be true for many cases, it was found that some material defects occurred a half year or even an entire year after the order was completed.

From the results it could be stated that the model could be used in practice, however it is far from perfect and reliable. An advantage of the model is that it generally does predict defects within acceptable standards, with the false negatives being around 8%. However, depending on the criticality of components or requirements set by the customer, missing 8% of the defects could be insufficient. Also, because the model has a high number of false positives it could be difficult to use the predictions to take preventive actions and cause an overreaction of preventive or corrective processes. In contrast, the model suggests about 6% of materials could be defective and inspecting just these materials is far less intensive than inspecting all materials.

8 Discussion

The following section explains the findings of this research and suggests how the results should be interpreted. The chapter first explains what information was gained from literature and business understanding and continues to explain what was found with statistical analyses. Finally, the results from the prediction model are analyzed and suggestions about the practicality of using procurement process deviation data for predicting material defects are made.

8.1 Business Understanding & Literature

When comparing the views of the business and domain experts at ASML, and how the business is run to the information from literature, both sources provided similar views on the relationship between material quality and the purchasing process. At ASML, material quality is measured using the basic form of a defect rate, which is also most commonly suggested in literature as an effective indicator. Therefore, using the data concerning material notifications is a valid and relevant indicator about material defects according to both experts and literature. In literature, little is mentioned about the influence of process deviations on material quality and this can be recognized in the knowledge and understanding of business experts on the topic. The experts consider process deviations as a plausible influencer of material quality but do not consider it to be the best indicator of future defects, generally stating an interest to understand if there is a relationship. Through logical reasoning the business experts consider that an increase of production pressure on the suppliers due to quantity changes or delivery date changes could indeed have influence on the supplier's quality. This was somewhat supported by literature, which states that order quantity, delivery dates, time windows, schedule flexibility, and demand uncertainty are all factors that have influence on supplier quality. However, literature does not state pressure as the reason for this, nor could this be measured or proven based on the information from ASML. Interestingly, the business experts did not expect price of materials to be of influence on the quality of their suppliers, while literature clearly states this is a factor. This discrepancy could be because ASML generally has close ties to their suppliers and sets fixed material prices for materials annually. Price deviations are therefore not very common, and not considered. Factors such as decentralized manufacturing, monitoring supplier with IT support, and performing root cause analyses on defects were also not considered by experts but are stated as influencing factors in literature. Here the discrepancy occurred because these processes or actions are already standardized within ASML, therefore the business experts did not consider to mention them. As mentioned, the business experts do consider process deviations, also related to procurement, to have some influence on material quality. They also agree that it is likely that more deviations result in a higher risk of material defects, which was also suggested by literature. A high number of deviations could also be considered an indicator of poor communication and unstandardized processes, which is very relevant for quality according to literature and to some degree by the business experts. Finally, all business experts suggested that material maturity and complexity would be a very good indicator of quality. This was also supported by literature and could possibly be used to improve material quality predictions. Unfortunately, the material maturity and complexity are out of this research scope and were not investigated beyond business understanding and literature. However, because the business experts were adamant that these factors could be very important, it would be highly advisable to investigate the relationship in future research. The findings from the business understanding and the information collected from literature complemented each other quite well when comparing the findings. It was therefore expected that, based on only the literature and insights from business experts, a statistical and predictive relationships between purchasing process deviations and material quality would be identified when analyzing the available data.

8.2 Statistical Analysis

The information that was collected from literature, business experts and documentation was next applied to understanding the data ASML holds and how to inspect it with statistics. It was critical for the research that there would be some relationship between the occurrence of material defects and the occurrence of procurement process deviations, because this would help support ideas from literature. It was found that there is a one-sided relationship between the two, in which material defects only occurred in orders where at least one deviation was logged. This finding seems to suggest that material defects could be a direct consequence of process deviations, which was somewhat expected when testing the relationship. It is however not realistic to state process deviations will always lead to material defect or that defects are a direct consequence of the process deviations only. This is because, in the analyzed data, only 2% of all orders is not associated with any process deviations. In turn, this indicates that cases without a deviation are actually too uncommon to make the conclusion that deviations will lead to defects. Furthermore, because just 1.5% of all orders actually result in a material defect, the probability of a defect occurring without any process deviations is also very small. Critically, the results of the first statistical test seem promising and help support the research but are not conclusive and do not provide a sufficient base to make claims on.

To further understand the relationship between process deviations and material defects, a statistical analysis inspecting the relationship between the total number of deviations per order and defect occurrence was conducted. While the results of this test cannot be relied on with certainty, because the distribution of number of defects was not by definition normally distributed, the results do clearly show that the total number of deviations is statistically significant when compared to the occurrence of defects. It was expected that a higher number of deviations would result in a higher probability or occurrence of defective materials. This was mainly because more deviations potentially result in a greater variety of deviation types and more critical deviations which could have had influence on multiple quality influencers. However, the results do not show that a high number of process deviations is always related to a defect. Nevertheless, using histograms to compare the number of deviations pers class, it was observed that, on average, the number of deviations is slightly higher in cases with a material notification than without. Together with the statistical significance, this is interpreted as an indication that the total number of deviations is of influence on the quality of materials. It should however also be stated that the total number of deviations are also strongly influenced by just a few specific deviation types that occur frequently. Most of these specific deviations later proved to be statistically relevant to material defects separately. This could suggest the total number of deviations is not quite as relevant as the tests show, but rather based on just a few highly frequent deviations.

To better understand the relationship between specific deviation types and material notifications, their statistical relationships was tested using the T-test. Out of the 18 logged changes that were considered to be relevant according to literature or business experts, 9 were statistically related to the occurrence of a material defect. As was expected, both deviations linked to the delivery date, the rescheduling, and reconfirmation of the new date, were statistically significant. Both experts and literature suggested this would be the case and both deviation types were logged quite frequently in the change data. Changing the order quantity or rescheduling the quantities to be delivered also proved to be statistically significant to defects, as was expected from both literature and the business experts. Unexpectedly, the reconfirmation of the delivery quantity was not statistically significant, even though it closely resembles the order change quantity. No definitive explanation for this was identified, but it was observed that a lack of confirmation by the supplier was the case in multiple orders where a quantity change was requested from ASML. This could suggest that quantity changes were desired, but not manageable by the suppliers and potentially resulted in defects. Reapproving the order was also marked as a relevant indicator of defects. This was not suggested in literature or by business experts but could be a result of other big changes occurring on an order. These big changes in turn require the procurement department to reapprove the order. This deviation could also be linked to the reconfirmation of the order from the supplier's side, which was also significant. Less surprising was to find that deleted orders did not result in a material defect. This would be expected because the occurrence of the order deletion change indicates material are never delivered and can therefore not be defective. Finally, as was expected by literature, but not by business experts, price changes proved to be statistically relevant to material defects. In some cases, the price was not yet determined when the order was placed, which means the supplier would start producing materials for the customer without setting a price. Interestingly, these cases never resulted in a material defect. This is somewhat unexpected but could perhaps be explained by understanding that the supplier can then set the price. This could suggest that the supplier can claim a higher value of the product but is therefore also more inclined to produce better materials. Changing the price of an order could result in the supplier getting less value out of an order and therefore lowering the priority which could result in poorer quality. It could also be suggested that price changes are also related to quantity changes, which in turn could lead to the expected increase in production pressure and other issues related to quantity changes.

The other statistically insignificant changes also contained some surprises that were not anticipated. Firstly, changing the remaining delivery lead time for the supplier did not prove to be significant to material defects. This is probably because lead time changes only occurred in 7 orders, of which none resulted in a material notification. Because delivery date changes do influence material quality, it is expected that the analyzed supplier's data does not provide a complete picture of the influence of this specific deviation. This could also be said for changing the delivery window, which is not statistically relevant in this dataset, but also only occurs just once. Three other types of deviations also occurred rarely and are possibly not significantly related to material defects because of this. These are changing order manager, closing an order before the invoice has been completed, and reclosing an order once it was initially marked as completed already. These changes are generally rare according to the business experts but therefore it would be interesting to see if there would be a relationship if they were to occur more frequently. However, according to literature these deviations types were not considered to be of influencers on quality and it is therefore no surprise that they are considered insignificant according to the analysis. Showing that frequency was not always relevant for statistical significance, the single deviation that was most common, resending the purchase order to the supplier, was not considered statistically significant. This deviation can be considered to be a kind of indicator of communication level, having to resend the order more often implies communication about the order was maybe poor. However, this is not a very reliable indicator and could even be considered as opposite, resending orders more frequently means there is closer communication between customer and supplier. The deviation was inspected because it was the only custom logged deviation at ASML and perhaps of interest but did not prove to be relevant. Finally, the more general 'last changed on' indicator was not considered statistically relevant. This indicator specifically monitors when a change was made to the general purchase order data table in SAP. The indicator is similar to the total number of deviations per order, but only includes one data table. It was expected that this deviation would be significant because it is so similar to the statistically significant total deviations field. The disparity between the two fields could suggest that the general PO changes are less significant for material notifications. This can be supported by the fact that the table does not count delivery date or quantity changes either.

8.3 Prediction Results

Prior to the research there was little information about the influence of process deviations on material quality. In this case the research looked specifically at the effect of the procurement process on supplier caused defects, about which no previous research could be identified. Because of this, the research was started with little expectations and with the main goal of exploring the potential possibilities. Observations from literature, business experts, and the statistical analyses suggested that predicting material defects using data concerning procurement process deviations could be a possibility. It was therefore expected that certain prediction models would perform reasonably well, however the initial prediction results were far better than expected. Each of the six models provided at least moderately good predictions based on the training data, suggesting that predictions of material defects based on the frequency of deviations in the purchasing process are a serious possibility.

Comparing the six prediction models, it was observed that the gradient boosted trees outperformed the other tested models with a significant margin. Using the accuracy, Cohen's Kappa, and F-score to compare model performances, it was shown that the model is the best for this case, and it produces a very high valued output when trained and tested on balanced data. It was expected that this model would outperform the decision tree, because it is an extension of this model, and would outperform the random forest, because this was frequently observed in scenarios where the two were compared. Furthermore, gradient boosted trees have frequently proven to be very powerful predictors and are a common algorithm of choice when making prediction models. Expectations from logistic regression were moderate, because it was not expected that the problem could be approached so simply. Somewhat surprisingly, the model still performed reasonably and worked better than expected. It is expected that the model's decent performance is related the earlier identified statistical relationships, but this was not further investigated. The use of neural networks in the PNN and MLP models also showed decent predictive capabilities. This was also somewhat expected from previous findings and the expectation that there could be specific patterns that increase the defect risk. However, none of the models were tuned extensively to identify the best parameter settings for optimal results. This means that it is possible that the tested models could outperform each other if tuned to produce better results. This was not done in this research due to scoping and time limitations. The objective was to identify if prediction is possible based on procurement process deviations only, therefore turning the models to perform optimally was not a primary objective.

The good performance of the prediction models should also be considered critically and not simply accepted as proof that predicting material defects is possible using only procurement process deviations frequencies Firstly, the training data was quite significantly manipulated to ensure it could be used for training the models. Due to the imbalance of cases with and without material notifications, a large chunk of the majority class, without material notifications, was randomly undersampled and discarded. This could have serious implications because critical rows could have been deleted in the process and the model is trained on a less diverse and realistic dataset. Furthermore, also to tackle the imbalance, the minority class was oversampled using SMOTE. This imposes the risk of creating ambiguous cases that could overlap too much with other classes. The synthetic and rebalanced data was then bootstrapped to produce the final dataset. All these actions have influence on both bias and variability of the dataset, which could have significant influence on the output and reliability of the models. The training and test data also only consider one specific supplier that only delivers highly complex and low frequency materials. This means that the models may not work well on data from other suppliers, or perhaps data from other suppliers cannot even be used to make accurate predictions. Simply looking at one supplier and customer does provide insights into the possibilities and potential deviation types but cannot be used to prove prediction is always possible. Finally, there is a risk of overfitting the models to this dataset which would in turn also result in a high bias when tested on other data. To ensure this is not the case, different data sources, supplier types, and maybe even material types would also need to be investigated. Considering all the limitations of this research, it would be advisable to further investigate the relationship of previously mentioned limitations on supplier material quality.

To test if the model is relevant on real data despite the intensive transformations, the model was validated on the complete dataset that was used for the statistical analysis and from which the training data was generated. The prediction results show that the model is far less capable of making accurate predictions on real data, which is must larger than the training set and highly unbalanced. This is probably caused by multiple factors such as overfitting and model bias. Most likely this is a result of the data balancing actions and the true relevance or impact of process deviations on material defect occurrences. This can be somewhat observed in the number of false positives the model produces. Both in absolute and percentage of total, the model produces more false positives than false negatives, which suggest it is somewhat biased toward marking orders with material notifications. Based on the models scores, it could still be considered an acceptable model and clearly provides predictive insights, but the error predictions do make practical application of the model more difficult. It could be considered desirable for the model to be slightly biased towards false positives, because this would mean the risk of missing probable defects is smaller, unfortunately the model is still not capable of accurately predicting all material defects. Therefore, in practice the model would probably still miss critical defects and at the same time trigger too many defect warnings to be able to actually investigate the predictions. If the action taken by the customer would be to physically inspect materials that were considered to be at risk, this would likely be a somewhat too costly approach to actually implement. However, if physical inspections were to be an option, the model ensures less than 7% of the order items would need to be checked, and only 9% of the defects would not be captured. A more practical model would preferably capture all material defects, which would provide certainty that all defects can be identified, even though a costly approach is still required. Another approach to this would be to include confidence or probability scores to indicate how high the risk of a defect is. This would allow ASML to decide if inspections are desired based on the level of risk they are willing to take. The model does provide a confidence indicator of each prediction, but this was not analyzed due to time constraints.

Based on the findings and output of the prediction model, the research suggests that predicting material defects is possible based on process deviations in the procurement process. With acceptable accuracy levels, most material defects can be detected before they the components are implemented in the factory. However, based on reliability requirements of the customer, applying these predictions to real world scenarios, more investigation is probably required. Particularly parameter tuning, identification of other relevant data, and generalization of the model would result in a more reliable and secure prediction.

9 Conclusion

In the following section the research is concluded. This is done by briefly addressing the major findings of the research and relating these to the main goal and research questions. The chapter also reflects on the implication for research and practice and finally, the chapter states limitations of the research and suggests future research.

9.1 Major Findings

The goal of this research is to identify if purchasing process deviations can be used as a predictor of supplier material quality. The research shows that, to an acceptable degree, data concerning procurement process deviations can be used to predict supplier material defects. To derive this output, the research set out to answer five research questions to support the research goal. First, valid indicators of supplier material quality and the factors influencing these indicators were researched. It was identified that supplier quality can be categorized into four types, specifically product, service, process, and organizational quality. Because the goal is to predict material quality, the research focused on predicting a form of the material defect rate, which is the most common indicator of product quality. The defect rate can be calculated as the number of defective materials out of the total number of delivered materials. Therefore, predicting material defects can be a good way to predict supplier material quality. The research also investigated what factors have influence on supplier quality in general and, more specifically, if process deviations could have influence on this quality. The research shows there are multiple factors that have influence on supplier quality, ranging from people, product, or process factors. Secondly, identifying the influence of purchasing process deviations was to be investigated. The influence of process deviations, particularly related to procurement, was not previously researched extensively and would therefore still require further research. However, based on a combination of literature and business expert knowledge, it was recognized that there should be a relationship between process deviations and material quality. Question three and four focused on proving the relationship between specific deviation types and the total number of deviations per order to material defects. Two specific factors, material quantity and delivery date changes were identified to be particularly relevant in literature and according to business experts. These also proved to be significant in statistical analyses measuring the relation between material defects and process deviations. Furthermore, based on the decision tree prediction model, the delivery date was again marked as a relevant deviation type. Besides quantity and delivery date, from the available data, price changes and document reapprovals were identified as relevant indicators. Due to data limitations, not every type of influencing factor identified in literature could be observed in the data, therefore only specific deviations that occur in the collected dataset could be tested. The total number of deviations was also identified as a significant indicator of defects. In general, orders with a slightly higher average number of deviations seemed to have a somewhat higher risk of resulting in a defect. Finally, question five would be answered by identifying if a reliable prediction of material defects could be derived from process deviation data. Using the logged deviations from orders placed at a supplier, multiple prediction models were used to test if they could predict material defects based on the logs. Initial prediction proved promising, particularly gradient boosted trees scored very high on the training data. When testing the model on real and highly unbalanced data the model performed far less accurately, but still made acceptable predictions. Combining the research findings, it can be stated that purchasing process deviations can be used to make an acceptable prediction of supplier material quality, however, to improve the reliability of the output will require more research.

9.2 Implications for Research

Research about predicting material quality and about the influence of process deviation on material quality is very limited. This research provides some insight into the influence of process deviations, particularly of the purchasing process, on supplier material quality. Furthermore, the research clearly shows there is both a statistical and predictive relationship between purchasing process deviation frequencies and supplier material quality. It identifies what deviation types are possibly related to material defects and shows that change logs of the purchasing process can be used to train prediction algorithms and make acceptable predictions. By doing this, the research shows that the frequency and combination of purchasing process deviations have influence on supplier product quality. It can also be used to propose that other processes may also have influence on material quality. It should also be mentioned that the literature review provides an overview of quality measurement indicators and influencing factors commonly used in recent literature. When conducting the literature review few sources providing such an overview could be identified, therefore the tables could also be used in other research concerning supplier quality indicators and influencing factors. The literature review further provides an initial approach to linking research about the key topics together and shows there are still gaps in the literature that must be filled.

9.3 Implications for Practice

The research shows that deviating from the standard purchasing process does in fact have influence on material quality and therefore should be taken into consideration when purchase orders are changed. This does not imply that customers should never make changes, because it was also shown that, by far, most process deviations do not result in a defect. Also, it is suspected that deviations predominately only occur when strictly required by customer or supplier. However, the impact of frequently making deviations should be considered to improve quality. More specifically, frequently adjusting purchase orders could result in a higher risk of defective materials being delivered by the supplier, particularly when these changes are related to quantity, delivery date, price, or when multiple changes occur in one instance and reconfirmation of orders is required. Using change logs to predict material defects is a serious possibility, but depending on the accuracy and reliability requirements of the customer, the predictions will prove to be of practical use or not. The model is incapable of producing near perfect predictions and therefore will likely make false predictions. In the research case data specifically, the model could be used for prediction but would certainly miss some materials at risk and frequently mark materials that are not defective. Testing show less than 9% of defects would be missed and less than 7% of non-defective orders would be marked as at risk using the case data. Therefore, basing predictions of defects solely on purchasing process changes will probably not make a significant enough contribution to predicting supplier quality in every situation. If suppliers were interested in implementing a similar prediction tool but require it to be more reliable, it would be advisable to first identify other data that could be used to improve the accuracy of predictions. Expectations are that adding material characteristics, such as complexity or maturity, could further improve the model's output. Also, using data from more than one supplier should be considered, and maybe even different prediction models for different supplier types could be relevant when accurate predictions are desired. Suppliers that successfully implement such a prediction model that meets accuracy requirements could use the output to, for example, apply extra inspections of materials at risk, or base safety stock levels on product risk levels.

9.4 Limitations & Future Research

The research shows promising and interesting results, however there are still multiple limitations to the research which could be addressed in future research. First, and most obviously, the discrepancy between the output of the model based on the balanced training data and the unbalanced real data shows the model can be improved upon. When training and testing the model on the transformed and balanced dataset, the prediction output is near perfect and far more reliable than when the model is tested on the original and unbalanced dataset. In future research it would perhaps be useful to further tune the prediction model to produce better results or maybe use more data and more characteristics. Examples of this are that the model is trained and tested on data from just one supplier which only produces a certain type of material. Using data from multiple suppliers would ensure the model would be more generalizable and potentially highlight or reduce the impact of certain process deviations. Furthermore, using data from different suppliers could ensure a broader spectrum of material types is included in the prediction. Currently, primarily highly complex and medium-low production volume materials are included in the dataset. Because the data is limited to one supplier and a specific subgroup of material, it would be advisable for future research to broaden the scope and use more varied data for research. Perhaps investigating if different suppliers are affected by different types of deviations and if different material types have influence on the predictability are valid research objectives. Both are also backed by literature and business experts in this research. Another limitation of this research is that it does not consider information from the supplier side. While this does ensure the prediction can be made solely based on information the customer has, it also reduces the insight that can be gained from the predictions. If the supplier would be included in the research, perhaps other underlying reasons for material defects could be identified or even confirmed. Perhaps in future research it would be interesting to investigate the suppliers' opinions on the impact of (purchasing) process deviations. This information could then be used to validate if the process deviations did indeed lead to production complications or pressure that influence material quality. Also, the research is limited to the deviations that could be identified from the available data and does not necessarily include ever factor that could have influence on quality according to literature. This means that other types of procurement process deviations could also have influence but were not present in this dataset. This issue could probably partially be resolved by taking more data from multiple suppliers. Finally, the research does not investigate if particular process patterns are of influence and if other processes could also be relevant for predicting material quality. Therefore, it remains unsure if the sequence or order of changes could be of influence and if other parallel process could also be of influence. Furthermore, the research only uses frequency counts of purchasing process deviations to make a prediction. Because of this, the research does not provide insights into the effect of specific changes, for example, if moving the delivery date forward or backward has a different relationship to quality. To summarize, future research should investigate if the process sequence of logged changes and the details of each change are a factor when predicting supplier material quality. Research should also include a broader range of suppliers and materials that are used to base the prediction on, and include more data containing a greater range of deviations. Perhaps other parallel processes are also of influence and could be investigated. Combining all these improvements, the model will likely become more accurate when tested on real and unbalanced data.

10 References

Ang, C.-L., Davies, M., & Finlay, P. N. (2001). An empirical study of the use of information technology to support total quality management. *Total Quality Management*, 145-157.

ASML. (2021). *Supplier Handbook: Your guide to working with ASML.* Veldhoven: ASML.

Baird, K., Hu, K. J., & Reeve, R. (2011). The relationships between organizational culture, total quality management practices and operational performance. *International Journal of Operations & Production Management*, 789-814.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 - A step-by-step data mining guide.* SPSS.

Chen, C.-C., & Yang, C.-C. (2003). Total-costs based evaluation system of supplier quality performance. *Total Quality Management & Business Excellence*, 325-339.

Chen, C.-C., Yeh, T.-M., & Yang, C.-C. (2004). Customer-focused rating system of supplier quality performance. *Journal of Manufacturing Technology Management*, 599–606.

Chen, K. L., Chen, K. S., & Li, R. K. (2005). Suppliers capability and price analysis chart. *International Journal of Production Economics*, 315–327.

Chin, K. S., Yeung, I.-K., & Pun, K. F. (2006). Development of an assessment system for supplier quality management. *International Journal of Quality & Reliability Management*, 743-765.

Chuah, P. L., Wong, W. P., Ramayah, T., & Jantan, M. (2010). Organizational context, supplier management practices and supplier performance: A case study of a multinational company in Malaysia. *Journal of Enterprise Information Management*, 724-758.

Clemons, R., Baddam, S. R., & Henry, R. M. (2020). Factors influencing swift and effective resolution of supplier problems. *International Journal of Productivity and Performance Management*, -.

Cooper, H. M. (1984). *The integrative research review: A systematic approach.* Beverly Hills, CA: Sage.

Dijkman, R., Turetken, O., IJzendoorn, G. R., & de Vries, M. (2019). Business Process Exceptions in Relation to Operational Performance. *Business Process Management Journal*, 908-922.

Glas, A. H. (2018). The impact of procurement on supplier satisfaction: service, communication, and speed. *International Journal of Integrated Supply Management*.

González-Benito, J., & Dale, B. (2001). Supplier quality and reliability assurance practices in the Spanish auto components industry: a study of implementation issues. *European Journal of Purchasing & Supply Management*, 187-196.

Hand, D., & Christen, P. (2017). A note on using the F-measure for evaluating record linkage algorithms. *Statistics & Computing*, 539-547.

Hawkins, T. G., Gravier, M. J., Berkowitz, D., & Muir, W. A. (2015). Improving services supply management in the defense sector: How the procurement process affects B2B service quality. *Journal of Purchasing & Supply Management*, 81-94.

Huo, B., Zhao, X., & Lai, F. (2014). Supply Chain Quality Integration: Antecedents and Consequences. *Transactions on Engineering Management*, -.

Immonen, M., Hallikas, J., & Pynnonen, M. (2016). Antecedents of system purchasing in B2B services. *Journal of Purchasing & Supply Management*, 205-213.

Kaynak, H., & Hartley, J. L. (2008). A replication and extension of quality management into the supply chain. *Journal of Operations Management*, 468–489.

KNIME. (2021, January 1). *PNN Learner Node*. Retrieved from KNIME Hub: https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.mine.bfn.radial.RadialBasisFunctionLearnerNodeFactory

KNIME. (2021, January 1). *RProp MLP Learner*. Retrieved from KNIME Hub: https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.mine.neural.rprop.RPropNodeFactory2

Landis, R. J., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *International Biometric Society*, 159-174.

Lee, Y.-C., Yen, T.-M., & Tsai, C.-H. (2008). The Study of an Integrated Rating System for Supplier Quality Performance in the Semiconducter Industry. *Journal of Applied Sciences*, 453-461.

Lin, C., Kuei, C.-H., & Chai, K.-W. (2013). Identifying critical enablers and pathways to high performance supply chain quality management. *International Journal of Operations & Production Management*, 347-370.

Maestrini, V., Luzzini, D., Caniato, F., Maccarrone, P., & Ronchi, S. (2018). The impact of supplier performance measurement systems on supplier performance. *International Journal of Operations & Production Management*, 2040-2061.

Mellat-Parast, M. (2015). An institutional theory of quality outcomes in strategic supply chain partnership. *International Journal of Quality & Reliability Management*, 346-360.

Munir, M., Jajja, M. S., Chatha, K. A., & Farooq, S. (2020). Supply chain risk management and operational performance: The enabling role of supply chain integration. *International Journal of Production Economics*.

Noshad, K., & Awasthi, A. (2015). Supplier quality development: A review of literature and industry practices. *International Journal of Production Research*, 466-487.

Noshad, K., & Awasthi, A. (2018). Investigating critical criteria for supplier quality development. *nternational Journal of Management Science and Engineering Management*, 215-224.

Park, S., & Hartley, J. L. (2002). Exploring the Effect of Supplier Management on Performance in the Korean Automotive Supply Chain. *The Journal of Supply Chain Management*, -.

Picaud-Bello, K., Johnsen, T., Calvi, R., & Giannakis, M. (2019). Exploring early purchasing involvement in discontinuous innovation: A dynamic capability perspective. *Journal of Purchasing and Supply Management*.

Quigley, J., Walls, L., Demirel, G., MacCarthy, B. L., & Parsa, M. (2018). Supplier quality improvement: The value of information under uncertainty. *European Journal of Operational Research*, 932-947.

Randolph, J. J. (2009). A Guide to Writing the Dissertation Literature Review . *Practical Assessment, Research & Evaluation*, -.

Salimian, H., Rashidirad, M., & Soltani, E. (2020). Supplier quality management and performance: the effect of supply chain oriented culture. *Production Planning & Control*, -.

Shin, H., Benton, W. C., & Jun, M. (2009). Quantifying suppliers' product quality and delivery performance: A sourcing policy. *Computers & Operations Research*, 2462-2471.

VDL ETG Almelo. (2021, February 1). *Het Bedrijf*. Retrieved from VDL ETG Almelo: https://www.vdletg.com/en/het-bedrijf/locations/almelo

Wang, H., & Fang, Z. (2019). Supply Chain Quality Coordination Strategies for Complex Product under the Manufacturer Being at a Disadvantage. *Association for Computing Machinery*, -.

White, T. E., Iammartino, R., & Fossaceca, J. M. (2019). Assessing the impact of requirements review on quality outcomes. *Quality and Reliability Engineering International*, 882–893.

Xu, L., Peng, X., Pavur, R., & Prybutok, V. (2020). Quality management theory development via meta-analysis. *International Journal of Production Economics*, -.

Yoo, S. H., & Cheong, T. (2018). Quality improvement incentive strategies in a supply chain. *Transportation Research Part E*, 331–342.

Yu, Y., & Huo, B. (2019). The impact of relational capital on supplier quality integration and operational performance. *Total Quality Management & Business Excellence*, 1282-1301.

Zhang, M., Guo, H., Huo, B., Zhao, X., & Huang, J. (2019). Linking supply chain quality integration with mass customization and product modularity. *International Journal of Production Economics*, 227–235.

Zsidisin, G. A., Petkova, B., Saunders, L. W., & Bisseling, M. (2016). Identifying and managing supply quality risk. *The International Journal of Logistics Management*, 908-930.

Zu, X., & Kaynak, H. (2012). An agency theory perspective on supply chain quality management. *International Journal of Operations & Production Management*, 423-446.

11 Appendix

A Literature Review Approach

Search restrictions used to filter relevant literature

Search Fields	**Search Input**
Research Terms (Title, Abstract, or Keywords)	["supplier quality" AND management AND performance]
Publish Date	[2000 – 2020]
Subject Area	[Business OR Engineering OR Decision Science OR Computer Science OR Economics OR Social Science OR Material Science OR Mathematics OR Multidisciplinary]
Document Types	[Review OR Research OR Case]
Language	[English]

Search output and data collection overview, number of articles identified/deleted

Search Results	Science Direct	Scopus	ACM Digital Library	IEEE	Web of Science
Initial search *(1)*	12	70	12	1	48
Article types	Review (0) Research (12) Case (0)	Review (6) Research (64) Case (0)	Review (0) Research (12) Case (0)	Review (0) Research (1) Case (0)	Review (3) Research (45) Case (0)
Exclusions post topic & keywords filtration *(2)*	2	18	8	0	18
Exclusions post abstract filtration *(3)*	2	30	1	0	11
Exclusions of identified duplicates *(4)*	0	8	0	0	9
Exclusions post discussion & conclusion filtration *(5)*	1	4	2	0	0
Total identified unique articles	7	10	1	1	10

B Table & Field Changes Descriptions (Recorded Process Deviations)

Table Name	Field Name	SAP Description	Scope
EINKBELEG	F01 E	PO header text field	Out
EINKBELEG	F24 E	PO header text field	Out
EINKBELEG	F02 E	PO line item text field	Out
EINKBELEG	F03 E	PO line item text field	Out
EINKBELEG	F04 E	PO line item text field	Out
EINKBELEG	F05 E	PO line item text field	Out
EINKBELEG	F06 E	PO line item text field	Out
EINKBELEG	F80 E	PO line item text field	Out
EINKBELEG	F92 E	PO line item text field	Out
EKES	EINDT	Delivery Date of Vendor Confirmation	In
EKES	MENGE	Quantity as Per Vendor Confirmation	In
EKES	KEY	Vendor conformation creation indicator	Out
EKES	ERDAT	Creation Date of Confirmation	In
EKES	UZEIT	Delivery Date Time-Spot in Vendor Confirmation	In
EKES	XBLNR	Reference Document Number	Out
EKET	KEY	Delivery Schedule creation indicator	Out
EKET	EINDT	PO line item delivery date	In
EKET	MENGE	Scheduled Quantity	In
EKET	SLFDT	Statistics-Relevant Delivery Date	Out
EKKN	KEY	Account assignment creation indicator	Out
EKKO	INCO2	Incoterms (Part 2)	Out
EKKO	KEY	PO header creation indicator	Out
EKKO	PROCSTAT	Purchasing document processing state	Out
EKKO	EKGRP	Purchasing Group	In
EKKO	FRGKE	Release Indicator: Purchasing Document	In
EKKO	FRGZU	Release status	Out
EKKO	RLWRT	Total value at time of release	Out
EKKO	FRGSX	Release Strategy	Out
EKKO	INCO2_L	Incoterms (Local)	Out
EKKO	KONNR	Number of Principal Purchase Agreement	Out
EKKO	TELF1	Vendor's Telephone Number	Out
EKKO	VERKF	Responsible Salesperson at Vendor's Office	Out
EKPA	KEY	Partner role creation indicator	Out
EKPO	ZCOMM_ZONE_RPRNT	Commitment zone reprint	In
EKPO	BSTAE	Confirmation Control Key	In
EKPO	LOEKZ	Deletion indicator in purchasing document	In
EKPO	EFFWR	Effective value of item	Out
EKPO	BRTWR	Gross order value in PO currency	Out
EKPO	SCHPR	Indicator: Estimated Price	In
EKPO	STAPO	Item is statistical	Out
EKPO	NETWR	Net Order Value in PO Currency	Out
EKPO	AEDAT	PO line item Change Date	In
EKPO	KEY	PO line item creation indicator	Out
EKPO	NETPR	PO line item Net Price	In
EKPO	MENGE	Purchase Order Quantity	In
EKPO	ADRN2	Number of delivery address	Out
EKPO	ELIKZ	"Delivery Completed" Indicator	In
EKPO	EREKZ	Final Invoice Indicator	In
EKPO	KONNR	Number of principal purchase agreement	Out
EKPO	KTPNR	Item number of principal purchase agreement	Out
EKPO	LGORT	Storage location	Out
EKPO	PLIFZ	Planned Delivery Time in Days	In
EKPO	PRIO_REQ	Unknown	Out
EKPO	PRIO_URG	Unknown	Out
EKPO	REPOS	Invoice receipt indicator	Out

C Output Statistical Analyses

Output Chi-Square Tests (Complete Data)

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
ProcessDeviation * MN_occurence	11229	100.0%	0	0.0%	11229	100.0%

ProcessDeviation * MN_occurence Crosstabulation

			MN_occurence		
			0	1	Total
ProcessDeviation	0	Count	229	0	229
		Expected Count	226.0	3.0	229.0
		% of Total	2.0%	0.0%	2.0%
	1	Count	10853	147	11000
		Expected Count	10856.0	144.0	11000.0
		% of Total	96.7%	1.3%	98.0%
Total		Count	11082	147	11229
		Expected Count	11082.0	147.0	11229.0
		% of Total	98.7%	1.3%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3.101[a]	1	.078		
Continuity Correction[b]	2.153	1	.142		
Likelihood Ratio	6.098	1	.014		
Fisher's Exact Test				.078	.047
Linear-by-Linear Association	3.101	1	.078		
N of Valid Cases	11229				

a. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 3.00.
b. Computed only for a 2x2 table

Output Independent Samples T-test (Complete Data)

Independent groups t-test

Group Statistics

	Group	N	Missing Count	Missing Count (Group Column)	Mean	Standard Deviation	Standard Error Mean
Total Deviations	0	11082	0	0	32.731	24.072	0.2287
Total Deviations	1	147	0	0	39.8707	18.3061	1.5099

Levene Test
The Levene Test is used to test for the equality of variances.

	F	df 1	df 2	p-Value
Total Deviations	14.3094	1	11227	0.0002

Independent Groups Statistics
Confidence Interval (CI) Probability: 95.0%
Differences are reported of the groups: 0 - 1

	Variance Assumption	t	df	p-value (2-tailed)	Mean Difference	Standard Error Difference	CI (Lower Bound)	CI (Upper Bound)
Total Deviations	Equal variances assumed	-3.5823	11,227	0.0003	-7.1397	1.9931	-11.0465	-3.233
Total Deviations	Equal variances not assumed	-4.6754	152.7733	6.40E-6	-7.1397	1.5271	-10.1567	-4.1228

Output Independent Samples T-test (Balance Prediction Data)

Independent groups t-test

Group Statistics

	Group	N	Missing Count	Missing Count (Group Column)	Mean	Standard Deviation	Standard Error Mean
Total Deviations	0	1158	0	0	32.2211	23.1815	0.6812
Total Deviations	1	606	0	0	40.0768	18.0086	0.7315

Levene Test
The Levene Test is used to test for the equality of variances.

	F	df 1	df 2	p-Value
Total Deviations	38.5072	1	1762	6.78E-10

Independent Groups Statistics
Confidence Interval (CI) Probability: 95.0%
Differences are reported of the groups: 0 - 1

	Variance Assumption	t	df	p-value (2-tailed)	Mean Difference	Standard Error Difference	CI (Lower Bound)	CI (Upper Bound)
Total Deviations	Equal variances assumed	-7.2722	1,762	5.30E-13	-7.8557	1.0802	-9.9744	-5.737
Total Deviations	Equal variances not assumed	-7.8587	1,513.9044	7.31E-15	-7.8557	0.9996	-9.8165	-5.8949

Output T-test Change Logs Compared to MN Occurrence

Levene's Test

Test Column	test statistic (Levene)	df 1	df 2	p-value (Levene)
EKKOEKGRP	2.090983	1	11227	0.148199
EKKOFRGKE	59.57304	1	11227	1.28E-14
EKESEINDT	141.4569	1	11227	0
EKESERDAT	0.053064	1	11227	0.817819
EKESMENGE	1.93163	1	11227	0.164608
EKESUZEIT	0.053064	1	11227	0.817819
EKETEINDT	11.0683	1	11227	0.000881
EKETMENGE	490.7097	1	11227	0
EKPOAEDAT	28.92804	1	11227	7.66E-08
EKPOBSTAE	51.40518	1	11227	7.99E-13
EKPOELIKZ	0.053064	1	11227	0.817819
EKPOEREKZ	0.159278	1	11227	0.68983
EKPOLOEKZ	56.19619	1	11227	7.05E-14
EKPOMENGE	1580.723	1	11227	0
EKPONETPR	181.3902	1	11227	0
EKPOPLIFZ	0.372052	1	11227	0.5419
EKPOSCHPR	5.886304	1	11227	0.015275
EKPOZCOMM_ZONE_RPRNT	25.52769	1	11227	4.43E-07

Descriptive Statistics

Test Column	Group	N	Missing Count	Missing Count (Group Column)	Ignored Count (Group Column)	Mean	Standard Deviation	Standard Error Mean
EKKOEKGRP	0	11082	0	0	0	0.003519	0.059221	0.000563
EKKOEKGRP	1	147	0	0	0	0	0	0
EKKOFRGKE	0	11082	0	0	0	0.447573	1.213513	0.011528
EKKOFRGKE	1	147	0	0	0	1.278912	1.62507	0.134034
EKESEINDT	0	11082	0	0	0	1.63743	2.764779	0.026263
EKESEINDT	1	147	0	0	0	4.761905	5.010036	0.413221
EKESERDAT	0	11082	0	0	0	9.02E-05	0.009499	9.02E-05
EKESERDAT	1	147	0	0	0	0	0	0
EKESMENGE	0	11082	0	0	0	0.012001	0.209727	0.001992
EKESMENGE	1	147	0	0	0	0	0	0
EKESUZEIT	0	11082	0	0	0	9.02E-05	0.009499	9.02E-05
EKESUZEIT	1	147	0	0	0	0	0	0
EKETEINDT	0	11082	0	0	0	1.424562	1.900167	0.01805
EKETEINDT	1	147	0	0	0	3.44898	2.25189	0.185733
EKETMENGE	0	11082	0	0	0	0.011099	0.131507	0.001249
EKETMENGE	1	147	0	0	0	0.14966	0.528033	0.043551
EKPOAEDAT	0	11082	0	0	0	15.03781	11.13471	0.105772
EKPOAEDAT	1	147	0	0	0	16.20408	7.301659	0.602231
EKPOBSTAE	0	11082	0	0	0	0.081483	0.274576	0.002608
EKPOBSTAE	1	147	0	0	0	0.006803	0.082479	0.006803
EKPOELIKZ	0	11082	0	0	0	9.02E-05	0.009499	9.02E-05
EKPOELIKZ	1	147	0	0	0	0	0	0
EKPOEREKZ	0	11082	0	0	0	0.000271	0.016452	0.000156
EKPOEREKZ	1	147	0	0	0	0	0	0
EKPOLOEKZ	0	11082	0	0	0	0.082476	0.288863	0.002744
EKPOLOEKZ	1	147	0	0	0	0	0	0
EKPOMENGE	0	11082	0	0	0	0.00397	0.064308	0.000611
EKPOMENGE	1	147	0	0	0	0.14966	0.528033	0.043551
EKPONETPR	0	11082	0	0	0	0.147085	0.532344	0.005057
EKPONETPR	1	147	0	0	0	0.605442	0.990073	0.08166
EKPOPLIFZ	0	11082	0	0	0	0.000632	0.025126	0.000239
EKPOPLIFZ	1	147	0	0	0	0	0	0
EKPOSCHPR	0	11082	0	0	0	0.011009	0.111052	0.001055
EKPOSCHPR	1	147	0	0	0	0	0	0
EKPOZCOMM_ZONE_RPRNT	0	11082	0	0	0	13.82981	11.44615	0.10873
EKPOZCOMM_ZONE_RPRNT	1	147	0	0	0	13.26531	7.69658	0.634803

Independent Group Statistics T-test

Test Column	Variance Assumption	t	df	p-value (2-tailed)	Mean Difference	Standard Error Difference	Confidence Interval Probability	Confidence Interval (Lower Bound)	Confidence Interval (Upper Bound)
EKKOEKGRP	Equal variances assumed	0.720458	11227	0.471258	0.003519	0.004885	0.95	-0.00606	0.013094
EKKOEKGRP	Equal variances not assumed	6.255734	11081	4.1E-10	0.003519	0.000563	0.95	0.002417	0.004622
EKKOFRGKE	Equal variances assumed	-8.20922	11227	2.47E-16	-0.83134	0.101269	0.95	-1.02984	-0.63283
EKKOFRGKE	Equal variances not assumed	-6.17966	148.1677	5.91E-09	-0.83134	0.134528	0.95	-1.09718	-0.5655
EKESEINDT	Equal variances assumed	-13.414	11227	1.03E-40	-3.12447	0.232926	0.95	-3.58105	-2.6679
EKESEINDT	Equal variances not assumed	-7.54605	147.1819	4.31E-12	-3.12447	0.414055	0.95	-3.94273	-2.30621
EKESERDAT	Equal variances assumed	0.115168	11227	0.908314	9.02E-05	0.000784	0.95	-0.00145	0.001626
EKESERDAT	Equal variances not assumed	1	11081	0.317332	9.02E-05	9.02E-05	0.95	-8.7E-05	0.000267
EKESMENGE	Equal variances assumed	0.693774	11227	0.487838	0.012001	0.017299	0.95	-0.02191	0.04591
EKESMENGE	Equal variances not assumed	6.024036	11081	1.76E-09	0.012001	0.001992	0.95	0.008096	0.015907
EKESUZEIT	Equal variances assumed	0.115168	11227	0.908314	9.02E-05	0.000784	0.95	-0.00145	0.001626
EKESUZEIT	Equal variances not assumed	1	11081	0.317332	9.02E-05	9.02E-05	0.95	-8.7E-05	0.000267
EKETEINDT	Equal variances assumed	-12.7987	11227	3.03E-37	-2.02442	0.158173	0.95	-2.33447	-1.71437
EKETEINDT	Equal variances not assumed	-10.8485	148.7707	1.45E-20	-2.02442	0.186608	0.95	-2.39316	-1.65567
EKETMENGE	Equal variances assumed	-11.6013	11227	6.09E-31	-0.13856	0.011944	0.95	-0.16197	-0.11515
EKETMENGE	Equal variances not assumed	-3.18024	146.2403	0.001797	-0.13856	0.043569	0.95	-0.22467	-0.05245
EKPOAEDAT	Equal variances assumed	-1.26629	11227	0.205435	-1.16627	0.921014	0.95	-2.97162	0.639076
EKPOAEDAT	Equal variances not assumed	-1.90739	155.1443	0.058318	-1.16627	0.611449	0.95	-2.37411	0.041566
EKPOBSTAE	Equal variances assumed	3.295545	11227	0.000985	0.074681	0.022661	0.95	0.030261	0.119101
EKPOBSTAE	Equal variances not assumed	10.25045	192.0271	6.07E-20	0.074681	0.007286	0.95	0.060311	0.089051
EKPOELIKZ	Equal variances assumed	0.115168	11227	0.908314	9.02E-05	0.000784	0.95	-0.00145	0.001626
EKPOELIKZ	Equal variances not assumed	1	11081	0.317332	9.02E-05	9.02E-05	0.95	-8.7E-05	0.000267
EKPOEREKZ	Equal variances assumed	0.199494	11227	0.84188	0.000271	0.001357	0.95	-0.00239	0.002931
EKPOEREKZ	Equal variances not assumed	1.732207	11081	0.083265	0.000271	0.000156	0.95	-3.6E-05	0.000577
EKPOLOEKZ	Equal variances assumed	3.461595	11227	0.000539	0.082476	0.023826	0.95	0.035773	0.129179
EKPOLOEKZ	Equal variances not assumed	30.057	11081	7.1E-191	0.082476	0.002744	0.95	0.077097	0.087855
EKPOMENGE	Equal variances assumed	-19.9879	11227	2.3E-87	-0.14569	0.007289	0.95	-0.15998	-0.1314
EKPOMENGE	Equal variances not assumed	-3.3449	146.0575	0.001046	-0.14569	0.043556	0.95	-0.23177	-0.05961
EKPONETPR	Equal variances assumed	-10.2088	11227	2.31E-24	-0.45836	0.044898	0.95	-0.54637	-0.37035
EKPONETPR	Equal variances not assumed	-5.60227	147.1219	1.01E-07	-0.45836	0.081816	0.95	-0.62004	-0.29667
EKPOPLIFZ	Equal variances assumed	0.304787	11227	0.760534	0.000632	0.002072	0.95	-0.00343	0.004694
EKPOPLIFZ	Equal variances not assumed	2.646468	11081	0.008145	0.000632	0.000239	0.95	0.000164	0.0011
EKPOSCHPR	Equal variances assumed	1.201863	11227	0.229442	0.011009	0.00916	0.95	-0.00695	0.028964
EKPOSCHPR	Equal variances not assumed	10.43577	11081	2.23E-25	0.011009	0.001055	0.95	0.008941	0.013077
EKPOZCOMM_ZONE_RPRNT	Equal variances assumed	0.596157	11227	0.551082	0.564508	0.946912	0.95	-1.2916	2.420621
EKPOZCOMM_ZONE_RPRNT	Equal variances not assumed	0.8765	154.6904	0.382117	0.564508	0.644048	0.95	-0.70776	1.836772

D Prediction Model Parameter Settings KNIME

Decision Tree Learner

Tab	**Setting**	**Set to**	**Comment**
Option	Class Columns	MN_occurrence	Target, nominal attribute.
	Quality Measure	Gini Index	Used to calculate the split. Gini is common and easy when target is categorical.
	Pruning Method	Minimal Description Length (MDL)	Select pruning method. Pruning is used to reduces tree size, overfitting, and increases generalization performance. This helps to improve prediction quality.
	Reduce Error Pruning	Yes	Default setting, Yes.
	Min number of records per node	50	Set stopping criteria for model. Selected because it results in a good performance, but tree is still readable.
	Average split point	Yes	Default setting, Yes.
	Number threads	8	Default setting, 8 (available processors).
	Skip columns without domain information	Yes	Default setting, Yes.
	Force root split column	No	Default setting, No.
	Binary nominal splits	No	Default setting, No. Tested with Yes, no significant output changes.

Logistic Regression Learner

Tab	Setting	Set to	Comment
Settings	Target Column	MN_occurrence	Target, nominal attribute.
	Reference Category	0	Category for which the probability is obtained. Probability is not evaluated therefore not specifically set.
	Use order from targeted columns domain	No	Default setting, No. Sorting is performed lexicographically.
	Solver	Stochastic average gradient (SAG)	Implements a variant of stochastic gradient descent, it tends to converge faster than stochastic gradient descent. Also, produced better score than alternative solver.
	Feature Selection	Change Log Counts + Total Deviations + Dev. Occurrence	Specify the independent columns that should be included in the regression model. Numeric and nominal data can be included.
	Use order from column domain	No	Default setting, No. Sorting is performed lexicographically.
Advanced	Perform calculations lazily	Yes	Default setting, Yes.
	Calculate statistics for coefficients	Yes	Default setting, Yes.
	Maximal number of epochs	200	Default value, 200. Max number of epochs is not reached.
	epsilon	1.5E-5	Default value, 1.5E-5.
	Learning rate strategy	Fixed	Fixed strategy works well with the SAG solver.
	Step size	0.1	Default value, 0.1. Indicates learning rate.
	Regularization Prior	Uniform	Default setting, Uniform.
	Hold data in memory	Yes	Default setting, Yes.

PNN Learner (DDA)

Tab	Setting	Set to	Comment
Options	Missing Value	Incorp	Default setting, Incorp. Not changed because no missing values in the data.
	Shrink after commit	Yes	Default setting, Yes. Reduces conflict rules.
	Use class with max coverage	Yes	Default setting, Yes.
	Maximum no. Epochs	No	Default setting, No.
Target Columns	Include Target Column	MN_occurrence	Target prediction.
PNN	Theta Minus	0.2	Default value, 0.2. Defines the upper boundary of activation for conflicting rules.
	Theta Plus	0.4	Default value, 0.4. Defines the lower boundary of activation for non-conflicting rules.

Gradient Boosted Trees Learner

	Setting	Set to	Comment
Options	Target Column	MN_occurrence	Target, nominal or probability distributions.
	Attribute selection	Change Log Counts + Total Deviations + Dev. Occurrence	Attributes on which the model should be learned.
	Limit number of levels	Yes	Default setting, Yes.
	Tree depth	10	Default values, 4-10. Set to 10 because it performs better than other default values.
	Number of models (boosting)	100	Default value, 100. Higher values are used for smaller datasets with few target values but risk overfitting if set too high.
	Learning rate	0.1	Default value, 0.1.
Advanced Options	Use midpoint splits	Yes	Default setting, Yes.
	Use binary splits for nominal columns	Yes	Default setting, Yes.
	Missing value handling	XGBoost	Default setting, XGBoost. No missing values in data, therefore not inspected.
	Data Sampling Rows (bagging)	No	Function is not used because input data is already bootstrapped.
	Attribute sampling (bagging)	Sample (square root)	Default value, Square Root. The square root of the total number of attributes is used as sample size. Most commonly used.
	Attribute selection	Use same set of attributes for entire tree	Set value is not relevant for Gradient Boosted Trees and does not affect output.

Random Forest Learner

Tabs	Setting	Set to	Comment
Options	Target Column	MN_occurrence	Target
	Attribute selection	Change Log Counts + Total Deviations + Dev. Occurrence	Attributes on which the model should be learned.
	Split Criterion	Gini Index	Default Gini, also used in forest.
	Limit number of levels (tree depth)	No	Default No. Depends on computing time.
	Minimum node size	1	Default 1, perfect fit. Identical rows could cause a problem
	Number of models	100	Default between 100-500, should result from parameter tuning

RProp MLP Learner

Tabs	Setting	Set to	Comment
Options	Max number of iterations	100	Number of learning iterations
	Number of hidden layers	3	Provided best result with number of neurons.
	Number of hidden neurons per layer	5	Provided best result with number of layers.
	Class column	MN_occurrence	Target, nominal or numerical values.
	Ignore missing value	No	Default setting, No. There are no missing values in the input data.