

Production Planning & Control

The Management of Operations

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/tppc20>

Value proposition of predictive discarding in semiconductor manufacturing

Geert van Kollenburg, Mike Holenderski & Nirvana Meratnia

To cite this article: Geert van Kollenburg, Mike Holenderski & Nirvana Meratnia (26 Jul 2022): Value proposition of predictive discarding in semiconductor manufacturing, Production Planning & Control, DOI: [10.1080/09537287.2022.2103471](https://doi.org/10.1080/09537287.2022.2103471)

To link to this article: <https://doi.org/10.1080/09537287.2022.2103471>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 26 Jul 2022.



Submit your article to this journal [↗](#)



Article views: 894



View related articles [↗](#)



View Crossmark data [↗](#)

Value proposition of predictive discarding in semiconductor manufacturing

Geert van Kollenburg , Mike Holenderski  and Nirvana Meratnia 

Interconnected Resource-aware Intelligent Systems, Eindhoven University of Technology, Eindhoven, the Netherlands

ABSTRACT

Optimizing semiconductor manufacturing processes is needed to solve the current shortage of computer chips. Discarding unfinished chips based on data-driven predictions models can significantly reduce time and resources otherwise spent on finishing faulty chips. The current paper presents the value proposition of predictive discarding at different stages in the manufacturing process, by combining model performance metrics with costs and benefits related to false and correct discards. While applied to the chip manufacturing process in this paper, predictive discarding is a generic methodology to minimize wasted resources by predicting product quality from process data. Through sensitivity analysis, we show that even with weak predictors, predictive discarding can still be beneficial, from both economic and sustainability perspectives. The proposed method is illustrated by analysing an empirical benchmark data set from the semiconductor manufacturing domain.

ARTICLE HISTORY

Received 11 March 2022
Accepted 12 July 2022

KEYWORDS

Prescriptive analytics;
predictive discarding;
machine learning;
semiconductor manufacturing;
Industry 5.0

1. Introduction

Computer chips are essential to our digital society. Currently there is a big shortage of these semiconductor chips. To solve the shortage, significant investments, like the ones made through the European Chips Act (The European Commission 2022), are needed. Next to such investments, manufacturing processes need to be optimized. The manufacturing of chips is done in batches on so-called wafers and comprises hundreds of steps. If too many chips on a wafer are faulty, all chips from that wafer will be discarded. Because chips can take up to three months to produce (Gupta et al. 2006), it is paramount that the wafers are of sufficient quality so that no resources (i.e. raw materials, time, energy) are wasted on manufacturing faulty wafers. Process monitoring and quality control are therefore crucial to avoid wasting resources on producing faulty wafers (May and Spanos 2006).

Throughout the wafer manufacturing process, systems are in place to measure, monitor and control the production process (den Boef 2016). Quality control systems are in place to ensure that faults affecting end-product quality are detected (Qin et al. 2006). While small faults may be directly reworked, severe faults may result in (unfinished) wafers being discarded (di Bella et al. 2019). Many faults can be detected by human experts, and measurements related to such faults can be hard coded into halting the production (May and Spanos 2006). However, combinations of smaller (within-spec) imperfections may also affect the end-product quality in unforeseeable ways (Melhem et al. 2015). It is infeasible for human operators to program all possible faults, process all available

data in real-time or identify complicated combinations of imperfections that lead to such faults.

Data-driven machine learning approaches may therefore be well-suited to predict end-product quality from process data collected during production (Yang and Lee 2012; Chou, Wu, and Chen 2010). If a prediction model indicates that the wafer does not, or will not, reach the required end-product quality, the wafer can be discarded on time. If this is done for unfinished products, where some or most of the measurements are not yet available, we call this methodology 'predictive discarding' and showed its potential for improved quality control in our previous work (van Kollenburg et al. 2022). In that work, classification models were trained to predict the outcome of an electrical test from misalignment measurements of a single layer on a wafer. There we showed how data of an unfinished wafer, collected during one of the many steps in the production process, could be used to predict the outcome of a critical electrical test which is done on finished wafers. The specific relations that were found between misalignments and end-product quality were not previously known and could help the decision-making process whether to continue or abort the manufacturing process. While it was shown that predictive discarding had the potential to reduce the consumed resources which would otherwise be spent on finishing faulty wafers, the work did not systematically study the conditions under which predictive discarding would be beneficial.

The current paper improves upon the predictive discarding framework by (i) proposing a methodology for calculating the benefits of predictive discarding, (ii) presenting a sensitivity analysis to determine the minimal conditions, in

terms of model performance and process parameters, under which predictive discarding may be beneficial and (iii) illustrating the benefits of predictive discarding using a publicly available empirical data set of a wafer manufacturing process. The rest of the paper is structured as follows. In Section 2, the basic methods to determine the benefits of predictive discarding are discussed. In Section 3, a sensitivity analysis is presented that studies the conditions under which predictive discarding are beneficial. Section 4 presents the empirical data analysis of the SECOM data set. In Section 5, all results are discussed and the paper ends with some conclusions in Section 6.

2. Calculating benefits of predictive discarding

The amount of resources (i.e. raw materials, energy consumption and time) that can be saved by predictive discarding depends on the production stage at which the discarding decision is made. The sooner a discard takes place, the more resources are saved that would otherwise be spent on manufacturing a faulty wafer. Saving resources can also lead to economic savings for wafer manufacturers.

In economic terms, the cost of resources already spent up until a certain point are called 'sunk cost' (Mankiw 2004). Sunk costs should not influence a decision-making process (such as whether to discard a wafer). The cost of resources required to finish a product are called 'avoidable costs' (Garrison et al. 2003). These avoidable costs are saved if a correct decision is made and are therefore crucial to the calculation of the benefits offered by predictive discarding. For the sake of generality and ease of calculation, we set the marginal costs of the resources required to produce a single wafer to 1 and denote it as C_{prod} . The relation between (marginal) avoidable costs C_{avoid} and (marginal) sunk costs C_{sunk} can then be defined as:

$$C_{\text{avoid}} = C_{\text{prod}} - C_{\text{sunk}} = 1 - C_{\text{sunk}} \quad (1)$$

For example, if 25% of costs have already been incurred before the decision to discard the unfinished wafer is made, $C_{\text{sunk}} = .25$ and $C_{\text{avoid}} = .75$.

When an unfinished wafer is incorrectly discarded, the avoidable costs will be saved, but the wafer will not obtain its end-product value. This will lead to lost sales, and hence, lost revenue. The loss related to a false discard can be quantified as:

$$\text{Loss} = \text{Value} - C_{\text{avoid}}, \quad (2)$$

where

$$\text{Value} = C_{\text{prod}} + \text{Profit margin} = 1 + \text{Profit margin} \quad (3)$$

The value of a wafer is assumed to be fixed with respect to the production cost. If a product has a marginal profit margin of, say 6%, then the value of the end-product is defined to be 1.06. If $C_{\text{avoid}} = .5$ (i.e. half of the production costs can still be avoided if discarded at that moment), then the loss related to a single wrong discard will be $\text{Value} - C_{\text{avoid}} = 1.06 - .5 = .56$. A profit margin 0 in the calculations ensures that the decision to implement predictive discarding is

based on resource consumption, and not on lost sales. Some intermediate products, especially in prefab, can have negative profit margins as individual components may require an investment of resources, making predictive discarding even more valuable.

The gains and losses explained above hold for each unfinished wafer. The benefits of predictive discarding over all produced wafers can be computed knowing (with slight abuse of notation) (i) the proportion of faulty wafers, denoted by $P(\text{faulty})$, (ii) the probability that the predictive model correctly rejects those faulty wafers, denoted by $P(\text{discard}|\text{faulty})$ (also known as recall) and (iii) the probability that a good wafer is wrongly discarded, denoted by $P(\text{discard}|\text{not faulty})$ (also known as false positive rate, FPR). Figure 1 illustrates how the probabilities, gains, and losses relate to four possible scenarios when deciding on any item. Both the recall ($P(\text{discard}|\text{faulty})$) and FPR ($P(\text{discard}|\text{not faulty})$) are model specific and can only be determined when a prediction model is fitted to data. If $P(\text{faulty})$ is not known, it can be approximated by the proportion of faulty wafers in the available data. Note that the value proposition described in this section also applies to decisions made based on other types of models, like first-principles models (which are not trained on data) if they are available.

The benefits of implementing predictive discarding in a manufacturing process can then be calculated as:

$$\text{Benefits} = P(\text{faulty}) * P(\text{discard}|\text{faulty}) * C_{\text{avoid}} \quad (4)$$

$$- P(\text{not faulty}) * P(\text{discard}|\text{not faulty}) * (\text{Value} - C_{\text{avoid}}).$$

The benefits are calculated as a reduction in resources required to finish the same number of good/non-faulty wafers, compared to a state in which all wafers are finished. If the profit margin is 0, a benefit of .05 indicates that 5% of the total production costs can be saved compared to running the entire process without predictive discarding. The interpretation will be less straightforward if the profit margin is not 0.

Exact production costs and profit margins are often confidential information of manufacturers. It is therefore difficult to determine directly under which conditions predictive discarding is beneficial. For example, false positives are more detrimental if profit margins are high. Higher profit margins lead to more lost sales than low profit margin. On the other hand, if avoidable costs are high, false negatives may be more costly as resources are spent on faulty products. Some processes may benefit from early discards even if there are false positives because valuable resources (raw materials and time) are not spent on manufacturing faulty products. In any case, it is important to evaluate how each variable in Equation (4) affects the benefits of predictive discarding. If the relevant process information is known, it is possible to determine what performance is required from a prediction model. If the information is not available, a sensitivity analysis may help to investigate the conditions under which predictive discarding may be beneficial.

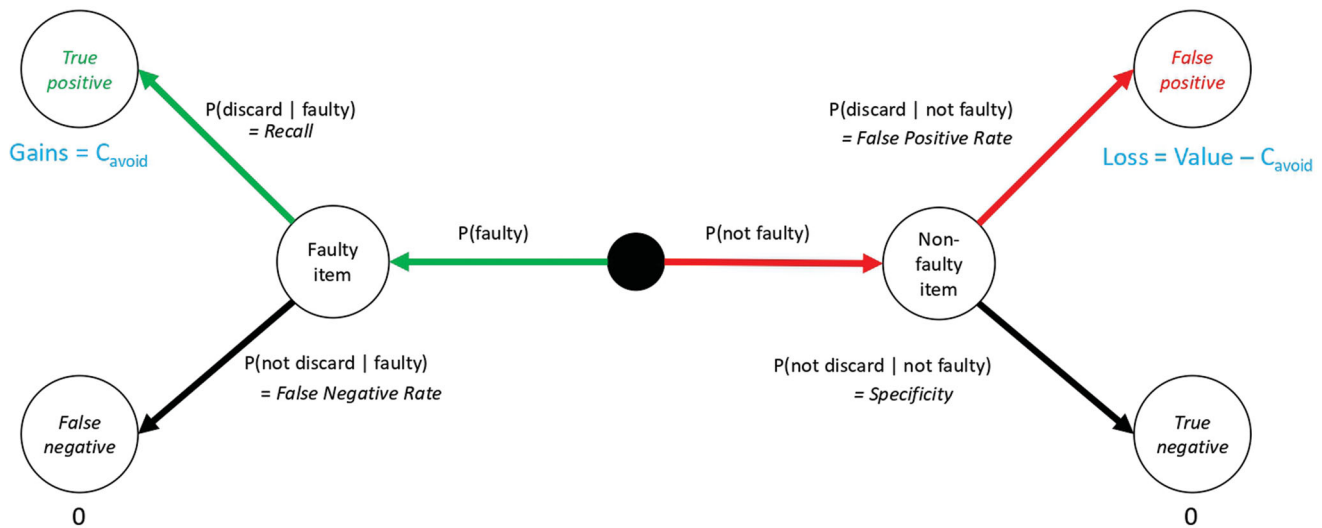


Figure 1. Schematic representation of the possible scenarios for predictive discarding, with the respective gains and losses provided under each outcome.

3. Sensitivity analysis

There are five variables in Equation (4), three of which relate to the process (and may be confidential) and two that relate to the prediction model performance. We performed sensitivity analyses by fixing three variables, changing the value of the remaining two variables, and calculating the outcome of Equation (4) for the chosen values. This allows a detailed investigation into the combinations of prediction model performance and process specifics for which predictive discarding may be beneficial. The sensitivity analyses were performed by:

- fixing profit margin and process parameters $P(\text{faulty})$ and C_{avoid} , and varying recall and FPR
- fixing model recall, $P(\text{faulty})$ and C_{avoid} , and varying FPR and profit margin

The Python code for this sensitivity analysis is available as supplemental material and can be used to explore other conditions than those presented below. Manufacturers can use known process parameters to calculate the model performance required to achieve benefits of predictive discarding.

3.1. Recall versus FPR

Our first sensitivity analysis is meant to evaluate how Recall and FPR relate to benefits of predictive discarding. This was studied for several combinations of process parameters, namely proportions of faulty wafers ($P(\text{faulty})$) of either .2 or .05 (i.e. 20% or 5% faulty wafers, respectively), and avoidable costs (C_{avoid}) of either .75 or .5 (indicating implementation at a quarter or halfway through the production, respectively). These values were chosen somewhat arbitrarily. The profit margin is fixed to 0% (to avoid including lost sales). The presented benefits are therefore upper bounds of the benefits,

calculated at the break-even point where the value of the finished wafer is exactly equal to its marginal production costs.

For the $2 \times 2 = 4$ combinations of process values, the benefits of predictive discarding are calculated from Equation (4) for several recall and FPR values. The results are shown in Table 1, in which each sub-table corresponds to one of the four combinations of process values. The results are based on a break-even profit margin of 0%. Sub-table (a) provides the condition with the lowest expected benefits and going to a sub-table the right or down is related to more possible benefits of predictive discarding (given the process parameters).

Within each sub-table, the top-left element is the worst-case condition evaluated in the study, where a prediction model correctly identifies 25% of all faulty wafers (Recall = .25) but discards 50% of the non-faulty wafers (FPR = .50). Each element to either the right or down is related to better model performance. The bottom-right element in each sub-table relate to a perfect classification model. The value of each element shows the benefits of predictive discarding under that respective condition, as calculated from Equation (4). Positive benefits indicate that resources can be saved by predictive discarding, negative values indicate that more resources are needed with predictive discarding (and should therefore not be implemented).

If predictive discarding is implemented when $C_{\text{avoid}} = .5$ and $P(\text{faulty}) = .05$ (Table 1(a)), predictive discarding is only beneficial if there are no false positives. An increase in $P(\text{faulty})$ to .2 (Table 1(b)) increases the benefits more than an increase in C_{avoid} to .75 (Table 1(c)). When the values of both parameters increase (Table 1(d)), there is a positive multiplicative (i.e. interaction) effect on the benefits. In the latter combination of process values, predictive discarding may even be beneficial with a relatively weak prediction model having a recall of 55% and FPR of 30%.

Table 1. Results of the recall versus FPR sensitivity analysis into the benefits of predictive discarding for four combinations of process parameters.

	Recall							Recall					
	25%	40%	55%	70%	85%	100%		25%	40%	55%	70%	85%	100%
(a) $P(\text{faulty}) = .05, C_{\text{avoid}} = .50$							(b) $P(\text{faulty}) = .20, C_{\text{avoid}} = .50$						
FPR							FPR						
50%	-.23	-.23	-.22	-.22	-.22	-.21	50%	-.18	-.16	-.15	-.13	-.12	-.10
40%	-.18	-.18	-.18	-.17	-.17	-.16	40%	-.14	-.12	-.11	-.09	-.08	-.06
30%	-.14	-.13	-.13	-.12	-.12	-.12	30%	-.10	-.08	-.06	-.05	-.03	-.02
20%	-.09	-.08	-.08	-.08	-.07	-.07	20%	-.05	-.04	-.02	-.01	.01	.02
10%	-.04	-.04	-.03	-.03	-.03	-.02	10%	-.01	.00	.02	.03	.05	.06
0%	.01	.01	.01	.02	.02	.02	0%	.02	.04	.06	.07	.08	.10
(c) $P(\text{faulty}) = .05, C_{\text{avoid}} = .75$							(d) $P(\text{faulty}) = .20, C_{\text{avoid}} = .75$						
50%	-.11	-.10	-.10	-.09	-.09	-.08	50%	-.06	-.04	-.02	.00	.03	.05
40%	-.09	-.08	-.07	-.07	-.06	-.06	40%	-.04	-.02	.00	.02	.05	.07
30%	-.06	-.06	-.05	-.04	-.04	-.03	30%	-.02	.00	.02	.04	.07	.09
20%	-.04	-.03	-.03	-.02	-.02	-.01	20%	.00	.02	.04	.06	.09	.11
10%	-.01	-.01	.00	.00	.01	.01	10%	.02	.04	.06	.08	.11	.13
0%	.01	.02	.02	.03	.03	.04	0%	.04	.06	.08	.10	.13	.15

The profit margin was set to 0 in these calculations, such that the values in the cells indicate the proportion of the total resources that can be saved (positive benefits) or are required (negative benefits) to produce the same number of good quality wafers compared to a process without predictive discarding.

Table 2. Results of the FPR versus profit margin sensitivity analysis into the benefits of predictive discarding for four combinations of process parameters.

	Margin							Margin					
	25%	20%	15%	10%	5%	0%		25%	20%	15%	10%	5%	0%
(a) $P(\text{faulty}) = .05, C_{\text{avoid}} = .75$							(b) $P(\text{faulty}) = .20, C_{\text{avoid}} = .75$						
FPR							FPR						
25%	-.10	-.09	-.08	-.06	-.05	-.04	25%	-.02	-.01	.00	.00	.02	.03
20%	-.08	-.07	-.06	-.05	-.04	-.03	20%	-.01	.00	.01	.02	.03	.04
15%	-.05	-.05	-.04	-.03	-.02	-.02	15%	.02	.02	.03	.03	.04	.05
10%	-.03	-.02	-.02	-.01	-.01	.00	10%	.04	.04	.04	.05	.05	.06
5%	.00	.00	.00	.00	.00	.01	5%	.06	.06	.06	.06	.06	.07
0%	.02	.02	.02	.02	.02	.02	0%	.08	.08	.08	.08	.08	.08
(c) $P(\text{faulty}) = .05, C_{\text{avoid}} = .90$							(d) $P(\text{faulty}) = .20, C_{\text{avoid}} = .90$						
25%	-.06	-.05	-.04	-.03	-.01	.00	25%	.02	.03	.04	.05	.06	.07
20%	-.04	-.03	-.02	-.02	-.01	.00	20%	.03	.04	.05	.06	.07	.07
15%	-.03	-.02	-.01	-.01	.00	.01	15%	.05	.05	.06	.07	.07	.08
10%	-.01	-.01	.00	.00	.01	.01	10%	.06	.07	.07	.07	.08	.08
5%	.01	.01	.01	.01	.02	.02	5%	.08	.08	.08	.08	.08	.09
0%	.02	.02	.02	.02	.02	.02	0%	.09	.09	.09	.09	.09	.09

Model recall was set to .5. Positive values indicate profits, negative values indicate losses with respect to producing the same number of good quality wafers compared to a process without predictive discarding.

3.2. FPR versus profit margin

The second sensitivity analysis evaluates the effects of FPR and profit margins, given that a prediction model can correctly identify 50% of the faulty wafers (i.e. recall = .5). In such a situation, what FPR can be allowed before the benefits of predictive discarding disappear? And since the FPR leads to lost sales: for which profit margins does predictive discarding compensate for the missed revenue due to lost sales?

Again, $2 \times 2 = 4$ conditions were studied: the proportion of faulty wafers ($P(\text{faulty})$) was either .05 or .20, and the avoidable costs (C_{avoid}) were either .75 or .90. The values were chosen somewhat arbitrarily. The benefits under these conditions are shown in Table 2 for several FPRs (between 0 and 25%) and profit margins (between 25 and 0%). Within each sub-table, the top-left element is the worst-case condition where a profit margin of 25% is missed for every wrongly discarded wafer, and 25% of all good wafers are wrongly discarded (FPR = .25). Each element to the right is

related to a lower profit margin, each element down relates to better model performance.

If predictive discarding is implemented based on a prediction model with recall of .50, when $C_{\text{avoid}} = .75$ and $P(\text{faulty}) = .05$ (Table 2(a)), predictive discarding is only beneficial if there are no false positives. If a model reaches a recall of 50% earlier in the process, with $C_{\text{avoid}} = .90$ (Table 2(c)), a FPR of 10% can still lead to benefits if the profit margin is between 0 and 5%. The most benefits can be obtained if $P(\text{faulty})$ is higher (Table 2(b,d)). Even with a recall of 50%, the benefits of predictive discarding can have values of up to 9%, with respect to the total production costs.

4. Empirical application

The SECOM data set (McCann et al. 2010) consists of 591 process variables, collected during the manufacturing of 1567 wafers. The original data description states that the process variables are presented in the chronological order in

which they are measured during the wafer production. Each wafer has a label 'Pass' or 'Fail' based on a quality test. In total, 104 (approximately 7%) of the wafers in the data set failed the test.

Using a design-of-experiments setup, (Moldovan et al. 2017) studied whether the test results could be predicted from the process data using various combinations of preprocessing steps, feature selection algorithms, and several classification models. The best performing models, according to the authors, had recall values of over 90%, but at the same time had FPRs of 70% or higher. The results presented in that paper indicate that the available data may not be strongly causally related to the end-product quality. Discarding wafers based on such predictions will not be beneficial, especially because no resources could be saved for correctly identifying faulty wafers after the manufacturing process is finished and all data is available.

According to the original data description (McCann et al. 2010), there are fixed moments during manufacturing where decisions are made about whether to abort the production. Such decisions are made only based on subsets of the data that is available at that moment. Predictive discarding works in the same way. However, since the exact moments are not defined in the data set, we arbitrarily divided the 591 process variables into 19 blocks of 31 variables. This simulates that data was collected for 19 sub-steps of the manufacturing process and will ensure that prediction models can only use subsets of data (details follow in Section 4.1). By doing so, variables 1 through 31 are related to 'Step 1' of the manufacturing process, variable 32 through 62 are related to 'Step 2' and so on.

To the best of our knowledge, the current paper is the first to predict the outcome variable from strict subsets of the data. More importantly, this also allowed us to estimate the benefits of predictive discarding when implemented at different stages of the manufacturing process. It may be expected that models can make better predictions when more variables become available. At the same time the avoidable costs decrease after each step, lowering the possible gains for correct discards at each subsequent step. Yet, even without exact knowledge about the production costs, prediction model recall and FPR can be related to gains and losses to determine the value for a manufacturer to implement predictive discarding.

4.1. Data analysis

The SECOM data set has a substantial amount of missing data and required preprocessing before it could be analysed. First, all variables which had more than 15% of the observation missing were removed from the data. Variables which had only one unique value (i.e. that were constant) were removed. Also, variables with two unique values were removed because these in general had limited information, low variance, and/or provided single unique values for subsets of the data. Missing data in the remaining variables were imputed with sklearn's KNNImputer (Troyanskaya et al. 2001) using the default settings and three neighbours. Note

that extreme values were not removed because these values specifically may be (causally) related to whether a wafer fails the quality test. All data handling and analyses described below were performed in Python 3 (Van Rossum and Drake 2009) using Jupyter Notebooks (Kluyver et al. 2016). The pre-processed data and the Python code for preprocessing the original data set is available as supplemental material.

We use six readily available prediction models to classify the 'Pass/Fail' outcome, namely: Linear Discriminant Analysis (LDA) and Quadratic Discriminant analysis (QDA) (Hastie et al. 2009), K-nearest neighbours (KNN) (Cover and Hart 1967), Decision Tree (DT) (Breiman et al. 1984), Gradient Boosting Decision Tree (GB) (Friedman 2002), and Random Forest (RF) (Breiman 2001). LDA finds the best linear boundary in the parameters space that separates the observed groups. QDA finds the best quadratic boundary in the parameters space to separate the groups. KNN groups similar observations and classifies each new observation into the group it is most similar to. DT finds simple rules which can be used to predict an outcome variable from using the observed predictors in a sequential manner. GB is a special case of DT, where only regressions are used as rules for the tree, with a stochastic optimization procedure. RF creates multiple DTs based on sub-samples of the predictors and makes a final classification based on a majority-vote.

Each model was trained on every subset of the data. The manufacturing process was artificially split into $M = 19$ steps. At Step 1 only the process variables related to Step 1 were used. At Step 2, data relating to Steps 1 and 2 were used, and so on, until at Step 19 all available process variables were used. This simulates the empirical situation where data is only available for the finished parts of the manufacturing process. Then, at each stage, the corresponding data subset was split into training (70% of the data) and test set (30%) using sklearn's stratified shuffle split (Ojala and Garriga 2010) to ensure the training and test set had the same Pass/Fail ratio ($P(\text{faulty}) = 104/1567 \approx .066$ in the complete data).

At every step, the available variables were used as inputs in the prediction models in three ways. Firstly, the variables were scaled to unit variance and were all used as inputs. Secondly, variable selection was performed to select 25% of variables (rounded up to the nearest integer) which had the highest bivariate relation with the Pass/Fail outcome, assessed using sklearn's SelectKbest function (Ferri et al. 1994). Only those 25% were used as model inputs. Thirdly, principal components were computed that together contained 90% of the variance in the available process variables and these components were then used as model inputs.

Since there were 19 steps and the available data was used in 3 ways at every step, each model had to be trained 57 times. To find the best model parameters under each condition, a grid search cross-validation was performed on the training data to optimize the recall. A full description of the parameter spaces used for each model can be found in Appendix A. Since this is an illustrative example, the grid search was rather confined. The interested reader can use the available code to analyse the data in detail further or

Table 3. Benefits of implementing predictive discarding within the SECOM process. Only conditions which resulted in positive benefits are reported.

Stage	$\$C_{\text{avoid}}\$$	Model	Input	# predictors	Recall	FPR	Benefit at 0 profit	Max profit margin
1	0.95	LDA	All	17	0.29	0.28	0.60%	2%
1	0.95	KNN	All	17	0.06	0.28	0.20%	2%
1	0.95	GB	All	17	0.1	0.06	0.40%	6%
1	0.95	KNN	Select	5	0.1	0.06	0.40%	6%
2	0.89	GB	Select	10	0.1	0.02	0.40%	15%
3	0.84	KNN	All	52	0.13	0.05	0.10%	2%
3	0.84	KNN	PCA	27	0.13	0.05	0.10%	2%
4	0.79	KNN	All	67	0.26	0.05	0.40%	9%
4	0.79	KNN	PCA	32	0.26	0.05	0.50%	9%
5	0.74	KNN	PCA	46	0.23	0.04	0.10%	1%

The column 'Benefits at 0 profit' provide the (estimated) percentage of total resource consumption that could be saved when implementing predictive discarding. Note that the number of variables for input 'All' is less than Stage*31 due to variable removal in preprocessing.

specify other optimization objectives. The data-analysis pipeline is summarized in Algorithm 1.

Algorithm 1 Algorithmic description of the performed data analysis

```

1: Remove variables with many missing values
2: Impute remaining missing values
3: procedure Predictive discarding
4:   for  $m$  in  $1, \dots, M$  do
5:     Collect all variables available at Step  $m$ 
6:     Split into training and test
7:     for Inputs in ['all', 'select', 'pca'] do
8:       Find model parameters through (grid search) CV on training data
9:       Determine model performance in the test set.
10:      for each profit margin do
11:        Calculate the benefits for the test set using Equation 4
12:      end for
13:    end for
14:  end for
15: end procedure

```

In the first presentation of the SECOM dataset, McCann et al. (2010) already showed that the (causal) relation between available process data and test outcomes was limited.

Here, subsets of the already weak set of predictors are used. Model performance (recall and FPR) was therefore expected to be below what is generally deemed 'good' for classification models (Hastie et al. 2009). But since the value proposition of predictive discarding includes the losses and gains related to the classification, based on the sensitivity analyses done before, we still expected to find overall benefits in some conditions, especially those where the avoidable costs are high.

4.2. Results

The simulated implementations of predictive discarding which resulted in positive benefits are reported in Table 3. The first five columns indicate the sub-step after which the prediction is made, the respective avoidable costs, the classification model that enabled the benefits, how the variables were used as model inputs (All available, feature selection or based on PCA) and the number of predictors in the model, respectively. Columns six and seven provide the recall and FPR of each model. Column eight gives the benefits of predictive discarding when calculated with 0 profit margin (i.e. the break-even) and the last column provides the maximum

profit margin for which predictive discarding would still be beneficial, given the recall, FPR and C_{avoid} .

The results in Table 3 show that predictive discarding can be beneficial even when models have a low recall. In most classification problems, a recall of .10 would be deemed insufficient. In the current context, however, any correct identification of a faulty wafer may lead to significant benefits compared to current practice of finishing each within-spec wafer. The column 'Benefits at break-even' (i.e. with a profit margin of 0) indicates the percentage of the total resource consumption that can be saved when implementing predictive discarding, given the model performances and avoidable costs given in the other columns. The observed benefits ranged from 0.1% to 0.6%. While these percentages may be small, they relate to a large amount of resources in absolute term. If we incorporate also the revenue missed through lost sales, we can evaluate the limits for which predictive discarding is still economically viable. The maximum profit margin for which this is the case can be found in the last column of Table 3. The maximum profit margin ranged from 1% to 15%.

4.3. Monte Carlo validation

To ensure the results in Table 3 were not just chance findings based on a specific training/test split, we performed a Monte Carlo study to validate the results. For process steps 1 through 5, the analysis as described in Algorithm 1 was replicated 100 times, each time on different training/test splits. To limit the computational burden, only KNN and GB were considered. KNN was chosen because it most often resulted in positive benefits and GB because it allowed for the highest profit margin (cf. Table 3). The KNN and GB models were retrained, and the benefits were calculated for each of the 100 test sets. If a model showed positive benefits in significantly more than half of the test sets (as tested with a z-test for proportion), this was an indication that the results were not merely chance findings.

KNN resulted in positive benefits with any type of input after the first stage, and with feature selection after the second stage. The number of test sets resulting in positive benefits and the average benefits across the Monte Carlo replicates can be found in Table 4. Positive results of implementing predictive discarding after the first step were found

in over 80% of the test sets, irrespective of how the variables were used as inputs. The full distributions of the 100 calculated benefits are shown in Figure 2. A one-sided z-test was performed to test whether the proportion of test sets with positive benefits was larger than half (i.e. 50). For all types of inputs, the number of positive results were significantly different from what would be expected if the benefits were 0 or negative. The individual data sets in this MC study were

Table 4. Conditions where KNN leads to positive benefits through predictive discarding in more than half of the 100 Monte Carlo replications.

Stage	Cavoid	Input	# positive benefits	Mean
1	0.95	All	86**	.4%†
1	0.95	Select	82**	.3%†
1	0.95	PCA	81**	.3%†
2	0.89	Select	62*	.1%†

*Significant 95% confidence level.

**Significant at 99% confidence level.

†p value < .001.

not stored, but the results can, to a degree of randomness, be replicated using the code available as supplemental material.

Even though the majority of test sets presented in Figure 2 had positive benefits, there were still some test sets in which the benefits were negative. This observation reflects the importance of validation studies, especially if the data is unbalanced and/or the predictors are not strongly related to the outcome. Even then, it is important to check the results in more detail. For example, while the KNN results in Stage 2 using variable selection showed a statistically significant number of positive benefits, the distribution of the benefits, as shown in Figure 2, is not convincing from a substantive point of view. Lastly, the Monte Carlo simulations for GB did not show any statistically significant difference in benefits, suggesting that the results in Table 3 for the GB classifier were merely chance findings.

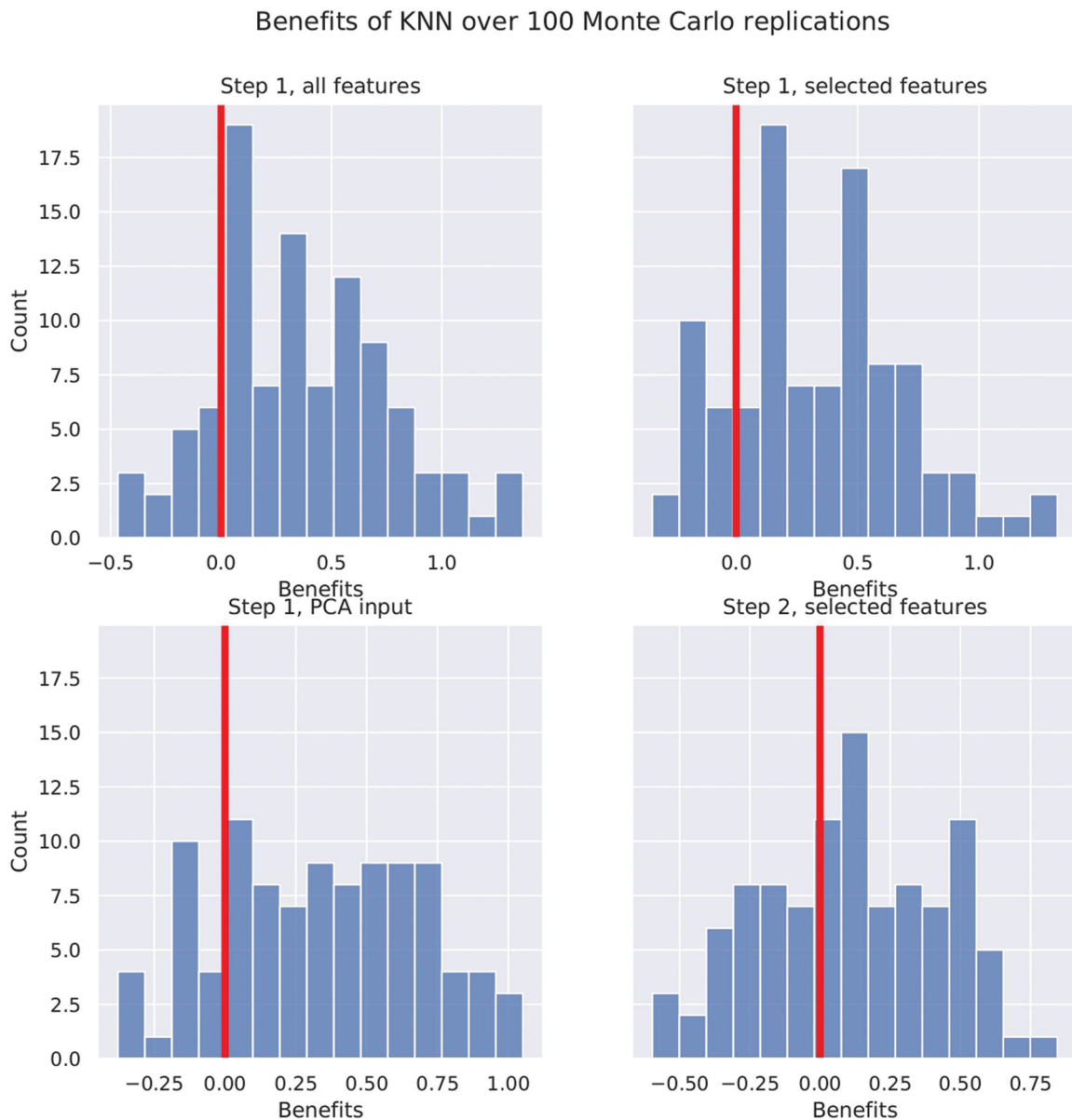


Figure 2. Distributions of the benefits of predictive discarding, calculated across 100 Monte Carlo replicates per condition. The vertical line is positioned at 0.

5. Discussion

This work illustrates how the value of predictive discarding can be determined by combining cost savings and missed revenue with the performance of data-driven prediction models. The research originated from an application in the field of wafer manufacturing, but the methodology provided in this paper is generic and can be used for any manufacturing process.

Certain simplifications were made in the illustrations to ensure explainability of the value proposition. For example, the calculations assume that the state of a wafer does not change after the moment of discarding. In practice, a wafer may become faulty by faults that occur in the time after the decision is made. This was not taken into account explicitly in the calculations of the benefits. The probability that a wafer becomes faulty later could be incorporated in the definition of the avoidable costs in future applications.

The optimal trade-off between recall and FPR depends on the profit margins and stage of production. A better approach in the future may be to train a model with weights related to the gains and losses associated with correct or incorrect classification. These weights may then be varied with respect to the stage at which predictive discarding is to be implemented. In this way, multiple sources of information can be included in the optimization of model predictions.

Furthermore, integration of all parameters that influence the benefits of predictive discarding may be challenging in practice. For practical implementation of predictive discarding, separate econometric models may be required to determine production costs, profit margins, etc. It may be, or may become, possible to forecast manufacturing costs and profit margins from external information sources like raw material availability and market fluctuations. Model optimization will then become an even more challenging exercise. The optimization of benefits of predictive discarding in such complex situations may be solved by formulating the problem as a Partially Observable Markov Decision Process (POMDP) (Egorov et al. 2017).

In this work, the value of predictive was shown under the assumption that there is complete trust in the model predictions. In practice, the benefits could be much higher if used complementary to existing expert knowledge. By using input from experts (i.e. overruling decisions), model predictions may be improved continuously (Werling et al. 2020). Vice versa, model predictions may help experts make better informed decisions. Future work on predictive discarding should evaluate the use of models that can be retrained based on (human) feedback provided on its predictions (Sjödin et al. 2021).

Predictive discarding only focuses on predictions of end-product quality. This does not directly lead to a better understanding of the manufacturing process itself and the relations between sub-processes. Research is currently being done to evaluate whether a path model called Process PLS (van Kollenburg et al. 2021) can be used to model relations between distinct parts of the semiconductor manufacturing process. Path modelling may allow for predictive process control when used to relate observations in one machine to

specific settings of another machine. For example, observing a mishap in one machine may be solved by changing the settings in a machine later in the process. Future work could focus to further integrate statistical models into existing control systems.

The environmental impact of data-driven models, like AI and machine learning may be significant. Training and optimizing readily available models can have significant carbon footprints (Dhar 2020; Strubell, Ganesh, and McCallum 2019). However, the costs for running exploratory analyses, like the ones presented in this paper, or implementing already trained prediction models, may be negligible (especially compared to the overall production footprint). The current work neither did extensive optimizations nor used complicated prediction models, which enable us to keep the carbon footprint of the analyses presented in this paper to a minimum. All analyses described in this paper had a combined carbon footprint of approximately .21 kg of CO₂, calculated from the CPU power usage (.140 kw/h), computation time (3 hours) and average footprint of electricity in the Netherlands (.505 kgCO₂/kWh). For comparison, this is equivalent to the average greenhouse gas emission of a passenger car driving 850 metres (i.e. just over half a mile).

6. Conclusions

Incorporating process information into the evaluation of predictive model performance makes predictive discarding an asset for the manufacturing industry. Feasibility studies, like the sensitivity analyses presented in this paper, require limited resources and can be performed using readily available historical data. It was shown that, even with simple classifiers and weak predictors, predictive discarding can result in a reduction in resource consumption, saving valuable raw materials, energy, and time. Future implementations of predictive discarding will give more concrete insights in which resources can be saved.

Predictive discarding not only improves the availability of good quality wafers and computer chips, which have become crucial to human liveability, but also improves the sustainability of the wafer manufacturing process itself. This also holds for other manufacturing processes. Investments in AI to complement human efforts in manufacturing processes has become a core part of the transition to Industry 5.0. As this work shows, using data-driven methods within decision-making strategies may also significantly improve production processes, both sustainably and profitably. Predictive discarding may therefore become standard practice in Industry 5.0, where AI and humans work together to achieve sustainable production.

Acknowledgements

The authors would like to express their gratitude to Patrizia Vasquez, Daniele Pagano, Francesco Io-Piano, Paola Giuffrè, Daniele Li-Rosi and Giuseppe Di Martino for their valuable input.

Author contributions

Geert van Kollenburg: conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, validation, writing – original draft. Mike Holenderski: funding acquisition, data curation, investigation, methodology, software, visualization, writing – reviewing and editing. Nirvana Meratnia: supervision, writing – reviewing and editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

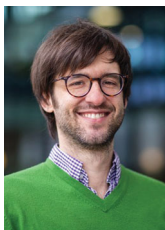
This work was supported by the ECSEL Joint Undertaking (JU) under grant agreement No 826589.

Notes on contributors

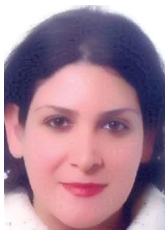


expertise in statistics and machine learning have led to novel approaches for process analysis.

Geert van Kollenburg is a researcher at Eindhoven University of Technology. He has worked together with national and international partners on data-driven solutions for sustainable industry and has done research on multivariate statistical process control, structural equation modelling, chemometrics, handheld spectroscopy, water pollution and social science applications. His focus on sustainability and



Mike Holenderski is an assistant professor at the Department of Computer Science and Mathematics at the Eindhoven University of Technology in the Netherlands. He did his PhD in Computer Science on the topic of multi-resource management in embedded real-time systems. His current research focuses on reliable and trustworthy machine learning for process optimization and control and hybrid data/knowledge driven modelling of high-dimensional data.



Nirvana Meratnia is Professor of Pervasive Computing at the Eindhoven University of Technology in the Netherlands. Her research covers various aspects of computing incl. distributed machine learning and AI, embedded/edge AI, data-driven networking and smart sensor systems. She has been involved in several national and international projects addressing (distributed) computation and intelligence in the context of Internet of

Things and Cyber Physical Systems creating societal and economic impacts.

ORCID

Geert van Kollenburg <http://orcid.org/0000-0002-4985-5283>

Mike Holenderski <http://orcid.org/0000-0002-5406-4793>

Nirvana Meratnia <http://orcid.org/0000-0002-8379-770X>

Data availability statement

All code to reproduce the results in this paper is publicly available at <https://doi.org/10.17632/bnthfntfrm> (van Kollenburg and Holenderski 2022). The publicly available SECOM data set is included in the previous reference, but was originally published at <https://archive.ics.uci.edu/ml/datasets/SECOM>.

References

- Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.
- Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. 1984. *Classification and Regression Trees*. Boca Raton, FL: Chapman & Hall/CRC.
- Chou, Pao-Hua, Menq-Jiun Wu, and Kuang-Ku Chen. 2010. "Integrating Support Vector Machine and Genetic Algorithm to Implement Dynamic Wafer Quality Prediction System." *Expert Systems with Applications* 37 (6): 4413–4424. doi:10.1016/j.eswa.2009.11.087.
- Cover, Thomas, and Peter Hart. 1967. "Nearest Neighbor Pattern Classification." *IEEE Transactions on Information Theory* 13 (1): 21–27. doi:10.1109/TIT.1967.1053964.
- den Boef, Arie J. 2016. "Optical Wafer Metrology Sensors for Process-robust CD and Overlay Control in Semiconductor Device Manufacturing." *Surface Topography: Metrology and Properties* 4 (2): 023001.
- Dhar, Payal. 2020. "The Carbon Impact of Artificial Intelligence." *Nature Machine Intelligence* 2 (8): 423–425. doi:10.1038/s42256-020-0219-9.
- di Bella, Roberto, Diego Carrera, Beatrice Rossi, Pasqualina Fragneto, and Giacomo Boracchi. 2019. "Wafer Defect Map Classification Using Sparse Convolutional Networks." In *International Conference on Image Analysis and Processing*, 125–136. Cham, Switzerland: Springer.
- Egorov, Maxim, Zachary N. Sunberg, Edward Balaban, Tim A. Wheeler, Jayesh K. Gupta, and Mykel J. Kochenderfer. 2017. "POMDPs. jl: A Framework for Sequential Decision Making under Uncertainty." *The Journal of Machine Learning Research* 18 (1): 831–835.
- Ferri, Francesc J., Pavel Pudil, Mohamad Hatef, and Josef Kittler. 1994. "Comparative Study of Techniques for Large-Scale Feature Selection." In *Machine Intelligence and Pattern Recognition*, Vol. 16, 403–413. Amsterdam, The Netherlands: Elsevier.
- Friedman, Jerome H. 2002. "Stochastic Gradient Boosting." *Computational Statistics & Data Analysis* 38 (4): 367–378. doi:10.1016/S0167-9473(01)00065-2.
- Garrison, Ray H., Eric W. Noreen, Peter C. Brewer, and Rania Uwaydah Mardini. 2003. *Managerial Accounting*. New York: McGraw-Hill.
- Gupta, Jatinder N. D., R. Ruiz, J. W. Fowler, and S. J. Mason. 2006. "Operational Planning and Control of Semiconductor Wafer Production." *Production Planning & Control* 17 (7): 639–647. doi:10.1080/09537280600900733.
- Hastie, Trevor, Tibshirani Robert, Jerome H. Friedman, and Jerome H. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Vol. 2. New York, USA: Springer.
- Kluyver, Thomas, Benjamin Ragan-Kelley, Fernando P'erez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, et al. 2016. "Jupyter Notebooks – A Publishing Format for Reproducible Computational Workflows." In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, edited by F. Loizides and B. Schmidt, 87–90. Amsterdam, The Netherlands: IOS Press.
- Mankiw, N. Gregory. 2004. *Principles of Microeconomics*. Mason, OH: Thomson/South-Western.
- May, Gary S., and Costas J. Spanos. 2006. *Fundamentals of Semiconductor Manufacturing and Process Control*. Hoboken, NJ: John Wiley & Sons.
- McCann, Michael, Yuhua Li, Liam Maguire, and Adrian Johnston. 2010. "Causality Challenge: benchmarking Relevant Signal Components for Effective Monitoring and Process Control." In *Causality: Objectives and Assessment*, edited by Isabelle Guyon, Dominik Janzing, Bernhard Schölkopf, 277–288. Whistler, Canada: PMLR.
- Melhem, Mariam, Bouchra Ananou, Mohand Djaziri, Mustapha Ouladsine, and Jacque Pinaton. 2015. "Prediction of the Wafer Quality with

- Respect to the Production Equipments Data." *IFAC-PapersOnLine* 48 (21): 78–84. doi:[10.1016/j.ifacol.2015.09.508](https://doi.org/10.1016/j.ifacol.2015.09.508).
- Moldovan, Dorin, Tudor Cioara, Ionut Anghel, and Ioan Salomie. 2017. "Machine Learning for Sensor-Based Manufacturing Processes." In *2017 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, edited by Rodica Potolea, Radu Razvan Slavescu, 147–154. Danvers, MA: IEEE. doi:[10.1109/ICCP.2017.8116997](https://doi.org/10.1109/ICCP.2017.8116997).
- Ojala, Markus, and Gemma C. Garriga. 2010. "Permutation Tests for Studying Classifier Performance." *Journal of Machine Learning Research* 11 (6): 1833–1863.
- Qin, S. Joe., Gregory Cherry, Richard Good, Jin Wang, and Christopher A. Harrison. 2006. "Semiconductor Manufacturing Process Control and Monitoring: A Fab-wide Framework." *Journal of Process Control* 16 (3): 179–191. doi:[10.1016/j.jprocont.2005.06.002](https://doi.org/10.1016/j.jprocont.2005.06.002).
- Sjödén, David, Vinit Parida, Maximilian Palmié, and Joakim Wincent. 2021. "How AI Capabilities Enable Business Model Innovation: Scaling AI through Co-evolutionary Processes and Feedback Loops." *Journal of Business Research* 134: 574–587. doi:[10.1016/j.jbusres.2021.05.009](https://doi.org/10.1016/j.jbusres.2021.05.009).
- Strubell, Emma, Ananya Ganesh, and Andrew McCallum. 2019. "Energy and Policy Considerations for Deep Learning in NLP." *arXiv preprint: 1906.02243*, Accessed 14 February 2022. doi:[10.48550/arXiv.1906.02243](https://doi.org/10.48550/arXiv.1906.02243).
- The European Commission. 2022. "EU Chips Act." The European Commission. Accessed 5 January 2022. https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-chips-act_en
- Troyanskaya, Olga, Michael Cantor, Gavin Sherlock, Pat Brown, Trevor Hastie, Robert Tibshirani, David Botstein, and Russ B. Altman. 2001. "Missing Value Estimation Methods for DNA Microarrays." *Bioinformatics* 17 (6): 520–525. doi:[10.1093/bioinformatics/17.6.520](https://doi.org/10.1093/bioinformatics/17.6.520).
- van Kollenburg, Geert, Roel Bouman, Tim Offermans, Jan Gerretzen, Lutgarde Buydens, Henk-Jan van Manen, and Jeroen Jansen. 2021. "Process PLS: Incorporating Substantive Knowledge into the Predictive Modelling of Multiblock, Multistep, Multidimensional and Multicollinear Process Data." *Computers & Chemical Engineering* 154: 107466. doi:[10.1016/j.compchemeng.2021.107466](https://doi.org/10.1016/j.compchemeng.2021.107466).
- van Kollenburg, Geert, and Mike Holenderski. 2022. "Data and Code of: Value Proposition of Predictive Discarding in Semiconductor Manufacturing." *Mendeley Data*. doi:[10.17632/bnthfntfrm](https://doi.org/10.17632/bnthfntfrm).
- van Kollenburg, Geert, Mike Holenderski, Patrizia Vasquez, and Nirvana Meratnia. 2022. "Predictive Discarding of Wafers Based on Power Leakage Predictions from Single Layer Misalignment Data." *Procedia Computer Science* 200: 1508–1515. doi:[10.1016/j.procs.2022.01.352](https://doi.org/10.1016/j.procs.2022.01.352).
- Van Rossum, Guido, and Fred L. Drake. 2009. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.
- Werling, Joris, Claude Yugma, Ameer Soukhal, and Thierry Mohr. 2020. "An Agent-based Simulation Model with Human Resource Integration for Semiconductor Manufacturing Facility." In *2020 Winter Simulation Conference (WSC)*, 1801–1812. Orlando, FL: IEEE. doi:[10.1109/WSC48552.2020.9383873](https://doi.org/10.1109/WSC48552.2020.9383873).
- Yang, Lei, and Jay Lee. 2012. "Bayesian Belief Network-based Approach for Diagnostics and Prognostics of Semiconductor Manufacturing Systems." *Robotics and Computer-Integrated Manufacturing* 28 (1): 66–74. doi:[10.1016/j.rcim.2011.06.007](https://doi.org/10.1016/j.rcim.2011.06.007).

Appendix A. parameter space

Table A1. Parameter space used in the grid search cross-validation.

Linear discriminant	'solver': ['svd', 'lsqr']
Quadratic discriminant	defaults
K-nearest neighbours	'n_neighbours': [2, 5]
	'leaf_size': [1, 5]
	'p': [1, 2]
Gradient boosting	'max_depth': [3, 5]
	'subsample': [.3, .6]
Decision tree	'ccp_alpha': [.01, .001]
	'max_depth': [5, 7]
Random forest	'min_samples_split': [3, 5]
	'n_estimators': [100, 1000]