Accurate Predictions, Invalid Recommendations: Lessons Learned at the Dutch Social Security Institute UWV

Marcus Dees, Massimiliano de Leoni, Wil M. P. van der Aalst, and Hajo A. Reijers

1 Introduction

Process-aware recommender systems ("PAR systems" hereafter) are a new breed of information systems that predict how the executions of processes will evolve in the future and determine those that are most likely to fail to meet desired levels of performance (e.g., costs, deadlines, customer satisfaction). Recommendations are provided regarding effective contingency actions that should be enacted to recover from risky executions. PAR systems are expert systems that run in the background and continuously monitor the execution of processes, predict their future, and sometimes provide recommendations. Conforti, de Leoni, Rosa, van der Aalst, and ter Hofstede (2015) and Schobel and Reichert (2017) discuss examples of PAR systems.

A substantial body of research on process monitoring and prediction includes the surveys by Teinemaa, Dumas, Rosa, and Maggi (2017) and Márquez-Chamorro, Resinas, and Ruiz-Cortés (2018), but as Márquez-Chamorro et al. (2018) indicate,

M. Dees (✉)
Uitvoeringsinstituut Werknemersverzekeringen (UWV), Amsterdam, The Netherlands
e-mail: marcus.dees@uwv.nl

M. de Leoni
University of Padova, Padova, Italy
e-mail: deleoni@math.unipd.it

W. M. P. van der Aalst
RWTH Aachen University, Aachen, Germany

Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: wvdaalst@pads.rwth-aachen.de

H. A. Reijers
Utrecht University, Utrecht, The Netherlands
e-mail: h.a.reijers@uu.nl

J. vom Brocke et al. (eds.), *Business Process Management Cases Vol. 2*,
https://doi.org/10.1007/978-3-662-63047-1_13

"little attention has been given to providing recommendations. [p. 17]" In fact, how process participants would use these predictions to recover from those executions that cause problems is often overlooked. It seems that process participants are tacitly assumed to make the "right decision" about the best corrective actions for each case. This view also holds for approaches based on mitigation/flexibility "by design" (Lhannaoui, Kabbaj, & Bakkoury, 2013). Unfortunately, the assumption that an effective corrective action will be chosen is not always met in reality. Interventions are often selected based primarily on human judgment, which naturally relies on the subjective perception of the process instead of objective facts.

In particular, an organization will profit from using a PAR system only if the system makes accurate decisions and the organization bases its decisions on the system's decision. Considerable attention is paid to making accurate predictions—particularly the proper use of data, measuring accuracy, and so on—but we show that the approach to making effective decisions is just as important as making accurate predictions. Both parts are essential ingredients of an overall solution.

This case study reports on a field experiment that we conducted at UWV, a Dutch governmental agency. The case study is inspirational in that it outlines how to enrich BPM programs through PAR systems (vom Brocke, Mendling, & Rosemann, 2021), and it also serves to study challenges in leveraging the value of PAR systems for decision-making.

UWV provides, among other things, financial support to Dutch residents who lose their jobs and seek new employment. Several subjects ("customers" hereafter) unintentionally receive more unemployment benefits than they are entitled to. Such errors are eventually detected—usually after several months—triggering a "reclamation" event. To retrieve the amount of excess support from a customer is difficult, time-consuming, and, often unsuccessful—not to mention upsetting for the customer, who may not have the necessary funds, given his or her joblessness. In this context, an effective recommender system should be able to detect the customers who are most likely to cause a reclamation event, thus providing operational support for UWV in preventing the provision of excess benefits.

Research at UWV has shown that the main cause of a reclamation event lies with the customer making a mistake when informing UWV about his or her other income. Therefore, interventions should be aimed at preventing customers from providing incorrect income information. To follow up on this idea, we developed a predictor module that relies on machine-learning techniques to monitor and identify the subjects who are most likely to receive excess support.

UWV's stakeholders considered several interventions to prevent reclamation events and chose to send an e-mail to the subjects who were suspected of being at higher risk as a field experiment. The contents of the e-mail were determined by previous research at UWV showing the types of errors customers make filling in the form. The e-mail is sent just before the customers would need to provide information about their income to UWV. The results showed that risky customers were detected well, but no significant reduction in the number of reclamations occurred, so the intervention did not achieve the desired effect of preventing reclamation events. Our findings show the importance of conducting research on the interventions themselves, even before they are tested.

2 Situation Faced

UWV is the social security institute of the Netherlands and is responsible for the implementation of a number of employee-related types of insurances. One of the processes that UWV executes is the unemployment benefits process. When residents in the Netherlands become unemployed, they file a request with UWV, which then decides whether they are entitled to benefits. When requests are accepted, these customers receive monthly benefits until they find new jobs or the maximum period for their entitlements is reached. However, sometimes the customers make errors on their other-income forms, resulting in overpayment of benefits until the errors are discovered.

The process for paying unemployment benefits is bound by legal rules. UWV's customers and employees are required to perform certain steps for each benefit month. Figure 1 depicts a typical scenario of a customer who receives benefits for a particular benefit month, along with the steps that are executed in each month until the process is complete and any errors are discovered. Before a customer can receive a payment of benefits for a benefit month, he or she must send an income form to UWV that specifies whether the customer has received income of any kind for that benefit month and, if so, how much. The benefits are adjusted as a function of such income, up to no benefits if the customer's income exceeds the amount of benefits to which he or she is entitled.

Figure 1 shows that, in August, when any error related to the benefit month of June is discovered, 2 months of benefits (June and July) have already been paid based on erroneous information. Although the amounts to be reclaimed are often comparatively small—usually, a few hundred Euros—about 12,000 customers per month are subjects of reclamations, leading to significant losses if the overpayments are not reclaimed.

The main cause of reclamation events is customers' failing to fill in the correct amount of other income earned on the income form. If the customer has not received a payslip, he or she estimates this income. Even with a payslip, customers often make mistakes, as the required amount is the social security wages, not the gross salary or the salary after taxes. Customers who receive their payslips every 4 weeks, rather than every month, may also err in calculating the monthly amount, which,

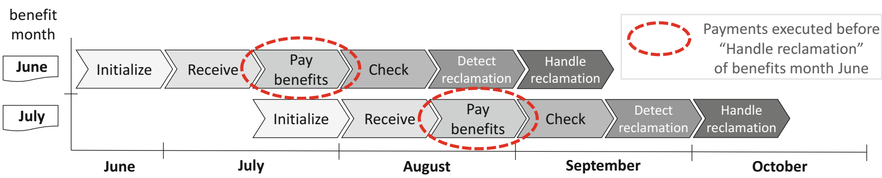

Fig. 1 An example of the activities UWV executes related to providing unemployment benefits to a customer for the months of June and July. (The year is irrelevant.) Each row is related to the activities that must be performed to handle an income form for the month for which benefits are to be paid ("benefit month"). The work for each benefit month takes several months to complete; for example, the work for the benefit month of June is finally completed in September

except for February in most years, is higher than the amount for 4 weeks. Other factors, like vacation money or pension-related income, can also make it difficult to determine the correct amount.

Since the reclamations are caused by customers filling in income forms incorrectly, UWV's only recourse is to prevent customers from making mistakes when they fill in the income form. UWV determined that targeting only the customers who were most likely to make an error on the form was the most fiscally responsible approach, along with the approach least likely to irritate or over-communicate with customers who were unlikely to make errors. Therefore, a recommender system that could identify customers who had a high risk of making an error and could propose interventions would be helpful.

3 Action Taken

Figure 2 shows our approach to the development and testing of a PAR system for UWV. The approach is in line with the BPM Lifecycle Dumas, Rosa, Mendling, and Reijers (2018) presented. Steps 1a and 1b of the approach are to analyze and identify the organizational issue (i.e., process analysis), which, as described in Sect. 2, was related to needless reclamation events.

The second step was to develop a recommender system consisting of a predictor module (Step 2a) and a set of interventions (Step 2b) (i.e., process redesign). The predictor module identifies the cases on which the intervention should be applied, that is, those with the highest likelihood of leading to a reclamation event. (Section 3.1 describes the predictor module setup.) Then a set of interventions is selected in concert with stakeholders to ensure the stakeholders' support, not only for the intervention but also for the changes necessary to implement it. Of several possible interventions put forward, one was chosen (Step 3) that did not strain the limited availability of resources at UWV for executing an experiment. Section 3.2 elaborates on the process of collecting ideas for interventions and selecting the intervention for the field experiment.

Step 4 was to design a field experiment (i.e., process implementation). The field experiment was set up as an A/B test (Kohavi & Longbotham, 2017).

In an A/B test, one or more interventions are tested under the same conditions, to find the alternative that best delivers the desired effect. In our field experiment, the intervention could be tested in the natural setting of the process environment. The objective of the field experiment was to determine the effect of applying an intervention to cases at a specific risk level with respect to whether a customer triggers a reclamation event (the process metric). All other factors that could play a role in the field experiment were controlled as far as possible in our business environment. Under these conditions, the field experiment would show whether a causal relationship existed between the intervention and the change in the values of the process metrics. Section 3.3 describes the setup for the UWV study.

Analysis of the results of the field experiment (Step 5) determined whether the intervention had the desired effect (i.e., process monitoring). Sections 4.1 and 4.2

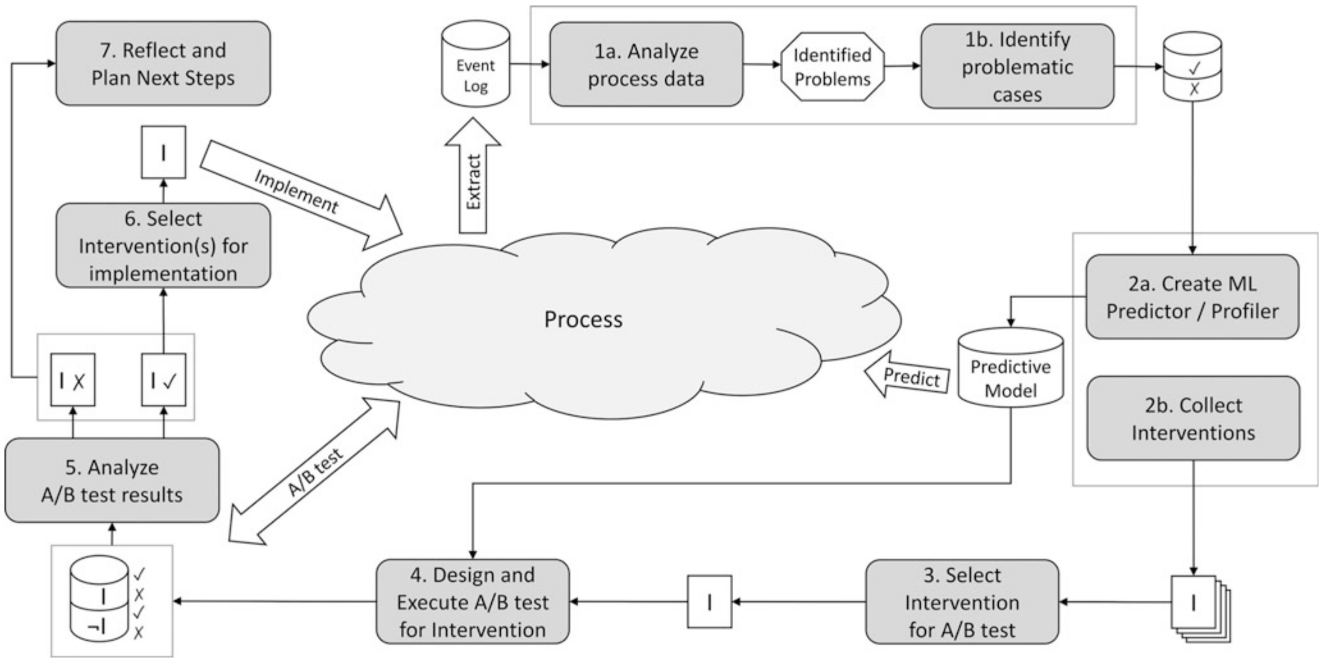

Fig. 2 Overview of the steps that make up the research method for one improvement cycle. “I” refers to “Intervention”

contain the analysis of the intervention and the predictor module, respectively. If the intervention had an effect, then both the direction of the effect (i.e., whether the intervention leads to better or worse performance) and the size of the effect must be calculated. If the intervention had the desired effect, it would be implemented as part of the process (Step 6).

The final step (Step 7) is the reflective phase in which the lessons learned from the execution of the approach are discussed. Section 5 contains the lessons learned for the UWV case.

The interventions, along with the predictor module from Step 2a, make up the PAR system. After the decision to implement an intervention, the PAR system's predictor module must be updated. Changing the process also implies that the situation under which the predictions are made has changed, so a period of time after the change takes effect should be reserved to gather a new set of historic process data on which the predictor module can be retrained.

This research method requires that many choices be made, such as which organizational issues will be tackled and which interventions will be tested. Prior to making a choice, the research participants should be aware of any assumptions or biases that could influence their choices.

3.1 Building the Predictor Module

The prediction is based on training a predictor module that uses historical data, for which we used the data-mining techniques Logistic Regression and ADA Boost. The choice to use these two techniques was based on experience with these techniques at UWV. Both techniques were tuned through hyper-parameter optimization (Claesen & Moor, 2015). To this end, UWV historical data was split into a training set with 80% of the cases and a test set with 20% of the cases. The models were trained with a fivefold cross-validation that used several configurations of the algorithm's parameters. The models were tested on the set of 20% of the cases. The Receiver Operator Characteristic (ROC) curve (Fawcett, 2006), which measures the model's quality, shows for every percentage of false-positives the percentage of true positives that the model finds. The area under the ROC curve (AUC) summarizes a model's total performance. The models are ranked using the AUC, and the models with the highest AUC value were selected for the predictions to be used for the experiment.

The most important features of the predictive models such as the number of occurrences of a first income form and whether the most recent payment was done automatically or manually are process-related. Leaving out all process-related features would reduce the AUC value by 8%, so ignoring the process nature of cases would be detrimental to the predictive models' quality.

The predictor module, implemented as a stand-alone application in Python, leveraged the sci-kit learn library (Pedregosa et al., 2011) to access the data-mining functionality. In the UWV case, the historical data was extracted from the company's systems related to the execution of every activity for 73,153 customers who concluded the reception of unemployment benefits between July 2015 and July

2017. Space limitations prevent us from providing details on how the prediction module was built, but details are available in the technical report (Dees, de Leoni, van der Aalst, & Reijers, 2019) that accompanies this case description.

3.2 Collecting and Selecting the Interventions

After three brainstorming sessions with 15 of UWV's employees and two of its team managers, the choice of the intervention was made based on the experience and expectations of these stakeholders. The goal of the intervention was to prevent customers from putting the wrong amount on the income form. The sessions initially put forward three types of interventions based on the actors who are involved in the intervention (i.e., the customer, the UWV employee, the last employer):

1. An intervention that supports the customer in advance of filling in the income form
2. An intervention in which the UWV employee verifies the information provided by the customer on the income form, and, if necessary, corrects it after contacting the customer
3. An intervention in which the last employer of the UWV customer is asked to supply relevant information more quickly so the UWV can promptly verify the accuracy of the information the customer provides on the income form

An intervention can be executed only once a month, between receipt of the two income forms for two consecutive months. In the final brainstorming session, the stakeholders opted for option 1—supporting the customer in fill out the income form correctly. In the stakeholders' experience, their support with filling out the form helps customers reduce the chance of incurring reclamations.

Only one intervention was selected for the experiment as a result of the limited availability of resources at UWV that arose from temporary under-staffing and an uneven capacity demand throughout the month. The selected intervention entails proactively providing customers, whom the predictor model identified as being most likely to make errors, with information about how errors are often made in filling out the income form. The UWV employees indicated that most mistakes were made regarding topics related to the definition of social security wages, the definition of unemployment, and receiving pay every 4 weeks instead of monthly.

Next, the medium through which the customer would be informed had to be determined: a physical letter, an e-mail, or a phone call from a UWV employee. To keep costs low, sending the information by e-mail was chosen. An editorial employee of UWV determined the phrasing, and the e-mail contained hyperlinks to pages on the UWV website to allow customers to access additional information if they needed it. The customers who received the e-mail were not informed about the experiment. A tool UWV uses to send e-mails to large numbers of customers at the same time provided the functionality to determine whether the e-mail was received—that is, without bouncing—and whether the e-mail was opened. Since the timing of sending the message can influence its success, it was sent on the day preceding the last workday of the calendar month in which the predictor module marked the customer as risky. This approach increased the likelihood that the customer would read the message before filling in the income form for the subsequent month.

3.3 Setting Up and Executing the Field Experiment

The field experiment was used to determine whether using the PAR system's design for prediction and intervention would reduce the number of reclamation events. Specifically, we first determined the number and the nature of the customers who were to be monitored. Then these customers were split into two groups: one on which the PAR system was applied (the experimental group) and one on which the PAR system was not applied (i.e., the control group).

We conducted the experiment with 86,850 cases handled by UWV's Amsterdam branch. These customers were receiving benefits and were not the 73,153 customers who were used to train the predictor module. Of the 86,850 cases, 35,812 were part of the experimental group. The intervention was executed on 30 August 2017, 28 September 2017, and 30 October 2017 by sending the e-mail. The predictor was used to compute the probability of a reclamation event for the 35,812 cases of the experimental group. The probability was higher than 0.8 for 6747 cases, and the intervention was executed for those cases.

4 Results Achieved

The intervention did not have a preventive effect, even though the risk prediction was reasonably accurate. Sections 4.1 and 4.2 describe the results achieved.

4.1 The Intervention Did Not Have a Preventive Effect

Figure 3 shows the results of the field experiment in terms of the percentage of reclamations observed in each group. The number of reclamations did not significantly decrease when the system was used, as they decreased from 4.0% without the intervention to 3.8% with the intervention. The effectiveness of the system as a whole is therefore 0.2%.

The PAR system deemed 6747 cases as risky and were sent the e-mail. Out of these 6747 cases, 4065 (60%) received and opened the e-mails and 2682 (40%) either did not receive it (i.e., the e-mail "bounced") or did not open it. Since there were almost no bounces, most of the 40% did not open it. Among the customers who did receive the e-mail, only 294 clicked on the links and accessed UWV's website, and 10.9% of those had a reclamation event in the subsequent month, which is more than 2.5 times the average and around 1.7 times the frequency among customers who

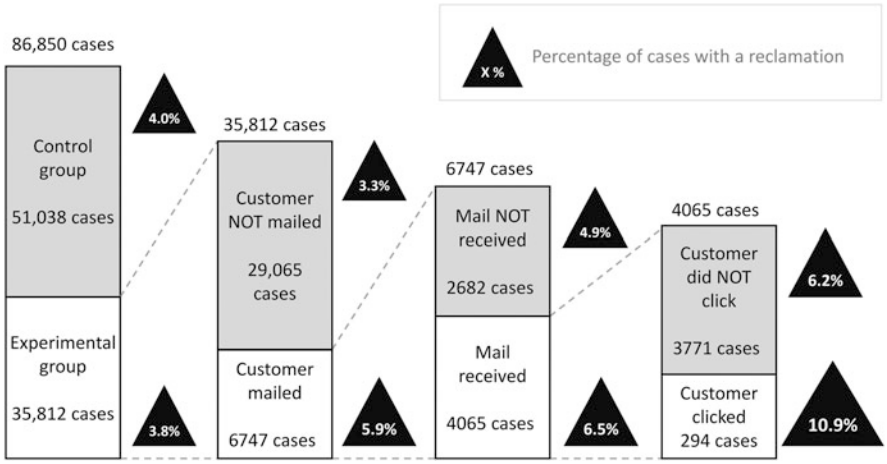

Fig. 3 The number of cases and the percentage of cases that had a reclamation event for all groups. The results show that risky customers were identified, but the intervention did not solve the problem

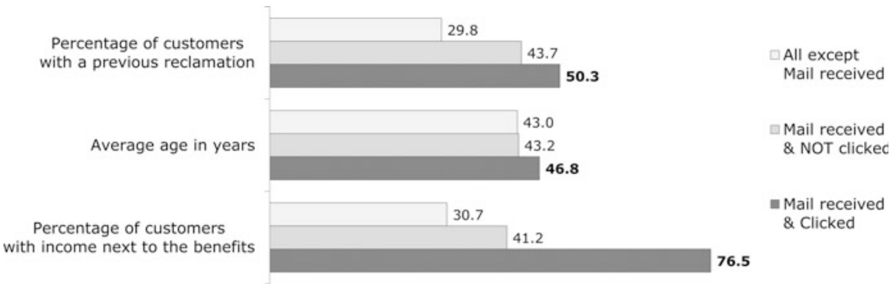

Fig. 4 Comparison of the characteristics of customers who did not receive the e-mail, those who received it but did not click the link, and those who accessed UWV's website through the e-mail's link

opened the e-mail but did not click the links. Therefore, accessing additional information seemed to increase the chances of error!

We conducted a comparative analysis among the customers who did not receive the e-mail, i.e., the e-mail bounced, those who received it but did not click the links, and those who clicked on the links and reached the website. The results of the comparative analysis, shown in Fig. 4, indicate that 76.5% of the customers who clicked the e-mail's links had income in addition to the benefits. (Recall that customers can receive benefits even while employed if their income is reduced, in which case they receive benefits for the difference.) Our results are reasonable, as mistakes are more frequent when filling the income form is more complex, as when there is some other income instead of none. Among the customers who clicked on the e-mail's link, 50.3% had a previous reclamation, and they are on average

3.5 years older than those who did not click on the link, which is a statistically significant difference.

The results suggest that e-mailing is counterproductive or at least that there was a positive correlation between exploring the additional information provided and being involved in a reclamation in the subsequent month. To a smaller extent, 6.2% of the customers who received the e-mail but did not click the links had reclamations versus a mean of 3.8–4.0%. Clearly, the intervention did not achieve the intended goal.

4.2 The Predicted Risk Was Reasonably Accurate

The analysis shows that the intervention did not lead to improvement. We sought to determine whether this result was caused by inaccurate predictions, an ineffective intervention, or both. This section reports on our analysis of the quality of the predictor module, for which we use the cumulative lift curve (Ling & Li, 1998). We chose this measure because of the imbalance in the data, as Ling and Li (1998) advised. As mentioned in Sect. 2, every month only 4% of UWV's customers are involved in reclamation events.

In cases of unbalanced data sets (e.g., between customers who experience reclamation events and those who do not), precision and recall are generally unsuitable for assessing the quality of predictors. In addition, because of the low cost of sending an e-mail, false-negatives, referring to customers with undetected reclamations during the subsequent month, occur more often than false-positives, that is, customers who are incorrectly predicted as having reclamations during the subsequent month.

Figure 5 shows the cumulative lift curve for the case study at UWV. The rationale is that, in a set of *x*% of randomly selected customers, one expects to observe *x*% of the total number of reclamations. The figure shows that, in our case, the predictions are better than random. For example, the 10% of customers with the highest risk of having a reclamation accounted for 19% of all reclamations, which is roughly twice what could be expected in a random sample.

In summary, although the prediction technique can certainly be improved, the prediction was reasonably effective (cf. Sect. 3.1). However, because the system as a whole did not bring about a significant improvement, we conclude that the lack of a significant effect was likely caused by the ineffectiveness of the intervention.

5 Lessons Learned

The experiment proved to be unsuccessful. While the predictions were reasonably accurate, the intervention of sending an e-mail to predicted high-risk customers did not reduce the number of reclamations. In fact, the group of customers who received the e-mail and clicked on its link to more information had twice as many reclamations as the average population. Section 5.1 elaborates on the reasons the

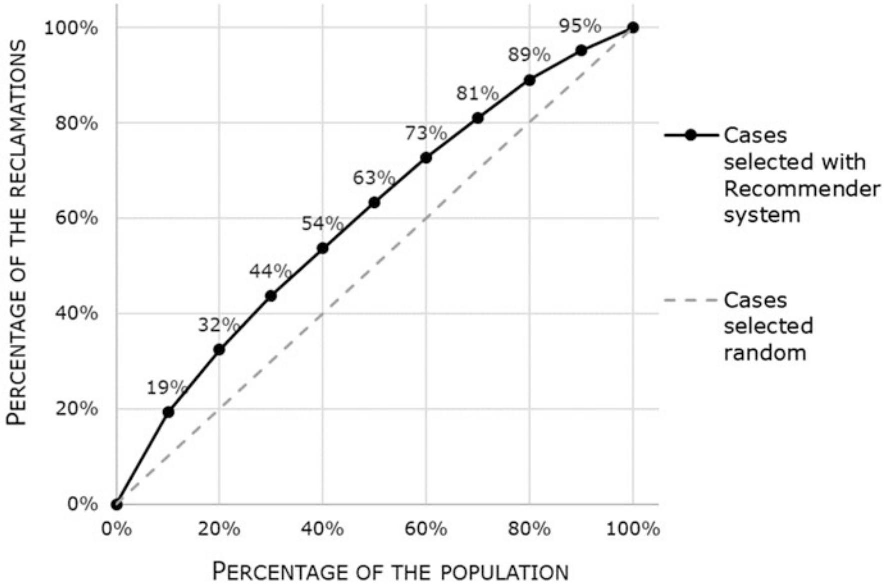

Fig. 5 The cumulative lift curve shows that using the recommender system is superior to using a random selection of cases to predict reclamation events

intervention did not work, while Sect. 5.2 focuses on the lessons learned and explains how the research methodology should be updated.

5.1 Why Did the Intervention Fail to Work?

One of the reasons the intervention was not successful could be related to the timing of sending the e-mail, as different timing during the month could have been more appropriate. However, timing does not explain why only 294 of the 6747 cases acted on the e-mail by clicking the links. Other reasons may be that the customers found the e-mail's message unclear or that the links in the e-mail body pointed to confusing information on the UWV website, as among the group of 294 cases who clicked the links to get more information, reclamations actually occurred 2.5 times as often.

The communication channel could also be part of the cause. Sending the message by letter or by calling the customer might have worked better. We heard several comments from stakeholders to the effect that they did not expect the failure because, for example, "after speaking to a customer about how to fill in the income form, almost no mistakes are made by that customer." In this regard, the UWV case sets a very valuable example that technology does not deliver value to a process per se but that it needs to fit the context to support specific objectives and that it needs close alignment with other capability areas (vom Brocke et al., 2021).

5.2 What Should Be Done Differently Next Time?

We learned that the A/B testing is beneficial in assessing the effectiveness of interventions. The involvement of stakeholders and other process participants, including UWV's customers, helped to achieve our goal of testing an intervention, even though the results did not achieve the expected results. We learned a number of lessons about how to adjust our approach that we will put in place for the next round of testing:

1. Creating a predictor module requires the selection of independent features as inputs to build the predictive model. Our analysis of the reasons for the intervention's failure will help us to derive new features to be incorporated when training the predictor. For instance, the features presented in Fig. 4 can be used to train a better predictor, such as, for the UWV case, a Boolean feature for whether a customer has income other than the benefits.
2. The insights derived from the analysis can also be useful in putting forward other possible interventions. For instance, an intervention could be a manual check of the income form when a customer has had a reclamation in the previous month. This intervention is derived from the feature representing the number of executions of Detect Reclamation, as discussed in Sect. 4.1.
3. The interventions for the A/B test (Step 3 in Fig. 2) should be pre-assessed. The intervention used in our experiment provides information to the customers related to filling the income form, but, before running the test, we could have checked historical event data on whether the average number of reclamations decreased when customers were provided information and support in completing the income form. If we had done this, we could have avoided running a test that was destined to fail.
4. Since a control group was compared with an experimental group on which the system was employed, and the comparison was measured end-to-end, it is impossible to state the reason for the intervention's failure beyond just observing it. For instance, we should have used questionnaires to assess the reasons for the failure: the customers who received the e-mail should have been asked why they did not click on the links or, if they did click on them, why they still made mistakes. Clearly, questionnaires are not applicable to all kinds of interventions, but some methods could have been used to acquire the information needed to analyze reasons for the intervention's ineffectiveness.
5. It is unlikely that the methodology in Sect. 3 provided complete results after one cycle. The methodology should be reiterated in multiple cycles. In fact, this finding is compliant with the principle of action research, which is based on the idea of continuous improvement cycles (Cronholm & Goldkuhl, 2003; Rowell, Riel, & Polush, 2017).
6. Although multiple cycles are useful, just one cycle took a few months to be carried out, so the cycle needs to be repeatable at a higher speed with multiple interventions tested at each cycle. In addition, if an intervention is clearly ineffective, its testing should be stopped without waiting for the cycle to end.

All the lessons learned share one theme: *Accurate predictions are crucial, but their effect is nullified if it is not matched by effective recommendations that are based on evidence from historical and/or experimental data*.

5.3 Conclusion

When one builds a process-aware recommender system, both the predictor module and the recommender parts of the system must be effective for the whole system to be effective. In our case, the predictor module was sufficiently accurate, but the intervention did not have the desired effect. The lessons learned from the field experiment have been used in building an updated research method that uses high-speed iterations with multiple interventions. Systematic support will be needed for each step of the approach to meet these requirements.

We plan to improve the predictor module by using different techniques and leveraging contextual information about the customers and their history. Our analysis showed that, for example, the presence of some monetary income next to the benefits is strongly related to reclamations. We want to use evidence from the process executions and insights from building the predictor module, to select one or more interventions to be tested in a new experiment.

We aim to devise a new technique that finds the best intervention based on the specific case. Different cases might require different interventions, and the choice of the best intervention should be automatically derived from the historical facts recorded in the system's event logs. In other words, the system will rely on machine-learning techniques that (1) reason about past executions to find the interventions that have generally been effective in the current case and (2) make recommendations accordingly.

References

Claesen, M., & Moor, B. D. (2015). *Hyperparameter search in machine learning*. CoRR abs/1502.02127.

Conforti, R., de Leoni, M., Rosa, M. L., van der Aalst, W. M. P., & ter Hofstede, A. H. (2015). A recommendation system for predicting risks across multiple business process instances. *Decision Support Systems, 69*, 1–19. https://doi.org/10.1016/j.dss.2014.10.006.

Cronholm, S., & Goldkuhl, G. (2003). Understanding the practices of action research. In *The 2nd European conference on research methods in business and management*.

Dees, M., de Leoni, M., van der Aalst, W. M. P., & Reijers, H. A. (2019). *What if process predictions are not followed by good recommendations?* arXiv e-prints arXiv:1905.10173, 1905.10173.

Dumas, M., Rosa, M. L., Mendling, J., & Reijers, H. A. (2018). *Fundamentals of business process management* (2nd ed.). New York: Springer. https://doi.org/10.1007/978-3-662-56509-4.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters, 27*(8), 861–874.

Kohavi, R., & Longbotham, R. (2017). *Online controlled experiments and a/B testing* (pp. 922–929). Boston, MA: Springer US.

Lhannaoui, H., Kabbaj, M. I., & Bakkoury, Z. (2013). Towards an approach to improve business process models using risk management techniques. In *2013 8th international conference on intelligent systems: Theories and applications (SITA)* (pp. 1–8).

Ling, C. X., & Li, C. (1998). Data mining for direct marketing: Problems and solutions. In *Proceedings of the fourth international conference on knowledge discovery and data mining (KDD-98)* (pp. 73–79). New York City, New York, USA, August 27–31.

Márquez-Chamorro, A. E., Resinas, M., & Ruiz-Cortés, A. (2018). Predictive monitoring of business' processes: A survey. *IEEE Trans Services Computing, 11*(6), 962–977.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Rowell, L. L., Riel, M. M., & Polush, E. Y. (2017). *Defining action research: On dialogic spaces for constructing shared meanings* (pp. 85–101). New York: Palgrave Macmillan US.

Schobel, J., & Reichert, M. (2017). A predictive approach enabling process execution recommendations. In *Advances in intelligent process-aware information systems – Concepts, methods, and technologies* (pp. 155–170). Cham: Springer International. https://doi.org/10.1007/978-3-319-52181-7.

Teinemaa, I., Dumas, M., Rosa, M. L., & Maggi, F. M. (2017). Outcome-oriented predictive process monitoring: Review and benchmark. CoRR abs/1707.06766, 1707.06766.

vom Brocke, J., Mendling, J., & Rosemann, M. (2021). Planning and scoping business process management projects and programs with the BPM Billboard. In J. vom Brocke, J. Mendling, & M. Rosemann (Eds.), *BPM cases. Digital innovation and business transformation in practice* (Vol. 2). New York: Springer.