Prescriptive Process Monitoring

Major Project
8th Semester
in
Information Technology

Submitted by
Dhanush Vasa (IIT2019208)
Mitta Lekhana Reddy (IIT2019204)
Under the guidance of
Dr. Ranjana Vyas

Department of Information Technology
Indian Institute of Information Technology,
Allahabad 211015, (U.P.) INDIA

Indian Institute of Information Technology, Allahabad

Candidates Declaration

I hereby declare that the work presented in this report entitled **Double Machine learning Prescriptive Process Monitoring**, submitted towards the fulfillment of BACHELOR'S THESIS report of Information Technology at Indian Institute of Information Technology Allahabad is an authenticated record of our original work carried out under the guidance of Dr. Ranjana Vyas. Due acknowledgments have been made in the text to all other material used. The project was done in full compliance with the requirements and constraints of the prescribed curriculum.

Dhanush Vasa
(IIT2019208)

Mitta Lekhana Reddy
(IIT2019204)

Date:
Place: Prayagraj

Indian Institute of Information Technology, Allahabad

Certificate from Supervisor

This is to certify that the statement made by the candidate is correct to the best of my knowledge and belief. The project titled **Double Machine Learning Prescriptive Process Monitoring** is a record of candidates' work carried out by him under my guidance and supervision. I do hereby recommend that it should be accepted in the fulfillment of the requirements of the Bachelor's Thesis at IIIT Allahabad.

Dr. Ranjana Vyas
(On final examination and approval of the thesis)

Date:
Place: Prayagraj

Indian Institute of Information Technology, Allahabad

Certificate of Approval

The forgoing thesis is hereby approved as a creditable study carried out in the area of Information Technology and presented in a manner satisfactory to warrant its acceptance as a pre-requisite to the degree for which it has been submitted. It is understood that by this approval the undersigned does not necessarily endorse or approve any statement made, opinion expressed or the conclusion drawn therein, but approves the thesis only for the purpose for which it is submitted.

Committee on final examination for the evaluation of thesis:

1. Dr. Ranjana Vyas
2. Dr. Rahul Kala
3. Dr. Pragya Singh

(Dean A & R)

Date:
Place: Prayagraj

Plagarism Report

ORIGINALITY REPORT

8%	7%	7%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	link.springer.com Internet Source	5%
2	library.oapen.org Internet Source	1%
3	"Process Mining Workshops", Springer Science and Business Media LLC, 2022 Publication	1%
4	"Business Process Management Workshops", Springer Science and Business Media LLC, 2019 Publication	1%
5	proceedings.mlr.press Internet Source	<1%
6	Kateryna Kubrak, Fredrik Milani, Alexander Nolte, Marlon Dumas. " Prescriptive process monitoring: ? ", PeerJ Computer Science, 2022 Publication	<1%
7	ceur-ws.org Internet Source	<1%
	deepai.org	

Contents

Abstract

Utilising corrective measures, prescriptive process monitoring aims to enhance business processes. These methods include monitoring the prescribed process. There is no upper limit on the number of interventions, but prescriptive process monitoring solutions disregard this reality. Interventions maximize resource utilization. An intervention at any stage of the loan origination process could involve the preparation of a new loan offer with the expectation that the applicant will approve it. This intervention requires a credit representative. Consequently, this action cannot always be performed promptly. This study describes an observation technique. This strategy optimizes a cost function with constrained resources. Using predictive modeling and causal reasoning, this strategy identifies hazardous situations. Case outcomes are also affected by interventions. These calculations determine treatment costs in order to optimize a cost function. Initial research indicates that the technique outperforms a predictive (non-causal) baseline.

Keywords: Double Machine Learning, Orthogonal Random Forest (ORF) , Prescriptive Process Monitoring (PrPM), Resource Allocator · comprehensibility

Chapter 1

Introduction

The goal of the discipline of process mining known as prescriptive process monitoring, also known as PrPM[2], is to improve the efficiency of business processes by using strategies derived from predictive analytics and business process management. PrPM seeks to detect possible adverse outcomes that might have an effect on the performance of a process and then prescribe or trigger measures to avoid or minimise the effects of such adverse events.

Logs of events and other information regarding the execution of a business process are the foundation of PrPM. These logs include a plethora of information that may be mined for trends and utilised to forecast future actions. To anticipate and prevent problems that may arise over the course of a process's execution, predictive analytics approaches are used to the event logs.

PrPM makes recommendations or initiates actions to avoid or minimise probable adverse consequences after such outcomes have been recognised as being possible. These interventions may take a variety of forms, such as rearranging the order in which operations are performed, allocating more resources, or modifying the characteristics of the process. The purpose of this endeavour is to optimise the performance of the process by lowering the frequency of undesirable results while simultaneously raising the bar for the process's level of efficacy and efficiency.

Process Monitoring (PM)[3] is a relatively young discipline that has been receiving a lot of attention in recent years owing to its potential to enhance the efficiency and effectiveness of corporate operations. It is used in many fields to boost productivity, save expenses, and satisfy customers in healthcare, finance, manufacturing, and logistics, to name a few.

In real-world situations, the capability of these resources must be taken into account when triggering actions due to their scarcity. For instance, in the loan origination process, only so many loan officers may approach a potential borrower with a second loan offer designed to boost the chances of a positive response. Thus, activating an intervention would be contingent on a loan officer being available, and it would not be possible to do so at any given moment.

And it takes more than just individuals to carry out an intervention. For instance, interventions may call for specialised equipment, which may be in short supply. In light of this, it is imperative that all relevant resources, not just human ones, be considered when deciding when and which intervention to initiate.

To overcome this shortcoming, we present a novel method we call Double Machine Learning Prescriptive Process Monitoring (DPrPM)[1] that use predictive modelling and causal inference to calculate the potential benefit of triggering an intervention in each individual situation. This method takes into consideration both the resources needed for each intervention and the resources that are already available to determine which patients should be prioritised for treatment.

In this context, interventions are used throughout a case's execution to improve the process's overall efficiency. However, there is a certain amount of resources available, and each action demands part of them. As a result, it's crucial to weigh the benefits against the costs before deciding to initiate an intervention.

To solve this problem, the strategy takes into consideration a gain function that incorporates both the cost of a bad result and the cost of an intervention. By utilising a predictive model to estimate the likelihood of a bad case result and an Orthogonal Random Forest (ORF)[1] to estimate the impact of triggering an intervention on that likelihood, we may maximise the gain function. Using these estimations, the method chooses which instances to treat given the available resources and the benefit of triggering an intervention in each.

This method may be put into practise for any number of business procedures with limited resources and the need for intervention to boost productivity. The method has several potential applications, including healthcare, where it may be used to decide whether or not to initiate an intervention throughout the course of a patient's treatment in order to maximise the result, taking into account factors like the availability of staff, equipment, and facilities.

This research proposes using Double Machine Learning Prescriptive Process Monitoring (DPrPM) to overcome the problem of few intervention resources in PrPM methods. To evaluate the likelihood of a poor case result and the impact of triggering an intervention, we plan to employ both predictive and double machine learning models. To prioritise the utilisation of scarce treatment resources, this assessment is utilised to calculate the benefit from activating an intervention in each instance. A real-world event log will be used to compare the method to one that just uses prediction models.

The paper is structured into several sections. Section 2 provides background concepts and related work on PrPM techniques, predictive models and causal models. Section 3 explains the proposed methodology in detail, including the use of predictive and causal inference models. Section 4 in-dept experimental setup and evaluation metrics.

Chapter 2

Background and Related Work

2.1 Predictive Process Monitoring

When it comes to estimating the likelihood of adverse case outcomes, Prescriptive Process Monitoring (PrPM) approaches are quite similar to Outcome Oriented Predictive Process Monitoring (PPM) techniques. An outcome-oriented PPM method's input is an event log that depicts the running of a business process. Each trail is made up of a string of occurrences. In one event, you may read about the completion of one specific occurrence of an action. A time stamp, an activity label, and a unique case identifier are the three components that make up an event. Other possible event characteristics include the presence of the resource (the people actually carrying out the action). Some other qualities exist, and they might fall into either the case or event category. The case attributes are those whose values remain constant inside the scope of the case, whereas the event attributes are those whose values do change.

Outcome-oriented PPM techniques use the (incomplete) case history to forecast the final resolution of an open case. In the common binary PPM approach, a case's result might be either positive or negative. A priori knowledge of the

PPM technique and past case outcome data are used to arrive at the idea of case, the result of a case. Each trace in an event log that has been labelled indicates the result of a specific instance.

2.2 Prescriptive Process Monitoring

When the chance of a case whose conclusion is leading to an unfavourable consequence is beyond a threshold value (for example, 70%), then a single or several alerts are issued, as suggested by Fahrenkrog et al. [1]. Initiating the intervention at the time of alarm generation decreases the likelihood of a bad case result. The threshold has been experimentally optimised with regard to a gain function using this method.

To calculate the likelihood of a bad case result and to reliably estimate the impact of the intervention on this outcome, Metzger et al. use [3]. They set up and introduced policy-based reinforcement learning to initiate proactive process adaptation. This study focuses on the challenge of understanding when to activate an intervention rather than posing the issue of whether or not to do so.

Both the Metzgers et al. [3] and the Fahrenkrog et al. [1]papers make the prior assumption that the number of interventions is infinite and that they may be activated at a certain time. The limitations of available resources are contrasted with in this work.

In order to maximise a certain performance metric, Weinzerl et al. [4] presented a PrPM approach in which the next action is suggested for an active instance.

In this analysis, intervention is not formally explored. Neither the price of intervention nor the fact that an intervention may be triggered only if a resource is available to carry it out were taken into account in this article.

Chapter 3

Proposed Methodology

The fundamental goal of the strategy is to maximise the overall benefit by selecting the best time and treatment for each individual patient. This is accomplished via the approach's two stages: the learning stage and the resource allocation stage.

Two machine learning models are constructed using an event log that is created throughout the learning phase. The first model calculates the odds of a bad thing happening. The goal of this prediction model is to provide light on how likely unfavourable consequences are in a particular circumstance. The second model is a causal one, which calculates the expected result of a certain action. The goal of this prescriptive model is to illuminate the impact of various interventions on a case's final resolution. This approach may assist decide whether an intervention is likely to enhance an outcome by calculating the intervention effect.

The net benefit of an intervention is calculated during resource allocation by weighing the projected risk of adverse outcomes against the estimated therapeutic effect. Decisions on whether or not to treat a specific case, and, if treatment is indicated, when to act in order to maximise the overall benefit, may be made

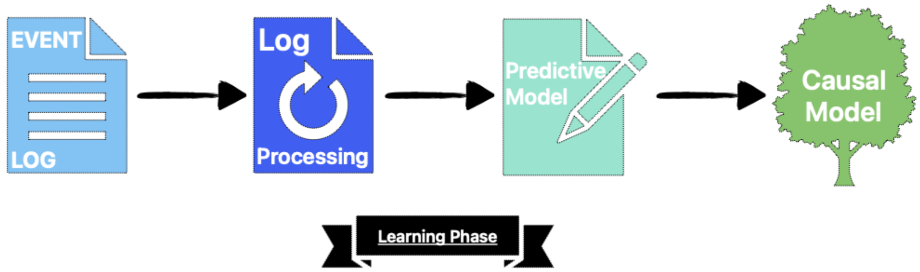

Figure 3.1: Proposed Methodology For Work-Flow

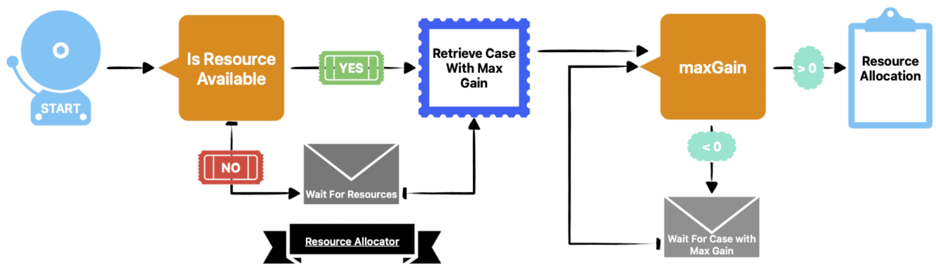

Figure 3.2: Proposed Methodology For Resource Allocator

with the use of this data.

This method integrates predictive and prescriptive modelling to help users make better choices about how to proceed in a given scenario. Providers may improve outcomes while minimising unneeded interventions by evaluating the risk of poor outcomes and the possible effect of actions, and then designing resolution strategies accordingly.

3.1 Data-set Log Preprocessing

A vital part of any strategy for analysing data is the preparation of logs. Data preparation includes checking for inconsistencies and errors, finding useful patterns, and encoding the information so it can be analysed. For our purposes, we use the strategy for cleaning data, extracting k-prefixes, and encoding prefixes suggested by Teinemaa et al. [**prepocessingpaper**]. This makes it more likely that the data is in a usable form, allowing us to draw useful conclusions from our examinations.

After the data has been cleaned, we can pinpoint the cause of the result and the measures taken to mitigate its occurrence. Since this is a process-dependent stage, we must first learn the purpose of the business process. A financial analysis's business process purpose, for instance, may be to cut down on the amount of problems that crop up during financial transactions.

We analyse the log data to find patterns and trends that may point to actions that might lower the likelihood of adverse outcomes. We may try to identify patterns in the occurrence of undesirable results, such the use of a certain set of interventions. Once we have this data, we may make adjustments to the action plan or step up our case monitoring to lower the likelihood of undesirable results.

Our method for analysing data relies heavily on the preprocessing of logs and the identification of outcomes and actions. If we take a methodical approach to analysing the log data, we may discover opportunities to enhance company procedures and lessen the likelihood of unfavourable results.

3.2 Predictive Model

The method of developing a model to forecast the likelihood that an undesirable result will occur in a given set of events. To determine whether to evaluate the treatment impact and define benefits, a threshold is utilised that is optimised empirically based on the calculated probabilities.

The first step in developing the prediction model is to create a prefix log by extracting prefixes of length k from each trace. This extraction method guarantees that the training log is representative of the test log. For simplicity's sake, let's say a full trace has seven events; by extracting prefixes up to five events, we get five partial traces ranging from one event to five.

After that, we'll encode every trace prefix into a single, uniformly-sized feature vector. This is a crucial stage in preparing the log data for machine learning. The process is captured in the feature vectors, which are then used to train the machine learning model.

The encoded log is then used to teach a machine learning system to predict the likelihood of an undesirable result. This phase entails sorting the procedure into good and bad categories using classification methods. By running examples through the trained classification method, we may get a predictive model that provides an estimate of the likelihood of an undesirable result.

Predictive process monitoring is a machine learning classification issue with a strong emphasis on outcomes. The resulting predictive model may be used to improve efficiency and lessen the likelihood of unfavourable events.

3.3 Orthogonal Random Forest (ORF)

To estimate treatment effects or Conditional Average Treatment Effects (CATE), the causal model known as an Orthogonal Random Forest (ORF) is implemented. ORF is effective in situations when there are numerous event characteristics with categorical values, which may lead to a feature explosion, while other causal models, such as linear regression or propensity score matching, struggle with high-dimensional feature spaces.

ORF excels at high-dimensional data processing without compromising accuracy or interpretability, which is its fundamental benefit. This is due to the fact that ORF employs a sophisticated machine learning method called a random forest approach, which can handle massive volumes of data and capture complicated correlations between features. Even when there are confounding factors or other sources of bias, ORF is intended to offer estimates of treatment effects that are free from bias.

Another benefit of ORF is that it works well with data that has a mix of continuous and categorical variables, which is common in many real-world contexts. ORF may also deal with missing data, which is prevalent in observational data like as event logs.

In situations with high-dimensional feature spaces, such as event logs, ORF is a helpful tool for predicting treatment effects or CATE. It is a versatile and potent instrument for causal inference since it can deal with continuous and categorical parameters, as well as missing data.

There is a two-stage procedure for calculating CATE using ORF. A propensity score, or the likelihood of getting the therapy given the reported variables, must first be estimated. To estimate the propensity score, a distinct model is used; for example, logistic regression is used and then trained on the training data. The propensity score is then used to weight the observations in the treatment and control groups until their covariate distributions are equal.

The CATE is then estimated using the ORF in the second stage. In order to predict the treatment impact for each observation, the ORF is trained on the weighted data. The difference between the treatment and control outcomes for a specific observation is then used to derive the CATE estimate for that observation.

Overfitting and loss of interpretability are avoided when using ORF to estimate the CATE, which is one of its main benefits. To prevent bias from confounding factors or selection bias, ORF can capture complicated interactions between characteristics and provide reliable estimates of the treatment impact.

The precision of the CATE estimate is affected by the precision of the propensity score estimate and the ORF model, as well as the quality of the data. As a result, it is crucial to use suitable methods for validating the model's performance, such as cross-validation, and to pick the variables carefully for use in estimating the propensity score and in the ORF model.

The input data for the Orthogonal Random Forest (ORF) technique to estimate the Conditional Average Treatment Effect (CATE) should be in the form of a collection of tuples

$$\{(T_i, Y_i, W_i, X_i)\}_{i=1}^{n}$$

for n occurrences. In every case i, the acceptable therapy (T_i) and the measured result (Y_i) should be specified. The preprocessing stage yields T_i and Y_i, which might vary with the intended outcome of the procedure. Potential confounding factors are represented by W_i, while information that achieves heterogeneity is stored in X_i. In this research, we examine an outcome-oriented approach to the loan application process in an effort to improve the approval rate for new loans without disrupting existing ones. We presume that the therapy is known in advance and that the intervention improves the proportion of approved applications. From the encoded log data, we may derive X and W. Case and event characteristics such activity names, resources, and timestamps make up X, a feature vector. We hypothesise that all of these features act as possible confounding variables in the study and help us achieve heterogeneity in the intervention's effect. However, a subject matter expert may spot irrelevant details that should be eliminated from W in order to boost performance.

3.4 Resource Availability Allocator

Decisions on whether or not to treat a given case and the best time to intervene may be made with the help of a combination of a predictive model, a causal model, and a resource allocator. But there's a price to pay for automatically intervening in instances, so that's something to think about. If the predictive model predicts a likelihood greater than some threshold, then comparing the expenses with and without the intervention will help us evaluate the entire benefit of the intervention.

Indirect expenses, such as lost work time while undergoing treatment, are also included part of an intervention's total cost. Direct costs include things like the price of pharmaceuticals, medical equipment, and staff. In contrast, the costs of doing nothing take into account everything that may happen as a result of not intervening, such as further medical bills, time off work, and expensive care needs down the road.

Decisions on whether or not to intervene become more pressing when the expense of doing so exceeds the expected benefits. The CATE of the intervention and the likelihood of the undesirable event must both be taken into account here. The CATE of implementing an intervention in a specific scenario may be estimated with high precision when the predictive and causal models are used together.

In general, we are able to make educated judgements regarding when and whether to administer therapy by carefully comparing the costs of intervention in a specific situation against the advantages it would provide. This allows us to maximise the overall gain while minimising the expenditures.

Empirical optimisation of the threshold, using methods like cross-validation and grid search, is one option when it is not feasible to determine the optimal threshold in advance. To attain the greatest feasible benefit for the instances, we need to choose the cutoff that maximises our unique objective function. In order to guarantee a high possibility of the undesirable result in each given instance, the threshold is used.The threshold is adjusted to improve the chances of detecting situations that might lead to undesirable outcomes.

3.4.1 Interventionless Cost

$$cost(case_{id}, T_{i=0}) = prob_{uout} * cost_{uout} \tag{3.1}$$

The cost in the event that the intended result is not attained and the intervention is not implemented may be computed using the equation 3.1. This is represented by i having a value of 0 (zero). The equation has two variables, $prob_{uout}$ representing the expected likelihood of the undesirable result and $cost_{uout}$ representing the cost associated with its occurrence. The equation, in other words, calculates the opportunity cost of doing nothing when a predictive model indicates a high possibility of an adverse result.

3.4.2 Intervention Cost

$$cost(case_{id}, T_{i=1}) = (prob_{uout} - CATE_1) * cost_{uout} + case_{T_1} \quad (3.2)$$

To solve this issue, we need to weigh the anticipated causal impact ($CATE_1$) of a potential intervention (T_i) against the expenses involved with implementing that intervention ($cost_{T_1}$ and $cost(case_{id}, T_i = 0)$). i = 1 in equation 2 represents the cost incurred if the intervention is carried out and the case concludes with an unfavourable result. The likelihood of the undesirable result is also known to us as the estimated ($prob_{uout}$).

In order to decide whether or not to implement the intervention, we need to assess the benefit of doing so by applying $T_{i=1}$ to the $case_{id}$ with the largest savings. Equation 3 demonstrates how to achieve this with the help of the 3.1 and 3.2 equations. The first component of our issue is whether or not to handle $case_{id}$, and this is determined by the gain.

Given the estimated causal effect ($CATE_1$), associated costs ($cost_{T_1}$ and $cost(case_{id}, T_{i=0})$, and estimated probability of an undesirable outcome ($prob_{uout}$), we need to determine the gain from applying $T_{i=1}$ to $case_{id}$ that results in the greatest cost reduction, and use that gain to determine whether or not to treat $case_{id}$.

3.4.3 Gain

$$gain(case_{id}, T_{i=1}) = cost(case_{id}, T_{i=0}) - cost(case_{id}, T_{i=1}) \quad (3.3)$$

Different distributions for treatment time are desirable because of their potential effects on resource allocation's efficacy and efficiency. In circumstances when the therapy is conventional, a predetermined length of time may be appropriate, but in many others, the length of time needed to achieve the desired effect may vary from patient to patient. In certain circumstances, it may be inefficient to use a constant time.

Treatment time T_{dur} may be more accurately represented by a normal or exponential distribution. A bell-shaped curve, with most instances clustered around the mean and fewer outliers, is assumed in the normal distribution. The probability distribution of the duration is assumed to decrease exponentially with increasing duration in the exponential distribution.

Depending on the given circumstances and data, a different distribution may be chosen for T_{dur}. Consider using a normal distribution if you have information on the range of possible treatment times, for instance. But an exponential distribution may be more suitable if there is evidence that the duration may follow one, such as if the medication is known to lose efficacy with time.

In conclusion, a normal or exponential distribution may better represent the variability in treatment duration and increase the efficiency and efficacy of resource allocation than adopting a set duration, which may be the simplest approach to describe T_{dur}.

Taking into account the available means of implementing $T_{i=1}$, we provide a ranked list of the highest possible profits for each active instance $case_{id}$. As soon as a resource becomes available, it is immediately reserved for T_{dur} and given to the case with the greatest potential benefit.

Basically, we keep a stack of active instances with a maximum gain connected with each one. In the event of a resource being available, we prioritise the case in the queue that stands to benefit the most from its allocation. This guarantees that the most promising cases will be given top consideration.

After we have assigned a resource to an appropriate case, we will hold it for the period of time indicated by the variable T_{dur}.. As a result, this precludes other instances from utilising the same resource during that period, which guarantees that each case has exclusive access to the resource it requires.

Overall, this method improves our efficiency and effectiveness by making the most of every available asset. We can make the greatest use of our resources and get the best results if we prioritise cases based on the profits they can provide and then allocate those resources appropriately.

Chapter 4

Proposed Experimental Setup and Evaluation Metrics

4.1 Dataset

We evaluated our methodology by utilising a real-life event log, more precisely the $BPIC2017$ log, which is representative of the loan origination process. Each entry in the log represents a separate request for a loan, and each request has a result—either one that was sought or one that was not desired. The result that is wanted is for the bank to give the client a loan, and for the client to accept the offer and sign the paperwork; the outcome that is undesirable is for the bank to cancel the application, or for the customer to refuse the offer. There are 31,413 apps and 1,202,267 events included inside the event log.

We employed all of the available qualities in the log as input in order to gain the greatest performance possible from the predictive and causal models. This included characteristics such as the number of offers and event numbers, as well as temporal information such as the hour of the day, day of the month, and month. We recovered prefixes with lengths that were less than or equal to the 90th percentile of the case lengths in the log. We then encoded these prefixes using aggregate encoding to produce fixed-size feature vectors in order to eliminate any potential bias caused by lengthy cases.

Additionally, we carried out extra preprocessing stages in order to determine the outcomes of instances in accordance with the activities that were completed. instances that conclude with "A Pending" events are seen as having favourable results, but instances that conclude with "A Denied" or "A Cancelled" events are regarded as having undesirable consequences that call for action. We determined the necessary intervention to minimise failed loan applications based on the report submitted by the winner of the BPIC challenge [**ChallengeWinner**]. According to the findings of the analysis, the likelihood of reaching an end state labelled "A Pending" improves as more offers are made to customers. As a result, circumstances in which there was only a single offer were taken into consideration (T = 1), whereas cases in which there was more than one offer were disregarded (T = 0).

4.2 Proposed Experimental Setup

We suggest employing an XGBoost model to estimate the likelihood of negative case outcomes. our model has demonstrated promising results in a variety of classification issues [8], including outcome-oriented PPM [7], so we believe it would be useful in our application as well. The learning rate, the subsample, the maximum tree depth, the colsample by tree, and the minimum child weight are some of the hyperparameters that may be changed in this model. We are going to utilise the orthogonal random forest (ORF) that is included in the EconMl package to estimate the CATE, and then we are going to modify its settings, such as the minimum leaf size, maximum depth, subsample ratio, and lambda regularisation.

We will split the dataset into a 60-20-20 training, validation, and testing split, respectively, using timestamps to simulate real-life situations. Assuming that the estimated CATE is accurate, we will allocate resources to decrease a case's probability of a negative outcome. We will compare our approach to a purely predictive baseline proposed in [1], where resources are allocated based on the highest $Prob_{uout}$ instead of the cases with the maximum gain, considering the CATE as the new gain achieved from treating cases.

To summarize, our proposed approach involves the following steps:

- In order to predict the likelihood of negative case outcomes, it is possible to train an XGBoost model and fine-tune its hyperparameters using the training data.
- To estimate the CATE, we may utilise the ORF from the EconMl package and fine-tune its settings with the help of the validation set.
- Consider the predicted CATE as the additional benefit from treating patients and allocate resources accordingly.
- Compare our method, which prioritises the greatest $Prob_{uout}$ when allocating resources, to the predictive baseline suggested in [**basepaper**].
- Using timestamps to model real-world scenarios, divide the dataset into three equal parts for training, validation, and testing.

Chapter 5

Results

Keeping in mind the T_{dur} variability, we present the results of our proposed method, which examined the effects of available resources on aggregate gain and the proportion of treated cases. The variation in total benefit and proportion of treated cases as resource availability (C1) increases is depicted. Both indicators improve when additional resources become available. If the available resources exceed 50%, however, the total benefit nearly doubles. Comparing cases in which the resource rate is above and below 50%, for instance, it is evident that the overall gain increases considerably when the resource rate is above 50%. This is because once more than half of the available resources are made available, a larger number of cases can be treated.

We conduct experiments using the fixed, normal, and exponential distributions for T_{dur}. As demonstrated, the fixed distribution yields a greater net gain than the normal and exponential distributions because there is less fluctuation in the allocation of resources across instances that need intervention. This is in contrast to the normal and exponential distributions, which both exhibit a decreasing amount of variability. As a consequence, the diversity in treatment time has a significant impact on the net gain.

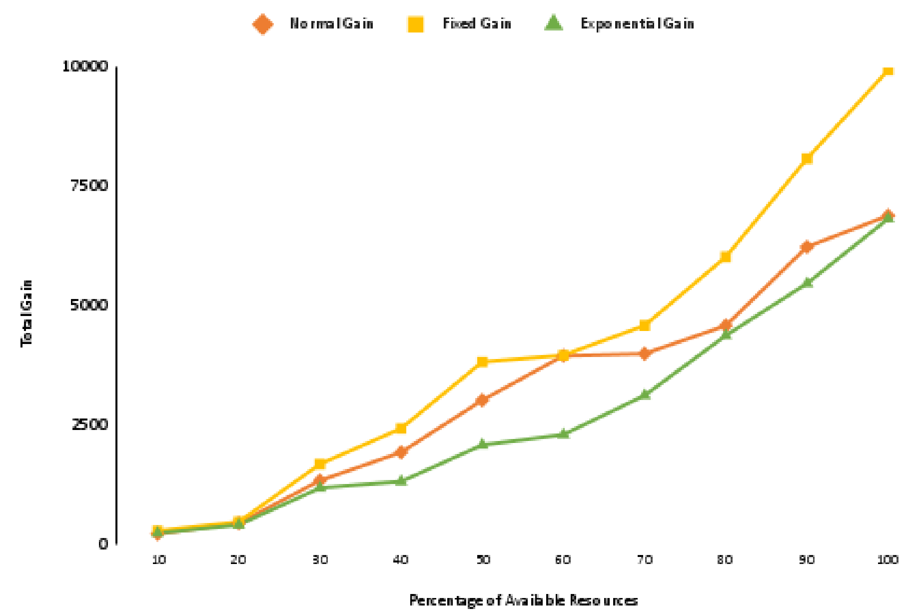

Figure 5.1: C1 AND C2 : AVAILABLE RESOURCES VS TOTAL GAIN

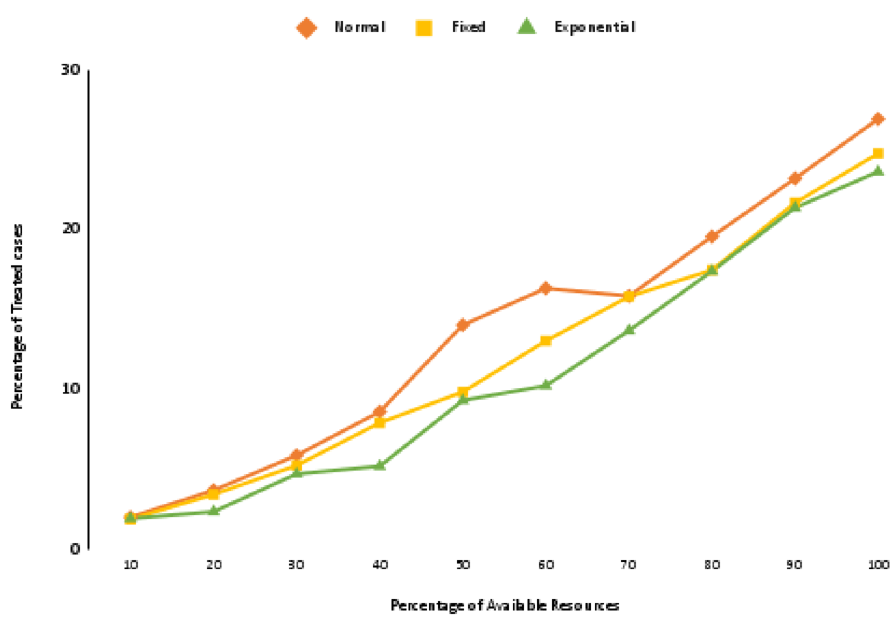

Figure 5.2: C1 AND C2 : AVAILABLE RESOURCES VS TREATED CASES

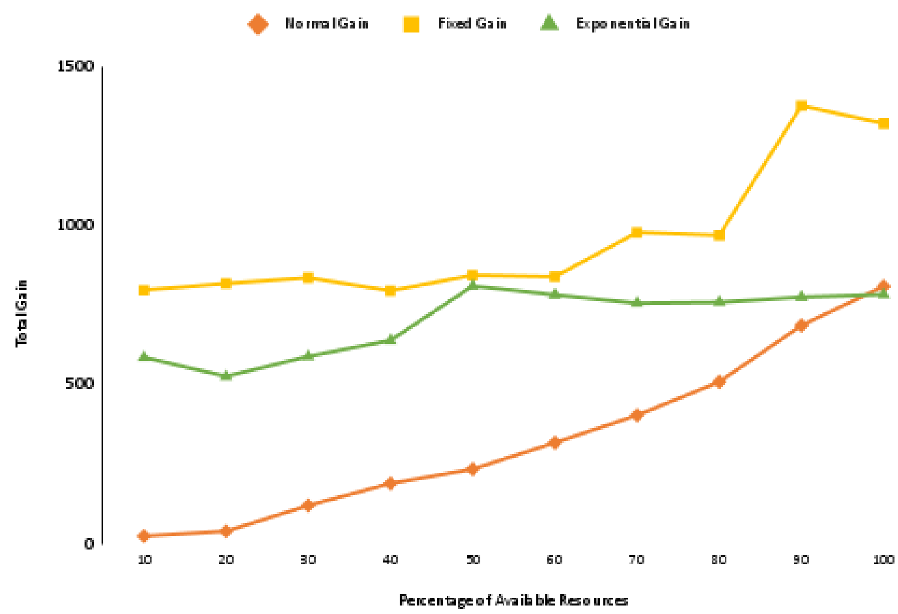

Figure 5.3: C3 : AVAILABLE RESOURCES VS TOTAL GAIN

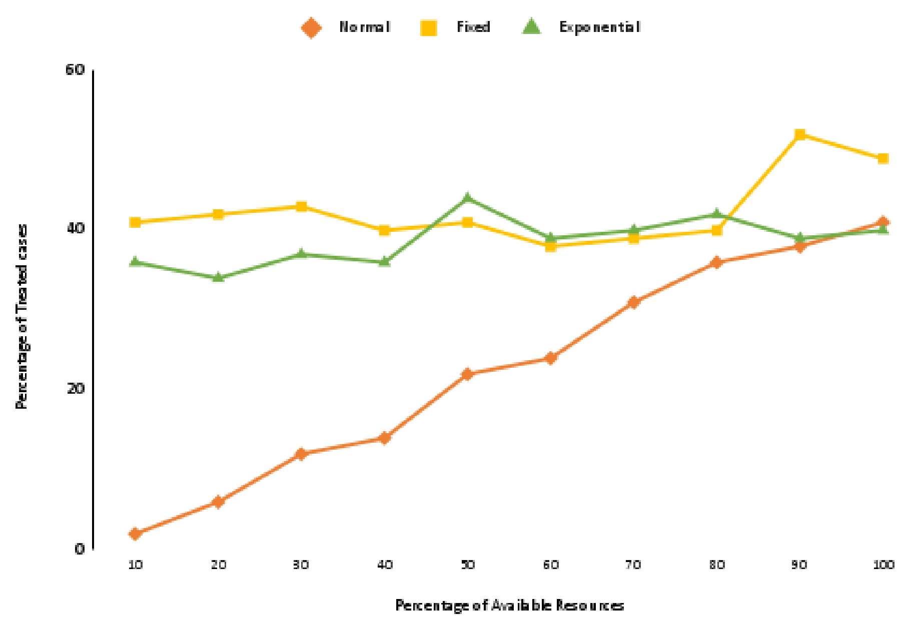

Figure 5.4: C3 : AVAILABLE RESOURCES VS TREATED CASES

In order to provide a solution, we distribute our efforts among the situations that have the greatest $prod_{uout}$ rather than those that have the largest gain. The CATE is something that we believe to be a new gain that comes from treating instances. As a result, we need a threshold in order to decide whether or not to act based on the $prod_{uout}$. Setting a threshold may be done in two different ways: first, depending on a specific threshold, such as 0.5, we trigger an intervention if there are available resources and the undesirable consequence is over the provided threshold. The second choice is to use an empirical threshold. In this method, the authors calculate an ideal threshold based on historical data. We experimented with a number of different thresholds, as indicated. However, when the T_{dur} distribution is taken into account, the findings are very different. When it is equal to 0.5, the normal distribution provides a greater overall benefit than other thresholds do. Even though it is equal to 0.6, the exponential distribution results in a greater overall benefit. In addition, when it equals 0.7, the winner is the fixed distribution. You can see the results of optimizing the threshold in the supplemental material, where we also illustrate how the overall benefit varies with respect to various thresholds and different T_{dur} values.

When compared to a baseline that is solely predictive, we find that our methodology always results in a greater net benefit while using the same amount of resources as the other option. For instance, in the context of a constant distribution, addressing just 25% of instances with our technique results in a net gain of 10,000, but in the predictive method, treating twice as many cases as our approach does (50% of cases) only results in a net gain of 1,400. As a result of this, it seems that using causal inference with predictive modelling may be able to improve the effectiveness of prescriptive process monitoring techniques.

Chapter 6

Conclusions and scope for further work

We designed a prescriptive monitoring system in order to optimize a net gain function with limited resources. Predictive models identify the situations most likely to fail (and incur high costs), whereas causal models identify the conditions most likely to benefit from an intervention. Case interventions are funded according to the projected benefits of the models. It outperforms the current gold standard, which relies solely on a prediction model, in terms of resource allocation and processing fewer occurrences, according to preliminary research.

If the suggested approach indicates that treating this specific case will result in a greater net benefit than treating other similar cases, an intervention will be initiated. Consequently, treatment may need to be delayed. When it comes to loan origination, it is often preferable to engage a consumer two days rather than one day after they make an offer. The activity may not be sufficient to accomplish the desired outcome on its own. Timing of interventions is crucial. As a result, postponing action until conditions alter can reduce anxiety and possibly increase reward. If the projected benefit is positive and there is no greater net gain, our technique would initiate the intervention "call customer" one day after the offer if the projected benefit is positive. When allocating resources, it may be necessary to weigh the predicted future benefit of acting on a case against the estimated net benefit of acting on the same case immediately. The proposed strategy may be used in conjunction with a method that optimizes the precise timing of a response to a specific event. Time limits for case interventions may warrant further investigation. It is a waste of time to engage a customer about a loan offer they have already rejected or have not received.

A further deficiency of the proposed strategy is that it necessitates a specific type of intervention. A second loan offer, phone call, etc. may be necessary. Numerous interventions are incorporated into prospective employment pathways.

Chapter 7

References

[1] Fahrenkrog-Petersen, S.A., Tax, N., Teinemaa, I., Dumas, M., de Leoni, M., Maggi, F.M., Weidlich, M.: Fire now, fire later: alarm-based systems for prescriptive pro- cess monitoring. arXiv preprint arXiv:1905.09568 (2019).

[2] Metzger, A., Kley, T., Palm, A.: Triggering proactive business process adaptations via online reinforcement learning. In: BPM. pp. 273–290. Springer (2020)

[3] Weinzierl, S., Dunzer, S., Zilker, S., Matzner, M.: Prescriptive business process monitoring for recommending next best actions. In: BPM Forum. vol. 392, pp. 193–209. Springer (2020)

[4] van Dongen, Boudewijn (2017): BPI Challenge 2017. 4TU.ResearchData. Dataset. https://doi.org/10.4121/uuid:5f3067df-f10b-45da-b98b-86ae4c7a310b. ChallengeWinner

[5] Povalyaeva, E., Khamitov, I., Fomenko, A.: Bpic 2017: density analysis of the interaction with clients. BPI Challenge (2017).

[6] Teinemaa, I., Dumas, M., Rosa, M.L., Maggi, F.M.: Outcome-oriented predictive process monitoring: Review and benchmark. ACM TKDD 13(2), 1–57 (2019).

[7] Fernandez-Delgado,M.,Cernadas,E.,Barro,S. ,Amorim,D.:Doweneedhundreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. 15(1), 3133–3181 (2014).

[8] Kubrak, K., Milani, F., Nolte, A., Dumas, M. (2022, September 29). Prescriptive process monitoring: Quo vadis? PeerJ Computer Science, 8, e1097. https://doi.org/10.7717/peerj-cs.1097

[9] Bozorgi, Z. D., Teinemaa, I., Dumas, M., Rosa, M. L., Polyvyanyy, A. (2023, March). Prescriptive process monitoring based on causal effect estimation. Information Systems, 102198. https://doi.org/10.1016/j.is.2023.102198

[10] Weytjens, H., De Weerdt, J. (2022, August). Learning uncertainty with artificial neural networks for predictive process monitoring. Applied Soft Computing, 125, 109134. https://doi.org/10.1016/j.asoc.2022.109134

[11] Kim, H. (2023). Double/Debiased Machine Learning for Static Games with Incomplete Information. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4377695

[12] Han, S., Kim, H., Lee, Y. S. (2020, July 2). Double random forest. Machine Learning, 109(8), 1569–1586. https://doi.org/10.1007/s10994-020-05889-1

[13] Chang, N. C. (2020, February 4). Double/debiased machine learning for difference-in-differences models. The Econometrics Journal, 23(2), 177–191. https://doi.org/10.1093/ectj/utaa001

[14] CONG, Z., Liu, J., Manchanda, P. (2021). The Role of “Live” in Livestreaming Markets: Evidence Using Orthogonal Random Forest. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3878605

[15] Xu, J., He, J., Gu, J., Wu, H., Wang, L., Zhu, Y., Wang, T., He, X., Zhou, Z. (2022, January 15). Financial Time Series Prediction Based on XGBoost and Generative Adversarial Networks. International Journal of Circuits, Systems and Signal Processing, 16, 637–645. https://doi.org/10.46300/9106.2022.16.79

[16] Park, G., Song, M. (2023). Optimizing Resource Allocation Based on Predictive Process Monitoring. IEEE Access, 11, 38309–38323. https://doi.org/10.1109/access.20