TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

From predictions to recommendations

Citation for published version (APA):
Verdaasdonk, M. J. A., & de Carvalho, R. M. (2022). From predictions to recommendations: Tackling bottlenecks and overstaying in the Emergency Room through a sequence of Random Forests. *Healthcare Analytics*, *2*, Article 100040. https://doi.org/10.1016/j.health.2022.100040

Document license:
CC BY

DOI:
10.1016/j.health.2022.100040

Document status and date:
Published: 01/11/2022

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 14. Jul. 2024

Healthcare Analytics 2 (2022) 100040

Contents lists available at ScienceDirect

Healthcare Analytics

journal homepage: www.elsevier.com/locate/health

From predictions to recommendations: Tackling bottlenecks and overstaying in the Emergency Room through a sequence of Random Forests

Mike J.A. Verdaasdonk, Renata M. de Carvalho *

Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands

ARTICLE INFO

Keywords:
Process-aware recommendations
Bottlenecks identification
Inter-case features
Random Forest
Healthcare

ABSTRACT

One of the goals to improve the quality of care in hospitals is to set a maximum of four hours for patients to be diagnosed and/or receive acute care in the Emergency Room (ER). Unfortunately, this is not always true and some patients *overstay*. The aim of this work is threefold: (1) to identify which patients will overstay during their admission to the ER; (2) to identify which (pair of) activities might heavily influence the time spent in the ER; and (3) to recommend actions to reduce such time. For that, a sequence of insightful supervised prediction models for generating recommendations is proposed. The method provided makes it possible to generate useful/actionable recommendations for problematic patients based on activities. State of the art techniques did not manage to generate recommendations at the arrival of the patient and/or did not take the interplay between patients into account.

1. Introduction

Every year, there are about 2 million visits to the Emergency Room (ER) in the Netherlands [1]. During an ER visit, patients are diagnosed and might receive acute care. In order to improve the quality of care, it is expected for patients to not spend more than 4 h in the ER [2]. Patients that need to receive more attention and/or that would need to spend time for a treatment should be transferred to a specialized department. Although this 4-hour rule is applied, there are still situations were patients stay longer than 4 h in the ER. This phenomena is called overstaying and therefore these patients are called overstaying patients. Overstaying can be a result of potential bottlenecks in hospital settings [3]. Such bottlenecks might happen due to limited capacity, knowledge and facilities for caring for patients, and/or because it is extremely difficult to predict and plan activities that should take place [3].

Process Mining [4] is a technique that tries to combine the knowledge of Business Process Management with Data Science. It has been applied to healthcare to analyze and solve different problems within this domain. Senderovich et al. [5,6] proposed a method to take into account inter-case features for predicting patients throughput time. de Leoni et al. [7] explored a similar approach with the goal to recommend next activities for patient to get the lowest throughput time possible. For this approach, an initial set of activities for a specific patient (a partial trace) should be known before the recommendation can be made.

Notwithstanding, to advance in a solution for the overstaying patients, this paper proposes a method that will consider both inter-case features as bottleneck activities. In short, we would like to answer the following research question: "Which method can be used for formulating effective recommendations for patients whose are likely to overstay?". As inter-case features, for each arriving patient, we consider both the whole set of patients present in ER at that moment, and the patients present in the ER that are of the same specialism. For detecting bottlenecks, we make an analysis on the data to identify them. Furthermore, we also consider the input from the domain knowledge, through interview with the ER staff. As said, the goal is not only to identify (or predict) the overstaying patients, but also to make recommendations that would reduce their ER-time. For that, we focus our recommendations on activities/aspects that can be influenced by the hospital staff.

In this research, we used a dataset from a Dutch hospital containing information about patients that visited the ER from January 2019 until September 2020. The dataset contains personal anonymized information, as well as events that occur throughout the whole stay of the patient in the ER. We consider that patients leave the ER when they are either sent back home or when they are moved to another department.

The remainder of this paper is structured as follows. Section 2 describes the required preliminaries to understand the report. Section 3 compares the related work. Section 4 introduces the method. Sections 5 and 6 introduce the two different aspects implemented in the used

* Corresponding author.
E-mail addresses: m.j.a.verdaasdonk@student.tue.nl (M.J.A. Verdaasdonk), r.carvalho@tue.nl (R. M. de Carvalho).

https://doi.org/10.1016/j.health.2022.100040
Received 21 October 2021; Received in revised form 16 February 2022; Accepted 8 March 2022

2772-4425/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

method while Section 7 evaluates this method. Section 8 concludes the paper.

2. Background

In this section, the required preliminaries are provided. We will provide the definitions of *inter-case features* and *bottlenecks*, shortly introduce the prediction model *Random Forest*, and present the *Response* constraint over activities, from the Declare modeling language [8].

2.1. Inter-case features

In process mining, a case defines the perspective to which we look at the process. Each event is linked to a case and all the events of a case are ordered in time to form a trace, denoted by $\langle a, b, c\rangle$ to indicate that three activities (a, b and c) happened to this case in this order. To be useful, a prediction model needs information about a case. This information can be delivered as a number of variables (a.k.a. features). There are two types of features, namely *intra-case* and *inter-case*. Intra-case features [9] only focus on the case (e.g. patients) itself, e.g., the features *age* and *diagnosis*. Inter-case features [10] focus on the context of cases running at the same moment, e.g., a feature defining the number of other cases of the same type at some moment in time. Thus intra-case features describe the case in an isolated setting, where inter-case features describe the context of other cases, which may influence the case in question.

2.2. Bottlenecks

Bottlenecks are related to the time of specific components of a case [4]. Bottlenecks might be caused by machines, people or activities. This paper focuses only on bottlenecks caused by activities. Activities which deviate a lot in execution time and therefore can cause a too long throughput time for different cases are seen as bottleneck activities. Furthermore, activities that take most of the time of a case can be seen as bottleneck activities [4].

2.3. Decision tree and Random Forest

A decision tree [11] is a machine learning technique that can be used for both classification and regression. A decision tree consists of a tree with decisions in order to reach a specific result. Such decisions are based on features/attributes and conditions from the data. The decision tree algorithm is easy to understand and interpret. However, a single tree might not be sufficient to produce effective results for a complex problem.

Random Forest [12] is a tree-based supervised learning model that can be used for prediction. This prediction model tend to find the feature with the best split of groups recursively. These groups are based on the resulting value of the supervised learning method, e.g., resulting time in the process of a patient. Furthermore, Random Forest is based on a combination of decision trees and it looks for a majority vote to find the best feature per split. Another characteristic of the Random Forest is the de-correlation of the trees, where at each split the tree only looks at a random sample of features instead of the total number of features. These two characteristics results in reduction of variance of the final prediction model. In this paper, we use Random Forest as predictor model due to the easy understandability and interpretability of the final decision tree.

2.4. Response constraint

The response constraint, from the Declare modeling language [13], restricts the occurrence of two activities in a timely manner. For every case, a response constraint can be satisfied or not. The response constraint is satisfied if, for two chosen activities, one chosen activity is always eventually followed by the second chosen activity. For instance, consider the response constraint *response(a,b)* (formalized in Linear Temporal Logic as $\square(a \rightarrow \diamond b)$). For every occurrence of a, it is expected that a b will eventually follows. The traces $\langle a, c, b\rangle$ and $\langle b, a, c, b\rangle$ would satisfy the response constraint, as for each a in the trace, there is eventually a b after it. In contrast, the cases $\langle a, c\rangle$ and $\langle a, c, b, a\rangle$ would not satisfy the response constraint. Therefore, the response constraint will be used as a reference to bottlenecks that relate two activities.

3. Related work

The emergency department overcrowding is a patient safety issue across countries and a major contribution to such issue is the long waiting times for inpatient hospital admission or discharge. Long waiting times can also affect patient satisfaction and quality of care. In this context, several researches [14–16] try to predict the waiting time of patients in the emergency department as a way to reduce the negative impact on operational efficiency, patient safety and quality of care. Such researches rely on the fact that patients satisfaction is improved with patients receiving accurate information about waiting times, and/or on actions from the staff to mitigate the problem (e.g., increasing medical resources, improving hospital bed access).

Predicting hospital admission at the time of triage [17–20] is also considered a way to reduce the time spent in the emergency department, as the hospitalization process can be initialized in parallel to the process within the emergency department. It can also be used to prioritize patients that should receive care [21] or as a way to indicate which patients have a clear outcome and could quickly be either discharged or sent to another specialized department [22].

In Process Mining, Senderovich et al. [5,6] illustrated the importance of inter-case features for predicting patient duration in a hospital setting. They make use of distances based on time and the order of activities to define an inter-case feature per patient. Where the inter-case features used in this paper, are related to the total number of patients and the number of patients with the same specialism when a patient arrives. Specialism are defined when patients arrives at the Emergency room. Both papers take the order of activities into account for predicting patient duration, although there is no explicit focus on activities which can be seen as bottlenecks in [5] and [6]. This article has a focus on bottleneck activities.

de Leoni et al. [7] used, same as in this paper, a complete process-aware system for predictive and prescriptive analytics. They make use of techniques to predict and recommend potential best next activities to be taken based on previous activities and key performance indicators scores. de Leoni et al. [7] applied their technique on running cases, where in this paper the prediction and recommendations can be made at the arrival of the patient. Furthermore, they make use of a complex transition system, which make it hard to illustrate which activities can be seen as problematic. Therefore it is hard to find the activities to focus on, for improving the process of problematic cases. Where in this paper the prediction model Random Forest is used to get insights in the importance of features.[1] de Leoni et al. [7] predicted the next activities based on only intra-case features, where this article takes into account both intra- as inter-case features for prediction.

4. Overview of the approach

As briefly mentioned in the introduction, we aim at making a prediction of overstaying for patients at the start of the process, i.e., without any activity performed in the Emergency department except from triage color determination, which is done at the start of the process. Table 1 summarizes the features used in this paper (referred as "patient data" throughout this paper).

[1] Note that activities are presented as features in this paper.

Table 1
Features present in patient data.

Type of features	Features
Identification	AdmissionCode, PatientCode
Patient demographics	Age, Gender, BMI
Registration	RegistrationDate, RegistrationTime, ArrivalDate, ArrivalTime, TriageDate, TriageTime
Arrival	ComingFrom (e.g., GP referral, accident site), TransportationMode
Triage	TriageColor, Specialism, Complaints, PulseRate, RespiratoryRate, OxygenLevel, BloodPressure, Temperature

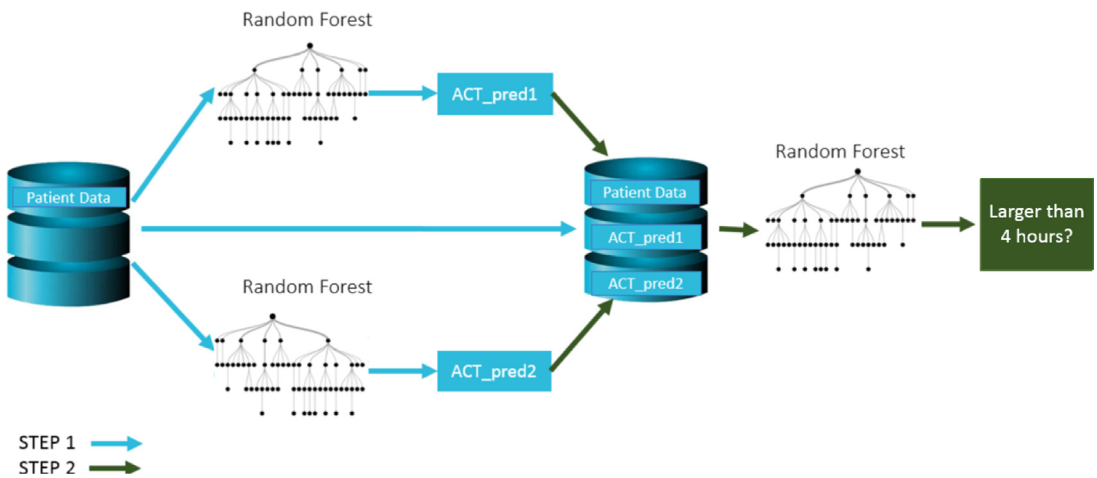


Fig. 1. Method illustrated.

In addition, we also have data about the activities that happened to each patient while in the emergency department (e.g., treatments that were started, lab tests that were required/performed, medications prescribed). This event data is first analyzed to determine potential bottleneck activities, i.e., activities which cause bottlenecks in throughput time. These analyses are done in Disco[2] and discussed with hospital personnel. The domain knowledge is used in this initial phase to validate the bottlenecks identified while analyzing the event data. Then, for each of these identified bottleneck activities and other influences of the throughput time of patients, such as hospitalization, a different Random Forest model is trained with patient data as the only input in order to predict if the activity in question will happen to a certain patient.

Then, these predictions are combined with the patient data. This "extended" data is used to train a new Random Forest model for predicting overstaying of patients. These steps are illustrated in Fig. 1. Note that, in this figure only two bottleneck activities predictions are made, however this method is applicable for less or more predictions. Finally, recommendations are made for patients which are predicted as overstaying patients. These recommendations are based on the activities which are predicted to be performed on a patient during his/her stay in the Emergency department. Note that, we assume that recommendations that focus on activities performed by the hospital will decrease the running time of that activity.

4.1. Prediction method used

This paper makes use of Random Forests to get insights in which features and/or activities are important for predicting activities and overstaying. Note that, prediction models like Neural Networks could give better results, however they are black box models, which makes it hard to gather useful insights on input features and therefore useless for making recommendations in our method. Moreover, Random Forest is used because of its characteristic of low bias and its characteristic of de-correlation, which results in lowering the variance.

All Random Forests in this paper make use of balanced weight, because of the difference in number of patients per output variable group, e.g., there are 8688 patients with activity *consult SEH* in their execution trace (the trace of all activities performed from the admission of the patient until his/her discharge) and 41395 patients without this activity.

4.2. Feature selection

For the selection of features used for the prediction model, we make use of an analysis and domain knowledge. First, a Random Forest with patient features (e.g., *age*) as input and binary output feature overstaying is trained to check which variables are most important for such prediction. For this, we used feature importance to check which variables (features) have influence on overstaying. Second, these variables are discussed with the hospital for validation to inject domain knowledge in the feature selection step. Note that domain knowledge is mostly used for taking into account extra features and less on removing features. Pointing out unimportant features for prediction is a hard task for domain experts.

4.3. Prediction method measures

For measuring the quality of the prediction model, i.e., Random Forest, two different measures were used. For predicting activities, pairs of activities, and if a patient will be hospitalized the accuracy is measured. Accuracy is used to measure the percentage of correct predictions. This measure is chosen above others because there is no difference in impact of false positives and negatives. For predicting overstaying, we make use of the measure Negative prediction value (NPV), which measures the percentage of true negatives against the total number of predicted negatives. This measurement is used because false negatives are more problematic than false positives; false positives result in focusing on patients who eventually are not overstaying, where

[2] https://fluxicon.com/disco/

false negatives results in not focusing on patients, who eventually overstay and therefore who are not in line with the research problem. The negative predictive value illustrates how well we can predict if a patient is not going to be overstaying.

5. Identifying bottleneck activities

Overstaying can be a result of different aspects. Fabrizio et al. [3] said that potential bottlenecks in hospital setting can be from a limited capacity of beds and medical personnel, planning and activities. This paper focuses on activities which cause bottlenecks.

5.1. Method

This section focuses on the activities that can influence the overstaying of patients. These activities are used for prediction of overstaying and therefore, pointing out the patients who are likely to stay longer than 4 h in the Emergency department. To find these type of activities, we made use of analyzing techniques combined with domain knowledge. First, the process of patients that stayed longer than 4 h is compared with the process of patients who stayed less than 4 h in the emergency department. This is done using the tool Disco. This tool makes it possible to get visible insights in the process steps and duration of specific activities or pair of activities. Second, the findings are discussed and validated with the hospital to inject domain knowledge to potential bottlenecks activities or pair of activities. We focus on finding activities which influence the possibility of overstaying.

5.2. Bottleneck activity selection

The selection of bottlenecks starts with analyzing the aggregated process of both the group of patients staying above and below 4 h. During the analysis, the focus is to find activities which have a difference in occurrence and/or duration between both different groups. Those activities can be seen as bottleneck activities [4]. The comparison between duration is focused on the median of activity duration, because this measure is better prevented against outliers than the mean.

5.3. Pairs of activities as bottlenecks

As mentioned, we also discussed the findings about bottlenecks with domain experts. From them, we understood that some pairs of activities might also cause bottlenecks. Analyzing them further, we say that a pair of activities (e.g., A, B) is a bottleneck when the *response(A,B)* constraint from Declare is satisfied and the time difference between A and B is significantly higher for overstaying patients. In this approach, the pair of activities A and B satisfies to the response constraint if and only if activity A is eventually followed by activity B.

For a very few number of patients, we noted that activity A needed to be executed again (after B), but with no need to re-execute B. For these cases, although they do not satisfy the constraint *response(A,B)*, we still considered them as "satisfying" and we considered the time between the first execution of A and the execution of B. For example, if we have trace σ: $\langle a, b, c, a \rangle$, the response constraint $\square(a \rightarrow \diamond b)$ would be violated for trace σ, because the second a is not followed by a b. However, according to our domain experts, we considered the second a as a way to correct mistakes in the first a, rather than another different execution. According to this rationale, these cases would still be considered as satisfying the response constraint.

Each bottleneck (pairs of) activities will be predicted by a different Random Forest model, receiving the patient data as input. This is depicted as "STEP 1" in Fig. 1.

6. Considering inter-case features

State of the art techniques for prediction of process-related outcomes, e.g., running time of a patient, usually only make use of features related to the case itself, e.g., the feature *age* [5]. These features are called intra-case features. However, these prediction methods are missing the interplay between cases in many situations [5]. For example, if the limited capacity of beds in a hospital is exceeded because of the number of patients in a hospital, it is likely that a new patient has to wait longer for being allocated to an available bed. Where in the same situation with less patients, it is more likely that the same patient is being allocated to an available bed faster. When a patient is allocated to a bed faster and follows the same process, the throughput time of that patient is shorter. Same holds when there is no doctor available because of the large number of patients in the hospital.

6.1. Inter-case features used

In this paper intra- and inter-case features are taken into account to ensure the interplay between cases in the hospital. Two different inter-case features are used in this approach. The first inter-case feature is defined as the total number of patients in the hospital when a new patient arrives. For this feature, we check the arriving time of the new patient and check how many other patients have arrived and are still in the ER. The second inter-case feature is defined as the number of patients with the same specialism in the ER. Same as the previous feature, we check the arriving time of the new patient and check how many other patients have arrived and are still in the ER. However, the predefined specialism (which is registered at the arrival of the patient) is also taken into account. Therefore, only patients with the same specialism as our new patient are added to this inter-case feature.

7. Evaluation

In this paper, we aim at answering the research question: "Which method can be used for formulating effective recommendations for patients who are likely to overstay?". We will check if this question is successfully tackled with the use of the Negative Prediction Value (NPV) score, which is related to the quality of prevention for missing overstaying patients. Furthermore, we will test the effectiveness of possible recommendations by checking how many patients (out of the test set) will be successfully transformed from an overstaying patient to a non-overstaying patient when recommendation would be applied. This is done by comparing the means of (pair of) activities for overstaying and non-overstaying groups. Then, these differences will be used for decreasing the throughput time of predicted overstaying patients for the activities which are true positives (i.e., activity is eventually performed and was predicted to be performed). Finally, the resulting throughput time (after decreasing) of all overstaying patients will be checked against the threshold of 4 h.

7.1. Data explanation and pre-processing

The data used in this paper come from the a Dutch hospital and consists of patient data and activities performed on patients in the Emergency department from January 2019 until September 2020.

The used patient data consists of the features presented in Table 1. The timestamps are processed as *yyyy-mm-dd hh:mm, e.g., 2020-09-30 14:33*. Furthermore, the arrival time of a patient is transformed into an integer and is called *presentation time*, e.g., arrival time 8:30 is transformed into the integer 3, which is [07:12:00–09:36:00]. Note that the date is not relevant for transforming the arrival time into an integer. Fig. 2 shows the categories we created in the x-axis and the number of patients in each category in the y-axis.

The activities are process-related events to the patient treatment in the ER, e.g., *Consult op de SEH*, which means that a consult is made

Table 2
All scores of (sequence of) Random Forest(s).

	Baseline(Base)	*Base + IC*	*Base + pred.act.*	*Base + IC + pred.act.*	*Base + IC + perf.act.*
NPV	0.80	0.80	0.80	0.80	0.85
Accuracy	0.62	0.62	0.59	0.61	0.68
Execution Time (in sec)	330	373	1693	1994	390

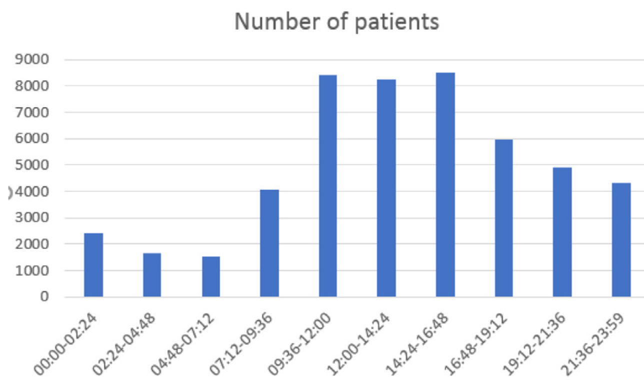


Fig. 2. Arrival pattern of patients over time.

for a patient. All these activities are provided with a starting time, also processed as *yyyy-mm-dd hh:mm*. All activities with corresponding timestamps and registrations ID of the patient are used for a different analyses.

Our dataset has 68184 cases and 1048012 events. In discussion with the hospital, only patients with the *age* of 18 years and older are selected for this approach. Furthermore, the original datasets consist of timestamps in two different columns, i.e., Date (yyyy-mm-dd 00:00:00) and Time (hh:mm). Therefore, we have to pre-process these columns into one column (yyyy-mm-dd hh:mm).

7.2. Setup

All Random Forests shown in Table 2 are tuned on minimal number of samples required to split an internal node. Each Random Forest chooses the number of samples which gave the best corresponding measure of that model. Furthermore, the dataset is split into 30 percent test data and 70 percent training data. All different prediction models shown in Table 2 are performed on a computer with 8 GB of RAM.[3]

7.3. Results

The scores for the Negative Prediction Value (NPV) and its corresponding accuracy of all combinations of (sequence of) Random Forest(s) for predicting overstaying are shown in Table 2. The baseline model considers only patient data (no events related to patients) to make the predictions. We use "+ IC" to denote that the inter-case features described in Section 6 were added to the model; and "+ pred.act." to denote that the predicted values for the bottlenecks, as discussed in Section 5 were added. The last column of Table 2 shows the measures for our approach when the real values for the performed (or not) bottlenecks were used (rather than the predicted ones). Note that, this last column shows the ability of the model if the previous bottlenecks prediction would have been perfect.

Although the results in Table 2 do not show improvement for the model trained with the predicted bottleneck activities, we will later discuss (c.f. Section 7.3.3) how this model can help on providing actionable recommendations that might reduce the time spent in the emergency room.

[3] The code for the whole methodology and experiments can be found at: https://github.com/MikeVerdaasdonk/SEQ-RF.

7.3.1. Activities

Table 3 shows the difference in percentage of overstaying patients between different groups. Groups are based on activities or pairs of activities which are performed or not. All activities (*Consult-SEH* and *Labaanvraag-SEH*), pair of activities (*RAD-TO-EXT*) and hospitalization (*Opname*) used as features for the sequence of Random Forests are shown in Table 3. For instance, *Opname=1* represents a patient which is eventually hospitalized in another department, where *Opname=0* represents a patients which is dismissed from the hospital or dismissed to another hospital. *Labaanvraag-SEH* and *Consult-SEH* represent if the activities Labaanvraag and Consult took place at the SEH (Spoedeisende hulp (=Emergency department)), respectively. *RAD-TO-EXT* represents that the activity 'Radiologie' is eventually followed by the activity 'Externe verslaglegging' and therefore satisfies the response constraint (see Section 2) for a patient. Note that all these features have a significant difference in percentage for overstaying patients.

Fig. 3 shows the correlation between activities, pair of activities, hospitalization and overstaying. Note that in this figure all features used for predicting overstaying have a correlation above 0,15 with overstaying.

The accuracies of the Random Forests for predicting activities, pair of activities and hospitalization, which are part of the sequence of Random Forests of the method, are shown in Table 4.

Note that there is a significant difference in accuracy between the Random Forest predictions and there is no perfect prediction in any case, i.e., accuracy = 1.0.

Looking back at Table 2, the Negative Prediction Value (NPV) of predicted activities, in columns "Base + pred.act." and "Base + IC + pred.act.", does not seem to increase compared to the baseline in column "Baseline (Base)". Note that the last column, where we use the real value for performed activities instead of prediction, the NPV improved by five percent and has a significant increase of the overall prediction (accuracy). Although there is no increase, we will discuss how the predictions are still useful.

7.3.2. Inter-case

The use of both inter-case features does not seem to significantly increase the Negative Prediction Value as shown in columns "Base + IC" and "Base + IC + pred.act." in Table 2. Furthermore, Fig. 4 shows the feature importance ranking, and there, both inter-case features are not ranked very high in the feature ranking of the final Random Forest. Note that, the higher the feature is ranked, the better it can divide different groups based on the output variable (i.e., overstaying). For instance, *CHI* is ranked higher than *NEU* in Fig. 4, therefore the specialism CHIrurgie has a more unequal distribution of overstaying and non-overstaying patients for patients with specialism CHIrurgie and/or patients without specialism CHIrurgie than the specialism NEUrologie. Thus it is easier to predict overstaying patients with the specialism CHIrurgie over the specialism NEUrologie.

7.3.3. Recommendations

We use the final Random Forest to make recommendations that serve to reduce the time a patient spends in the ER. Although we have seen in the previous two sections that the NPV value does not increase when considering the predicted bottlenecks and/or the inter-case features, they are used in this step to provide the recommendations.

Whenever a new patient arrives to the ER, the patient data during admissions is considered. From this input, the inter-case features are calculated and all possible bottlenecks are predicted for this patient

Table 3
Comparison of means between groups.

	Opname = 0	Opname = 1	z-test	p-value
Perc. Overstaying	24%	40%	−36,8	0
	Consult_SEH = 0	Consult_SEH = 1	z-test	p-value
Perc. Overstaying	26%	45%	−31,7	0
	RAD_TO_EXT = 0	RAD_TO_EXT = 1	z-test	p-value
Perc. Overstaying	20%	45%	−58,6	0
	Labaanvraag_SEH = 0	Labaanvraag_SEH = 1	z-test	p-value
Perc. Overstaying	12%	36%	−65,8	0

Correlation matrix	Labaanvraag_SEH	Rad_to_ext	Consult_SEH	Opname	Overstaying
Labaanvraag_SEH	1,00	0,13	0,08	0,27	0,24
Rad_to_ext	0,13	1,00	0,08	0,09	0,27
Consult_SEH	0,08	0,08	1,00	0,16	0,15
Opname	0,27	0,09	0,16	1,00	0,17
Overstaying	0,24	0,27	0,15	0,17	1,00

Fig. 3. Correlation matrix.

Table 4
Accuracy of predicting activities, hospitalization and pair of activities.

Activity	Accuracy Random Forest
Lab aanvraag	0.81
Radiologie eventually followed by externe verslaglegging	0.68
Consult SEH	0.8
Opname	0.7

```
Feature ranking:
1. Rad_to_ext (0.285404)
2. Labaanvraag_SEH (0.246124)
3. Opname (0.112453)
4. INT (0.079999)
5. Consult_SEH (0.071426)
6. CHI (0.034994)
7. Triage color (0.022603)
8. CAR (0.021341)
9. Age (0.020754)
10. nr_patients_total (0.015247)
11. Eigen vervoer (0.014813)
12. Presentation_time (0.012769)
13. MDL (0.011898)
14. Ambulance (0.011398)
15. NEU (0.008659)
16. History_bin (0.007816)
17. nr_patients_same_spec (0.005128)
```

Fig. 4. Feature ranking of Random Forest for predicting overstaying.

in order to know whether this patient is likely to go through such bottleneck or not. All this information is combined and used to navigate the final Random Forest model. When navigating the final tree, for each bottleneck that this navigation goes through, a recommendation is generated. Fig. 5 gives an example of three new patients and the recommendations generated for them.

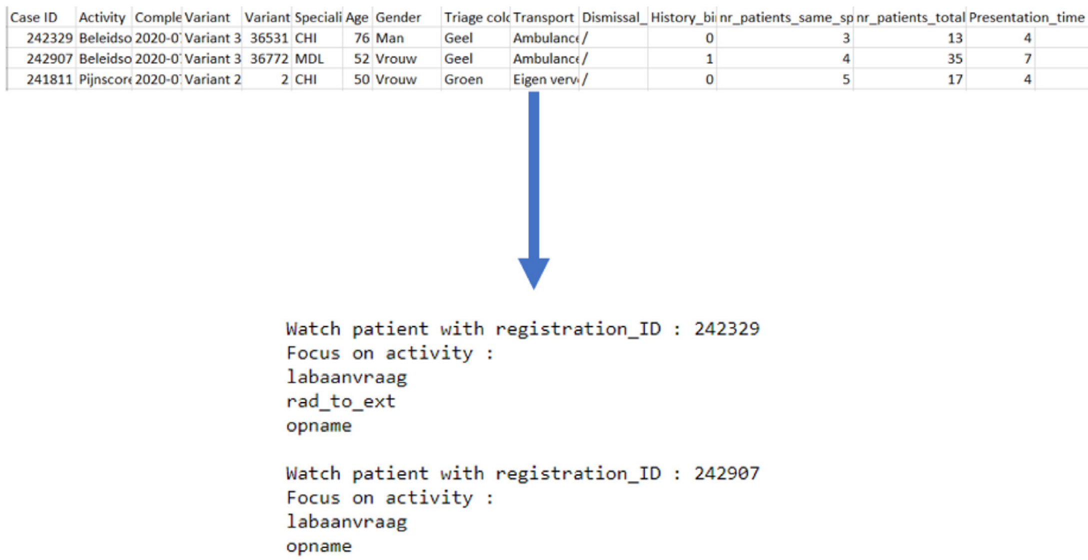
Algorithm 1 gives an overview of how the recommendations are generated. We start with the final decision tree (line 1 in Algorithm 1) generated as outcome of the Random Forest that predicts if a patient is overstaying or not ("STEP 2" in Fig. 1). For each new patient that comes to the ER, the patient data (in line 3) is retrieved and used to navigate the tree. Lines 5 to 15 in Algorithm 1 shows the process of navigating the decision tree. We start checking if the node is related to a bottleneck activity, when this is the case, we add it to the *activities* list (line 7) initialized as empty in line 4. Furthermore, to navigate the tree, the decision in the node is evaluate and the process either continues to the left side of the tree or to the right one. This is repeated until we get to the leaf of the tree, which indicates the class of the prediction (either "overstaying" or not). In case the patient is predicted as overstaying, then the recommendation to this patient is generated and the bottleneck activities that appear when navigating the tree are presented as activities to pay attention to.

Note that, it is not sufficient to predict the bottlenecks and generate recommendations for each of them, not all of them might be relevant. There might still be cases where patients are predicted to go through a bottleneck but this is not really relevant for the patient to overstay the 4 h at the ER.

Algorithm 1 Generating recommendations

1: $finalTree \leftarrow GetFinalTreeFromRandomForest()$
2: $node \leftarrow GetRootOfTree(finalTree)$
3: $newPatient \leftarrow GetNewPatientInfo()$
4: $activities \leftarrow$ empty list
5: **while** $node$ is not a leaf **do**
6:     **if** $node$ is activity **then**
7:         add $node$ to $activities$
8:     **end if**
9:     $decision \leftarrow EvaluateNodeDecision(node, newPatient)$
10:     **if** $decision$ is go left **then**
11:         $node \leftarrow leftNode$
12:     **else**
13:         $node \leftarrow rightNode$
14:     **end if**
15: **end while**
16: **if** $node$ is from class OVERSTAYING **then**
17:     $GenerateRecommendation(newPatient, activities)$
18: **end if**

When recommendations would be applied, 3% (500 of 15000 patients) of all patients in the test set could be changed from an overstaying patient (i.e., more than four hours in the hospital) to a non-overstaying patient (i.e., less than four hours in the hospital). This results in a decrease of about 10% (i.e., $\frac{3983-4440}{4440}$) overstaying patients. The boxplots representing all overstaying patients with and without recommendations are shown in Fig. 6. Note that, these boxplots are based on potential decrease of activities duration. Due to Covid-19, we were not able to apply the methodology in practice yet. Hence, these potential decreases are based on a simulation. Whenever a recommendation was generated for a predicted overstaying patient, we replaced



Case ID	Activity	Comple	Variant	Variant	Speciali	Age	Gender	Triage colc	Transport	Dismissal_	History_bi	nr_patients_same_sp	nr_patients_total	Presentation_time
242329	Beleidso	2020-0	Variant 3	36531	CHI	76	Man	Geel	Ambulance	/	0	3	13	4
242907	Beleidso	2020-0	Variant 3	36772	MDL	52	Vrouw	Geel	Ambulance	/	1	4	35	7
241811	Pijnscore	2020-0	Variant 2	2	CHI	50	Vrouw	Groen	Eigen verv	/	0	5	17	4

```
Watch patient with registration_ID : 242329
Focus on activity :
labaanvraag
rad_to_ext
opname

Watch patient with registration_ID : 242907
Focus on activity :
labaanvraag
opname
```

Fig. 5. Predict new patient; input and output.

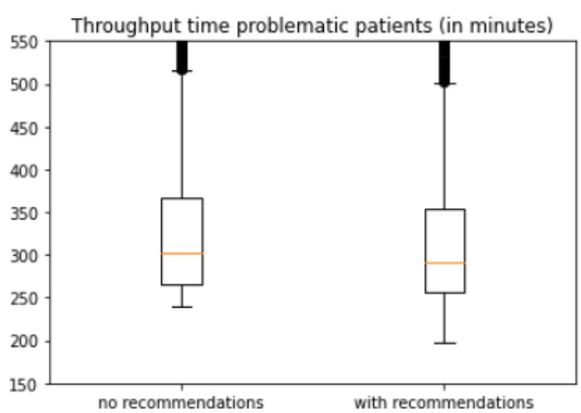


Fig. 6. Boxplot overstaying patients throughput time with and without recommendations.

the actual time it took for such activity(ies) to take place by the mean duration of the same activity(ies) for non-overstaying patients.

Recommendations for all patients would be really time-intensive and are not in line with the research question of this paper. Therefore, use is only made of recommendations for predicted overstaying patients.

7.4. Interpretation

As shown in Table 3 and Fig. 3 all chosen activities, pair of activities and hospitalization used for the sequence of Random Forests have a significant impact on predicting overstaying.

Although there seems to be no significant decrease or increase in Table 2 for the NPV, the method of implementing (bottleneck) activities still gives a good score. Furthermore, this method gives the possibility to make recommendations based on activities for the hospital because of the high feature importance of the bottleneck activities, as seen in Fig. 4. Moreover, as shown in Table 2, when activities would be predicted better, there is a possibility to increase the NPV value against the baseline and therefore implementing bottleneck activities for predicting overstaying seems effective.

The non significant increase or decrease for using inter-case features as shown in Table 2, may well be a result of the fairly good prediction for the number of patients at the hospital which makes it possible to improve the planning of hospital personnel and beds. Fig. 2 showed the arrival pattern of patients over time on a daily basis. This figure shows that there are possibilities for the hospital to make a good personnel planning over time. The inter-case features could be more influential to the score for hospitals if this existing prediction for the number of patients over time is not well done.

8. Conclusion

This section gives a brief overview of the accomplished results against the objectives presented earlier. The method described in this paper is compared with state of the art methods. Furthermore, limitations and possible future work are described.

8.1. Against state of the art

We accomplished to make it possible to generate a PAR (Process-Aware Recommendation) system for generating useful recommendations for most of the overstaying patients at the start of the process or at any moment in time. Table 2 showed that the method gives good results for predicting overstaying patients at the beginning of the process. Figs. 4 and 6 show that recommendations based on activities could be useful because of the decrease in throughput time of overstaying patients and the high feature ranking of activities.

Compared to state of the art techniques, this method prevents a black box model to get useful insights from the process. Furthermore, this method combines a PAR system with inter-case features to inject the interplay between cases in the prediction models. The method also makes it possible to make predictions and recommendations at the beginning of the process (i.e., at the arrival of the patient).

Different from the other techniques in the literature, which rely on actions from hospital staff when presented with the waiting time of patients only, we are also able to present to them which activities are supposed to take place and take longer than usual (for the usual values, we use the median time of that activity for patients that do not overstay). With our technique, the hospital staff are more guided on what they should pay attention and/or try to reduce the time of it. Moreover, our recommendation system only shows recommendations for the patients predicted to overstay, which also reduces the amount of information received by the hospital staff.

8.2. Limitations and shortcomings

There are two limitations that are worthy to discuss: the validation of our approach and the limitation in timestamps of the dataset.

As the first case of Covid-19 in the Netherlands appeared in February 27, 2020, and the first lockdown starting on March 12, 2020, it was not possible for us to validate our approach in practice. To be able to show the validity of our approach, we used the final months of our dataset. For that, we simulated new patients arriving to the ER, we checked when recommendations would be generated and we assumed that the hospital staff would be able to reduce the time it actually took for the recommended activities to become the median time for non-overstaying patients. Although we use data until September 2020, we still believe that this has no influence of Covid-19, because patients inflow were still common in the hospital. Only after this period, we could notice a drop in the inflow.

Finally, the dataset provided by the hospital only consists of starting times of activities. The missing of ending times of activities made it hard to measure and analyze activity duration. Note that, not registering ending time results in missing the possibility to distinguish between activity duration and waiting time until the next activity is performed. Therefore, it might be possible that the used bottleneck activities would have another difference in duration between different patient groups. Furthermore, it might be possible that we were able to find activities or pair of activities with larger waiting time between different patient groups. This would make it possible to generate recommendation based on these bottlenecks.

8.3. Future work

As possible improvements for the method, we can point to increasing the negative prediction value, and a more precise analysis of bottleneck activity or pair of activities duration. As described in the previous section, a more precise analysis of the activities duration would be possible if the ending times of activities are registered in the dataset of the hospital. An improvement of the quality of the final prediction model, i.e., NPV, is possible if we can better predict the bottleneck activities, pair of activities and hospitalization as seen in Table 4. Perfect predictions, i.e., accuracy is 1.0, results in a higher NPV, as seen in the last column of Table 2. Furthermore, improvement of the quality of the model would be possible if we analyze the specific groups in the test set and try to find groups where it is hard for the prediction model to have a good score for the NPV, e.g., patients with specialism MDL and are transported by ambulance have a NPV of 0.67. Based on a threshold formulated by the stakeholder, some groups can be filtered out of the system, which improves the NPV for the rest of the prediction model. We believe that these improvements would advance the methodology described in this paper, being able to achieve even better results.

In addition, we aim at investigating other effects that might influence the overstay of patients in the ER. For instance, researches [23–25] have been done to show that readmissions are not only common in the ER, but can also influence it.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Panneman, B. Blatter, Letsel informatie systeem representatief voor alle SEH's in Nederland? 2016, VeiligheidNL.
[2] I. Ashkenazi, L. Gefen, O. Hochman, E. Tannous, The 4-hour target in the emergency department, in-hospital mortality, and length of hospitalization: A single center-retrospective study, Am. J. Emerg. Med. 47 (2021) 95–100.
[3] F.F. Baldassarre, F. Ricciardi, R. Campo, Waiting too long: bottlenecks and improvements – a case study of a surgery department, TQM J. 30–2 (2018) 95–100.
[4] W. Aalst, Data Science in Action, Vol. 2, 2016, pp. 275–292.
[5] A. Senderovich, C.D. Francescomarino, F.M. Maggi, From knowledge-driven to data-driven inter-case feature encoding in predictive process monitoring, Inf. Syst. 84 (2019) 255–264.
[6] A. Senderovich, C. Di Francescomarino, C. Ghidini, K. Jorbina, F.M. Maggi, Intra and inter-case features in predictive process monitoring: A tale of two dimensions, in: J. Carmona, G. Engels, A. Kumar (Eds.), Business Process Management, Springer International Publishing, Cham, 2017, pp. 306–323.
[7] M. d. Leoni, M. Dees, L. Reulink, Design and evaluation of a process-aware recommender system based on prescriptive analytics, in: 2020 2nd International Conference on Process Mining, ICPM, 2020, pp. 9–16.
[8] C. Di Ciccio, M.L. Bernardi, M. Cimitile, F.M. Maggi, Generating event logs through the simulation of declare models, in: J. Barjis, R. Pergl, E. Babkin (Eds.), Enterprise and Organizational Modeling and Simulation, Springer International Publishing, Cham, 2015, pp. 20–36.
[9] P. Agarwal, D. Swarup, S. Prasannakumar, S. Dechu, M. Gupta, Unsupervised contextual state representation for improved business process models, in: A. Del Río Ortega, H. Leopold, F.M. Santoro (Eds.), Business Process Management Workshops, Springer International Publishing, Cham, 2020, pp. 142–154.
[10] A. Grinvald, P. Soffer, O. Mokryn, Inter-case properties and process variant considerations in time prediction: A conceptual framework, in: A. Augusto, A. Gill, S. Nurcan, I. Reinhartz-Berger, R. Schmidt, J. Zdravkovic (Eds.), Enterprise, Business-Process and Information Systems Modeling, Springer International Publishing, Cham, 2021, pp. 96–111.
[11] S. Shalev-Shwartz, S. Ben-David, Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press, 2014.
[12] M. Pal, Random forest classifier for remote sensing classification, Int. J. Remote Sens. 26 (1) (2005) 219.
[13] F.M. Maggi, R.P.J.C. Bose, W.M.P. van der Aalst, Efficient discovery of understandable declarative process models from event logs, in: J. Ralyté, X. Franch, S. Brinkkemper, S. Wrycza (Eds.), Advanced Information Systems Engineering, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 270–285.
[14] Q. Cheng, N.T. Argon, C.S. Evans, Y. Liu, T.F. Platts-Mills, S. Ziya, Forecasting emergency department hourly occupancy using time series analysis, Am. J. Emerg. Med. 48 (2021) 177–182, http://dx.doi.org/10.1016/j.ajem.2021.04.075.
[15] F. Kadri, A. Dairi, F. Harrou, Y. Sun, Towards accurate prediction of patient length of stay at emergency department: a GAN-driven deep learning framework, J. Ambient Intell. Humaniz. Comput. (2022).
[16] Y. Sun, K.L. Teow, B.H. Heng, C.K. Ooi, S.Y. Tay, Real-time prediction of waiting time in the emergency department, using quantile regression, Ann. Emerg. Med. 60 (3) (2012) 299–308, http://dx.doi.org/10.1016/j.annemergmed.2012.03.011.
[17] A. De Hond, W. Raven, L. Schinkelshoek, M. Gaakeer, E. Ter Avest, O. Sir, H. Lameijer, R.A. Hessels, R. Reijnen, E. De Jonge, E. Steyerberg, C.H. Nickel, B. De Groot, Machine learning for developing a prediction model of hospital admission of emergency department patients: Hype or hope? Int. J. Med. Inf. 152 (2021) 104496, http://dx.doi.org/10.1016/j.ijmedinf.2021.104496.
[18] C.A. Parker, N. Liu, S.X. Wu, Y. Shen, S.S.W. Lam, M.E.H. Ong, Predicting hospital admission at the emergency department triage: A novel prediction model, Am. J. Emerg. Med. 37 (8) (2019) 1498–1504, http://dx.doi.org/10.1016/j.ajem.2018.10.060.
[19] T. Goto, C.A. Camargo Jr., M.K. Faridi, R.J. Freishtat, K. Hasegawa, Machine learning–based prediction of clinical outcomes for children during emergency department triage, JAMA Netw. Open 2 (1) (2019) e186937, http://dx.doi.org/10.1001/jamanetworkopen.2018.6937.
[20] Y. Raita, T. Goto, M.K. Faridi, D.F.M. Brown, C.A. Camargo, K. Hasegawa, Emergency department triage prediction of clinical outcomes using machine learning models, Crit. Care 23 (1) (2019).
[21] J.-T. Lee, C.-C. Hsieh, C.-H. Lin, Y.-J. Lin, C.-Y. Kao, Prediction of hospitalization using artificial intelligence for urgent patients in the emergency department, Sci. Rep. 11 (1) (2021) 19472, http://dx.doi.org/10.1038/s41598-021-98961-2.
[22] R. van Delft, R.M.D. Carvalho, Using machine learning techniques to support the emergency department, in: Informatics Revolution for Smarter Healthcare, 2021.
[23] K. Jusmanova, C. Rice, R. Bourke, A. Lavan, C.G. McMahon, C. Cunningham, R.A. Kenny, R. Briggs, Impact of a specialist service in the emergency department on admission, length of stay and readmission of patients presenting with falls, syncope and dizziness, QJM: Int. J. Med. 114 (1) (2020) 32–38, http://dx.doi.org/10.1093/qjmed/hcaa261.
[24] N. Pecorelli, G. Guarneri, K. Alagol, M. Mazza, R. Quattromani, S. Partelli, S. Crippa, R. Castoldi, G. Balzano, M. Falconi, The impact of minimally invasive surgery on hospital readmissions, emergency department visits and functional recovery after distal pancreatectomy, Surg. Endosc. 35 (10) (2021) 5740–5751, http://dx.doi.org/10.1007/s00464-020-08051-z.
[25] M. Gwin, M. Saleki, H. Lampert, N. Meo, M. Bann, Emergency department visits and readmissions after COVID-19 hospitalization: a cross-sectional analysis, Int. Emerg. Med. 16 (6) (2021) 1715–1718, http://dx.doi.org/10.1007/s11739-021-02644-9.