Les Cahiers du GERAD ISSN: 0711–2440

Improving nurse scheduling using a random forest algorithm to predict employee well-being

S. Séguin, Y. Villeneuve, J. Maître, R. Grimard

G–2025–25

March 2025

Citation suggérée : S. Séguin, Y. Villeneuve, J. Maître, R. Grimard (Mars 2025). Improving nurse scheduling using a random forest algorithm to predict employee well-being, Rapport technique, Les Cahiers du GERAD G– 2025–25, GERAD, HEC Montréal, Canada.

Avant de citer ce rapport technique, veuillez visiter notre site Web (https://www.gerad.ca/fr/papers/G-2025-25) afin de mettre à jour vos données de référence, s'il a été publié dans une revue scientifique.

Suggested citation: S. Séguin, Y. Villeneuve, J. Maître, R. Grimard (March 2025). Improving nurse scheduling using a random forest algorithm to predict employee well-being, Technical report, Les Cahiers du GERAD G–2025–25, GERAD, HEC Montréal, Canada.

Before citing this technical report, please visit our website (https://www.gerad.ca/en/papers/G-2025-25) to update your reference data, if it has been published in a scientific journal.

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2025
– Bibliothèque et Archives Canada, 2025

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2025
– Library and Archives Canada, 2025

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

Improving nurse scheduling using a random forest algorithm to predict employee well-being

Sara Séguin [a, b]

Yoan Villeneuve [a, b]

Julien Maître [a]

Renaud Grimard [c]

[a] *Université du Québec à Chicoutimi, Saguenay (Qc), Canada, G7H 2B1*

[b] *GERAD, Montréal (Qc), Canada, H3T 1J4*

[c] *Timesphere, Saguenay (Qc), Canada, G7H 6P3*

sara.seguin@uqac.ca

March 2025
Les Cahiers du GERAD
G–2025–25

Abstract : This paper introduces a new approach to nurse scheduling that integrates employee well-being into the decision-making process. A random forest regressor is trained to estimate a well-being score for each nurse, leveraging data from previous work weeks and considering multiple factors related to past schedules. This score is incorporated into a mixed-integer linear programming model to guide the assignment of shifts, aiming to better align schedules with individual needs. Nurses with lower well-being scores are prioritized for reduced overtime and increased shift preferences, promoting a fairer distribution of workload. The proposed method generates schedules that balance operational requirements with employee health, potentially mitigating fatigue and absenteeism.

Keywords: Nurse scheduling, linear integer programming, optimization, random forest regressor, employee absenteeism, prediction

---

Acknowledgements: The authors would like to thank Maud-Christine Chouinard and Marie-Ève Poitras for the insights and collaboration in the early stages of the project as well as the CIUSSS Saguenay-Lac-St-Jean for providing the datasets for the project.

1 Introduction

Nurse scheduling is a difficult task for many reasons. In the province of Québec, one of the main drawbacks is that the nurses are unionized, thus the scheduling is rigid, and must follow many rules leaving very little opportunity to modify the current methodology. It is also well documented that there is currently a shortage of nurses to provide sufficient care to the patients. Of course, the COVID-19 pandemic has exacerbated the problem. In [7], the authors noticed that 25% of a 891 nurse sample were in a distress situation, caused by high pressure and a lack of resources. A recent paper [18] notes that 47.9% of nurses in their study suffered from anxiety troubles and burnout. The actual nurse shortage is caused by many factors, mainly by budget cuts and management restricted by rigid union rules.

These numbers show that nurse scheduling should be improved in order to satisfy the employees. In order to increase well-being, the authors of [19] explain that schedules should be more flexible. In the province of Québec, a schedule consists of 14 days on three shifts. The manager starts planning the schedule 8 weeks before it is available to the nurses. Therefore, nurses need to send their workdays preferences at least 6 weeks before the beginning of the new schedule, and the schedule is available 4 weeks before the first day. Creating the schedules is a tedious task, since most of the schedule is done manually and calling lists are used to fill in the blanks. Then, the overtime is planned, still using calling lists and the process is repeated until the schedule is conform to the collective agreement. Although the schedule is planned, nurses may not show up for their shift, leading to forced overtime. The risk of errors increase when the nurses work longer than their initial shift. In fact, authors of [16] observe that the risk increases after a 8.5 hour shift, and that the risk triples when they work more than 12.5 hours.

Nurse scheduling is widely studied in the literature, mainly by proposing mixed-integer linear programming models that build schedules around a set of constraints and different objectives. In [20], the objective function maximizes the number of days off. Nurses have different qualifications and can only work in certain areas. The proposed solution satisfies the nurse quotas for every shift and when compared to the manual schedule, the working hours are more fair. The authors of [2] used a branch-and-cut solver to generate nurse schedules for a one month horizon. Many constraints allow a violation and are used to avoid work overload. Results on real data from a hospital unit show that the schedule obtained by the solver considers constraints that are difficult to take into account manually, leading to a better planning. A two-phase approach is proposed in [8]. In the first phase, a schedule is proposed, then, the preferences for different shifts are taken into consideration in the second phase. Different levels of days off such as funerals, weddings, vacation and ordinary days off are considered and preferences are assigned given the different levels. The test case consisted of 20 nurses, and results show that most of the requests for days off were considered in the final planning.

As the scheduling problems contain many variables and constraints, many heuristics methods are proposed in the literature to obtain solutions faster. The authors of [11] build a schedule for regular nurses and nurses that do not have a fixed position. The heuristic, implemented easily in an Excel worksheet, allows to find a schedule that respects preferences and equity. Although obtaining a solution is quicker with a commercial optimization solver, in practice, Excel is a better solution for the managers since it does not require expertise in optimization. A multiobjective algorithm is proposed in [22] to assign nurses to operating rooms. The algorithms minimizes the makespan, which is the time length between the entry of the first patient and the output of the last patient, and maximizes the throughput, which is the total number of operations conducted. An ant colony algorithm [5] is used. Results are compared with two other studies [13, 21] and authors conclude that their solution proposed an optimal makespan and has a better balance in the allocated resources. Also, a comparable number of operations is conducted and yet, nurses have a reduced number of working hours.

The proposed studies in the literature are interesting, but all used either optimization model or heuristics to obtain solutions. Moreover, as noted by [7], better planning is necessary to propose better schedules to the nurses.

In this paper, we propose to predict the well-being of nurses using a Random Forest Regressor (RFR) algorithm. The scheduling is conducted using a mixed-integer linear problem, but a parameter of the optimization model is predicted using the RFR algorithm.

Machine learning tools are efficient to predict an output based on input data. In the prediction of absenteeism, machine learning tools have been widely used [1, 9, 10]. According to the literature on the absenteeism prediction, most of the papers dealing with that theme have been published after 2018 [14]. In addition, more than the half of the works exploited regression methods, one thirds used classification methods and the rest applied clustering methods [14]. For example, Lima et al. [12] attempted to predict absenteeism of public Brazilian security agents by exploiting deep models. More specifically, the objective was to identify workers prone to long-term absenteeism. This work is based on a dataset of 6 years of professional data. Also, the deep models that have been exploited are a MultiLayer Perceptron (MLP), Recurrent Neural Network and Long Short-Term Memory. It should be noted that those three models have been compared to a baseline Support-Vector Machines classifier. The best results were obtained with MLP and when the 6 years of data have been considered for training and testing the model. Indeed, MLP reached an accuracy of 78.42%.

Random Forest (RF) algorithms are a collection of decision trees. Each tree is a weak learner, but when combined, they form a strong learner. Moreover, using bootstrap sampling allows to create multiple input data sets in order to train a robust predictive model. RF are non-parametric supervised learning algorithms that rely on the divide and conquer paradigm, more precisely by branching on attributes until a leaf node is reached. RF can be used for classification or regression and has been used successfully for the prediction of absenteeism [9, 17].

The originality of this work is using the history of the past work weeks of employees to predict a well-being score, which is then used in the optimization model that computes the scheduling of the employees. Employees with a bad well-being score will have better chances to be assigned to their shifts of choice, for example.

The paper is organized as follows. Sections 2 details the methodology, more precisely the RF algorithm and the MILP optimization model. Results are presented in Section 3 and concluding remarks in Section 4.

2 Methodology

Since it is merely impossible to change the work schedules, caused by the collective agreement in place for the nurses, this project is concerned with providing tools that can help improve the scheduling, by respecting the current restrictions.

First, available data is analyzed and used to predict the well-being of the nurses, by using a random forest algorithm. Second, this metric is used as an input parameter for the mixed-integer linear program that assigns the shifts to the nurses, based on the collective agreement constraints. This Section explains in details the random forest algorithm and the optimization model.

2.1 Random forest algorithm

A random forest algorithm is used to predict the well-being score of nurses for the next two weeks, which is then used to generate the schedule by solving an optimization problem.

A random forest regressor is an algorithm that creates a machine learning model relying on and combining regression trees (base models). Hence, the resulting model is a forest of regression trees, where each tree provides an estimated value. It should be noted that a regression tree is a specific type of decision tree [4]. The difference between them relies in the way the target is determined. Indeed, the decision tree assigns the label of the majority class of instances in the leaf node, whereas a regression tree computes the average of the target values of instances in the leaf node. The number of trees in the forest is one of the hyperparameters that should be set by the user. Hence, since the random forest regressor creates multiple trees, this algorithm belongs to the family of ensemble learning methods and more precisely employs bagging, short for bootstrap aggregating [3]. The main advantage of using bagging is to reduce overfitting and to increase the stability and performances of predictions.

Figure 1 illustrates a random forest algorithm. It consists of n decision trees, each predicting the well-being score. Then, the average of the predictions from all the decision trees is taken as the predicted SC.

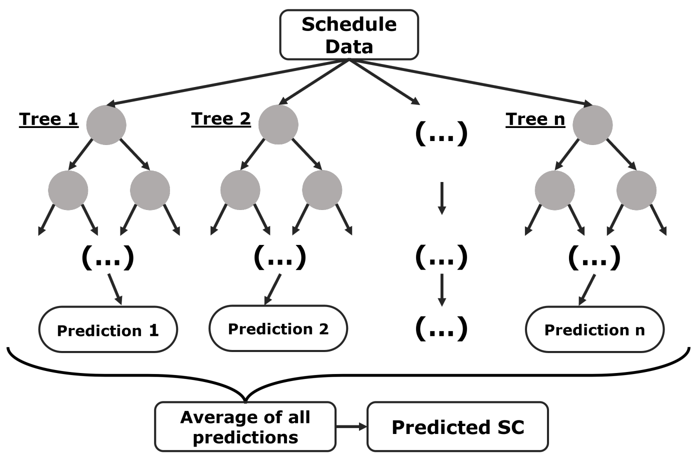

Figure 1: Random forest algorithm

The data provided to the RF algorithm, for this project, is detailed in Table 1.

Table 1: Input data to the RF algorithm

Name	Description
Directly from historical data	
#	Employee #
Week	Week number
REG	# of regular shifts primary affectation
SEC	# of regular shifts secondary affectation
O	# of overtime shifts
MO	# of mandatory overtime shifts
EU	# of shifts in another nursing unit
AI	# of shifts on sick leave
NA	# of declared normal absence shifts
U	# of shifts of unemployment
DI	# of shifts on disability insurance
Calculated from historical data	
CS	# of consecutive days worked, if over 5, variable is equal to - 5, if not 0
OVERLOAD	See Eq.(2)
ER	# of extra shifts

The goal of the RF algorithm is to predict the well-being (SC) of each nurse. To calculate this value, previous schedules are analyzed and relevant data is used to determine a score for each employee. The formula is created using knowledge in the literature, knowledge of the persons in charge of creating the schedules and data from the previous schedules. A high value of SC indicates that the employee is in a poor well-being state.

The formula used to predict the SC is given by:

$$SC = 10 - (10 \times F_1 \times F_2 \times F_3),$$

where F_1 is related to the primary affectation of the nurse, F_2 the risk of work overload and F_3 the risk related to the unplanned leaves. The three values of F are all given by values $\in [0, 1]$, therefore $SC \leq 10$. The details for each value are given below.

$F1$. **Relation with primary affectation** Each nurse has a primary affectation related to her work contract, and a secondary affectation. Technically, a nurse with a day contract should always be scheduled during the day. In reality, it is sometimes difficult to obtain the work quota, therefore nurses can be affected to their secondary affectation, but with a penalty. For this project, scheduling a nurse on the secondary affectation leads to a well-being decrease. The following formula is used to calculate F_1:

$$F_1 = \frac{\text{REG} - \frac{\text{SEC}}{3}}{\text{REG}}, \tag{1}$$

where REG are the number of primary affectation shifts and SEC the number of secondary affectation shifts. The value of F_1 ranges between $[0.67, 1]$, where 0.67 represents a schedule during which the nurse is always affected to her primary affection. The division by 3 is arbitrary and is chosen based on the input data. Multiple tests were conducted on the datasets and the value of 3 allows to diminish the impact of F_1 on the value of SC.

$F2$. **Risk of work overload** Many factors are considered for the risk of work overload: overtime (O), mandatory overtime (MO), external unit (EU), consecutive shifts (CS) and extra shifts on regular time (ER). All these factors decrease the well-being of the nurse, since they are not part of the initial work contract.

Therefore, the variable $OVERLOAD$ is defined:

$$OVERLOAD = O + MO + EU + CS + ER \tag{2}$$

All of the previous factors are considered in the formula to calculate $F_2 \in [0, 1]$:

$$F_2 = 1 - \frac{OVERLOAD}{20}. \tag{3}$$

The division by 20 is chosen since in the historical data available, the maximal score obtained for $OVERLOAD$ is 16. Therefore, it could be necessary to adapt this value given the input data.

$F3$. **Risk related to unplanned leave** The factors considered in the unplanned leaves are: normal absence (NA), unemployment (U), absence due to illness (AI) and disability insurance (DI). The formula used to calculate the risk $F_3 \in [0, 1]$ is:

$$F_3 = \left(1 - \frac{NA + U + AI + DI}{10}\right)^{1.5}. \tag{4}$$

The parameter 1.5 is used to increase the impact of unplanned leaves on the final SC. As for the division by 10, the value is chosen based on the input data, since the unplanned leaves are always less or equal to 10.

2.2 Optimization problem

The mixed-integer linear problem that allows to obtain a schedule to assign the nurses to the different work shifts is described in this section.

The **sets** are:

I:	set of nurses
$I^{int} \subset I$:	subset of nurses in the unit
$I^{ext} \subset I$:	subset of external nurses
$I^{inf} \subset I$:	subset of regular nurses
$I^{aux} \subset I$:	subset of auxiliary nurses
J_i:	number of days in the schedule $i \in \{1, \ldots, 14\}$
Q_j:	workshifts per day $j \in \{\text{night}, \text{day}, \text{evening}\}$

The **decision variables** are:

$$x_{ijq} = \begin{cases} 1, & \text{if nurse } i \in I \text{ is affected to shift } q \in Q \text{ of day } j \in J, \\ 0, & \text{otherwise.} \end{cases}$$

$$ts_{ijq} = \begin{cases} 1, & \text{if nurse } i \in I \text{ works overtime during shift } q \in Q \text{ of day } j \in J, \\ 0, & \text{otherwise.} \end{cases}$$

$penalty_i =$ Penalty associated to the maximum number of shifts

The **parameters** are:

$$Aff_{ijq} = \begin{cases} 1, & \text{if nurse } i \in I \text{ can be assigned to primary} \\ & \text{affectation shift } q \in Q \text{ on day } j \in J, \\ 0, & \text{otherwise.} \end{cases}$$

$$Aff^{sec}_{ijq} = \begin{cases} 1, & \text{if nurse } i \in I \text{ can be assigned to secondary} \\ & \text{affectation shift } q \in Q \text{ on day } j \in J, \\ 0, & \text{otherwise.} \end{cases}$$

$$Con_{ij} = \begin{cases} 1, & \text{if nurse } i \in I \text{ requests a day off on day } j \in J, \\ 0, & \text{otherwise.} \end{cases}$$

$$Exp_i = \begin{cases} 1, & \text{if nurse } i \in I \text{ has an expertise,} \\ 0, & \text{otherwise.} \end{cases}$$

$$Pref_{ijq} = \begin{cases} 1, & \text{if shift } q \in Q \text{ is a preference for nurse } i \in I \\ & \text{on day } j \in J, \\ 0, & \text{otherwise.} \end{cases}$$

$SC_i =$	**Value of the well-being metric predicted** by the random forest algorithm for nurse $i \in I$
$B^{inf} \in I^{inf}$ and $B^{aux} \in I^{aux} =$	Minimal number of shifts per nurse
$Max_i =$	Maximal number of shifts per nurse $i \in I$ according to contract.
$QQ_{jq} =$	Minimal number of nurses on shift $q \in Q$ on day $j \in J$.

The objective function consists in maximizing the number of nurses affected to the schedule, in order to schedule the nurses to all the shifts. The first term aims at giving preferences to nurses with a high well-being score. The second term is concerned with the secondary affectations, therefore a penalty is applied for every secondary affectation. The third term penalizes overtime shifts. The fourth term penalizes largely the model when external nurses to the unit must complete the work quota. Finally, the fifth term is related to the maximal number of shifts worked by a nurse.

$$\max_{x,t,penalty} \sum_{i \in I^{int}} \sum_{j \in J} \sum_{q \in Q} ((x_{ijq} \times Pref_{ijq} \times SC_i) - (x_{ijq} \times Aff^{sec}_{ijq} \times 100) - ts_{ijq} \times SC_i \times 400))$$
$$- \sum_{i \in I^{ext}} \sum_{j \in J} \sum_{q \in Q} (x_{ijq} \times 10000) - \sum_{i \in I} (penalty_i \times SC \times 35) \quad (5)$$

s.t.

$$\sum_{q \in Q} x_{ijq} \leq 1 \qquad \forall i \in I, j \in J, \quad (6)$$

$$x_{ijq} \leq Aff_{ijq} \qquad \forall i \in I^{int}, j \in J, q \in Q, \quad (7)$$

$$x_{ijq} \leq Aff^{sec}_{ijq} \qquad \forall i \in I^{int}, j \in J, q \in Q, \quad (8)$$

$$\sum_{j \in J} \sum_{q \in Q} x_{ijq} \leq Max_i - \sum_{j \in J} Con_{ij} \quad \forall i \in I^{int}, \quad (9)$$

$$x_{ijq} < Con_{ij} \qquad \forall i \in I^{int}, j \in J, q \in Q, \tag{10}$$

$$ts_{ijq} < Con_{ij} \qquad \forall i \in I^{int}, j \in J, q \in Q, \tag{11}$$

$$\sum_{j \in J} \sum_{q \in Q} x_{ijq} + \sum_{j \in J} con_{ij} \geq B^{inf} \qquad \forall i \in I^{inf}, \tag{12}$$

$$\sum_{j \in J} \sum_{q \in Q} x_{ijq} + \sum_{j \in J} con_{ij} \geq B^{aux} \qquad \forall i \in I^{aux}, \tag{13}$$

$$Max_i + 3 - \lfloor SC_i \rfloor + penalty_i \geq \sum_{j \in J} \sum_{q \in Q} x_{ijq} \qquad \forall i \in I^{int}, \tag{14}$$

$$x_{ij1} + x_{i(j+1)2} \leq 1 \qquad \forall i \in I, j \in J, \tag{15}$$

$$x_{ij1} + x_{i(j+1)0} \leq 1 \qquad \forall i \in I, j \in J, \tag{16}$$

$$x_{ij0} + x_{i(j+1)2} \leq 1 \qquad \forall i \in I, j \in J, \tag{17}$$

$$\sum_{i \in I} ((x_{ijq} + ts_{ijq}) \times Exp_i) \geq 1 \qquad \forall j \in J, q \in Q, \tag{18}$$

$$x_{ijq} + ts_{ijq} \leq 1 \qquad \forall i \in I, j \in J, q \in Q, \tag{19}$$

$$\sum_{j \in J} \sum_{q \in Q} ts_{ijq} \leq \left\lfloor \frac{8}{SC_i + 1} \right\rfloor \qquad \forall i \in I^{int}, \tag{20}$$

$$\sum_{i \in I} (x_{ijq} + ts_{ijq}) = QQ_{jq} \qquad \forall j \in J, q \in Q, \tag{21}$$

$$\sum_{q \in Q} x_{6jq} + x_{7jq} = 2 \times Fds_{i0} \qquad \forall i \in I, \tag{22}$$

$$\sum_{q \in Q} x_{0jq} + x_{13jq} = 2 \times Fds_{i1} \qquad \forall i \in I, \tag{23}$$

$$\sum_{q \in Q} x_{0jq} + x_{6jq} + x_{7jq} + x_{13jq} \leq 2 \qquad \forall i \in I, \tag{24}$$

$$(x_{6,0,i} \times x_{7,0,i}) + (x_{6,1,i} \times x_{7,1,i}) + (x_{6,2,i} \times x_{7,2,i}) \leq 2 \qquad \forall i \in I, \tag{25}$$

$$\begin{aligned} &\sum_{q \in Q} x_{ijq} + x_{i(j+1)q} + x_{i(j+2)q} + x_{i(j+3)q} \\ &+ x_{i(j+4)q} + x_{i(j+5)q} + ts_{ijq} + ts_{i(j+1)q} + ts_{i(j+2)q} \\ &+ ts_{i(j+3)q} + ts_{i(j+4)q} + ts_{i(j+5)q} \leq 5 \qquad \forall i \in I, j \in J, \end{aligned} \tag{26}$$

$$x_{ijq}, ts_{ijq} \in \mathbb{B} \qquad \forall i \in I, j \in J, q \in Q. \tag{27}$$

$$penalty_i \in \mathbb{Z} \qquad \forall i \in I. \tag{28}$$

Eq.(6) ensures that nurses work only one regular shift per day. Eq.(7)–(8) are used to assign shifts to the primary or secondary affectation of the nurse. Days off are taken into account with Eq.(9)–(13). Soft constraints, represented by Eq.(14), allow to change the maximal number of shifts assigned to a nurse, given the well-being score. The collective agreement requires that after each shift worked, two shifts must remain unassigned for rest, as seen in Eq.(15)–(17). The expertise of a nurse is considered with Eq.(18) and ensures that during each shift, there is at least one nurse with an expertise. Eq.(19) are used to force the model to assign a regular or an overtime shift, not both at once. Depending on the well-being score of the nurse, Eq.(20) limits the maximal number of overtime shifts that can be assigned. Quotas of nurses during each shift are given by Eq.(21). The collective agreement in Québec imposes to work every other week-end and this is given by Eq.(22)–(25). Finally, nurses can not work more than five consecutive shifts, either in regular or overtime, as explicited by Eq.(26). Variables' domain are given by Eq.(27)–(28).

3 Results

This Section details the numerical results. First, the case study is presented. Then, the performance of the random forest algorithm is presented. Finally, the quality of the schedules generated by the mixed-integer linear program is analyzed.

3.1 Case study

The methodology and results in this paper are based on historical data provided by the Centre Intégré Universitaire de Santé et de Services Sociaux du Saguenay-Lac-St-Jean (CIUSSS SLSJ). One year of data for a specific nursing unit is available from December 2018 until December 2019. The data represents 11,314 instances, where each instance represents a work shift for one nurse. The unit is composed of nurses and auxiliary nurses and the work quotas are different for each type of nurse.

There is a total of 28 regular nurses and 12 auxiliary nurses. Schedules are constructed on a two week horizon and must respect the collective agreement, such as regular shifts, overtime, consecutive shifts, week-ends, and so forth.

Nurses are affected to a a primary shift in their work contract, but can also be scheduled to a secondary affectation. Figure 2 exposes the primary and secondary affectations of each nurse, as they are indirectly related to the quotas.

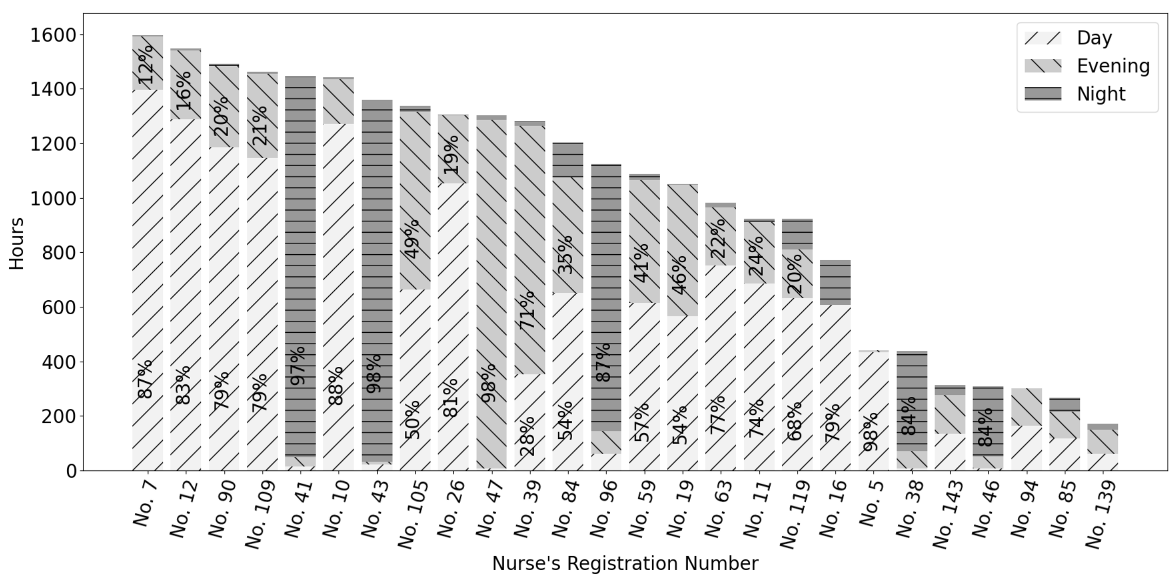

Figure 2: Primary and secondary affectations of the nurses

The analysis of the data shows that there is always less nurses available than the required quotas. Therefore, other nurses need to compensate, or the nursing unit is simply not meeting the quotas. Both situations are problematic since they increase the workload, but this is what happens in practice.

3.2 Random forest algorithm

The strategy leave-one-subject-out is chosen to create the training and testing sets. Therefore, all the nurses are part of the training set, except one, that is used as a testing set. This choice is motivated by the size of the datasets, which consist of 28 regular nurses and 12 auxiliary nurses. Also, as the goal of the model is to predict the well-being score of a nurse, the nurse that is analyzed is taken out of the dataset. There is a total of 535 instances available for the RF algorithm. The methodology used to train and validate the prediction of the well-being score is shown in Figure 3. As the schedule is planned for 2 weeks, the two last schedules, for a total of 4 weeks, are used to predict the well-being score.

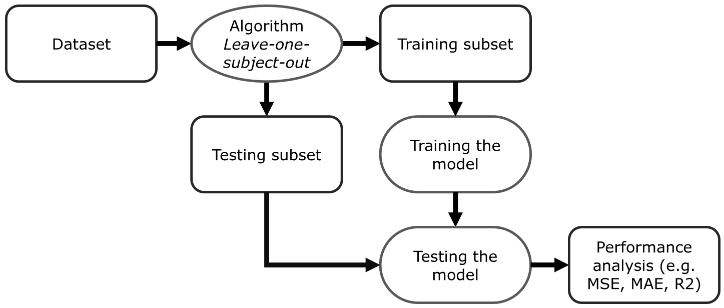

Figure 3: Training and testing sets for the RF algorithm

The Random Forest Regressor from the Sci-Kit learn library [15] is used to compute the results. The hyperparameters are also optimized using this library, and detailed in Table 2.

Table 2: Hyperparameters for the Random Forest Regressor

Parameters	Values
n_estimators	160
max_features	auto
max_depth	80
min_samples_split	10
min_samples_leaf	1
bootstrap	True

3.3 Results – Selected cases

Given the limitations on paper length, a single example for one nurse's well-being score is presented, as well as one schedule.

3.3.1 Predicted score

Figure 4 shows the calculated score from historical data on the solid line and the predicted well-being score on the dashed line, for nurse 11. This nurse was on work leave from week 19 until week 33, as shown from the SC score of 0 for these weeks. One of the main takeaways of this figure is that the evolution of the score actually follow the historical data, allowing an accurate prediction of the well-being score.

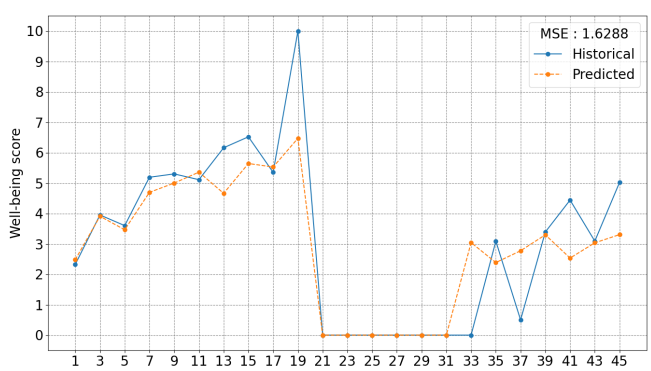

Figure 4: Nurse 11. Historical and predicted well-being (SC) score

These results show that the RF algorithm predicts correctly the tendency of the SC, even if there is a slight difference between the calculated score and the predicted score. The computation time to train the model is 11.31 seconds.

SC are calculated for all the nurses and are then used as input parameters in the optimization model.

3.3.2 Schedule

The optimization model is modeled with AMPL [6] and solved with Xpress.

In order to validate results, schedules generated with the RFR and the optimization model are compared with the actual historical schedules, to determine if the proposed approach allows to enhance the current schedules. Many other tests were conducted, but for the purpose of the paper, only one case is presented.

The parameters of the optimization problem are presented in Table 3. A simple case is chosen with only two possibilities for the SC to illustrate the effect of the score on the assignments.

Table 3: Parameters of the optimization model used to generate the schedule

Parameter	Description	Value
SC_i	Well-Being Score	6 nurses with a score of 6 8 nurses with a score of 1
I^{inf}	Number of Nurses	14
Con_i	Planned Leave	0
Exp_i	Expertise	0
$Pref_{ijq}$	Preference	Random based of their respective work contract
B^{inf}	Min. Number of Assignments	8

Table 4 presents the scheduled obtained for the 14 nurses. The schedule is computed for two weeks, from Sunday to Saturday and when a letter is present, it represents a shift for the nurse. Overtime shifts are designated with *.

Table 4: Schedule. D is for Day shift, N for Night and E for Evening. The * represents an overtime shift

Day	105	143	26	59	16	90	96	84	47	10	94	139	85	119
Sun.	E*	E	N*			D	N	D		D		E		D
Mon.	D			D	N	D		D	E		E	E	N	
Tue.	D	E	D		N			D	E		E		N	D
Wed.	D	E		D			N		E	D	E		N	D
Thu.			D	D	N	D	N	D	E		E	E		
Fri.		E	D		N	D	N	E		D		E		D
Sat.	D		D	D	N			E*	E	N*	E		D	
Sun.	D		D	D	N				E	N*	E		D	E*
Mon.	D		E		N	D	N			D	E	E		D
Tue.	D			D		D	N	D	E		E	E	N	
Wed.		E	D		N	D			E	D	E		N	D
Thu.		E	D	D	N				E	D		E	N	D
Fri.	D	E		D	N		N	D	E	D	E			
Sat.		E	E*			D	N	D		D		E		DN*

Results are presented in Table 5. The well-being score, and shift contract are presented, as well as the total number of shifts in the schedule (occurrence), the number of overtime shifts and the number of shifts of preference that are assigned, on a maximum of 10.

Results show that all the nurses with a well-being score of 6, have a lower number of assigned shifts, which is the desirable outcome. Most of the nurses also were assigned their preferred shifts. Also, the overtime shifts are not assigned to the nurses with the SC score of 6, which is also desirable. This shows that the optimization model is actually building the schedules by taking into account the well-being score of the nurses.

Table 5: Results

Nurse #	105	143	26	59	16	90	96	84	47	10	94	139	85	119
SC	1	6	1	6	1	6	6	1	1	1	6	6	6	1
Contract	D	E	D	D	D	N	D	E	D	D-E	E	D	D	D
Occurrence	8	8	8	10	8	8	8	8	10	8	10	8	8	8
Overtime	1	0	2	0	0	0	0	1	0	0	2	0	0	2
Preference	7	8	6	8	0	8	7	7	9	8	9	7	1	7

4 Concluding remarks

This paper presents a novel methodology to create nurse schedules. A well-being score for each nurse to be assigned is computed using the past work weeks, in order to account for the fatigue of the employees. To do so, a random forest regressor is used and considers many attributes based on the last actual schedule. The goal with this parameter is to influence the optimization model, a mixed-integer linear problem that computes the schedules, in order to assign more shift preferences and less overtime, for example, to the employees that have a lower well-being score. Results show that the schedules provide fair schedules that could lead to reduced absenteeism. Future work based on this project would require actually implementing the computed schedules in a work place to assess the methodology in practice.

References

[1] Amal Al-Rasheed. Identification of important features and data mining classification techniques in predicting employee absenteeism at work. International Journal of Electrical & Computer Engineering (2088–8708), 11(5), 2021.

[2] Mélisende Brazeau. Affectation des infirmières aux salles de l'unité d'endoscopie digestive du centre hospitalier universitaire de sherbrooke. Master's thesis, Université de Montréal, Avril 2013.

[3] Michael R Chernick. Bootstrap methods: A guide for practitioners and researchers. John Wiley & Sons, 2011.

[4] Vinícius G Costa and Carlos E Pedreira. Recent advances in decision trees: An updated survey. Artificial Intelligence Review, 56(5):4765–4800, 2023.

[5] Marco Dorigo and Christian Blum. Ant colony optimization theory: A survey. Theoretical computer science, 344(2-3):243–278, 2005.

[6] Robert Fourer, David M. Gay, and Brian W. Kernighan. Ampl: A mathematical programing language. In Stein W. Wallace, editor, Algorithms and Model Formulations in Mathematical Programming, pages 150–151, Berlin, Heidelberg, 1989. Springer Berlin Heidelberg.

[7] Marie-Annick Gagné. Étude transversale descriptive de l'expérience au travail des infirmières québécoises. PhD thesis, Université de Montréal, 2017.

[8] Yun-Cheng Huang, Ya-Hui Hsieh, and Fu-Yun Hsia. A study on nurse day-off scheduling under the consideration of binary preference. Journal of Industrial & Production Engineering, 33(6):363–372, 2016.

[9] Pradeep Kumar Kushwaha, Ajay Rana, Swapnil Srivastava, Aamir Saifi, Aryan Tavish, and Prateek Chaturvedi. Employee absenteeism prediction using machine learning. In 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), volume 10, pages 116–121. IEEE, 2023.

[10] Natalie Lawrance, George Petrides, and Marie-Anne Guerry. Predicting employee absenteeism for cost effective interventions. Decision Support Systems, 147:113539, 2021.

[11] Antoine Legrain, Hocine Bouarab, and Nadia Lahrichi. The nurse scheduling problem in real-life. Journal of medical systems, 39(1):160, 2015.

[12] Edival Lima, Thales Vieira, and Evandro de Barros Costa. Evaluating deep models for absenteeism prediction of public security agents. Applied Soft Computing, 91:106236, 2020.

[13] Arezou Mobasher, Gino Lim, Jonathan F. Bard, and Victoria Jordan. Daily scheduling of nurses in operating suites. IIE Transactions on Healthcare Systems Engineering, 1(4):232–246, 2011.

[14] Isabel Herrera Montano, Gonçalo Marques, Susel Góngora Alonso, Miguel López-Coronado, and Isabel de la Torre Díez. Predicting absenteeism and temporary disability using machine learning: A systematic review and analysis. Journal of medical systems, 44(9):162, 2020.

[15] Fabian Pedregosa and al. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.

[16] Ann E Rogers, Wei-Ting Hwang, Linda D Scott, Linda H Aiken, and David F Dinges. The working hours of hospital staff nurses and patient safety. Health affairs, 23(4):202–212, 2004.

[17] Luiz Henrique A Salazar, Valderi RQ Leithardt, Wemerson Delcio Parreira, Anita M da Rocha Fernandes, Jorge Luis Victória Barbosa, and Sérgio Duarte Correia. Application of machine learning techniques to predict a patient's no-show in the healthcare sector. Future Internet, 14(1):3, 2021.

[18] Andrea M Stelnicki and R Nicholas Carleton. Mental disorder symptoms among nurses in canada. Canadian Journal of Nursing Research, 53(3):264–276, 2021.

[19] Seyyed Abolfazl Vagharseyyedin, Zohreh Vanaki, and Eesa Mohammadi. The nature nursing quality of work life: An integrative review of literature. Western Journal of Nursing Research, 33(6):786–804, 2011. PMID: 20719995.

[20] Maya Widyastiti, Amril Aman, and Toni Bakhtiar. Nurses scheduling by considering the qualification using integer linear programming. Telkomnika, 14(3):933–940, 2016.

[21] Wei Xiang. A multi-objective aco for operating room scheduling optimization. Natural Computing, 16(4):607–617, 2017.

[22] Wei Xiang, Jiao Yin, and Gino Lim. An ant colony optimization approach for solving an operating room surgery scheduling problem. Computers & Industrial Engineering, 85:335–345, 2015.