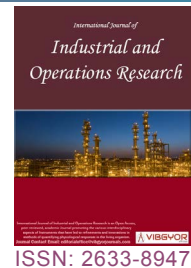


Predictive Analytics to Increase Roster Robustness in an Inbound Call Center



Lena Wolbeck* and Benedict Seyer

Information Systems Department, School of Business and Economics, Freie Universität Berlin, Germany

Abstract

Staff rostering is a crucial task in inbound call centers, as personnel costs usually account for the largest share of operating costs. Uncertainty of capacity, such as the presence of agents is often disregarded during rostering. This paper addresses the problem of uncertainty by using predictive analytics to predict agent absences and thus increase roster robustness. Operational data from four years of a call center serves as a basis for our use case. Predictors include characteristics of the service agents such as attendance history and regular working hours as well as other factors such as the weekday. Of the prediction algorithms tested, decision trees outperform other predictive modeling approaches. Evaluation based on an expected value framework shows that the predictive analytics approach performs best compared to the planned, unchanged roster and a general staff surcharge of 10%.

Keywords

Staff rostering, Uncertainty of capacity, Decision tree, Predictive modeling

Introduction

In many service industries, it is particularly important to avoid understaffing. One example is health care where a minimum number of nurses are required and a staff shortage, e.g., due to illness, has to be avoided. Another example is companies that provide full-service availability to customers, and in which an underestimation of demand or an unforeseeable absence of staff leads to a loss in revenue. We use an inbound call center offering secretary service via telephone as an example for the latter within this paper.

The call volume in an inbound call center varies highly between days as well as within a day. Ac-

cordingly, forecasting demand for service agents with specific skills is important in order to avoid unanswered calls and the resulting loss of revenue. Therefore, generating a roster that meets the work requirements is needed. Rostering is the task of assigning service agents to a predefined set of shifts in order to meet the demand taking into account contractual, labor and other rostering rules as constraints [1]. Commonly, cost minimization is the foremost objective for rostering since personnel costs account for the largest share of operating costs in a call center (around 60-70%) [2,3]. However, rosters are created a priori and during operations, further costs may arise due to roster disruptions. To avoid this, one can try to consider

***Corresponding author:** *Lena Wolbeck, Information Systems Department, School of Business and Economics, Freie Universität Berlin, Garystr. 21, Berlin 14195, Germany*

Accepted: September 10, 2020; **Published:** September 12, 2020

Copyright: © 2020 Wolbeck L, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Wolbeck and Seyer. *Int J Ind Operations Res* 2020, 3:007

Citation: Wolbeck L, Seyer B (2020) Predictive Analytics to Increase Roster Robustness in an Inbound Call Center. *Int J Ind Operations Res* 3:007

uncertainty already during staff rostering [4]. Three different types of uncertainty can be distinguished: Uncertainty of demand, arrival and capacity [5]. In this paper, we are looking at the uncertainty of capacity, as service agents represent human resources and their presence cannot be taken for granted. There are several strategies to deal with such uncertainty, including: Substitutions of probably absent service agents, reserve shifts for reactive substitution or a general increase in staff demand in order to cover possible absences [3,6-10]. Another more sophisticated option is using predictive analytics to estimate staff absences [3,9,11] and to adjust the roster accordingly. Various predictors such as attendance history, demographic or job characteristics can be used [12-19].

This study examines the use of predictive analytics to create more stable rosters in order to increase the robustness against short-term disruptions. Using predictive analytics, we aim at predicting staff absences and thus the target variable specifies whether an agent is absent or not. Real-world data of a European call center serves as the basis for a case study. Following the CRISP-DM process [20], the data provided are used to test various predictive models. We examine which model fits best for practical application of staff rostering and take accuracy as well as the true positive rate of the prediction results into account, because we have to deal with imbalanced data. In this paper, we discuss only the best performing model in more detail. In addition to the analysis for model selection, we evaluate the resulting costs of the best model using the expected value framework [21]. In our use case, it is the costliest case to predict an absent agent as present, since significant revenue losses are expected. Furthermore, the number of needed changes with and without using a predictive approach is compared in order to discuss the roster robustness. As our main contribution, we identify the use of predictive analytics to forecast employee absences to adjust an initial roster. The potential of such a modified roster results from a reduced number of unforeseen disruptions during roster operation. So far, an approach like this has not been discussed in the literature.

The remainder of this paper is organized as follows. In the next section, we introduce the context and the problem in more detail and shortly present the state of the art in the field of absence predic-

tion. In the third section, we discuss our use case of staff rostering in a European call center, the data provided and the process of data preparation. In section 4, we provide a description of our analysis and an evaluation of the best performing prediction model. Finally, we discuss our findings and give an outlook into future research as well as practical implications.

Problem Description and State of the Art

In order to get an overview on predictive analytics used in the area of staff availability, we examine different fields of scientific literature. First, we point out where the problem considered occurs in the staff rostering process in a call center. Regarding this, the focus is on strategies for dealing with uncertainties in capacity. Then, we discuss predictors for the absence of staff from the literature and give a brief introduction to predictive analysis.

Staff rostering in call centers

An inbound call center offers services via telephone for their own or external companies and deals with incoming calls from customers [3]. Due to variable call volume, forecasting the number of incoming calls is enormously important for a call center in order to optimize the staff efficiency and thus personnel costs. The whole staff rostering process can be divided into four sub problems [22]:

1. Forecasting Demand: Prediction of call volume in time intervals;
2. Determination of Work Requirements: Definition of flexible staff demand as (minimum) number of service agents;
3. Shift Scheduling: Generation and selection of shifts according to the needed qualifications such as language skill;
4. Rostering: Assignment of service agents to the defined shifts and creation of a roster.

These steps are usually executed sequentially, since the outcome of the previous step serves as input. An initial roster represents the overall result of staff rostering. This roster is adapted to the expected conditions and assumes a deterministic environment. However, in practice, roster disruptions such as short-term agent absences might disturb business operations and lead to changing requirements resulting in staff shortages. In this study, we focus on a subsequent step to initial rostering in order to

improve a roster in terms of robustness and to make roster operations more stable to agent absences.

Staff rostering is proven to be extremely difficult, especially when skills are taken into account and there is no fixed shift system [1,23]. Uncertainties are hardly considered in order to increase the roster's robustness and to reduce the number of shift changes during business operations [4,24]. The volatile volume of calls on the one hand and the unforeseeable presence of allocated staff on the other hand, who may be absent due to short-term illnesses, cause a high level of uncertainty in rostering [25]. Following Van den Bergh, Beliën [5], there are three types of uncertainty. For staff rostering in a call center, uncertainty of demand (length of a call) and uncertainty of arrival (time of a call) are important issues. These two aspects are often linked and are considered together in many publications, e.g. Green, Kolesar [26], Atlason, Epelman [22] and Avramidis, Chan [25]. In this paper, we focus on the third type - uncertainty of capacity (staff availability). Van den Bergh, et al. [5] identified only a few publications dealing with this kind of uncertainty - which is significantly less compared to the first two aspects. These include, among others, Bagatourova and Mallya [27] who use a heuristic combined with a simulation to incorporate uncertain employee presences in scheduling. Strategies for dealing with uncertainty in capacity are explained in the following section.

Solution approaches for dealing with uncertainty in capacity

Uncertainty of capacity refers to the deviation of allocated and actual present staff. There are various strategies for dealing with uncertainties in capacity in staff rostering. These include the use of reserve staff that serves as capacity buffer which is common in the airline industry [7]. Reserve agents are available for a short-term shift allocation, but this procedure is highly cost-intensive and therefore not applicable in every use case. Another proactive possibility is to consider staff substitutability during rostering and thus increase schedule flexibility [8]. In case of a shortage, rerostering is necessary to reassign vacant shifts to free agents. Such a reactive rerostering is time consuming [6], disrupts operations and can lead to service agent dissatisfaction [9]. In call centers, another option is to increase the staff demand by a surcharge [3] which also increases personnel costs [6]. Alfares [28], for example,

uses the company's standard surcharge of 10% on staff demand which is traced back to absences due to vacation, training or illness.

Another more sophisticated option is to adjust the roster, more precisely the demand or the allocation of service agents, in such a way that a staff shortage can be avoided even if an agent is absent. Better than a general surcharge are concrete buffers for service agents, which should be defined based on a case specific procedure. Ingels and Maenhout [10] examine how to define optimal buffers by including them in their optimization model to increase roster robustness. Doing this, they create a roster with greater stability against unforeseen disruptions. Predictive analytics is also suitable for this, since such models predict additional staff requirements [9,19]. In this way, possible disruptions are proactively addressed and an alternative roster can be generated that is more robust to short-term absences and minimizes the number of short-term shift changes. In the scientific literature, predictive analytics is not common in the field of uncertainty in capacity and prediction of absences. One example, however, is Wang and Gupta [19] who use the attendance history to predict employee presence in order to achieve an even distribution with regard to the probability of an unplanned absence within a scheduling heuristic. Further predictors and studies are presented in the following section.

Predictors for agent absences

The context in this study is an absent service agent whose presence is needed to cover staffing demand. In literature, various predictors for agent absences are identified. A basic model regarding staff attendance from Steers and Rhodes [13] is based on two main factors: motivation and the possibility to be present. Brooke [12] extends this model by removing demographic as well as motivational variables and adds health-related and organizational characteristics. The study from Steel and Rentsch [15] uses similar factors as in Steers and Rhodes [13], such as job satisfaction and demographic characteristics. Steel and Rentsch [15] figure out that gender and education are good predictors, and state that the length of the period over which absences are accumulated plays a major role. Additional determinants for agent absences are contract specific details and work attitude which seems to be more important than individual characteristics [16].

In a comprehensive study, 70 potential predictors for nurse absenteeism are categorized into eight groups: Attendance history, work attitude, duration of employment, burnout and stress, management traits, staff management practices as well as individual, work and job characteristics [17]. Wang and Gupta [19] investigate unplanned absences by individual characteristics such as attendance history and other factors like shift type, weekday, holiday and weather, of which personal history turns out to be the most suitable predictor. Expected workload is another factor influencing the presence of staff, examined by Green, Savin [18]. The factors that can be used as predictors in practice mainly depend on the data available.

The aim of predictive analytics is to forecast future events that are assessed as risk using historical data and advanced methods such as machine learning [11]. In predictive analytics, a distinction can be made between supervised and unsupervised learning. Supervised Learning regards predicting a specific target variable with historic data. Depending on the datatype, classification or regression are desired methods. Using predictive analytics, possible disruptions can be addressed proactively [9].

To the best of our knowledge, there is no study concerning the prediction of agent absences in a call center in order to increase roster robustness. However, predictive models are used to forecast varying demand for staff rostering in call centers, for example as presented in Gans, Koole [3] as well as in Wooff and Stirling [29]. In this paper, we aim at closing this gap. Therefore, we proceed according to a common approach for data analytics projects, the CRISP-DM process by Chapman, Clinton [20]. They suggest an iterative process consisting of six major steps which are described within the next section.

Call Center Data

In this paper, the staff rostering process of a European call center serves as use case. CRISP-DM is an iterative process consisting of various phases, starting with business and data understanding in order to determine the prediction goals. Then, the historical data received is prepared for the application of different models. These steps are described within this section. Afterwards, in the modelling phase (section 4), we generate different models and by comparing them, we determine the best of

them for prediction of agent absences. This model is then evaluated concerning the originally defined use case and prediction goals.

For the analysis, we use real-world data from a European call center that employs up to 400 agents. This call center handles secretary tasks for companies and is paid by these for each call answered. The revenues per call vary according to contractual agreements and call length. Since the call center provides services for companies all over Europe, for an incoming call a service agent with appropriate language skills has to be available. Therefore, not only the general call volume, but also language skills have to be considered when rostering. Unplanned absences lead to major disruptions in business operations and have a negative impact on revenue. Thus, the call center proactively tries to reduce the effects of such absences: Absences are predicted to allow for adjusting the demand for agents whenever there is a risk for understaffing.

Following Gans, Koole [3], a call center produces a large amount of data which can be divided into operational, marketing, human resources, and psychological data. We use two operational data sets (roster and log-in times) for extracting the executed shifts and absence notes during the period of 2014 to 2017. The roster data consists of 617,695 rows of data and five individual attributes about start and end times, task characteristics, and the assigned agent. Each row represents one specific task, for example, taking the phone calls with a specific language skill. Several tasks can add up to one shift. As a result, the data set contains 59,291 shifts based on 30-minutes intervals for the analysis. The log-in time's data contains information about when an agent was actively available for taking phone calls. The log-in times contain 3,138,303 rows with the same attributes as the roster data. Additionally, we consider a human resource data set with 15 attributes containing information about each agent's seniority, skills, contract agreements and working time schemes as well as demographic characteristics. These attributes can serve as predictors for the analysis in addition to shift properties such as day, time and required language skill.

To prepare the data for predictive analysis and enhance data quality, we perform several steps. First, exact duplicates are removed. Duplicates that can only be detected by domain knowledge are completely removed, since backtracking to the

original entry is not possible. Shifts are treated as duplicates if they affect the same agent in the same time period. Agents are considered duplicates if their ID is identical.

Second, we look at missing values. Some attributes with missing values are treated with specific domain knowledge. For the remaining missing values, their number is checked. Attributes with more than one third missing values are removed due to lack of explanatory power. There are different types of missing values and accordingly proper ways to handle them [30]. This study deals only with values missing completely at random and we replace cardinal scaled attributes with mean, ordinal with median and nominal with mode.

Third, the three data sets are merged to one data set. All information concerning the shifts are in the roster and log-in times data. Each roster shift is checked for a matching log-in and marked accordingly. Doing this, we receive a data set with the actually performed shifts and the planned shifts not performed due to an agent absence. Then, the human resource data adds 15 background attributes for each agent. The resulting data set serves as a basis for the predictive modeling.

Fourth, we remove absences contained in the data, which are known in advance. This analysis only considers several types of illness as unplanned absences, because planned absences such as vacation and overtime compensation need to be agreed upon and are not included in the data. However, the planned roster already contains absences due to sickness, which are mostly long-term illnesses or maternity leave periods. As we want to predict unexpected short-term absences, such cases are not considered in the prediction and thus deleted.

The next vital step is to prepare the target variable. It is therefore necessary to discuss how the target variable is modeled, even if modeling decisions are part of the modeling phase (see section 4). In the raw data, the target variable reveals several types of absences as well as further shift information. There are two options for preparing: Either the target variable is regarded as a logistic decision (whether an agent is absent or not for a certain shift) or the target variable is treated linear (to calculate a surcharge for a particular period of time). Linear modeling provides an aggregated method for the defined goal, but can still be calculated with

a logistic model. Choosing an appropriate time period for estimating the number of absent agents is a difficult for the linear decision. Since the roster is planned in 30-minute intervals and the staff demand may vary between the intervals, a time period of 30 minutes seems to be useful. But in this way a shift is regarded several times. In addition, all background information of the service agents affected must be combined to achieve the aggregated linear output. Therefore, the personal information taken into account in the literature could not be considered completely. For these reasons, we prepare the target variable for a logistic decision. The binary variable is created as follows: If an agent is sick all day and thus absent for the whole shift, the shift is obviously marked as absent. However, if a service agent leaves during a shift due to illness, the entire shift is marked as absent if more than 50% of the shift duration is affected. This method is adapted from practice. The developed target variable reveals 5,817 absent shifts out of a total of 57,359 shifts, which corresponds to 10.14%.

In the sixth step, the data for modeling is further improved by extracting additional information of certain attributes or combining given attributes into new information. In particular, we examine the date specification further for date characteristics such as weekdays or holidays. The combination of shift and agent information offers additional potential. Doing this, the final list of predictors includes 22 attributes:

- Time attributes: *Bridge day, day of the week, holiday, month, school holiday, shift length, year;*
- Demographic attributes: *Age, graduation, marriage status, number of children, severe disability, sex, state, tax class;*
- Contractual attributes: *Duration of employment, less than one week to retirement from the company, regular working hours;*
- Work-related attributes: *Attendance history, several lengths of language skill, several lengths of special shift tasks, weekly working hour difference.*

Before modeling, we make some further adjustments to the data. Categorical attributes are converted into dummy attributes, as well as a standardization of data for use in some models is done. In this analysis, a standard scaling is used by subtracting the mean and scaling to unit variance. A

problem for some models is imbalanced data - as in our case study: The majority of the data entries examined show normal presence without absence. Guo, Yin [31] mention different techniques to target this issue. Extensively discussed is the strategy of under-sampling and over-sampling [31-33]. We do not use under-sampling, since under-sampling bears the risk of losing important information [31], such as rare events like holidays. Likewise, we do not choose over-sampling, as it increases the probability of overfitting and causes a strong bias in the data set [31]. Another strategy is to manipulate the models by class weights. The analysis described in the following uses class weights along with further evaluation criteria as a strategy for dealing with imbalanced data. In addition, challenges arise in the evaluation of the prediction results due to imbalanced data, e.g., accuracy might not be the best criteria. Therefore, further information is often needed and the use of domain knowledge for intensive checks of the data is recommended [31].

Predictive Modeling and Model Selection

According to the steps of the CRISP-DM process [20], the prepared data is now used for modeling. The modeling is done with python (3.7.1) and the library of scikit-learn (SK-Learn, 0.21.2)₁ [34] - a standard library for data analysis. In order to identify a suitable model for predicting absences, we test different modeling approaches (section 4.1). Afterwards, the best prediction model is further evaluated in order to derive conclusions for the case study (section 4.2).

Predictive analytics for agent absences

In this study, the target variable to predict is binary, indicating whether an agent is absent or not. Based on this outcome, supervised learning techniques are used [35]. Of these, classification is the most appropriate based on the prediction intended. An important issue in classification is overfitting, which is why we pay special attention to it. Applying a cross-validation can lead to overfitting because the information might leak into the models [36,37]. Based on Stone [38], Leave-One-Out-Cross-Validation Nested-Cross-Validation is derived. This is further extended by Krstajic, Buturovic [39], who recommend repeating validations and ensuring equal class distributions (stratified splits). Following this, to avoid overfitting, a Repeated-Stratified-Nested-Cross-Validation (RS-NCV) is used. In this analy-

sis, the outer cross-validation is repeated once with three splits. Within each split on two parts an inner cross validation is executed to determine the best performing model on average, which we then validate using the third part. The inner cross validation is repeated ten times with three splits. Again, we use two splits to train the model and, for evaluation, we use the third split. For each modeling approach tested, the average of the evaluation criteria is calculated. Overall, the RS-NCV is performed twice: First, choosing the best model and, second, identifying the most suitable parameters of the best model.

There are various methods for modeling classification problems such as linear classifications, support vector classifiers, nearest neighbor analysis, decision trees and ensemble learning methods. We select a set of modelling techniques predicting classifications that meet the required criteria. Linear and quadratic discriminant analysis tries to separate the solution space with linear, respectively quadratic decision boundaries [35]. In addition to linear methods, the logistic regression tries to directly predict the class probabilities [35]. SK-Learn offers five different solvers for logistic regression. In this study, we use the solver saga presented by Defazio, Bach [40], because it is faster and offers the "l1" penalty. Using "l1"-penalty can lead to sparser solutions, less attributes are used for classification and the risk of overfitting is decreased. Support vector classifiers (SVC) can be considered as borderline cases as they produce non-linear boundaries by transforming the data set and creating linear boundaries within these transformed features [35]. Using SK-Learn, it is possible to vary between different kernels. This study compares a linear kernel, a radial basis function kernel (rbf.) as well as a polynomial kernel (poly) of third degree. In contrast to linear models, a memory based classifier is used - the nearest neighbor classifier [35]. That classifier maps classes with the help of the majority vote among the k nearest neighbors of the entry to predict [35]. SK-Learn automatically chooses the best algorithm (Brute-Force, K-D Tree [41] or Ball Tree [42]) for calculating the nearest neighbor. Another popular approach is the use of decision trees. Decision trees separate the solution space in dichotomous parts in several iterations [35]. SK-Learn adopts the CART algorithm of Breiman [43], which is quite similar to the C4.5 by Quinlan [44]. As decision criteria SK-Learn uses the Gini impuri-

ty. Furthermore, we exploit methods belonging to the group of ensemble learners. Ensemble learners combine several simpler models [35]. One of them is random forest like Breiman [45] proposes, that are constructed upon different trees. Unlike the original implementation, SK-Learn considers the probability average for estimating the class. Further ensemble learners that are tested are boosters, such as the ada boost of Freund and Schapire [46] and the gradient boosting classifier.

In addition to multiple kernels, we compare some other variations. For the k nearest neighbors' model, different values for k are used. Regarding the ada boost, a series of base estimators with different complexities are compared. Finally, for some models the possibility to use class weights is exploited to address imbalanced data. The use of this specification is marked with 'bal.'. Table 1 summarizes the ten best performing predictive models tested and an evaluation of those described in the following.

In order to select the best prediction model, we look at the models' accuracies. Since accuracy is not sufficient to evaluate the prediction based on imbalanced data [31], we also examine the true positive rate in the following. Due to the huge attendance-rate, the majority classifier's accuracy is 89.86%. We aim at identifying a model that has a higher accuracy than the majority classifier and a high true positive rate at the same time. This allows us to find a precise model, which avoids misclassifications that would be expensive in roster operations of our use case.

Table 1 shows the accuracy and true positive rate based on the inner cross-validation results of the ten best performing models. The SVC models each have a very high true positive rate, but due to their low accuracy with an average of 72% they are not suitable for a prediction. The other models all have a high accuracy (between 90.11 and 91.97%), which is higher than that of the majority classifier. Accordingly, the true positive rate is now decisive for the selection between these eight predictive models. The true positive rate varies greatly, the remaining models are sorted by ascending rate from the fourth row of Table 1. The best performing model is the decision tree without classifier weights, because it has an average true positive rate of 56.35%. The results show that even though the decision tree has the highest true positive rate, it lacks accuracy compared to ada boost or random forest. However, the true positive rate is much higher and thus the imbalanced data problem is solved appropriately.

The predictions based on the resulting decision trees may still suffer from overfitting [35]. In order to improve the generalizability and reduce complexity, we use the maximum tree depth, the maximum number of leaves, and the maximum number of attributes per split as stopping criteria. Using RS-NCV, we compare decision trees of varying complexity. The outer iteration of the RS-NCV is repeated twice, since the models are much more similar. The RS-NCV ensures that the reduced decision tree with the best generalization performs best. The best reduced tree is generated with a maximum

Table 1: Evaluation of predictive models.

	Validation Set 1		Validation Set 2		Validation Set 3	
	ACC	TPR	ACC	TPR	ACC	TPR
SVC (rbf, bal.)	71.21	75.28	70.89	75.48	72.67	73.93
SVC (poly, bal.)	71.77	74.78	71.13	75.17	73.47	72.98
Nearest Neighbor (3)	90.23	31.04	90.11	31.59	90.18	31.93
Random Forest (bal.)	91.67	31.17	91.58	31.32	91.58	31.35
Random Forest	91.79	32.80	91.58	32.21	91.67	32.58
AdaBoost (5)	90.91	35.11	90.80	34.87	90.79	34.79
AdaBoost (15)	91.97	44.15	91.78	43.50	91.86	43.67
AdaBoost (10)	91.76	45.13	91.66	45.40	91.58	44.55
Decision Tree (bal.)	90.79	53.96	90.65	53.71	90.54	53.93
Decision Tree	90.87	56.39	90.82	56.68	90.71	55.97
Best model	92.11	62.87	92.30	63.80	92.37	66.01

depth of 40, a maximum of 85 attributes and a maximum of 3,500 leaves. These maximum values are not reached by every decision tree, but more complex trees with less generalization are avoided. Stopping even earlier decreases the results with no increase in generalization as the RS-NCV shows.

Evaluation of the best performing decision tree

In the following, we take a closer look at the results of the best performing predictive model - a decision tree with the mentioned properties. In terms of feature importance, there is a distinct outcome regarding the impact on prediction. In Table 2 the feature importance of the ten highest ranked features with an importance greater than 1.0% is presented. The attributes *age* with 29.09% and *duration of employment* with 26.87% have the biggest impact, followed by the individual *attendance history* and *weekly working hour difference*. Overall, most attributes have an importance of up to 2%.

Table 3 shows a confusion matrix that presents

Table 2: Feature importance of the best performing decision tree.

Attribute	Feature Importance
Age	29.09%
Duration of employment	26.87%
Attendance history	9.80%
Weekly working hour difference	4.71%
Language: German	2.22%
Year	2.10%
Shift length	1.68%
Month: November	1.10%
Month: September	1.02%
Day of the week: Monday	1.01%

the most important evaluation criteria regarding a binary classifier of the best decision tree. The prediction of an agent's absence is shift-based, i.e. for each shift it is predicted whether the assigned agent will be absent or not. 1,259 cases of absence and 16,481 cases of presence are classified correctly. This is also shown in the accuracy of 92.78%. There are only 1,380 misclassifications with this model, which are quite equally spread between false negatives and false positives. With a true positive rate of 64.93%, significantly more than half of the agents' absences are predicted. At the same time, only 4.07% of actual presences are forecasted incorrectly.

To assess the economic consequences of the prediction, we use the expected value framework [21]. More precisely, we estimate the costs per hour for each of the cases in the confusion matrix in consultation with the call center's management. In the call center, costs are calculated per hour and not per shift, since the duration of the shifts varies greatly. Accordingly, the classified shifts with their respective duration in hours are considered in this evaluation step. The number of classified hours as well as the estimated additional costs per hour are summarized in Table 4. The costs correspond to the extra personnel costs due to additional allocat-

Table 4: Estimated costs and absolute (relative) number of classified hours using the best performing decision tree.

		Target Class	
		'absent'	'present'
Output Class	'absent'	12 €/h 7,865 h (6.43%)	12 €/h 4,499 h (3.63%)
	'present'	50 €/h 4,297.5 h (3.51%)	0 €/h 105,789.67 h (86.43%)

Table 3: Confusion matrix of prediction results using the best performing decision tree.

		Target Class		
		'absent'	'present'	
Output Class	'absent'	1,259 (6.58%)	700 (3.66%)	64.27% 35.73%
	'present'	680 (3.56%)	16,481 (86.20%)	3.96% 96.04%
		64.93% 35.07%	4.07% 95.93%	92.78% 7.22%

ed agents and, in the case of understaffing, to the expected decline in revenue. There are four cases:

1. There are no additional costs, if the baseline case of a present agent is predicted correctly.
2. If a replacement can be planned because an absence is predicted correctly, 12 €/h is charged for the extra agent.
3. Likewise, if an absence is incorrectly predicted, we calculate 12 €/h for double staffing.
4. If an absent agent is not detected and therefore no replacement is planned, additional costs in the amount of 50 €/h are assumed.

The expected value is then calculated based on the hours of each shift and its classification. For this model, the expected value reveals costs of 2.97€ per planned hour.

In order to evaluate this value and thus the benefit of the developed model, we also calculate the expected value per planned hour for two comparable cases. First, a baseline performance using the majority classifier is estimated. Due to the high number of hours in which an agent is actually present and thus the baseline prediction applies, the accuracy of the majority classifier is 89.86%. However, there are obviously no absent cases predicted correctly and 1,939 shift misclassifications result which corresponds to 12,162.5 hours. The expected value of the baseline model reveals costs of 4.97€ per planned hour. Second, we adapt a general staff surcharge of 10%, like it is practiced in the call center. This approach matches the identified absence-rate of 10.14% quite well. This second baseline model performs with a classification accuracy of 81.89% not as good as the other two. Many agents are predicted to be absent even though they are not. The true positive rate is 10%. Using such a general surcharge of agents results in 3,463.2 misclassifications and the expected value shows costs of 5.67€ per planned hour.

Comparing these three approaches, it becomes clear that across all evaluation criteria the decision tree developed in this study performs best. The prediction accuracy is higher and the expected costs as well as the number of misclassifications are lower. Hence the developed model is appropriate for its application.

Conclusion

A robust roster is essential for undisrupted business operations in an inbound call center. To increase robustness against short-term absences, staff rostering should take uncertainty in capacity into account. The presence of a service agent is a major uncertainty factor and therefore its consideration is important. This paper discusses whether the use of predictive analytics is suitable for predicting agent absences. Different predictive analytics approaches are examined. An evaluation based only on the classification accuracy is not sufficient for imbalanced data, which is why we additionally use the true positive rate to select the most suitable predictive model. A decision tree has proven to be the best model in our computational study. To increase generalization, stopping criteria for reduced decision trees are derived.

The best model is further evaluated and compared with two baseline cases: The majority classifier and a 10% staff surcharge. Using the expected value framework, we apply additional costs per planned hour from practice to get an insight into practical economic effects. Compared to the baseline cases, our prediction model leads to significantly lower additional costs. In addition, the number of disruptions that lead to understaffing can be reduced and thus the roster robustness increased. Our results show that using predictive analytics for prediction of agent absences is useful in order to achieve a more efficient staff rostering. This is meaningful both for research and practice.

In this paper, we do not consider possible correlations and the impacts of the prediction of absences on the entire staff rostering process. An inclusion of prediction models in the determination of work requirements and consecutive steps is an issue that could be addressed in subsequent studies. To validate the results and findings, predictive analysis based on further data is required. In addition, other methods such as neural networks and linear modeling may have potential for predicting agent absences.

References

1. Ernst AT, Jiang H, Krishnamoorthy M, Sier D (2004) Staff scheduling and rostering: A review of applications, methods and models. *European Journal of Operational Research* 153: 3-27.
2. Dietz DC (2011) Practical scheduling for call center

- operations. *Omega* 39: 550-557.
3. Gans N, Koole G, Mandelbaum A (2003) Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management* 5: 79-141.
 4. Wright PD, Mahar S (2013) Centralized nurse scheduling to simultaneously improve schedule cost and nurse satisfaction. *Omega* 41: 1042-1052.
 5. Jorne Van den Bergh, Jeroen Beliën, Philippe De Bruecker, Erik Demeulemeester, Liesje De Boeck (2013) Personnel scheduling: A literature review. *European Journal of Operational Research* 226: 367-385.
 6. Easton FF, Goodale JC (2005) Schedule recovery: Unplanned absences in service operations. *Decision Sciences* 36: 459-488.
 7. Ingels J, Maenhout B (2015) The impact of reserve duties on the robustness of a personnel shift roster: An empirical investigation. *Computers & Operations Research* 61: 153-169.
 8. Ingels J, Maenhout B (2017) Employee substitutability as a tool to improve the robustness in personnel scheduling. *OR Spectrum* 39: 1-36.
 9. Wolbeck L, Schlechter P, Schmitt D (2018) Nurse schedule evaluation through simulation with integrated rescheduling. *Proceedings of the Twenty-Second Pacific Asia Conference on Information Systems* 148: 1841-1848.
 10. Ingels J, Maenhout B (2019) Optimised buffer allocation to construct stable personnel shift rosters. *Omega* 82: 102-117.
 11. Mishra N, Silakari S (2012) Predictive analytics: A survey, trends, applications, oppurtunities & challenges. *International Journal of Computer Science and Information Technologies* 3: 4434-4438.
 12. Brooke PP (1986) Beyond the Steers and Rhodes model of employee attendance. *Acad Manage Rev* 11: 345-361.
 13. Steers RM, Rhodes SR (1978) Major influences on employee attendance: A process model. *Journal of Applied Psychology*.
 14. Harrison DA, Martocchio JJ (1998) Time for absenteeism: A 20-year review of origins, offshoots, and outcomes. *Journal of Management* 24: 305-350.
 15. Steel RP, Rentsch JR (1995) Influence of cumulation strategies on the long-range prediction of absenteeism. *Academy of Management Journal* 38: 1616-1634.
 16. Schalk R, Van Rijckevorsel (2007) Factors influencing absenteeism and intention to leave in a call centre. *New Technology, Work and Employment* 22: 260-274.
 17. Mandy M Davey, Greta Cummings, Christine V Newburn-Cook, Eliza A Lo (2009) Predictors of nurse absenteeism in hospitals: A systematic review. *Journal of Nursing Management* 17: 312-330.
 18. Green LV, Savin S, Savva N (2013) "Nurse vendor problem": Personnel staffing in the presence of endogenous absenteeism. *Management Science* 59: 2237-2256.
 19. Wang W-Y, Gupta D (2014) Nurse absenteeism and staffing strategies for hospital inpatient units. *Manufacturing & Service Operations Management* 16: 439-454.
 20. Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, et al. (2000) CRISP-DM 1.0. Step-by-step data mining guide. CRISP-DM consortium.
 21. Provost F, Fawcett T (2013) *Data science for business: What you need to know about data mining and data-analytic thinking*. Sebastopol, CA: O'Reilly Media, Inc.
 22. Atlason J, Epelman MA, Henderson SG (2008) Optimizing call center staffing using simulation and analytic center cutting-plane methods. *Management Science* 54: 295-309.
 23. Philippe De Bruecker, Jorne Van den Bergh, Jeroen Beliën, Erik Demeulemeester (2015) Workforce planning incorporating skills: State of the art. *European Journal of Operational Research* 243: 1-16.
 24. Bard JF, Purnomo HW (2005) Short-term nurse scheduling in response to daily fluctuations in supply and demand. *Health Care Management Science* 8: 315-324.
 25. Athanassios N Avramidis, Wyeon Chan, Michel Gendreau, Pierre L'Ecuyer, Ornella Pisacane (2010) Optimizing daily agent scheduling in a multiskill call center. *European Journal of Operational Research* 200: 822-832.
 26. Green LV, Kolesar PJ, Whitt W (2007) Coping with time-varying demand when setting staffing requirements for a service system. *Production and Operations Management* 16: 13-39.
 27. Bagatourova O, Mallya SK (2004) Coupled heuristic and simulation scheduling in a highly variable environment. *Proceedings of the 2004 Winter Simulation Conference* 1 and 2: 1856-1860.

28. Alfares HK (2007) Operator staffing and scheduling for an IT-help call centre. *European Journal of Industrial Engineering* 1: 414-430.
29. Wooff DA, Stirling SG (2015) Practical statistical methods for call centres with a case study addressing urgent medical care delivery. *Annals of Operations Research* 233: 501-515.
30. Little RJA, Rubin DB (2014) *Statistical analysis with missing data*. John Wiley & Sons, Hoboken, NJ.
31. Guo X, Yin Y, Dong C, Yang G, Zhou G (2008) On the class imbalance problem. *Proceedings of the Fourth International Conference on Natural Computation* 4: 192-201.
32. Japkowicz N (2000) Learning from imbalanced data sets: a comparison of various strategies. *AAAI Workshop on Learning from Imbalanced Data Sets* 68: 10-15.
33. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* 16: 321-357.
34. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, et al. (2011) Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12: 2825-2830.
35. Hastie T, Friedman J, Tibshirani R (2013) *The elements of statistical learning: Data mining, inference, and prediction*. (2nd edn), Springer.
36. Cawley GC, Talbot NL (2010) On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research* 11: 2079-2107.
37. Varma S, Simon R (2006) Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 7: 91.
38. Stone M (1974) Cross-validated choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)* 36: 111-133.
39. Krstajic D, Buturovic LJ, Leahy DE, Thomas S (2014) Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics* 6: 10.
40. Defazio A, Bach F, Lacoste-Julien S (2014) SAGA: A fast incremental gradient method with support for non-strongly convex composite objectives. *Advances in Neural Information Processing Systems* 27: 1646-1654.
41. Bentley JL (1975) Multidimensional binary search trees used for associative searching. *Communications of the ACM* 18: 509-517.
42. Omohundro SM (1989) *Five balltree construction algorithms*. International Computer Science Institute, Berkeley, California.
43. Breiman L (2017) *Classification and regression trees*. Routledge.
44. Quinlan JR (2014) *C4.5: Programs for machine learning*. Elsevier.
45. Breiman L (2001) Random forests. *Machine Learning* 45: 1-32.
46. Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55: 119-139.