A supervised learning model for medical appointments no-show management

Inês Ferreira and André Vasconcelos*

INESC-ID,
Instituto Superior Técnico,
R. Alves Redol 9, 1000-029 Lisbon, Portugal
Email: ines.d.ferreira@tecnico.ulisboa.pt
Email: andre.vasconcelos@tecnico.ulisboa.pt
*Corresponding author

Abstract: A no-show is a phenomenon that leads to an efficiency decrease in various sectors, including in the healthcare sector. This research proposes the usage of supervised learning techniques to predict medical appointments no-shows occurrence and to find patient replacements to fulfil last-minute vacancy slots. The prediction is performed using a classification algorithm that computes the probability of no-show for each patient based on features that have shown to influence his or her decision, such as the waiting time, the day of the appointment and the number of previous no-shows, among others. The features are extracted from two distinct healthcare datasets. In order to select the most suitable classification algorithm, a ten-fold cross-validation is used to perform a comparative analysis among the most used algorithms applicable to this type of classification problems. The gradient boosting algorithm proved to have the best performance in estimating no-shows.

Keywords: no-show; healthcare; supervised learning; classification algorithms; cross-validation.

Reference to this paper should be made as follows: Ferreira, I. and Vasconcelos, A. (2022) ‘A supervised learning model for medical appointments no-show management’, *Int. J. Medical Engineering and Informatics*, Vol. 14, No. 1, pp.90–104.

Biographical notes: Inês Ferreira received her Master’s degree in Computer Science and Computer Engineering at Instituto Superior Técnico (IST) of Lisbon University. She holds a BSc in Computer Science and Computer Engineering (Instituto Superior Técnico, Lisbon University). She developed a master thesis on ‘Last minute medical appointments no-show management’. Currently, she works at Accenture, architecting, implementing, and delivering state-of-the-art IT projects.

André Vasconcelos received his PhD in Computer Science and Computer Engineering (Instituto Superior Técnico, University of Lisbon). He also has an MBA (The Lisbon MBA) from Universidade Católica and Universidade Nova de Lisboa. He is currently an Assistant Professor in the Computer Science and Engineering, Instituto Superior Técnico, University of Lisbon and senior researcher at INESC-ID. He has developed strategic advice in the areas of information systems and business processes, in several organisations. In the scope of his professional activity, he develops specialised advisory activities in the field of enterprise architecture, in several projects of digital transformation. He has about 100 publications and has been involved in several international and national research projects related to bioinformatics, enterprise architecture, identification, and e-government including several European projects.

1 Introduction

The world is going through a phase of rapidly escalating costs which implies an efficient use of resources. However, the efficiency of various sectors is increasingly affected by no-shows. This research focus specifically in the healthcare sector, in which there are at least two negative effects whenever a scheduled patient misses an appointment without cancelling: firstly, the clinic's resources are wasted and secondly, medical service is denied to patients who could have benefited from the respective time slot. MedClick is an online platform that aims to help medical service providers to increase the efficiency of their practices. This research was developed in the context of that platform and it is focused on helping MedClick to achieve their goals by providing tools that, among other features, support the prediction of no-shows and also the fulfilment of 'last-minute' vacancy slots, by notifying patients whose needs and restrictions are best suited.

1.1 Objectives

The goal of this research is on the reduction of no-shows from patients in medical appointments in order to increase the productivity and the resource usage in healthcare services. To achieve the desired goal, this research provides a solution is able to:

- Minimise the occurrence of no-shows by using strategies to reduce their probability, such as reminder notifications.
- Build a supervised learning model capable of predicting no-shows based on a given set of features. For this step, several classifications algorithms are explored in order to find the most suitable for no-shows' problem.
- In the case of detecting a future no-show, the system must try to find a suitable replacement.
- Extract data from healthcare datasets and pre-process it until it is ready to be sent through the learning model and provide reliable predictions.

The above aims are expected to complement and improve the no-show algorithm structure that was previously implemented in MedClick (Sousa and Vasconcelos, 2020).

1.2 No-shows

Several studies have been developed, focused on detecting the origin of no-shows and finding possible solutions to this problem. Regarding the causes of no-shows, the most reported reason is when the patient forgets his or her appointment (Neal et al., 2005). Therefore, appointment reminders are commonly used to prevent that from happening (Leong et al., 2006). Several other reasons are reported for no-shows, such as: financial problems, lack of transportation, competing priorities, bad quality of the service and patient health status (Gany et al., 2011).

Besides some scheduling systems aimed at reducing no-shows, such as overbooking and open access (Cameron et al., 2010), there are some strategies such as patient education or patient sanctions. The first consists of providing all the important information in order to ensure that patients feel secure about their appointment. The second is used as an attempt to change the patient's behaviour (Guse et al., 2003). In addition, the field of supervised learning has been increasingly explored to reduce and predict no-shows. The previous solution implemented in MedClick uses a hybrid approach that combines logistic regression, as a population-based method, and Bayesian Inference, as an individual-based method. This approach has already been used for reducing no-shows in the healthcare sector (Alaeddini et al., 2015). There are many other studies targeted to predicting no-shows in different sectors, such as in airline companies (Lawrence et al., 2008) and in the hospitality sector (Antonio et al., 2017).

Ample literature is available discussing predictors of no-shows, which can be divided into two categories: patient's characteristics and appointment's characteristics. Regarding the first, several studies have demonstrated that no-show patients tend to be younger (Cashman et al., 2004), unmarried (Daggy et al., 2010), uninsured (Bennett and Baxley, 2009), with psychosocial problems (Compton et al., 2006) and finally, with prior no-show history. The second category includes the day of the scheduled appointment, the clinic's proximity and, finally, the waiting time, which corresponds to one of the major problems in healthcare services (George and Rubin, 2003). Although several studies proved the impact of these features, it is important to bear in mind that the results may vary depending on where the study is done.

1.3 Supervised learning algorithms

The idea of supervised learning is to analyse a set of training data and to train a function capable of predicting the output given new input data. Supervised learning problems can be further divided into regression and classification problems. The prediction of no-shows corresponds to a classification problem, in which a function must predict the class of a given observation. The effectiveness of this techniques depends on the performance of the chosen algorithm and, therefore, it is important to test and consider different approaches. For this research, four classification algorithms were considered, namely logistic regression, k-nearest neighbours (k-NN), random forests and gradient boosting. The logistic regression applies a logistic function that receives a set of features along with their respective coefficients and outputs the probability of no-show (Alaeddini et al., 2015). The coefficients are estimated during the training phase and their values are log odd ratios which may give information on the impact of each feature. Positive coefficients correspond to higher odds of occurring no-show and negative coefficients corresponds to lower odds. A feature with a coefficient near 0 has a low impact on the prediction. The k-NN is one of the simplest algorithms and it predicts the class of a given instance based on the classes of the k-NN (Altman, 1992). Random forest and gradient boosting are both ensembles of decision trees, in which the idea is to produce a strong learner by combining a group of weak learners. The first is an extension over bagging which consists of building multiple trees in parallel and combining them together to obtain a more stable and accurate prediction (Ho, 1995). The second is an extension over boosting, in which the learners are built in a sequential way and each tree corrects the classification error of the previous tree (Trevor et al., 2009).

1.4 Document outline

This paper is structured as follows: Section 2 describes the no-show approach that was previously implemented in MedClick. Section 3 describes the proposed solution.

Section 4 presents the evaluation tests that were performed along with the respective results and, finally, Section 5 concludes the paper.

2 MedClick previous solution

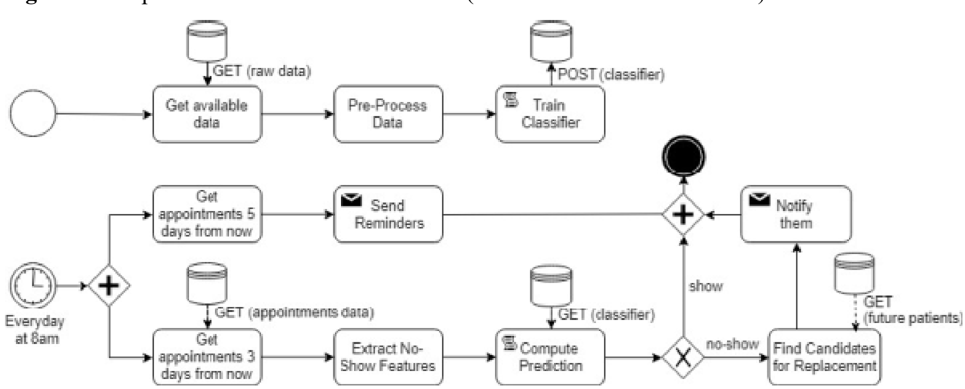
A no-show algorithm was already implemented in MedClick, in which the goal was to find patients interested in filling 'last-minute' vacancy slots (Sousa and Vasconcelos, 2020). After detecting a slot to be filled, the system starts by getting the filtered list of candidate patients, who would later be notified, from the least likely to miss the appointment to the one with the greatest probability of missing it. This requires a prior computation of the no-show probability associated with each candidate patient, which is performed using a hybrid approach that combines logistic regression, as a population-based method, and Bayesian Inference, as an individual based method (Alaeddini et al., 2015). After this step, patients are ordered accordingly to their no-show probabilities and the algorithm goes into a loop until it finds a replacement or until there are no more candidates left. At the end, if no replacement is found, the system notifies the healthcare centre that the algorithm is unable to fulfil the timeslot.

Despite the satisfactory results, there are some aspects that should have been considered in order to improve the quality of the system. One of the major limitations of the existing solution is that the algorithm that estimates the no-show probabilities is not being used to its full potential since it is only used to sort the candidate patients list according to their probabilities of missing the appointment. Instead of that, the algorithm could also have been leveraged to predict no-shows.

3 No-show management approach

In this research, a no-show module was developed primarily using NodeJS, except for some machine learning tasks that were implemented using Python due to its highly optimised libraries. The BPMN diagram in Figure 1 presents the workflow of the implemented module, in which two distinct tasks stand out.

Figure 1 Implementation of no-show module (see online version for colours)



The one that is represented at the upper part of the figure corresponds to the starting point of the module. It consists of pre-processing and using the available data to train the learning model using gradient boosting supervised learning algorithm (Trevor et al., 2009). The model is subsequently persisted into the data source in order to be used whenever the system needs to compute a no-show prediction.

In the lower part of the figure, it is represented the task that will be performed daily (e.g., every day at 8 AM), which consists of two sub-tasks. One is responsible for sending reminders to the patients with appointments five days from current date and the other one computes the probability of no-shows on the appointments three days from the current date, using the model that was previously trained. In the case of detecting a no-show, the system is responsible for notifying possible replacements, from the least likely to miss the appointment to the one with the greatest probability of missing it.

Both learning and prediction may use online or offline approaches. In this research, predictions are computed in real-time using new input data, which corresponds to an online approach. Regarding the learning, the offline approach consists of training the model once on historical data, remaining constant after being deployed to production. However, it is important to ensure that the model does not become unstable, which may happen very often. Considering this, it should be used a batch learning technique, combining both online and offline approaches (since using exclusively an online learning requires to constantly update the model as new data arrives, which is not viable). With a batch approach, the model is re-trained only after a certain number of observations have been inserted into the dataset.

3.1 Dataset

This research uses two distinct datasets. The first was provided by a Portuguese clinic, MD Clínica, and it was already considered in MedClick to test the previous algorithm (Sousa and Vasconcelos, 2020). The second was downloaded from Kaggle (https://www.kaggle.com/joniarroba/noshowappointments) and it contains data related to 110k medical appointments from Brazil.

Real-world data should not be sent through a model without first being pre-processed since it is often incomplete and it is likely to contain noisy and unreliable data. Considering that, the following sequence of pre-processing techniques was applied to both datasets:

- Removal of instances containing missing values.
- Removal of instances with inconsistent values, such as postal codes with less than seven digits, negative ages or scheduling days after the respective day of the appointment.
- Using the available data to extract features that already proved their negative impact on patient's no-show probability.
- Balancing the data by applying SMOTE (Bowyer et al., 2018) after splitting the data into training and test sets.

After the pre-processing, the list of features considered in each dataset is the following:

1 MD Clínica features: postal code, age, marital status, gender, insurance ID, number of previous appointments, number of prior no-shows, physician ID, and appointment's day.

2 Brazil features: age, gender, scheduling day, waiting time, handicap level, number of diseases, number of previous appointments, number of prior no-shows, physician ID, appointment's day.

4 Evaluation and results

There are several methods that can be used to evaluate the performance of the learning models. In this paper, a ten-fold cross-validation is used as it is one of the most efficient methods that allows model hyper parameters optimisation and it also evaluates the model performance with different subsets of data (Kohavi, 1995). The idea is to split the data into ten folds, one of which is used for testing the model and the remaining nine are used to train the model. Then, the process is repeated ten times so that each fold is used once as a test set.

The choice of the metrics depends on the type of problem. Regarding the problem of no-shows, it is known that there are typically more shows than no-shows and, therefore, the dataset used in this research is highly unbalanced since there is a negative majority class highly dominating over a positive minority class. This means that measuring only the accuracy would not be sufficient because, considering a hypothetical dataset in which 90% of the data belongs to one class, it is easy to create a classification model that gets an accuracy of 90% by simply assigning all data to the majority class. In order to get more reliable results, three additional metrics were also measured: precision, recall and F1-score (Powers, 2011).

4.1 Impact of pre-processing techniques

As mentioned in Section 3.1, raw data should not be sent through a model without first being pre-processed since it is often incomplete and likely to contain noisy and unreliable information. Considering that, several pre-processing steps were performed, and this section will cover the impact of those steps on the model's performance. The results presented in Table 1 and Table 2 describe several measures from ten-fold cross-validation, organised according to different pre-processing levels.

Initially, each model was evaluated with raw data and, as expected, the accuracy was the only measure getting good results due to the unbalanced data. This proves that accuracy is not a reliable measure in this type of problems since any model can get good results by simply assigning all data to the majority class (show), which happened in the logistic regression model, whose precision, recall and F1-score had values of 0.

As mentioned above, SMOTE technique was applied in order to balance the data. With this amendment, the models improved in general and their performance measures became more reliable, including the accuracy measure whose scores have decreased.

Finally, some features were extracted from the available data and others were discarded for being irrelevant to the problem. With this final step the models increased their performance, as shown in the lower part of both tables.

Table 1 Performance obtained at different levels of pre-processing (MD Clínica Data)

	Evaluation metric	*Classification algorithm*			
		Logistic regression	*k-nearest neighbours*	*Random forest*	*Gradient boosting*
Raw data	Accuracy	0.8278	0.8236	0.8160	0.8274
	Precision	0.0000	0.3211	0.2517	0.5178
	Recall	0.0000	0.0180	0.0315	0.0154
	F1-score	0.0000	0.0336	0.0540	0.0295
Raw data (resampled)	Accuracy	0.5846	0.5660	0.6824	0.7155
	Precision	0.2399	0.2021	0.2205	0.2720
	Recall	0.5293	0.5959	0.3330	0.3307
	F1-score	0.2943	0.3005	0.2639	0.2798
Processed data	Accuracy	0.7508	0.6527	0.7601	0.7681
	Precision	0.3552	0.2510	0.3618	0.3861
	Recall	0.3925	0.5049	0.4981	0.5098
	F1-score	0.3591	0.3334	0.4163	0.4328

Table 2 Performance obtained at different levels of pre-processing (Brazil Data)

	Evaluation metric	*Classification algorithm*			
		Logistic regression	*k-nearest neighbours*	*Random forest*	*Gradient boosting*
Raw data	Accuracy	0.7981	0.7605	0.7951	0.7981
	Precision	0.0000	0.2583	0.3356	0.5296
	Recall	0.0000	0.1024	0.0163	0.0010
	F1-score	0.0000	0.1429	0.0309	0.0022
Raw data (resampled)	Accuracy	0.6246	0.5761	0.6009	0.6369
	Precision	0.2621	0.2259	0.2531	0.2709
	Recall	0.4756	0.4579	0.5099	0.4803
	F1-score	0.2933	0.2991	0.3268	0.3222
Processed data	Accuracy	0.7304	0.7416	0.8071	0.8091
	Precision	0.4035	0.4015	0.5210	0.5285
	Recall	0.6304	0.5793	0.4373	0.5342
	F1-score	0.4820	0.4734	0.4732	0.5176

In addition, the importance of each feature was analysed. In contrast to Brazil values, the features from MD Clínica dataset, have shown a balanced distribution of importance (Figure 2). The features from Brazil that prove to have the biggest impact on predicting no-shows are the time that the patient had to wait to see their physician and the patient's age (Figure 3).

Figure 2 Feature importance (MD Clínica Data) (see online version for colours)

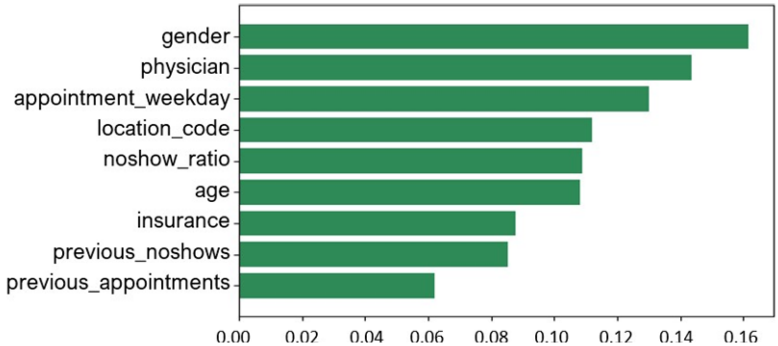
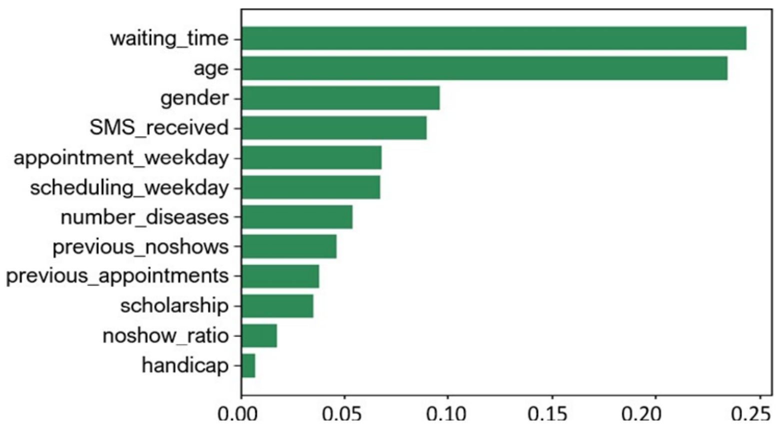


Figure 3 Feature importance (Brazil data) (see online version for colours)



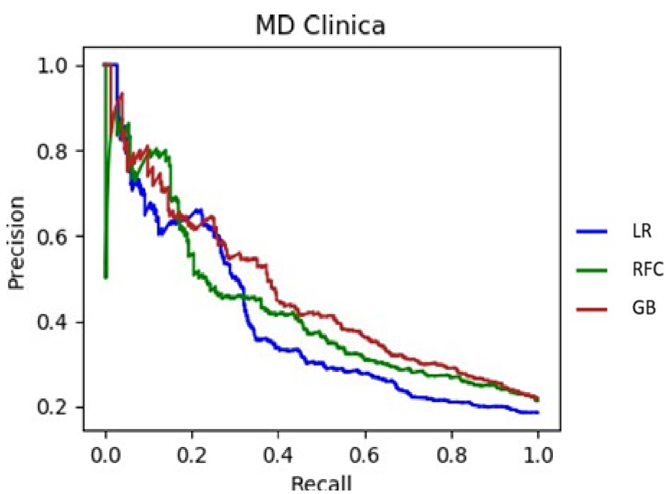
4.2 Choosing an optimal threshold

Except for k-NN algorithm that directly creates a class output (0 or 1), the remaining models return probability outputs that are subsequently converted into classes by using a threshold probability. The default value for this threshold is 0.5, which means that a probability above that value indicates positive class and a probability below indicates negative class. However, each problem must find their optimal threshold. When the data contains a negative majority class highly dominating over a positive minority class (such as in the no-show problem), the threshold must be chosen considering the precision-recall trade-off, since the precision does not depend on the number of true negatives. Precision and recall are inversely related, which means that decreasing the threshold leads to a decreased precision but to an increased recall. When choosing the optimal threshold for this research, it is important to consider the several clinics in which the solution will be applied since each approach may lead to different consequences. Hence, the threshold must be chosen considering three possible approaches:

- High precision and low recall: the model is not able to detect many no-shows, but it is highly trustable when it does. This means that the clinic's resources will continue to be wasted but, at least, there will be no overbooking since the system will not schedule replacement patients in slots whose original patient will not fail the appointment. This may be an advantage since it decreases the waiting lists and, consequently, does not decrease patient satisfaction.
- Low precision and high recall: most of the no-shows are detected but the model also classifies some shows as no-shows. This means that the clinic's resources will not be wasted with last-minute vacancy slots but, the system will accidentally overbook appointments since it will try to find replacements to slots in which the no-show was incorrectly detected. This may lead to long waiting lists and, consequently, to decreased patient satisfaction.
- Precision ‘ Recall: In this case, since both have a similar formula, saying that precision is equal to recall is the same as saying that the number of false positive (FP) is equal to the number of false negatives (FN). In other words, the number of no-shows that were incorrectly classified as show (FP) is equal to the number of shows that were incorrectly classified no-shows (FN).

When choosing the threshold value, it is recommended to use precision-recall curves, which uses different probability thresholds to summarise the trade-off between the precision and recall (Saito and Rehmsmeier, 2015). Figure 4 presents the precision-recall curve of the three algorithms over the MDClínica data.

Figure 4 Precision-recall curve (see online version for colours)



In this research, the thresholds were chosen as a way of getting similar precision and recall, which has resulted in a threshold of 0.5 for Brazil data and a threshold of 0.3 for MD Clínica data. As mentioned above, the value of this threshold must be adapted, in the future, according to each clinic approach.

4.3 Comparative analysis of classification algorithms

There is a wide range of classification algorithms and each particular problem must find their most suitable algorithm since the effectiveness of the solution will depend on their performance. For this reason, it is important to test and consider different options and as such, four different algorithms were tested in this research, namely logistic regression, k-NN, random forests and gradient boosting. As mentioned, before, the evaluation method was a ten-fold cross-validation. The mean of the results of each algorithm over the ten iterations is presented in Table 1 and Table 2, according to the considered dataset.

As shown in the lower part of both tables, the results obtained with MD Clínica data proved to be consistent with the results from Brazil data. From these results, despite the slight difference, it is clear that gradient boosting outperforms the remaining algorithms in each of the considered metrics. In general, the models achieved good accuracy results but the remaining metrics showed lower values, which might seem an indicator of a bad performance but it is important to consider that the human behaviour is extremely complex, which makes it hard to predict. Also, this learning model will be running as a part of a no-show algorithm that supports other strategies aimed at reducing no-shows, such as the reminders approach. Nevertheless, the algorithm that is preferable to implement in MedClick system is the gradient boosting whose recall results showed that around 50% of no-shows will be predicted, which leads to an increase in the efficiency of the clinic's resources.

Figure 5 Box plot interpretation (see online version for colours)

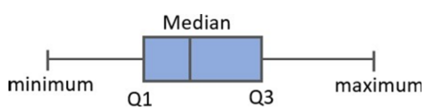
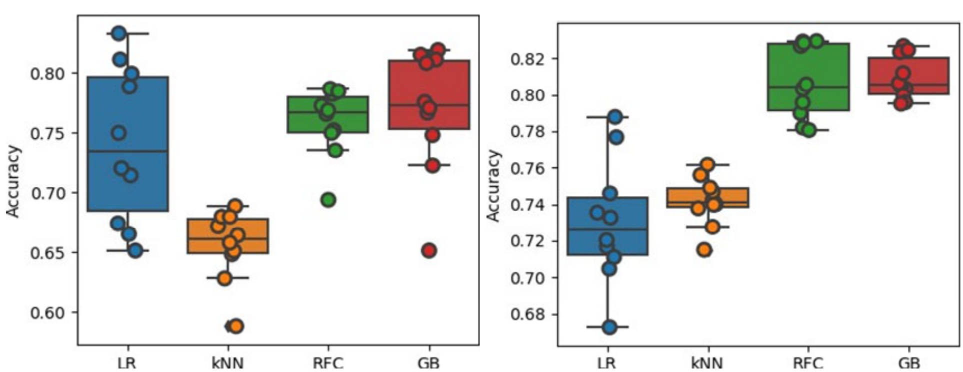


Figure 6 Accuracy results over ten iterations of cross-validation, (a) MD Clínica data (b) Brazil data (see online version for colours)



(a) (b)

Standard deviations are also computed, and their low values show that there is low variance among the different folds of cross-validation, which means that the algorithms would perform similarly with different data sets of the same clinic. In addition, several box plots were provided to illustrate the variability and dispersion of the results of each evaluation metric through the ten iterations of cross-validation. As shown on Figure 5, a box plot is a standardised way of displaying the distribution of results based on five values (Massart et al., 2005): minimum, first quartile (Q1), median, third quartile (Q3), and maximum: The Q1 value, also known as 25h percentile, is the middle result value between the smallest value and the median of the results, which, in turn, corresponds to the middle value of the results. Q3, also known as 75h percentile, is the middle value between the median and the highest result.

Figure 7 Precision results over ten iterations of cross-validation, (a) MD Clínica data (b) Brazil data (see online version for colours)

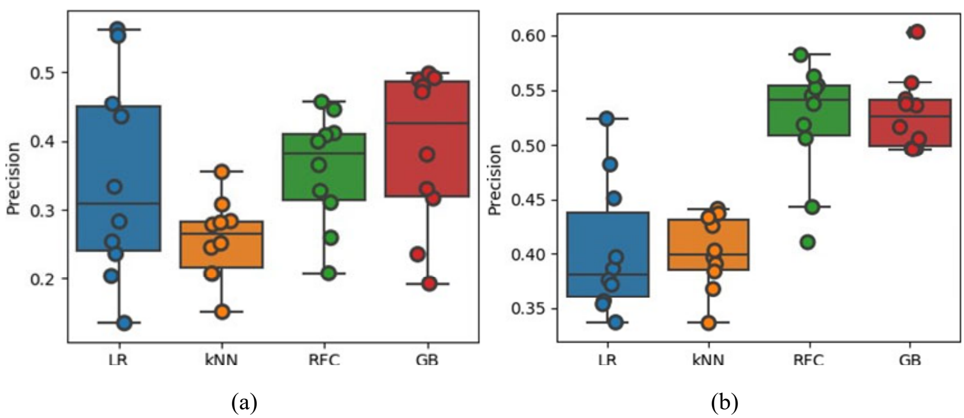


Figure 8 Recall results over ten iterations of cross-validation, (a) MD Clínica (b) Brazil data (see online version for colours)

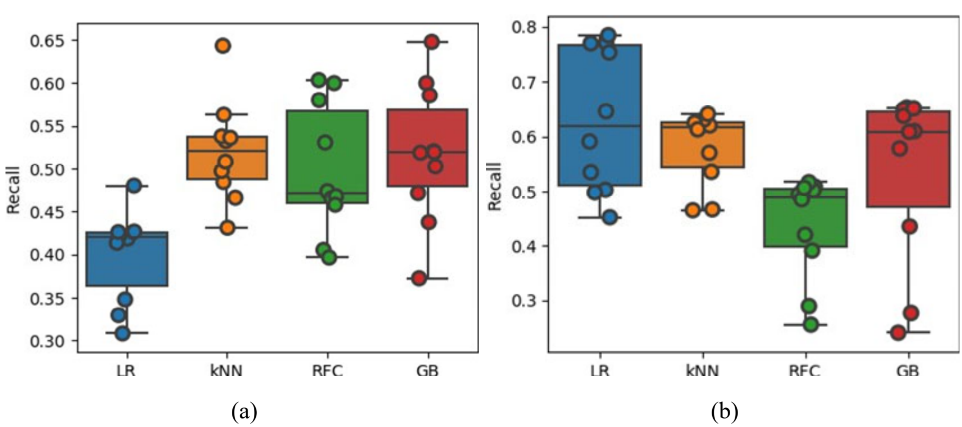
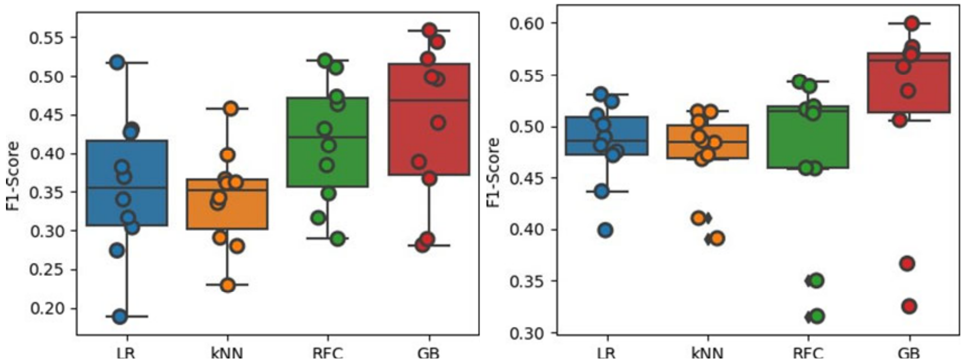


Figure 9 F1-score results over ten iterations of cross-validation, (a) MD Clínica (b) Brazil data (see online version for colours)



(a) (b)

Despite the results from this research, it is important to repeat these tests with the new data that will be provided by the Portuguese clinics since the choice of the algorithm depends on the given data.

4.4 Proposed solution vs. previous solution

This research was applied in the context of MedClick, in which a no-show approach had already been implemented (Sousa and Vasconcelos, 2020), which consists of using logistic regression, as a population-based method, and Bayesian inference, as an individual method. In this research, four classification algorithms were compared and the one that showed the best performance for the considered datasets was gradient boosting, an ensemble of decision trees.

Although both solutions had considered data from the same clinic (MD Clínica) during the evaluation processes, the provided data was different, not only in terms of quantity but also in terms of features. Also, the tests that were performed for the previous solution did not considered some important aspects, such as the fact that the data is unbalanced, which, despite the seemingly satisfactory results, lead to unreliable conclusions. All these factors make the comparison of both solutions a complex process, making it impossible to perform an objective analysis of the impact of new implemented strategies. However, performing the same tests for both solutions provides a better understanding of which solution can achieve the best performance according to the respective conditions. As such, some tests from the previous solution were repeated using the current available data over the gradient boosting algorithm, results of which as described next.

In this test, the data was divided into different portions of test and training sets, ranging from 10% of test data to 100%. In the latter case, all the data was simultaneously used to train and test the model. Table 3 presents the model accuracy for each split of the previous solution, which, despite the seemingly satisfactory results, lead to unreliable conclusions.

From 10% to 90% of test data, the previous model reaches around 70% of accuracy which can be easily achieved by assigning all the data to the majority class (show). As previously discussed, to avoid these unreliable results, other metrics should be measured, namely precision, recall and f1score. In order to compare both solutions, those tests were repeated with the gradient boosting model, which results are presented in Table 4. In order to ensure the reliability of the results, the accuracy was measured along with other metrics.

Table 3 Results from previous solution

Test data	*Accuracy*
10%	0.73
20%	0.72
30%	0.7
40%	0.7
50%	0.68
60%	0.67
70%	0.67
80%	0.68
90%	0.67
100%	0.78

Table 4 Results from proposed solution

Test data	*Accuracy*	*Precision*	*Recall*	*F1-score*
10%	0.79	0.43	0.50	0.46
20%	0.79	0.40	0.55	0.47
30%	0.77	0.40	0.46	0.43
40%	0.77	0.42	0.46	0.44
50%	0.75	0.38	0.47	0.42
60%	0.79	0.41	0.43	0.42
70%	0.74	0.34	0.55	0.42
80%	0.76	0.37	0.54	0.43
90%	0.73	0.32	0.50	0.39
100%	0.87	0.85	0.89	0.87

At first glance, comparing the accuracies from Table 3 and Table 4, one may assume that the model from this research is better than the previous one. However, in addition to the lack of evaluation metrics from the previous solution, the features that were considered in each solution were not the same, so the performance of both models should not be compared as a way of concluding which one is the best. In the previous solution, only four features were considered while the model from this research have considered ten features, which may justify the difference between both performances.

In both tables, it is possible to notice that the performance lowers slightly as the training data diminishes in size. Also, as expected, with 100% of data being simultaneously used to training and testing the model, the global performance increases.

5 Conclusions

This research is focused on no-shows in the healthcare sector and seeks to gather the necessary information to implement a solution capable of reducing no-shows and, consequently, increase the efficient use of clinic resources.

The proposed solution provides a contribution by improving existing Medclick classification algorithm. The previous solution was based on a hybrid approach (using both logistic regression for population based features and Bayesian inference for individual features). This paper performs a comparative analysis between four classification algorithms in order to choose the most suitable for the no-shows' problem. gradient boosting is the one with the best performance.

The research also uses new relevant features when computing a no-show. In the previous solution, only two features were considered relevant (patient's age and the day of the appointment). In order to provide more information to the model, this solution extracts the following features: patient's age, patient's gender, waiting time, day of the appointment, scheduling day, number of previous appointments, number of previous no-shows, number of patient diseases, scholarship status and finally, patient's handicaps.

The approach proposed in this paper uses the algorithm to detect no-shows. The previous solution was only using the classification algorithm to sort the candidates list, from the least likely to miss the appointment to the one with the greatest probability of missing it. This solution, in addition, leverages the algorithm to predict no-shows.

An additional contribution of this research is the implementation of strategies to reduce no-shows (including sending notifications before each appointment, in which the patient must confirm his or her presence). Finally, the solution also proposes an improvement in the method for selecting candidates for replacements. (By using a list that includes all patients who have already scheduled an appointment at a later date in the same healthcare centre and with the same health professional).

A relevant conclusion of this research is that the effectiveness of the solution depends on the performance of the chosen algorithm. In this paper, four different algorithms were tested, namely, logistic regression, k-NN, random forests and gradient boosting. The latter is the one presenting the best performance.

As future work, we suggest that the classification model is re-trained to prevent it from becoming unstable. Finally, throughout this research, several parameters were considered, to which default values were assigned, such as, the day of sending reminders, the day of predicting no-shows and the threshold for no-show probabilities. All these variables must be further explored in the future in order to find the most suitable values (which may be different among healthcare providers).

References

Alaeddini, A., Yang, K., Reeves, P. and Reddy, C. (2015) 'A hybrid prediction model for no-shows and cancellations of outpatient appointments', *IIE Transactions on Healthcare Systems Engineering*, Vol. 5, No. 1, pp.14–32.

Altman, N. (1992) 'An introduction to kernel and nearest neighbor nonparametric regression', *The American Statistician*, Vol. 46, No. 3, pp.175–185.

Antonio, N., Almeida, A. and Nunes, L. (2017) 'Predicting hotel booking cancellation to decrease uncertainty and increase revenue', *Tourism and Management Studies*, Vol. 13, No. 2, pp.25–39.

Bennett, K. and Baxley, E. (2009) 'The effect of a carve-out advanced access scheduling system on no-show rates', in *Family Medicine*, Vol. 41, No. 1, pp.51–46.

Bowyer, K., Chawla, N., Hall, L. and Kegelmeyer, W. (2011) 'SMOTE: synthetic minority over-sampling technique', CoRR, abs/1106.1813, 2011.

Cameron, S., Sadler, L. and Lawson, B. (2010) 'Adoption of open-access scheduling in an academic family practice', *Canadian Family Physician*, Vol. 56, No. 9, pp.906–911.

Cashman, S., Savageau, J., Lemay, C. and Ferguson, W. (2004) 'Patient health status and appointment keeping in an urban community health centre', *Journal of Health Care for the Poor and Underserved*, Vol. 15, No. 1, pp.474–88.

Compton, M., Rudisch, B., Craw, J., Thompson, T. and Owens, D. (2006) 'Predictors of missed first appointments at community mental health centers after psychiatric hospitalization', *Psychiatric Services*, Vol. 57, No. 4, pp.531–537, Washington, DC.

Daggy, J., Lawley, M., Willis, D., Thayer, D., Suelzer, C., DeLaurentis, P., Turkcan, A., Chakraborty, S. and Sands, L. (2010) 'Using no-show modeling to improve clinic performance', *Health Informatics Journal*, Vol. 16, No. 4, pp.246–259.

Gany, F., Ramirez, J., Chen, S. and Leng, J. (2011) 'Targeting social and economic correlates of cancer treatment appointment keeping among immigrant Chinese patients', *Journal of Urban Health*, Vol. 88, No. 1, pp.98–103.

George, A. and Rubin, G. (2003) 'Non-attendance in general practice: a systematic review and its implications for access to primary health care', *Family Practice*, Vol. 20, No. 2, pp.178–184.

Guse, C., Richardson, L., Carle, M. and Schmidt, K. (2003) 'The effect of exit-interview patient education on no-show rates at a family practice residency clinic', *The Journal of the American Board of Family Practice*, Vol. 16, No. 5, pp.399–404.

Ho, T. (1995) 'Random decision forests', in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, IEEE, Vol. 1, pp.278–282.

Kohavi, R. (1995) 'A study of cross-validation and bootstrap for accuracy estimation and model selection', *Proceedings of the 14th International Joint Conference on Artificial intelligence – Volume 2 (IJCAI'95)*. pp.1137–1143, Morgan Kaufmann.

Lawrence, R., Hong, S. and Cherrier, J. (2003) 'Passenger-based predictive modeling of airline no-show rates', in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '03*, pp.397–406, New York, NY, USA.

Leong, K., Chen, W., Leong, K., Mastura, I., Mimi, O., Sheikh, M., Zailinawati, A., Jenn Ng, C., Phua, K. and Teng, C. (2006) 'The use of text messaging to improve attendance in primary care: a randomized controlled trial', *Family Practice*, Vol. 23, No. 6, pp.699–705.

Massart, D., Smeyers-Verbeke, J., Capron, X. and Schlesier, K. (2005) 'Visual presentation of data by means of box plots', *LC-GC Europe*, Vol. 18, No. 4, pp.215–218.

Neal, R., Hussain-Gambles, M., Allgar, V., Lawlor, D. and Dempsey, O. (2005) 'Reasons for and consequences of missed appointments in general practice in the UK: questionnaire survey and prospective review of medical records', *BMC Family Practice*, Vol. 6, No. 1, p.47.

Powers, D (2011) 'Evaluation: from precision, recall and f-factor to roc, informedness, markedness correlation', *Mach. Learn. Technol.*, Vol. 2, No. 1, pp.37–63.

Saito, T. and Rehmsmeier, M. (2015) 'The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets', *PLOS ONE*, Vol. 10, No. 3, pp.1–21.

Sousa, D. and Vasconcelos, A. (2020) 'Last minute medical appointments no-show management'. *International Journal of Healthcare Information Systems and Informatics*, Vol. 15, No. 4, ISSN: 1555-3396, 2020 (to appear), DOI: 10.4018/IJHISI.

Trevor, H., Robert, T. and Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer-Verlag New York, ISBN: 978-0-387-84858-7.