



Using Different Models of Machine Learning to Predict Attendance at Medical Appointments

Luiz Henrique Salazar ¹, Anita Fernandes ^{1,2*}, Rudimar Dazzi ², Nuno Garcia ^{3,4}, Valderi R. Q. Leithardt ^{2,4,5}

¹University of the Itajai Valley, Specialization Course in Big Data, BRAZIL

²Master in Applied Computing, University of the Itajai Valley, Santa Catarina, BRAZIL

³Telecommunications Institute, IT Branch, Covilha, PORTUGAL

⁴Department of Informatics, University of Beira do Interior, Covilha, PORTUGAL

⁵COPELABS, Lusophone University of Humanities and Technologies, Lisboa, PORTUGAL

*Corresponding Author: anita.fernandes@univali.br

Citation: Salazar, L. H., Fernandes, A., Dazzi, R., Garcia, N. and Leithardt, V. R. Q. (2020). Using Different Models of Machine Learning to Predict Attendance at Medical Appointments. *Journal of Information Systems Engineering and Management*, 5(4), em0122. <https://doi.org/10.29333/jisem/8430>

ARTICLE INFO

Published: 30 Jul. 2020

ABSTRACT

Outpatient absenteeism is a recurring problem worldwide and in Brazil, it is a chronic problem. The number of appointments and exams scheduled and not performed, due to the non-attendance of patients, reaches significantly high rates and can be seen in all regions in the country and in different types of care and medical specialties. This practice generates waste of resources, disorganizes the offer of services, and limits the guarantee of care at different levels of assistance. In addition, it causes a series of dissatisfactions from users of the health system who really need and have not yet been able to access consultations and exams. This imbalance causes the misuse of the offer, an increase in the queue and waiting time, as well as a financial loss since it is paid for by a professional who is idle due to the absence of patients. It is necessary to understand the profile of these missing patients and to try to discover the reasons that lead this person to be absent, in order to predict a future absence for consultation. Thus, this work presents the study of different machine learning models in order to help predict whether or not the patient will attend the scheduled appointment. As a result, an end-to-end machine learning process was developed, considering exploratory data analysis, pre-processing, creation of machine learning models, analysis of results, and deployment of the most appropriate model in a web application. As a result, it was found that the Decision Tree algorithm represents an interesting choice as a final model for use in future observations.

Keywords: absenteeism, prediction, machine learning

INTRODUCTION

A recurring model in the area of public health is the high rate of patients who do not attend scheduled medical examinations and consultations. It is not difficult to find in Brazil today a user dissatisfied with the care provided by public health. Most of the time, the complaint concerns the delay in scheduling exams or procedures. Absenteeism compromises the ability to schedule appointments at a health unit, hinders the access of other users to the health system, leads to an increase in the waiting time for a particular consultation and generates financial expenses, given that the service is paid by the municipality even when the user does not attend a scheduled appointment. More than impairing the day of care, the unjustified absences of patients without any prior communication at appointments or scheduled exams compromise the ability to schedule appointments at a health unit, hinder the access of other users to the health system, leads to an increase in the waiting period for a specific consultation and generates financial expenses, given that the service is paid by the municipality even when the user does not attend an scheduled appointment.

More than damaging the day of care, the unjustified absences of patients without any prior communication in appointments or scheduled exams compromises the efficiency of the service. That is, the damage ends up affecting all users of the health system. In addition to not attending diagnostic support tests, absenteeism reaches a global prevalence of around 25% in specialized outpatient clinics (Silveira et al., 2018). The investigation of the main reasons for absenteeism in medical consultations is a fundamental factor for continued health care, for improving access to the public health system, for controlling waiting lines and improving the population's quality of life (Salazar et al., 2020).

Absenteeism from previously scheduled appointments compromises the effectiveness of medical care and creates a series of problems in public health systems around the world. Studies have identified long waiting times for health care as one of the main

challenges facing the system (AlRowaili, Ahmed and Areabi, 2016; Lenzi, Ben and Stein, 2019; Reid et al., 2015). A public opinion survey carried out in Canada (AlRowaili, Ahmed and Areabi, 2016), showed that 75% of respondents identified reducing the waiting list as a high priority action. Some studies recommend solving the problem by increasing resources as described by Santibáñez and Chow (2009) and Hoe (2007), while other studies show that the modernization of the process would lead to improvements, as described by Recht et al. (2013) and Lu, Li and Gisler (2011). Some authors suggest the need for organizational changes that facilitate adherence to consultations in the health service (Mohammadi et al., 2018).

The patient's failure to attend the medical appointment as scheduled is closely linked throughout the waiting time. Long waiting times lead to high non-attendance rates, which in turn leads to increased waiting times. The patient's failure to attend consultations has two direct effects. The first, obviously, involves the patients themselves, who postpone the chance of being treated by a doctor. The second affects health services, because the time lost due to the lack of care for one patient implies that another patient misses the opportunity to be seen by the doctor. This is called the opportunity cost (Elvira et al., 2018).

A better understanding of the topic makes it possible to create hypotheses to explain the reasons for its occurrence, thus contributing to the management and planning of services (Mohammadi et al., 2018; Salazar et al., 2020). The underutilization of medical consultations is a paradox in the face of constant complaints of excessive demand on the part of professionals and lack of supply from the perspective of users. The reorganization of agendas is a subject that is currently being discussed and aims to establish a balance between supply and demand, reduce waiting times, end the reserve of vacancies and, consequently, reduce absenteeism rates (Mohammadi et al., 2018; Salazar et al., 2020).

Understanding the reasons why health absenteeism is so high can provide social and economic benefits. In Brazil, for example, more precisely in the state of Santa Catarina, the unexcused absence from scheduled medical appointments caused a financial impact of at least 3.8 million Euros in 2016 for 20 units under the responsibility of the state government and municipalities with more than 100 thousand inhabitants that account for 45% of the state's population (Salazar et al., 2020).

Much research has been done on outpatient absenteeism by health researchers, but little has been done in relation to understanding the reasons that lead the patient to miss the appointment. Examples of these studies in Brazil are presented by Silva (2013), Cavalcanti (2013), Canelada et al. (2014), Monken and Moreno (2015), Bittar et al. (2016), and Silva et al. (2017).

Silva (2013), carried out an evaluation on how to reduce absenteeism in appointments made at a diabetes reference service. In his study, Silva pointed out the following actions as a solution to the problem: reduction in scheduling appointments from five to four appointments/day, in a single shift (morning or afternoon); distribution of a newsletter with the operating rules in the diabetes reference service; sending a message through the Short Message Service (SMS), one week before the scheduled date, remembering the date of the appointment and requesting that, if they are unable to attend, call ahead to cancel; optimization of the appointment scheduling system, avoiding delays in booking by telephone; monthly opening of medical diaries, avoiding accumulation of patients for appointment (virtual queue); "Informational pills" that are projected, through slides, on the television located in the waiting room of the outpatient clinic, during the period when patients are waiting for the call for consultations.

Cavalcanti et al. (2013) presented a study on the absenteeism of specialized consultations in the public health system: the relationship between causes and the work process of family health teams in João Pessoa, Paraíba (Brazil). As a solution to reduce absenteeism, they focused on home visits by community health workers as a mechanism for coping with absenteeism, through community awareness.

Canelada et al. (2014), carried out a study in a Medical Specialties Clinic, in the interior of the state of São Paulo and concluded that they should make a manual optimization of the doctors' agenda, calling each of the patients.

Monken and Moreno (2015), presented the use of control alerts as a tool for the loyalty of pediatric patients in a public clinic. The use of control alerts (red flags) in the scheduling system triggered measures to engage the mother in attending the consultations. The work was developed in the specialty of Neonatology of Newborn High Risk, in which the rate of absenteeism in April was 38%, and after the implementation of control alerts in May, it reduced to 11%.

Bittar et al. (2016) presented a study on absenteeism in outpatient care for specialties in the state of São Paulo. As conclusions, they point out that the causes of absenteeism in outpatient care must be studied individually in each specialty, given that there may be specific factors such as disease severity, availability of professionals and equipment.

Silva et al. (2017) presented a study on outpatient absenteeism in the postoperative period of orthopedic patients at a teaching hospital in São Paulo, however the result was inconclusive since it was not possible to establish reasons for absenteeism.

As you can see, the studies done focus on alerting patients through calls or text messages, but there is no study focused on trying to discover the real reasons for their absence. The main justification for this is the amount of data to be analyzed, therefore, the application of data science and data mining techniques for this problem becomes an interesting solution.

Within this context, it is necessary to analyze the topic in a scientific way so that the health area can react appropriately (Dogruyol and Sekeroglu, 2020). For this purpose, machine learning algorithms can serve as efficient tools to assist in the understanding of the patient's behavior in relation to his presence in the medical consultation. Some studies show the use of machine learning algorithms for predicting absenteeism in other contexts, such as the absence of workers in their jobs (Nelson, 2019; Priyanka and Nayak, 2020; Wahid, 2019).

Predictive models can be used in the health area to estimate the risk of a certain outcome occurring, given a set of socioeconomic, demographic characteristics related to life habits and health conditions. Its results, when combined with public health measures applied at the population level, can have positive implications for reducing costs and the effectiveness of interventions, such as treatments and preventive actions (Santos et al., 2019).

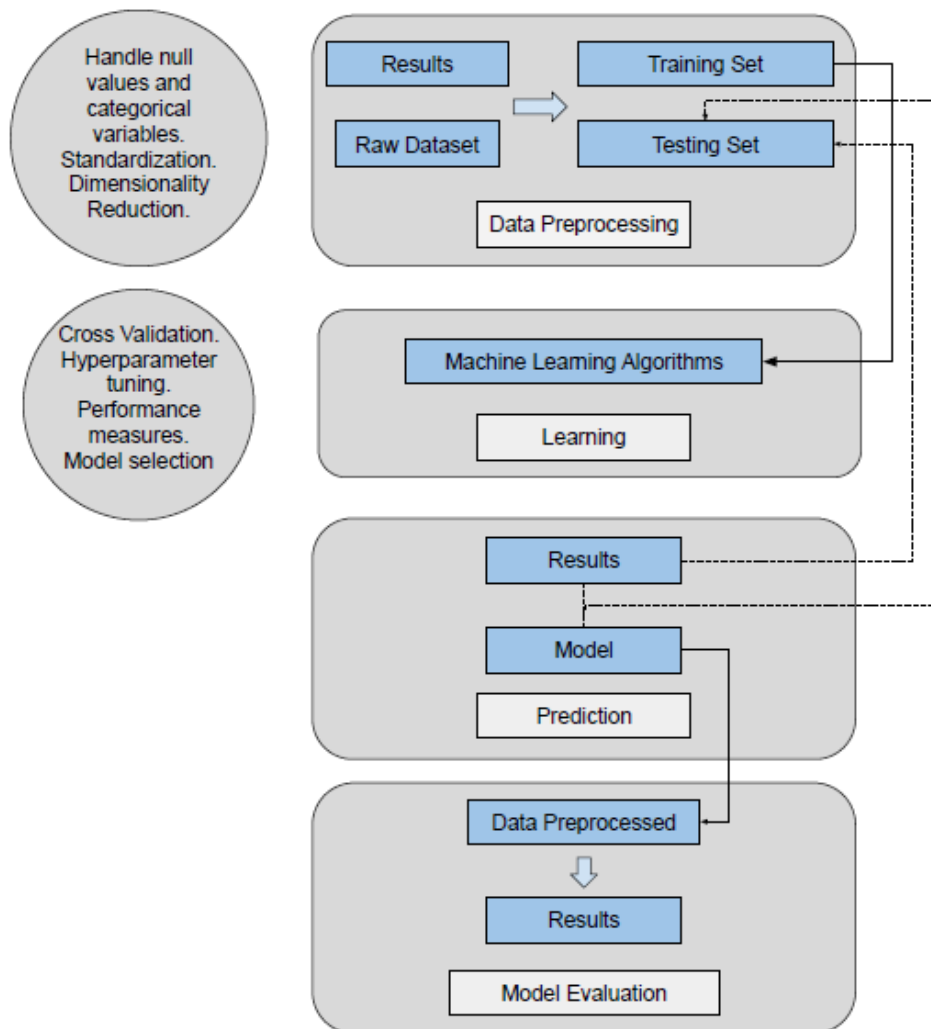


Figure 1. Roadmap for applying Machine Learning algorithms in predictive analysis (Raschka and Mirjalili, 2017)

Therefore, this work also intends to identify patterns in patients' behavior in order to predict whether or not to attend scheduled appointments, in order to provide subsidies to managers regarding the management of consultations. The data used were extracted from an open database available on the Kaggle platform and refer to medical appointments scheduled in public hospitals in the city of Vitoria in the state of Espirito Santo - Brazil. The data refer to 2015 and 2016. The target variable chosen refers to the patient's attendance or not at a medical appointment, and the predictors are represented by 14 variables related to the patient's demographic, socioeconomic and health profile. The following steps were carried out throughout the study: dividing the data into two sets (training - 70% and tests - 30%), pre-processing, learning and evaluation of the prediction models created.

In the learning stage, three algorithms were used in order to adjust the models: Logistic Regression, Random Forest and Decision Tree. For the selection of the best models, the hyper parameters of the algorithms were optimized by 10-fold cross-validation. For each algorithm, the best model was evaluated in test data using the AUC (Area Under the ROC curve) of the ROC (Receiver Operating Characteristic) curve, and other measures relevant to the model's performance. All models had AUC ROC greater than or equal to 0.6.

For a better understanding and understanding, the methodology used for the development of this study, the construction of the predictive model, the results obtained and the conclusions are presented.

METHODOLOGY

Predictive analysis consists of applying algorithms to understand the structure of existing data and generate prediction rules (Covington, 2019). These algorithms can be used in an unsupervised or supervised scenario. In unsupervised problems, only predictors (covariates) are available in the data set and in supervised cases, in addition to the predictors, there is also a variable of interest responsible for guiding the analysis (Abbott, 2014).

In the present work, Machine Learning algorithms are used in a supervised scenario in which the main objective is to adjust a model that relates the response to the predictors, in order to predict new events in future observations as described in Covington (2019) and Abbott (2014). The adjustment of predictive models can be represented by the steps of dividing the data set in training and testing, pre-processing, learning and model evaluation (**Figure 1**).

Table 1. Main Features of the Used Algorithms

Algorithm	Main Features
Logistic Regression	The response variable is categorical and the estimate is made using the maximum likelihood method. They are commonly used in classification problems, in situations where the dependent variable is of a dichotomous or binary nature, where the independent variables can be categorical or not (Montgomery, Peck and Vining, 2012).
Random Forest	It aims to combine predictions from a set of complex classifiers (decision trees with many divisions), applied to the bootstrap of the training set. The difference is in the random selection of predictors to be used, in order to reduce the correlation between them that will be aggregated to produce the final prediction.
Decision Tree Classifier	They are commonly used in classification tasks but it is also possible to find in the literature their use in regression tasks. It stands out for its ability to reduce the complexity of the decision process in a smaller collection of simple decisions, providing a solution that is easy to understand. The amount of calculations required to perform a classification is generally much less than the amount spent on neural networks, so its execution time is more efficient (Goodfellow, Bengio and Courville, 2016)

The sample is divided into training and test data to verify whether a model performs well, not only in data used for its adjustment (training), but also the ability to generalize to new observations (tests). The most used divisions are 60:40; 70:30; or 80:20, depending on the size of the data set (Fenner, 2019). In general, the greater the number of observations, the greater the proportion of the initial set used for training. In the present study, 70% of the data were used for training the algorithms (74,021 records) and 30% for testing the predictive performance of the adjusted models (36,459 records).

According to Raschka and Mirjalili (2017) and (Fenner, 2019), a classification problem is defined when there is a sample with independent observations $(X_1, Y_1), \dots, (X_n, Y_n) \sim (X, Y)$ and the objective is to build a function $f(x)$ that can be used to predict new observations $(X_{n+1}, Y_{n+1}), \dots, (X_{n+m}, Y_{n+m})$, that is, it is intended that:

$$f(X_{n+1}) \approx Y_{n+1}, \dots, f(X_{n+m}) \approx Y_{n+m}$$

In this way, Machine Learning algorithms are used in order to estimate f with the smallest possible error. Among the various algorithms available, some are not very flexible (less complex), but interpretable - as is the case with Decision Tree and Logistic Regression. On the other hand, there are more flexible (more complex) approaches, which makes it more difficult to understand how each predictor is individually associated with the response of interest - as in the case of Artificial Neural Networks, according to Fenner (2019), Covington (2019), Raschka and Mirjalili (2017) and Abbott (2014).

It is noted that less complex models can perform better than those that are more flexible, mainly because they are less subject to overfitting and, consequently, resulted in more accurate predictions in new observations. However, there is not a single algorithm capable of performing well in all applications, as described by Covington (2019) and Raschka and Mirjalili (2017). It is extremely important to compare some algorithms with different characteristics in order to select the one that results in a model with satisfactory predictive performance for the problem in question (Shalev-Shwartz and Ben-David, 2014).

In general, the algorithms can be grouped into the following categories: linear, non-linear and based on decision trees (James et al., 2017). In the present study, three algorithms were selected for comparison: decision tree, logistic regression and random forest (based on decision trees). The algorithms were selected based on the existing data sample, the flexibility and complexity of the algorithms and the previous knowledge of the classification algorithms addressed. **Table 1** describes the main characteristics of these algorithms.

CONSTRUCTION OF PREDICTIVE MODEL

The dataset used in this study consists of 110,527 appointments for medical appointments, each record being originally made up of 14 associated characteristics. The data were extracted from the "Medical Appointment No Shows" dataset of the Kaggle¹ website repository and refer to public hospitals in the city of Vitoria, in the state of Espírito Santo - Brazil. The data refer to schedules for the years 2015 and 2016. The characteristics are: PatientID; AppointmentID; Gender; ScheduledDay; AppointmentDay; Age; Neighbourhood; Government Grant (beneficiary of the Federal Government Program called Bolsa Família (1 or 0); Hipertension (1 or 0); Diabetes (1 or 0); Alcoholism (1 or 0); Handcap (patient with disability) (1 or 0); SMS_received (1 or 0); No-show (1 or 0).

In this initial stage, an initial analysis was also carried out in order to obtain a greater understanding of the existing data. Within the knowledge discovery process, one of the phases is the exploratory data analysis - Exploratory Data Analysis (EDA), whose main objective is to understand what data exist, what possible trends, and therefore, which statistical tests will be appropriate to use (Cox, 2017). The EDA combines advanced data visualization techniques and statistical models, which use well-designed graphs, allowing complex inferences and analysis to be made, even when the graphs are based on basic statistics (McCandless, 2014).

Some initial questions were defined to guide the analysis: what is the proportion of men and women in the sample?, which neighborhoods have the highest rate of absenteeism in consultations?, which month do patients miss the most visits?, if there is an influence on patient attendance in the consultation if he has previously received an SMS message, among others. Among the most relevant facts, it is noted that women attend more scheduled appointments than men and that the fact that the patient has a disability, receives help from the government's social program or suffers from alcoholism or hypertension has no direct relationship with cause of the patient not to go to the previously scheduled appointment. It is also observed that among the months observed, May and June are the months with the highest rate of dropout from consultations and that sending SMS to the

¹ <https://www.kaggle.com/datasets>

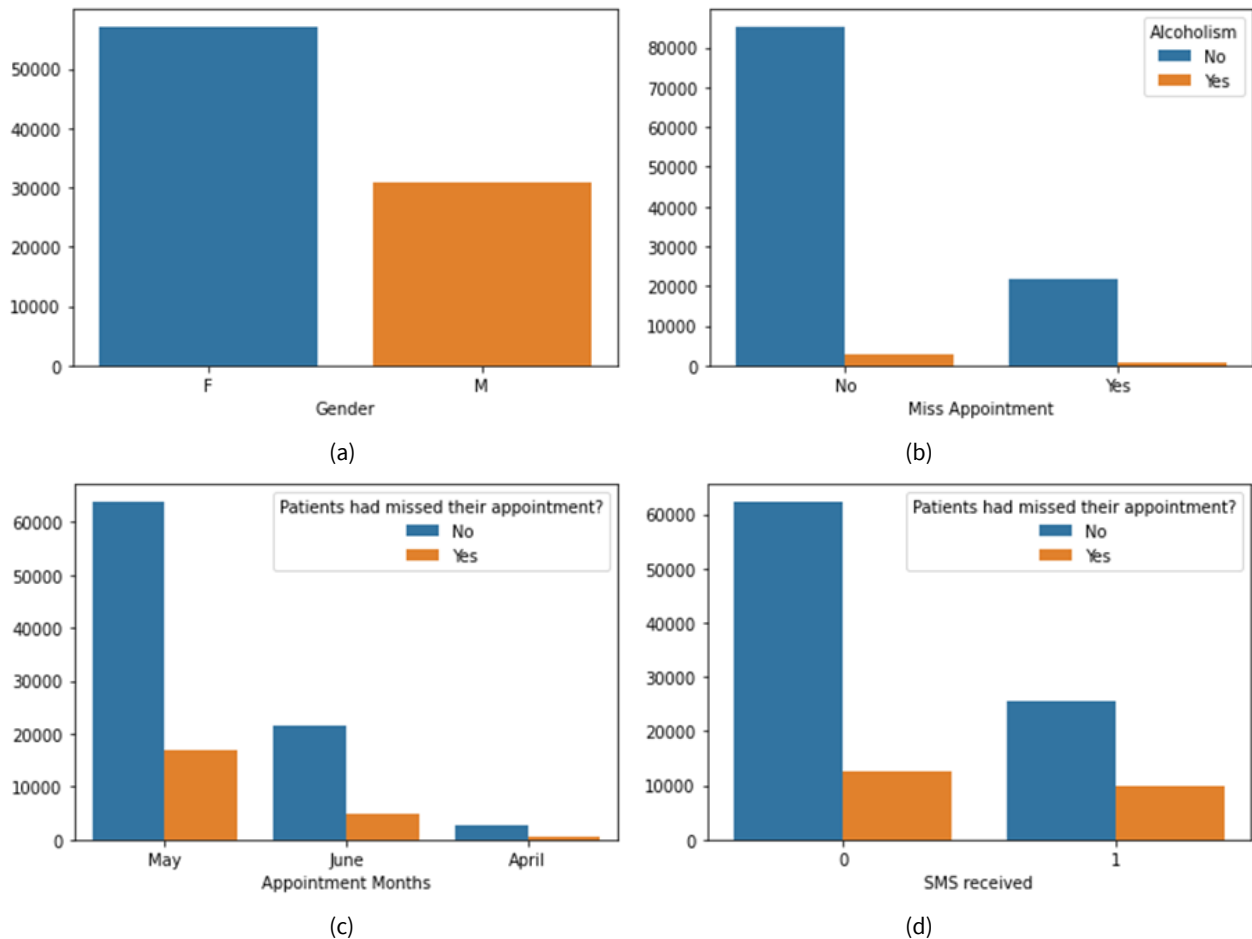


Figure 2. Analysis Results

patient before the appointment does not increase the likelihood of attending the appointment. **Figure 2** presents the results obtained for these questions, considering the gender that most frequents the scheduled appointments (a), the relationship between alcoholism and attendance at the consultation (b), months with the highest rate of non-attendance at the consultations (c) and the reception ratio of SMS previously with the attendance in the consultation schedule (d).

Pre-processing

Pre-processing is guided by the algorithms that will be used to adjust the predictive models described in Covington (2019) and Abbott (2018). In general, the following activities are performed during this stage: (i) transformation of quantitative variables (via standardization or normalization); (ii) reduction of the dimensionality of the data set (exclusion of highly correlated predictors or use of principal component analysis); (iii) exclusion of variables / observations with missing data or use of imputation techniques (mean, median or most frequent value in the case of numerical predictors); and (iv) organization of qualitative variables (decomposition of categorical variables into a set of indicator variables that will be used as predictors), according to Covington (2019) and Abbott (2018).

It is worth mentioning that the parameters estimated by pre-processing procedures, such as those resulting from the standardization of variables and the calculation of the value for the imputation of missing data, are obtained in training data and, later, applied to the test data (as well as new observations) before predictions are made. Such a procedure is adopted so that the performance of the test data is reliable to the real performance of the predictive model in future data (Fenner, 2019).

In the pre-processing stage, values for the attribute related to the patient's age were adjusted to only values between zero and 95 years ($n = 47$) - considering that it is currently very unlikely that people will live longer than 95 years. As for the variable that informs whether the patient has a disability or not, only values 0 and 1 were considered, eliminating the other values ($n = 199$). In order to add more information to the dataset, two new variables were calculated from the existing attributes. The first numeric variable was created by calculating the difference between the day of the appointment and the day of the appointment, resulting in the number of days the patient waited to be seen since the appointment was made. The other inclusion refers to a variable that indicates the patient's age group (young, adult or elderly).

Hyperparameter Optimization

In Machine Learning, there are two types of parameters to be estimated: the usual parameters of an algorithm, such as the weights of a logistic regression, and the hyperparameters, related to the control of the flexibility of the algorithm, according to Shalev-Shwartz and Ben-David (2014), and James et al. (2017). Hyperparameters are defined before training and are different from

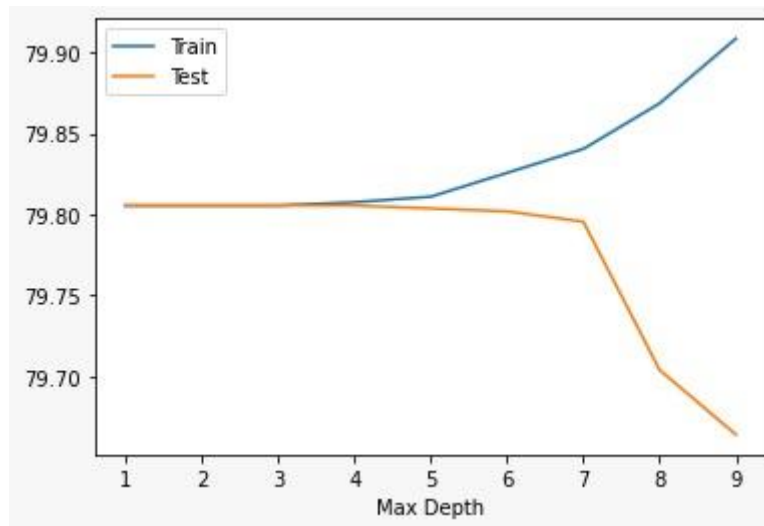


Figure 3. Performance comparison of the decision tree algorithm based on the maximum depth hyperparameter of the tree

the model's internal values, which are changed according to what the model is learning. For example, the maximum depth of a Decision Tree is considered a hyperparameter because it serves as a delimiter of the algorithm during the training process of the model based on Shwartz and Ben-David (2014), and James et al. (2017).

The control of the flexibility of an algorithm is dependent on the balance between bias and variance. The bias is related to the correspondence between the value predicted by a model for a given observation and the actual observed value, while the variance refers to the sensitivity of the predictions to the variability of the training observations, according to Shwartz and Ben-David (2014), and James et al. (2017). The use of Machine Learning algorithms becomes challenging from the moment that moderate flexibility is desired, that is, a flexible model has high variance, while a less flexible model has low variance but can present a high bias.

Thus, two main objectives are pursued when developing predictive models: selecting and evaluating the models. In the first case, for a given algorithm that has hyperparameters, the performance of different models, based on variations in the values for the hyperparameters, is evaluated to select the one that results in the best performance (balance between bias and variance). In the second, after defining the model, we try to estimate its prediction error in new observations (Hastie, Tibshirani and Friedman, 2017).

In the present work, the optimization of hyperparameters occurred through the use of the open source Machine Learning library for the Python language - Scikit learn. Initially, the Decision Tree algorithm was trained through the standard parameters of the library and, subsequently, a sequence of tests was performed in order to extract the best fit for the parameters of the algorithm. The maximum depth hyperparameter of the Decision Tree was tested in order to obtain the value with the best cost benefit, that is, to obtain satisfactory accuracy as well as reasonable computational performance. For that, according to the literature, with a value between 5 to 10 of maximum depth of the tree it is already possible to obtain a good performance in this type of algorithm (Abbott, 2014; Covington, 2019; Raschka and Mirjalili, 2017; Shwartz and Ben-David, 2014).

A concept linked to obtaining a model with good accuracy for the test data is the reduction of the impact of overfitting. Overfitting is observed when an algorithm performs almost perfectly with the training data but does not obtain a satisfactory result for the test data (James et al., 2017). In this way, an iteration was performed with 10 repetitions in order to find the ideal point at which the algorithm performs well for both the training model and the test (James et al., 2017) (Figure 3).

Thus, based on Figure 3, it can be seen that the maximum depth of the tree used was 3 levels. The other algorithms chosen to be used (Logistic Regression and Random Forest), as well as the other hyperparameters of the Decision Tree (*min_samples_leaf* and *min_samples_split*) were maintained with their default values.

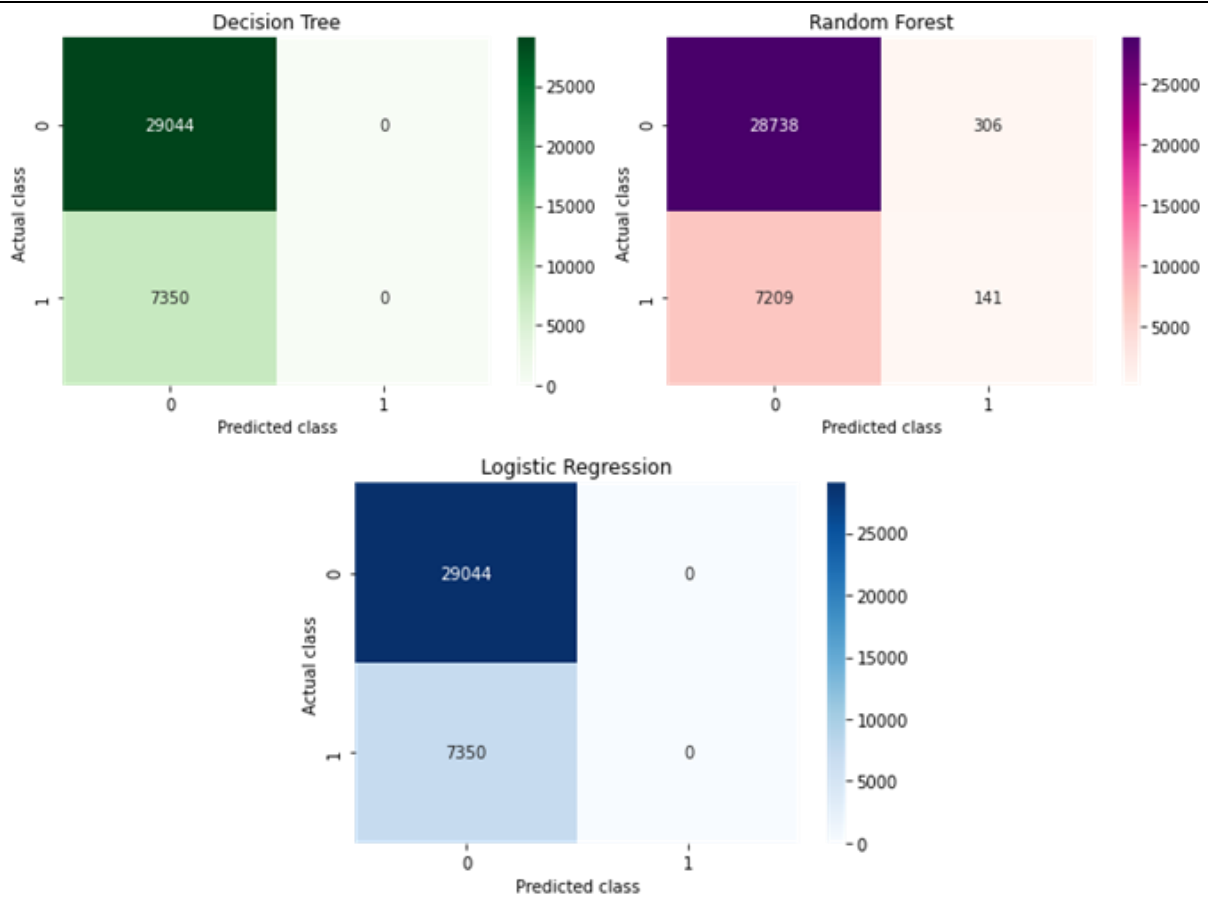
Holdout and Cross Validation

Among the re-sampling techniques, there are two main approaches: holdout and cross-validation. The holdout method, also called percent split, receives as input the information regarding the percentage in which the data will be divided into training and test subsets (James et al., 2017). The cross-validation technique, on the other hand, consists of randomly dividing training data into k parts of equal size, in which $k-1$ will compose the training data for model fitting, and the other part will be reserved for estimating its performance, according to Covington (2019) and Abbott (2014). Using the Scikit-learn library for Python, the application of both techniques can be performed in a simplified way.

In the present work, the application of the holdout method occurred through the call of the *train_test_split* function, in which 30% of the data were separated for testing while the rest was kept for training. Due to the function's internal algorithm using an attribute responsible for defining the randomness in which the data is divided, a fixed value was defined that allows the same training and test set to be obtained in every execution of the algorithm. Another parameter used in the function was that of data stratification, that is, when dividing the data into training and testing, the same proportion of classes as the original dataset is maintained, thus preserving the initial characteristics of the data.

Table 2. Confusion Matrix for Classification Problems (Deng et al., 2016)

Predicted Response	Resposta Observada			a+b	
	Present	Presente	Ausente		c+d
		True Positive (a)	False Positive (b)		
Absent		False Negative (c)	True Negative (d)		
		a+c		b+d	

**Figure 4.** Confusion matrix obtained for decision tree, logistic regression and random forest algorithms

In the **k-fold** cross-validation technique, the process of dividing data between training and testing continues until all parties have participated in both training and model validation, resulting in **k** performance estimates, according to Covington (2019) and Abbott (2014). There is no precise rule for choosing **k**, although dividing the data into 5 or 10 parts is more common. As **k** increases, the size difference between the original training set and the re-sampled subsets becomes smaller, and as this difference decreases, the bias of the cross-validation technique also becomes smaller. On the other hand, the time required to obtain the final result of cross-validation becomes longer (Kuhn, 2018).

Another optimization performed in the cross validation was the use of the **group k-fold** technique, that is, to separate the training and test data with cross validation according to the existing groups. This prevents the separation of the data from including elements of the same group in both the training and test data, thus improving the performance of the estimator to predict future data not yet observed by the model described in Kuhn (2018). In the present study, cross-validation with **k** = 10 was used and the group was defined as the patient identification parameter.

Models Performance Evaluation

Once the **k** value is established, it is necessary to define a measure to estimate the performance of the adjusted models. Such measures are important both in the selection stage and in the evaluation of predictive models, as described by Hastie, Tibshirani and Friedman (2017), and Kuhn (2018). The use of the adjusted model in different scenarios allows the establishment of cut-off points and evaluation of the model in terms of sensitivity and specificity. One way to visualize these and other points of the evaluator model is through the confusion matrix (Deng et al., 2016).

The confusion matrix is represented by the cross tabulation of observed and predicted classes for the test data, as shown in **Table 2**. In this table, **a** and **d** denote cases with correctly predicted responses, and **b** and **c**, represent classification errors. Sensitivity [$a/(a + c)$] is the proportion of True Positives (TP) among all individuals whose response of interest has been observed, and specificity [$d/(b + d)$] refers to the proportion of True Negatives (TN) among those with missing interest response.

After training the Decision Tree model and applying the fit to the extracted test data, the confusion matrix shown in **Figure 4** was obtained. After calculating the specificity and sensitivity of the model, a value of 0 was obtained, 0.9889 and 0.01986,

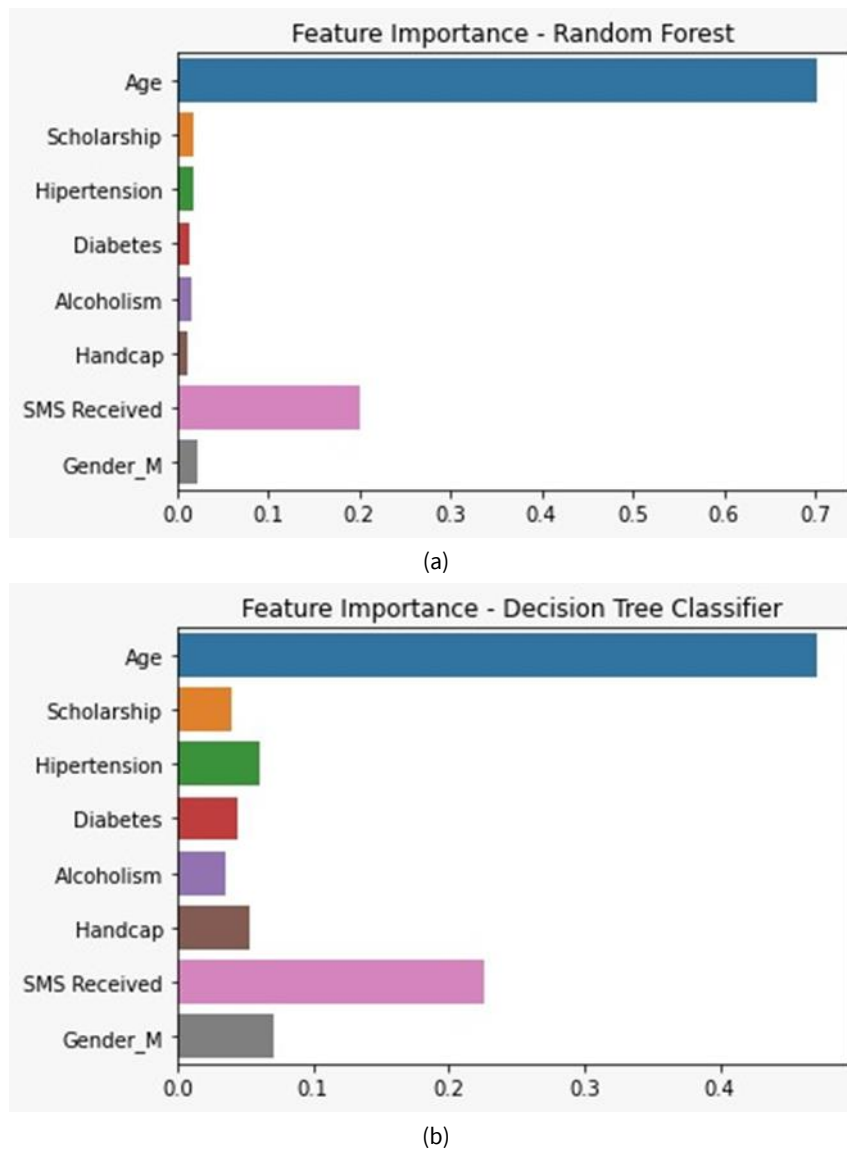


Figure 5. Importance of each attribute in the models generated from the Decision Tree (a) and Random Forest (b) algorithms


respectively. Thus, it is noted that the model chosen has a high degree of assertiveness when it predicts that the patient will attend the medical consultation, however, when it comes to predicting whether the patient will not attend, a high error rate is obtained. In these cases, a pertinent approach would be to perform the screening of the data in which the model predicted that the patient will not attend the consultation and perform the analysis by a specialist.


Another alternative to analyze the model is the ROC curve, which represents an adequate way to assess sensitivity and specificity, so that the overall performance of a classifier can be assessed by the area under the curve (AUC): the higher the AUC (closest of 1), the better the performance of the model based on Covington (2019) and Abbott (2014). AUC ROC can be useful when comparing two models with different predictors, different hyperparameters or even classifiers from different complementary algorithms. In the model selection stage, AUC ROC is also frequently used as a metric for optimizing hyperparameters in the *k-fold* cross-validation process (James et al., 2017).

RESULTS

After performing the data pre-processing and learning the models, the selection and evaluation of the predictive models was performed. The metric used for model optimization and selection during learning was the AUC ROC. However, the three models chosen achieved similar performances with the test data (AUC = 0.6). In the Decision Tree and Random Forest algorithms, after training and testing the model, the selection of the most important attributes for the model was performed, that is, those that exert the greatest influence on the prediction results (Figure 5).

According to Figure 3, in both models obtained through the Decision Tree and Random Forest algorithms, the attributes that most influence the result of the models are the patient's age and whether or not the patient received SMS before the consultation. Despite the weight of these two attributes being much greater than the others, even removing the other features from the dataset

Scheduled Date: 06/02/2020 

Appointment Date: 06/10/2020 

Age: 48

Gender: Female Male

Patient has scholarship from the government? Yes No

Patient has hipertension? Yes No

Patient is alcoholic? Yes No

Patient has any handicap? Yes No

Patient is diabetic? Yes No

Patient has received SMS before the appointment? Yes No

Probably, the patient will not attend to the appointment.

Figure 6. Example of using the developed web application

and generating a new model with the same algorithms, only with the most relevant attributes, the models had the same performance.

In order to obtain the accuracy of the models with a greater degree of confidence, the calculation of the accuracy interval with cross-validation was performed. It is known that 95% of the data density is between the mean and more or less two standard deviations. In this way, all the algorithms obtained an accuracy interval similar to the Decision Tree algorithm, between the values of 78% and 80%. Therefore, after the validation of the chosen predictive models, the Decision Tree algorithm represents an interesting choice as a final model for use in future observations.

Finally, in order to make the test of the generated model more easily, a web application² was developed that allows inserting attributes referring to the consultation and characteristics of the patient and predicting whether or not the patient will attend the consultation, through the machine learning model obtained. The Flask micro framework (Grinberg, 2014) was used to create the API (Application Programming Interface) following the REST (Representational State Transfer) architecture. In the frontend, the Vue.js framework (Filipova, 2016) was used to allow the user to insert data via the interface. The developed API has a predict method that receives requests with input parameters. The parameters are transformed to the data standard accepted by the model and are inserted as input into the developed Machine Learning model. **Figure 6** shows an example of using the application.

² <https://github.com/luizhsalazar/medical-appointment-classifier>

CONCLUSIONS

Considering the results obtained, it can be said that the information collected in the data set does not seem sufficient to build a solid predictive model. The improvement of the results, that is, the improvement of the capacities of the classifier presented in this work, seems to depend on an improvement in the amount of information available, both for patients and for consultations.

Patient information can be supplemented with more socio-demographic information. Likewise, with regard to consultations, supplementing information with data related to the procedures and processes to be performed on the patient, can provide the classifier with relevant information to better predict the classification.

Finally, it also seems reasonable to think that the severity of a disease and its consequences can be a significant variable in the patient's decision to attend an appointment or not. While it is true that these are very subjective concepts and each individual interprets them differently, health is something that the average individual generally takes very seriously. Therefore, providing this information from the patient's medical history can improve the model. In future work, it is intended to extend the model to include other variables and parameters not addressed in this work. It will also be necessary to compare new results and metrics in order to identify the best algorithm according to the amount of information.

ACKNOWLEDGEMENTS

This work is funded by FAPESC through Santa Catarina State - Brazil funds under FAPESC Public Call notice No. 06/2017 "Support for Research Groups of the ACADEMIC System Institutions and FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/EEA/50008/2020

REFERENCES

- Abbott, D. (2014) *Applied Predictive Analytics: principles, and techniques for the professional data analyst*. John Wiley & Sons, 2014. ISBN: 978-1-118-72796-6.
- AlRowaili, M. O., Ahmed, A. E. and Areabi, H.A. (2016) Factors associated with no-shows and rescheduling MRI appointments, *BMC Health Services Research*, 16, 679. <https://doi.org/10.1186/s12913-016-1927-z>
- Bittar, O. J. N. V., Magalhães, A.,Martines, C. M., Felizola, N.B. G. and Falcão, L. H. B. (2016). Absenteísmo em atendimento ambulatorial de especialidades no estado de São Paulo. *BEPA*, 13(152), 19-32. Available at: <http://attosaude.com.br/assets/conteudo/uploads/absenteismo-ambulatorial--art-original57eec18c360fb.pdf> (Accessed: 9 October 2019).
- Canelada, H. F., Levorato, C. D., Corte, R. I. A. S. and Diniz, E. E. S. (2014). Redução do Absenteísmo Através da Gestão da Agenda e do Trabalho em Rede. *Blucher Medical Proceedings*, 1(2), 145. <https://doi.org/10.5151/medpro-cihhs-10458>
- Cavalcanti, R. P., Cavalcanti, J. C. M., Serrano, R. M. S. M. and Santana, P. R. S. (2013). Absenteísmo de consultas especializadas nos sistema de saúde público: relação entre causas e o processo de trabalho de equipes de saúde da família, João Pessoa – PB, Brasil. *Rev. Tempus - Actas de Saúde Coletiva*. <https://doi.org/10.18569/tempus.v7i2.1344>
- Covington, D. (2019). *Analytics: Data Science, Data Analysis and Predictive Analytics for Business*, LuLu Press. ISBN: 978-0-359-82852-4.
- Cox, V. (2017). *Translating Statistics to Make Decisions: a guide for the non statistician*. Apress. 1st edition. <https://doi.org/10.1007/978-1-4842-2256-0>
- Deng, X., Liu, Q., Deng, Y. and Mahadevan, S. (2016) An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Information Sciences*, 340, 250-261. <https://doi.org/10.1016/j.ins.2016.01.033>
- Dogruyol, K. and Sekeroglu, B. (2020). Absenteeism Prediction: A Comparative Study Using Machine Learning Models. In: R. Aliev, J. Kacprzyk, W. Pedrycz, M. Jamshidi, M. Babanli and F. Sadikoglu (eds), 10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions - ICSCCW-2019. ICSCCW 2019. *Advances in Intelligent Systems and Computing*, vol 1095. Springer, Cham, 2020. https://doi.org/10.1007/978-3-030-35249-3_94
- Elvira, C., Ochoa, A., Gonzalez, J. C. and Mochón, F. (2017). Machine Learning Based no show prediction in outpatients visits. *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(7), 29-34. <https://doi.org/10.9781/ijimai.2017.03.004>
- Fenner, M. (2019) *Machine Learning with Python for Everyone*. Addison-Wesley Professional. ISBN: 978-0-134-84562-3.
- Filipova, O. *Learning Vue.js 2*. (2016) Packt Publishing, December, 14, 2016. ISBN: 978-1-78646-994-6.
- Goodfellow, I., Bengio Y. and Courville, A. (2016). *Deep Learning*. [S.l.]: MIT Press. ISBN: 978-0262035613.
- Grinberg, M. (2014). *Flask Web Development: Developing Web Applications with Python*. O'Reilly Media; 1 edition, April 28. ISBN: 978-1449372620.
- Grupo Hospitalar Conceição (GHC) e Fundação Oswaldo Cruz (FIOCRUZ), 2013. Available at https://www.arca.fiocruz.br/bitstream/icict/34822/2/mirian_silva_icict_espec_2013.pdf (Accessed: 9 July 2020).
- Hastie, T., Tibshirani, R. and Friedman, J. (2017). *The elements of statistical learning: data mining, inference, and prediction*. Springer, 9th edition. <https://doi.org/10.1007/978-0-387-84858-7>

- Hoe, J. W. M. (2007). Service delivery and service quality in radiology, *J Am Coll Radiol.*, 4(9), 643-651. <https://doi.org/10.1016/j.jacr.2007.04.013>
- James, G., Witten, D., Hastie, T. and Tibshirani, R. (2017). *An introduction to statistical learning: with applications in R*. New York: Springer; 7th edition. <https://doi.org/10.1007/978-1-4614-7138-7>
- Kuhn, M. and Johnson, K. (2018). *Applied predictive modeling*. Springer, 2nd edition. <https://doi.org/10.1007/978-1-4614-6849-3>
- Lenzi, H., Ben, A. J. and Stein, A. T. (2019). Development and validation of a patient no show predictive model at a primary care setting in southern Brazil, *PLoS ONE*, 14(4), e0214869. <https://doi.org/10.1371/journal.pone.0214869>
- Lu, L., Li, J. and Gisler, P. (2011). Improving financial performance by modeling and analysis of radiology procedure scheduling at a large community hospital. *Journal of Medical Systems*, 35(3), 299-307. <https://doi.org/10.1007/s10916-009-9366-6>
- McCandless, D. (2014). *Knowledge Is Beautiful*. Harper Design, 1th edition. ISBN: 978-0-062-18822-9.
- Mohammadi, I., Wu, H., Turkcan, A., Toscos, T. and Doebebling, B. N. (2018). Data analytics and modeling for appointment no show in community health cares. *J Prim Care Community Health*, 9, 2150132718811692. <https://doi.org/10.1177/2150132718811692>
- Monken, S. F. and Moreno, R. C. B. (2015). Utilização dos Alertas de Controle como Ferramenta para a Fidelização da Clientela de Pediatria em um Ambulatório Público. *Revista de Administração Hospitalar Inovação em Saúde*, 12(3), 94-105. <https://doi.org/10.21450/rahis.v12i3.2696>
- Montgomery, D. C., Peck, E. A. and Vining, G. (2012). *Introduction to Linear Regression Analysis*. John Wiley Sons, 5th edition. ISBN: 978-0-470-54281-1.
- Nelson, A., Herron, D., Rees, G. and Nachev, P. (2019). Predicting scheduled hospital attendance with artificial intelligence. *Digit. Med.*, 2, 26. <https://doi.org/10.1038/s41746-019-0103-3>
- Priyanka, D. and Nayak, J. (2020). Empirical Analysis of Absenteeism at Workplace Using Machine Learning. In: J. Nayak, V. Balas, M. Favorskaya, B. Choudhury, S. Rao and B. Naik (eds), *Applications of Robotics in Industry Using Advanced Mechanisms*. ARIAM 2019. *Learning and Analytics in Intelligent Systems*, vol 5. Springer, Cham. <https://doi.org/10.1007/978-3-030-30271-9>
- Raschka, S. and Mirjalili, V. (2017). *Python Machine Learning: machine learning and deep learning with Python, Scikit-learn and Tensorflow*, Packt Publishing; 2nd edition. ISBN: 978-1-78712-593-3.
- Recht, M., Macari, M., Lawson, K., Mulholland, T., Chen, D., Kim, D. and Babb, J. (2013) Impacting key performance indicators in an academic MR imaging department through process improvement, *Journal of the American College of Radiology*, 10(3), 202-206. <https://doi.org/10.1016/j.jacr.2012.08.008>
- Reid, M. W., Cohen, S., Wang, H., Kaung, A., Patel, A., Tashjian, V., Williams Jr., O. L., Martinez, B. and Spiegel, B. M. R. (2015). Preventing Patient Absenteeism: validation of a predictive overbooking model. *Am J Manag Care*, 21(12), 902-910. Available at: <https://www.ajmc.com/journals/issue/2015/2015-vol21-n12/preventing-patient-absenteeism-validation-of-a-predictive-overbooking-model?p=4> (Accessed: 23 June 2020).
- Salazar, L. H., Fernandes, A. M. R., Dazzi, R. L. S., Garcia, N. M. and Leithardt, V. (2020). Prediction of Attendance At Medical Appointments Based On Machine Learning. In 2020 *15th Iberian Conference on Information Systems and Technologies (CISTI)*, IEEE Xplore, ISBN: 978-989-54659-0-3. <https://doi.org/10.23919/CISTI49556.2020.9140973>
- Santibáñez, P., Chow, V. S. and French, J. (2009). Reducing patient wait times and improving resource utilization at British Columbia Cancer Agency's ambulatory care unit through simulation. *Health Care Management Science*, 12(4), 392-407. <https://doi.org/10.1007/s10729-009-9103-1>
- Santos, H. G., Nascimento, C. F., Izbicki, R., Duarte, Y. A. O. and Chiavigatto Filho, A. D. P. (2019). Machine learning for predictive analysis in health: an example of an application to predict death in the elderly in São Paulo, Brazil - Machine learning para análises preditivas em saúde: exemplo de aplicação para prever óbito em idosos de São Paulo, Brasil. *Cad. Saúde Pública*, 35(7), e00050818. <https://doi.org/10.1590/0102-311x00050818>
- Shalev-Shwartz, S. and Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press. ISBN-13: 978-1-107-05713-5.
- Silva, C. C., Andrade, D. C., Aplonário, J. H. D., Zamboni, C., Durigan, J. R. and Mercadante, M. T. (2017). Absenteísmo ambulatorial no pós operatório dos pacientes ortopédicos de um hospital de ensino de São Paulo. *Arq Med Hosp Fac Cienc Med Santa Casa São Paulo*, 62(2), 77-80. Available at: <https://pdfs.semanticscholar.org/bf75/cf68698956220c3a44b988f0607f5501f5ea.pdf> (Accessed: 27 April 2020).
- Silva, M. T. F. S. (2013). *Avaliação da Redução do Absenteísmo as Consultas Marcadas em um Serviço de Referência em Diabetes*. Projeto de pesquisa apresentado como requisito parcial para a obtenção do título de Especialista em Informação Científica e Tecnológica em Saúde, realizado pelo. Available at: <http://docs.bvsalud.org/biblioref/coleciona-sus/2013/31530/31530-752.pdf> (Accessed: 28 March 2020).
- Silveira, G. S., Ferreira, P. R., Silveira, D. S., and Siqueira, F. C. V. (2018). Prevalence of absenteeism in medical appointments at a basic health unit in southern Brazil – Prevalência de absenteísmo em consultas médicas em unidade básica de saúde do sul do Brasil. *Rev. Bras. Med. Fam. Comunidade*, 13(40), 1-7. [https://doi.org/10.5712/rbmfc13\(40\)1836](https://doi.org/10.5712/rbmfc13(40)1836)
- Wahid, Z., Satter, A. K. M. Z., Imran, A. A. and Bhuiyan, T. (2019). Predicting Absenteeism at Work Using Tree-Based Learners”, *ICMLSC 2019: Proceedings of the 3rd International Conf. on Machine Learning and Soft Computing*. <https://doi.org/10.1145/3310986.3310994>