

Prediction of hospital no-show appointments through artificial intelligence algorithms

Sarab AlMuhaideb,^a Osama Alswailem,^b Nayef Alsubaie,^b Ibtihal Ferwana,^a Afnan Alnajem^c

From the ^aDepartment of Computer Science, Prince Sultan University, Riyadh, Saudi Arabia; ^bHealth Informatics and Telecommunication Affairs, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; ^cDepartment of Biostatistics, Epidemiology & Scientific Computing, Princess Nora bint Abdulrahman University, Riyadh, Saudi Arabia

Correspondence: Dr. Sarab AlMuhaideb · Department of Computer Science, Prince Sultan University, SA Computer Science Department, CCIS, King Saud University, Saudi Arabia · T.: +966-11-8052836 · salmuhaideb@ksu.edu.sa

Citation: AlMuhaideb S, Alswailem O, Alsubaie N, Ferwana I, Alnajem A. Prediction of hospital no-show appointments through artificial intelligence algorithms. *Ann Saudi Med* 2019; 39(6): 373-381 DOI: 10.5144/0256-4947.2019.373

Received: July 1, 2019

Accepted: September 24, 2019

Published: December 5, 2019

Copyright: Copyright © 2019, *Annals of Saudi Medicine*, Saudi Arabia. This is an open access article under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND). The details of which can be accessed at <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Funding: None.

BACKGROUND: No-shows, a major issue for healthcare centers, can be quite costly and disruptive. Capacity is wasted and expensive resources are underutilized. Numerous studies have shown that reducing uncanceled missed appointments can have a tremendous impact, improving efficiency, reducing costs and improving patient outcomes. Strategies involving machine learning and artificial intelligence could provide a solution.

OBJECTIVE: Use artificial intelligence to build a model that predicts no-shows for individual appointments.

DESIGN: Predictive modeling.

SETTING: Major tertiary care center.

PATIENTS AND METHODS: All historic outpatient clinic scheduling data in the electronic medical record for a one-year period between 01 January 2014 and 31 December 2014 were used to independently build predictive models with JRip and Hoeffding tree algorithms.

MAIN OUTCOME MEASURES: No show appointments.

SAMPLE SIZE: 1 087 979 outpatient clinic appointments.

RESULTS: The no show rate was 11.3% (123 299). The most important information-gain ranking for predicting no-shows in descending order were history of no shows (0.3596), appointment location (0.0323), and specialty (0.025). The following had very low information-gain ranking: age, day of the week, slot description, time of appointment, gender and nationality. Both JRip and Hoeffding algorithms yielded a reasonable degrees of accuracy 76.44% and 77.13%, respectively, with area under the curve indices at acceptable discrimination power for JRip at 0.776 and at 0.861 with excellent discrimination for Hoeffding trees.

CONCLUSION: Appointments having high risk of no-shows can be predicted in real-time to set appropriate proactive interventions that reduce the negative impact of no-shows.

LIMITATIONS: Single center. Only one year of data.

CONFLICT OF INTEREST: None.

The term “no-show” refers to a patient’s failure to keep a scheduled appointment and make a cancellation. No-shows are one of the challenges faced by healthcare practices around the world.^{1,2} No-shows are a constant battle at King Faisal Specialist Hospital and Research Centre Organization (KFSHRC), despite several hospital-wide initiatives to improve attendance. Eighteen percent of appointments are no-shows.³ In comparison, the national average for no-show appointments in most specialties in the US⁴ is a median of 5%. Our no-show rate of 18% is a cause of great concern because of wasted capacity, underutilization of expensive resources, and it is a barrier to patient access.

Interventions range from simple telephone reminders and short message services (SMS)^{5,6} to deployment of a sophisticated scheduling methods that use machine-learning and artificial intelligence to decrease the rate of no show appointments.⁷⁻¹¹ By improving communications and increasing appointment flexibility, Mohamed et al decreased the no-show rate from 49% to 18% and sustained the no-show below a target of 25% for two years.¹² However, costs are associated with any intervention that reduces the no-show rate and the costs of interventions must be weighed against the reduction in the no-show rate, as uncertainty in patient attendance will always persist. Elvira et al suggested that a prediction system using machine-learning to reduce no-shows could be effective even if the results are not spectacular.¹⁰

Nelson and colleagues argued that further improvements in reducing no shows would require use of complex predictive models that better capture data on individual variability.¹³ As early as 1983, Bigby et al found that two-thirds of the cost-savings was generated from the patients that had a prior predicted probability greater than 20% of being no-shows.¹⁴ Therefore, a fundamental step in developing a predictive model is accurately predicting when patients inclined to miss appointments are most likely to not show up. The problem is too complex to rely on “simple linear, low-dimensional statistical models” that discard all but the most obvious predictive features. Simple approaches are usually applied to all patients rather than focusing on the no-show patients, who are relatively infrequent. No-shows exhibit a behavioral pattern that tends to repeat but may change over time with more recent behavior being more relevant.¹⁵ Complex approaches that capture this individual behavior will likely be more predictive of future behavior.

The aim of this research was to use machine-learning to build a model that predicts no-shows for individ-

ual appointments, based on the data in the electronic medical record. This model allows for the targeting of high-risk patients to reduce the cost of interventions aimed at reducing the no-show rate.

METHODS

A no-show occurs when a patient fails to attend his/her appointments without prior notice or on short notice (e.g., <24 hours).¹⁶ A minimum time threshold, usually 24 hours, is required for another patient to be scheduled into the cancelled time slot. A no-show at the ambulatory-care setting at KFSHRC is defined as a patient who cancels within 24 hours of the appointment, is 60 minutes late for a scheduled appointment, arrives after the clinic is closed, or does not show up at all.

In this study, JRip¹⁷ and Hoeffding trees¹⁸ were selected to model and classify patient appointments, as they are well-known algorithms in the machine-learning field. JRip is selected because it is characterized by compact size models,¹⁹ which facilitates understanding by humans.^{20,21} Hoeffding trees are selected because of their power in handling very large datasets with reasonable time and computational costs as well as their ability to adjust to the changes in concepts leading to no-shows. That is, Hoeffding trees does not assume that the data is drawn from a stationary distribution. Thus, even if the process leading to no show has changed over time, these algorithms have the ability to adapt to the new concept.

Experimental setup

In this study we used the open-source machine-learning software Weka (<http://www.cs.waikato.ac.nz/~ml/weka/>). We used different performance metrics to measure the effectiveness of a classifier with respect to a given dataset. The sensitivity, specificity, and predictive accuracy are among the most commonly used performance metrics.²² The predictive accuracy is measured as the proportion of correctly classified instances over all instances. Sensitivity measures the proportion of positive samples that are correctly classified as positive. Specificity measures the proportion of negative samples that are correctly recognized as negative. We also report the model size, which represents the number of JRip rules and the number of leaves in the Hoeffding trees. Another performance measure is the area under the curve index area under curve (AUC).²³ The AUC index values range from 0.5 (random behavior) to 1.0 (perfect classification performance). According to Hosmer et al²⁴ an AUC in the ranges (0.7, 0.8) and (0.8, 0.9) have acceptable and excellent discrimination, respectively. Finally, we also report the

model generation time in seconds. All measures are reported for the average performance of the algorithm on unseen data (test set). We conducted experiments on an Intel Core™ i7 -7500U CPU (2.9 GHz) with 16 GB RAM.

In the generation of the model, the machine “learns” from an initial set of data, usually called the “training” data, and with the model developed from that data, attempts to make predictions on a set of ‘test’ data. The test data is also known as the validation dataset for use in cross-validation, which measures how accurately the model will predict new data. We used a stratified 10-time, 10-fold cross-validation²⁵ procedure as it is considered the best error-estimation strategy.^{26,27} In this procedure, the whole dataset is partitioned into 10 disjoint folds, each of size 0.1 of the original dataset size. Cross-validation is done 10 times each using 9 folds for training the model and the one fold left out of the training phase is used as a test set. Each time a different fold is used as a test set. Results are then averaged over the 10 iterations. For preprocessing, we used IBM SPSS Modeler Version 18 (<https://www.ibm.com/us-en/marketplace/spss-modeler>) and Matlab R2015 (<http://www.mathworks.com/products/matlab/>).

Dataset

The data used in this study consisted of all historic outpatient clinic scheduling data for a one-year period between 1 January 2014 and 31 December 2014. The dataset contained 1 087 979 records; 11.3% (123 299) are records of no-show cases. Each record contains 11 predictor attributes (also called features or variables) in addition to the target class (the show-no-show variable, the same as the dependent variable in logistic regression). The predictor attributes include patient-specific data such as patient age, gender, nationality, geographic region, and the 12-month historic rate of no-show for each patient (noShow-rate), computed as the number of appointments considered no-shows for the patient divided by the total number of appointments for this patient (including attendance and no-shows). In addition, each record contains data about the visit including the appointment location (app-loc), day of the week (day-of-week), appointment time (time-slot), specialty, slot description (slot-des), registration type (reg-type), and no-show status (class). The last attribute (class) represents the target class that we aim to predict, in which the value is either (show) if the patient attended the appointment, or (no-show) otherwise. **Table 1**

Table 1. A complete list of predictor features and their information gain ranking.

Input	Type	No. Values/ range	Info gain ranking	Description
<i>noShow-rate</i>	Numeric	0.0 – 100.0	0.3596	The proportion of no-show visits among all visits per patient
<i>app-loc</i>	Nominal	32	0.0323	Appointment location
<i>specialty</i>	Nominal	24	0.025	Clinic specialty or medical service
<i>reg-type</i>	Nominal	18	0.0104	Type of patient as Ordinary, Employee, Dependent, Donor, etc.
<i>age</i>	Numeric	0-115	0.003	Patient age in years
<i>day-of-wk</i>	Nominal	7	0.003	Visit day of week (Sun, Mon, Tues, Wed, Thurs, Fri, Sat)
<i>slot-desc</i>	Nominal	4	0.002	Registration slot description, which can be new patient (NP), new follow-up (NF), follow-up patient (FU), or (Diagnostic)
<i>region</i>	Nominal	5	< 0.001	Geographic region of patient as Central, Eastern, Western, Northern, or Southern region
<i>time-slot</i>	Nominal	4	< 0.001	Time slot of visit as Early AM, AM, PM, or Night
<i>gender</i>	Nominal	2	< 0.001	Patient gender as Male or Female
<i>nationality</i>	Nominal	2	< 0.001	Patient nationality as Saudi or Non-Saudi
<i>class</i>	Nominal	2	—	Target class as either (show) if the patient attended the appointment, or (no-show) otherwise

shows information about the attributes included in the study. We ordered the attributes in the table according to the ranking of the information-gain ratio with respect to the target class (class).

The dataset was preprocessed as follows. We omitted records having slot-desc="Medication Refill" or "Chart Review" as these are not patient appointments. We also modified and minimized some data attributes. First, we minimized the nationality attribute so that it had only two categories: Saudi and Non-Saudi. Second, we minimized the slot-desc attributes to only four: follow-up (FU), new follow-up (NF), new patient (NP) and diagnostic. Any slot description that contained either FU, NF, or NP as part of its name was categorized according to the existing subscript. Diagnostic replaced any diagnostic-related descriptions. For example, we categorized any description that contains computed tomography (CT), magnetic resonance image (MRI), scan, etc. as Diagnostic. Third, the specialty attributes which are defined as the medical specialty providing care to the patient, contain 162 unique specialty categories. We grouped these into 24 unique categories. For example, we grouped "ENT-Otolaryngology", "ENT-Audiology", and "ENT-Head & Neck" under one attribute: "ENT". Last, we mapped appointment time (time-slot) to one of four intervals: Early AM (7-9 am), AM (9 am to 12 pm), PM (12-5 pm), or Night (5 pm to 7 am). As mentioned earlier, only 11.3% of the records in the original dataset related to appointments in which patients had attended (class=show) and the remaining 88.7% represented no-show cases (class=no show). The dataset is thus considered highly imbalanced, exposing the prediction performance to a considerable negative effect. Thus, we balanced the dataset with respect to the target-class attribute (class) so that approximately half of the records belong to the class = show category and the other half belonged to the class = no-show category. The resulting dataset contained 222 607 instances, of which 109 113 belonged to the no-show cases (49%). Mosaic plots were used to show the distribution of no-shows with respect to each of the top three indicator attributes: noShow-rate (historic no-show rate), app-loc (appointment location), and specialty. Mosaic plots are used because they are multidimensional plots that facilitate the illustration and recognition of different relationships between variables. The width of the bar is proportional to the number of observations within that category.

JRip

JRip is a competitive separate-and-conquer propositional rule learning algorithm that performs efficiently

on noisy datasets containing hundreds of thousands of records.¹⁷ It is an optimized version of the IRIP algorithm.²⁸ The algorithm starts with an empty rule set. The training dataset is split into a grow set (two-thirds of the training set) and a prune set (one-third of the training set). Using the grow set, and an information gain heuristic, the algorithm builds a model that actually overfits the data. Then, using the prune set, and pruning operators, the algorithm repeatedly applies a pruning operator to remove the condition or rule that yields the greatest reduction of errors on the prune set. To prune a rule, the algorithm considers the final sequence of conditions that maximize the value of v as in Equation 1, where P refers to the total number of instances in PrunePos, PrunePos, N refers to the total number of instances in PruneNeg, p is to the total number of instances in PrunePos that are covered by Rule, and n is the total number of instances in PruneNeg that are covered by Rule.

$$v(\text{Rule}, \text{PrunePos}, \text{PruneNeg}) = \frac{p + (N - n)}{P + N}$$

This process ends when applying any pruning operator would result in an increased error on the pruning set.

Table 2 shows the parameters used for JRip. The minimum total weight of the instances in a rule is minNo, optimizations refers to the number of optimization runs, and checkErrorRate refers to whether a check for an error rate $>1/2$ is included in the stopping criterion.

Hoeffding Trees

Several thousand appointments are scheduled daily in KFSHRC, which imposes high demands in terms of memory and computational requirements. Hoeffding trees²⁹ are considered one of the new generations of stream mining. Using these methods, learning is performed online from a stream of data through just one pass on the data. A very small constant time is required by these methods per example, which yields acceptable computational costs. Hoeffding trees, also known as Very Fast Decision Trees (VFDTs), rely on the assumption that the distribution of data does not change much, and thus, to find the best attribute to test at a given node, it is sufficient to consider a small subset of the training examples, rather than the entire training set as in conventional decision tree learning algorithms. Using this assumption, and given a stream of training instances, the first batch of instances are used to select the root test. The succeeding batches are passed down the tree to select the test for the next node in the tree, and so on recursively. The Hoeffding bound^{30,31} is defined as follows: Consider a real-valued random variable, $r \in R$. Assume that n independent observations of this variable are made with

Table 2. Parameter values used for the JRip.

Parameter	Value
numDecimalPlaces	2
minNo	2
optimizations	2
checkErrorRate	True
usePruning	True
batchSize	100
doNotCheckCapabilities	False

Table 3. Parameter values used for the Hoeffding tree.

Parameter	Value
batchSize	100
numDecimalPlaces	2
nmin	200
splitCriterion	Gini index
splitConfidence	1.0E - 7
τ	0.05
leafPredictionStrategy	Naive Bayes Adaptive

Table 4. Evaluation of the no-show model generated by the JRip and Hoeffding trees algorithms.

Model	Accuracy	(Sensitivity)	(Specificity)	AUC	Size	Time
JRip	76.44%	0.795	0.735	0.776	13	86.02
Hoeffding trees	77.13%	0.815	0.729	0.861	391	1.3

Table 5. The no-show model generated by the JRip algorithm.

(noShow-rate \geq 29.7872) \rightarrow class=no-show (57,470.0/7,828.0)
 (noShow-rate \geq 17.2414) \rightarrow class=no-show (41,815.0/14,369.0)
 (noShow-rate \geq 14.7727) \rightarrow class=no-show (12,214.0/5,432.0)
 (noShow-rate \geq 13.9241) and (specialty = SUR) and (age \geq 48) \rightarrow class=no-show (261.0/93.0)
 (noShow-rate \geq 12.3288) and (specialty = DTC) \rightarrow class=no-show (311.0/79.0)
 (noShow-rate \geq 12.1951) and (day-of-wk = Sunday) \rightarrow class=no-show (3,185.0/1,487.0)
 (noShow-rate \geq 12.1212) and (app-loc = Dental-R) and (age \geq 52) \rightarrow class=no-show (294.0/108.0)
 (noShow-rate \geq 10.9244) and (app-loc = Phy-Thpy-R) \rightarrow class=no-show (1,339.0/581.0)
 (noShow-rate \geq 13.0435) and (age \geq 67) and (gender = Female) \rightarrow class=no-show (253.0/107.0)
 (day-of-wk = Thursday) and (noShow-rate \geq 13.7931) and (noShow-rate \leq 14.5161) and (time-slot = EarlyAM) \rightarrow class=no-show (303.0/132.0)
 (noShow-rate \geq 12.7907) and (day-of-wk = Monday) and (noShow-rate \leq 13.7255) and (gender = Female) \rightarrow class=no-show (395.0/176.0)
 (noShow-rate \geq 10.4478) and (specialty = SUR) and (age \geq 19) and (app-loc = OB/GYN-R) \rightarrow class=no-show (44.0/1.0)
 default: class=show (104,723.0/21,622.0)

mean $r \in \mathbb{R}$. According to the Hoeffding bound, and with a probability of $1 - \delta$, the true mean of variable r is at least $r - \epsilon$, where ϵ is defined in Equation 2.

$$\epsilon = \sqrt{\frac{r^2 \ln \frac{1}{\delta}}{2n}}$$

The Concept Drifting Very Fast Decision Tree (CVFDT)¹⁵ is an improvement over the VFDT, which allows adapting the tree to the changes in the underlying process that generates the examples. In a nutshell, the CVFDT works on keeping the generated model consistent with a sliding window of examples by updating the required statistics at each node. However, if the current splitting test at a node no longer passes the Hoeffding test because an alternative attribute has a higher heuristic value, then a new subtree is grown with the alternate

splitting test at its root. If this new test becomes more accurate with the arriving data, it then replaces the old one.

Table 3 shows the parameters used for Hoeffding. The preferred number of instances for processing if batch prediction is being performed is batchSize. splitCriterion represents the split evaluation heuristic, such as information-gain or Gini index. The parameter SplitConfidence refers to the allowable error in a split decision. The parameter n_min is the number of instances a leaf should observe before re-computing the split evaluation heuristic. τ refers to the threshold below which a split is forced to break ties.

RESULTS

The most significant predictor, with an information-gain ranking of 0.3596, was the historic noShow-rate per per-

son (**Figure 1**). This predictor uses individual patients' no-show histories to predict the possibility of no-shows for future appointments. In this figure, the no-show rate (noShow-rate) is divided into six categories as follows. Category 0 refers to patients with a historic noShow-rate $\leq 10\%$, Category 1 is for a noShow-rate $\leq 20\%$, Category 2 is for a noShow-rate $\leq 30\%$, Category 3 is for a noShow-rate $\leq 40\%$, Category 4 is for a noShow-rate $\leq 50\%$, and Category 5 is for a noShow-rate $> 50\%$. The horizontal axis in the figure shows the six categories, whereas the vertical axis illustrates the actual class of the appointments (class) as show or no-show. As de-

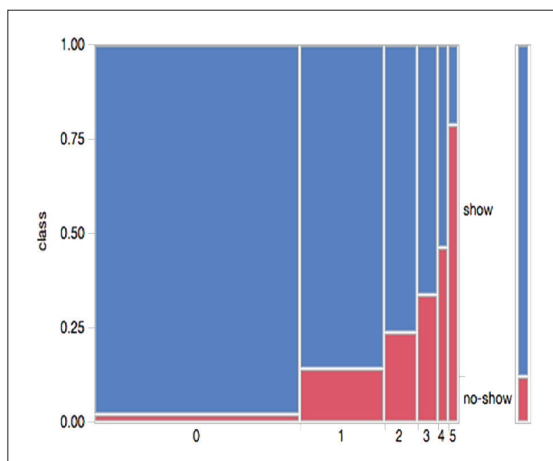


Figure 1. Distribution of no-show by the historic no-show rate (Example: patients who have a historic no-show rate greater than 50% [category 5] had the highest probability of not showing at the next appointment [79.1%]). Historic no-show rate was the most important predictor of future no-shows.

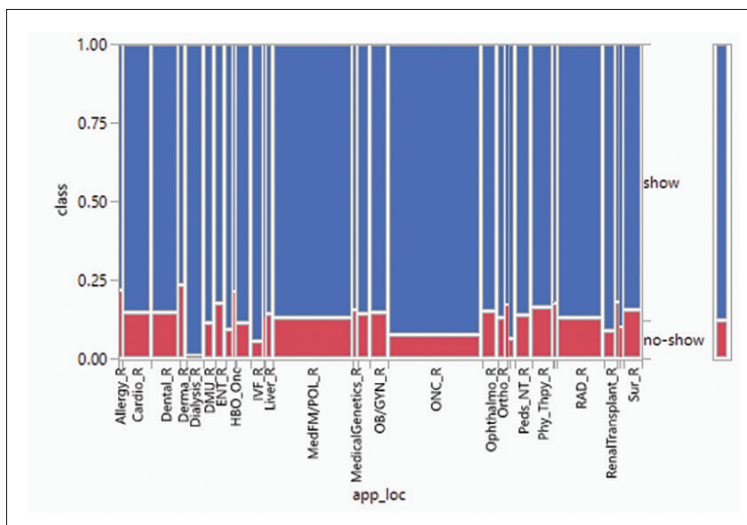


Figure 2. Distribution of no-show rate by appointment location.

picted in **Figure 1**, the majority of the patients (56.4%), the width of the category on the horizontal axis have a historic no-show rate in Category 0. Of the patients, 23% belong to Category 1, 9.1% are in Category 2, and 5.6% are in Category 3. On the other hand, the percentages of patients having no-show rates in Categories 4 and 5 are the least ($<3\%$ each). The figure also illustrates that the chances of a future appointment being a no-show increase in significant proportions for patients with multiple previous incidents of no-shows. Appointments for patients with Category 0 no-show rates yield a distribution of no-shows at only 2.34% (i.e., among appointments for patients with historic noShow-rates in Category 0, only 2.34% have class=no-show), whereas patients with Category 5 noShow-rates have a distribution of class=no-show at as high as 79.10%.

The appointment location predictor (app-loc), which has a much lower information-gain ranking (0.0323) than that of noShow-rate per person class is shown in **Figure 2**. Appointment locations serving patients with critical or unique conditions have a lower distribution of no-show (Dialysis at 1% and IVF at 5.5%) compared to less critical/unique conditions, such as Allergy at 21.7% and Dermatology at 23.5%.

Next, the specialty predictor which has an information-gain ranking of 0.025, as shown in **Figure 3** yielded a phenomenon similar to that of appointment location class. The no-show distribution is low for rare/critical specialties such as Urology (URO, 6.5%) and Oncology (ONC, 6.9%) compared to Surgery (SUR, 16.6%), Optometry (OPT, 23.8%), and Dietary (DTC, 27.23%).

The predictor reg-type which categorizes patients as permanent hospital employee, temporary hospital employee, locum hospital employee, hospital employee dependent, organ donor, ordinary patient (referred to the hospital), temporary patients (limited visits) and VIP has an information-gain ranking of 0.0104. Appointments related to locum tenens care providers prior to their employment at KFSHRC have very low no-show distributions (Pre-employment at 2.6% and Locum at 8.0%), as their employment decisions are based on the fulfillment of their medical evaluation. Patients who have easy access to appointments and can easily reschedule appointments have high distributions of no-shows: for example, dependents of physicians' at 24% and retired employees at 18.6%.

Both age and day-of-wk predictors have low information-gain rankings of 0.003. Although the distributions of no-shows among all age ranges are very similar, the distribution of no-shows for the 15-25 age group is as high as 51.53%. The lowest distribution of no-shows appears among patients of the 55-64 age group at 46.56%.

The predictor slot-desc, which has a low information-gain ranking of 0.002, attempts to predict no-shows by slot type. No-shows rates are lowest in the FU (Follow-up) slot type compared with NP (New Patient).

Predictor classes related to time-slot, gender and nationality all have information-gain rankings below 0.001, diminishing the prominence of no-show predictions.

The JRip and Hoeffding tree algorithms were independently applied to the preprocessed dataset as described above. **Table 4** shows the evaluation results for the generated models. It can be seen that the model generated using Hoeffding trees has a stronger prediction accuracy, sensitivity, and specificity. The AUC for the JRip model can be characterized as “acceptable”, whereas that for the Hoeffding tree model is “excellent discrimination”. Notably, the time required by JRip was much larger than that for Hoeffding trees. However, the size of the model generated by JRip is much smaller.

The model generated by JRip consists of 13 rules, which are shown in **Table 5**. This model resembles a decision list, represented by an ordered sequence of IF-THEN-ELSE rules. Each rule is in the form (IF) condition \rightarrow class (coverage/error). The numbers in the bracket stand for coverage/errors in the training data, which follows the standard convention of tree/rule induction. For example, (noShow-rate \geq 14:7541 \rightarrow class= no-show (112678.0/27674.0) means that the rule (noShow-rate \geq 14:7541) \rightarrow class= no-show covers instances with total weights of 112678.0, out of which there are instances with weights of 27674.0 misclassified. In this case, one instance corresponds to weight 1. The last rule is the default rule, providing the class applicable when all previous rules do not apply to the case under consideration. The model generated by the Hoeffding tree is not shown due to its prohibitive size.

DISCUSSION

The prediction of patient no-shows has been studied since the early eighties. Dove and Schneider showed that it is possible to accurately predict no-shows using a decision tree.³² Their study revealed that the most important predictor was the patient's previous appointment-keeping pattern. In another study, logistic regression was used to estimate the probability of patient no-shows at an outpatient clinic at a midwestern Veterans Affairs hospital in the US.³³ The study showed the existence of many indicators of no-show including being of a younger age, not having a cardiac problem, having depression and drug dependencies, not being married, and traveling long distances. However, the most important indicators were prior no-show rates and the number of hospital admissions.

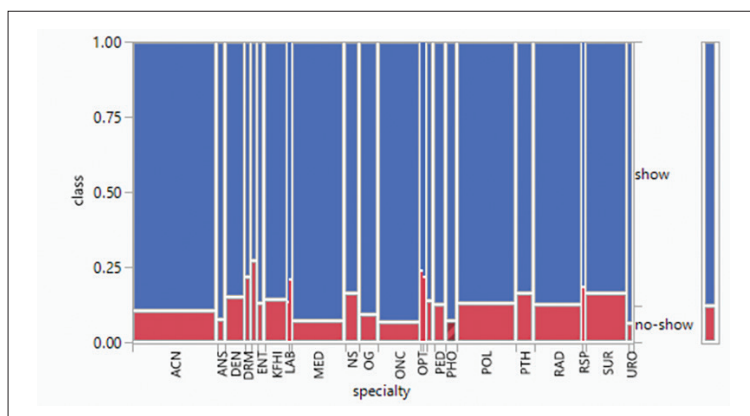


Figure 3. Distribution of no-show by specialty.

Lotfi and Torres used classification and regression trees with four different tree-growing criteria to predict the no-show of a patient.³⁴ The model was trained on 367 patient data and collected over five months from a physical therapy clinic within an urban health and wellness center situated in a public university. The study showed that gender, distance from the clinic, education, work status, days since last appointment, and attendance at prior visits are significant indicators. The resulting models were compared to models developed using artificial neural networks and Bayesian Networks. Classification and Regression Trees (CART) produced the most favorable results, with a predictive accuracy of 78%.³⁵

Having a vast amount of outpatient data in its electronic medical record and electronic scheduling systems, KFSHRC-Riyadh undertook this experiment utilizing two prominent machine-learning algorithms, JRip and Hoeffding trees, to predict no-shows. The resulting models can be used to help hospital administrators design strategy interventions to reduce the negative impact of no-shows.

Both of the algorithms used, JRip and Hoeffding trees, yielded reasonable degrees of accuracy of 76.44% and 77.13%, respectively, with an AUC index being at an acceptable discrimination power for JRip at 0.776, and at 0.861 with an excellent discrimination for Hoeffding trees. As for the outcomes, the JRip model not only gives a prediction, but also provides insight and reveals interdependence among the predictor attributes, demonstrating a more powerful tool. The information-gain ratio ranking shown in **Table 1** revealed that the no-show rate is related to the previous no-show behavior of the patient. The result for KFSHRC-Riyadh is consistent with that of earlier studies.³²⁻³⁸ In reference to **Table 5**, using this indicator alone, the prediction of

no-shows is as high as 88% for the noShow-rate \geq 29; as high as 74.39% for the noShow-rate \geq 17.2414; and 69.22% for the noShow-rate \geq 14.7727. The accuracy of this prediction is high, confirming that the higher the rate of a previous no-shows, the higher the chances that a patient will commit a no-show on subsequent visits. Patient education about the importance of showing up for an appointment is key to addressing the behavioral aspect of no-shows. Overbooking based on a sophisticated scheduling system with prediction capabilities can also mitigate no-shows related to patient behavior.

As an example of predictions from the model, the JRip algorithm predicted that 73% of appointments yielded no show when the patient was over 47 years old and with historic noShow-rate \geq 13.29241 and appointment specialty being surgery. Patient education may mitigate the risk of no-shows for these classes. On the other hand, with a noShow-rate \geq 12.32 in a non-physician clinic, dietician appointments have a higher predictability of no-shows at 79.74%. Similarity, with a noShow-rate \geq 10.9244 in a therapist-led clinic (physical therapy), appointments have a higher predictability of no-shows at 79.74%, according to the JRip model. These non-physician clinics are conducted in separate sessions on different days from the physician-led clinics that referred them to these non-physician clinics in the first place. This may explain the high no-show rate. The concept of one-stop shop with multidisciplinary clinics (to include dietician clinics) could address the no-show issue for these clinics.

Appointments on Sundays for patients with previous noShow-rate \geq 12.1951 yielded a 68% predictability of being a no-show. Since many patients come from outside Riyadh and may not get the right flights to reach clinics on Sundays due to the high demand on the airlines after the weekend, no-shows are predicted to be high. A sophisticated scheduling system with prediction capabilities to avoid the scheduling of out-of-town patients on Sundays would mitigate this risk. In addition, appointments on Thursdays Early AM for patients with noShow-rate between 13.7931 and 14.5161 yielded a 69% predictability of being no-shows. With Thursday being the day before the weekend, this may cause many patients to avoid coming for their appointments. Patient education may mitigate the risk of no-shows for these classes.

Appointments for patients over 51 years old with a noShow-rate \geq 12.1212 at the dental clinics yielded a prediction of no-show at 73. Dental appointments are scheduled so far in advance that for many patients, the likelihood of forgetting about the appointments is very high, which can explain this high no-show prediction.

Actively reaching out and a timely SMS reminder will mitigate this risk.

Appointments for female patients over 66 years with noShow-rate \geq 13.0435, yielded a prediction of no-show at 69%. Dependency on others to bring older women to their appointments may have a direct effect on no-show. Educating patient's relatives would lower the risk under these classes of patients. Appointments on Mondays for female patients with noShow-rates between 12.7907 and 13.7255, yielded a now show predictability of 69%. In general, women in Saudi Arabia depend on others for transportation, which have a direct effect on the no-show rate. No-shows for these classes of patients may change with the recent permission to drive granted to women.

It is worth noting that the JRip algorithm, despite its competitive performance and informative model, is not designed to deal with large sets of data. In a hospital with thousands of appointments every day like KFSHRC, data stream classification algorithms, like the Hoeffding tree algorithm, is more powerful in addressing the size issue. In addition, and as opposed to JRip, the Hoeffding tree algorithm builds the model incrementally from the data stream, allowing the generation of models that can adapt to concept-drifts. This study demonstrated the competency of the model generated by Hoeffding tree algorithm for the prediction of no-shows. The prediction of appointment no-shows can be done in real-time by integrating the Hoeffding tree algorithm into the scheduling system, so that it signals appointments that are at high risk of being a no-show. Accordingly, appropriate proactive interventions that reduce the negative impact of no-shows can be designed and integrated into the system for a timely action.

CONCLUSION

A history of multiple no-shows was the number-one predictor of future no-shows. Using data stream classification algorithms, like the Hoeffding tree algorithm, appointments having a high risk of no-show can be predicted in real-time. We recommended integrating such an algorithm in the health-information system to predict no-shows and mitigate the risk of no-show.

Acknowledgments

This study is approved by the Institutional Review Board at KFSH&RC. Special thanks to Dr. Edward B. Devol, Chairman of Biostatistics, epidemiology and Scientific Computing Department at KFSH&RC for his guidance and insightful recommendations. The authors would also thank the anonymous reviewers for their valuable and constructive comments.

REFERENCES

1. Alaeddini A, Yang K, Reddy C, Yu S. A probabilistic model for predicting the probability of no-show in hospital appointments. *Health Care Management Science*. 2011; 14(1):146157, DOI 10.1007/s10729-011-9148-9
2. Lacy NL, Paulman A, Reuter MD, Lovejoy B. Why we don't come: Patient perceptions on no-shows. *Annals of Family Medicine*. 2004; 2(6):541–545
3. Data warehouse information extract. Tech. rep., Healthcare Information Technology Affairs, King Faisal Specialist Hospital & Research Centre. 2016
4. Medical Group Management Association, Doing everything possible to prevent patient no-shows. 2017. URL <http://www.mgma.com/industry-data/polling/mgma-stat-archives/how-do-you-prevent-no-shows>
5. Downer SR, Meara JG, Costa ACD. Use of SMS text messaging to improve outpatient attendance. *The Medical Journal of Australia*. 2005;183(7):366–368
6. Hasvold PE, Wootton R. Use of telephone and SMS reminders to improve attendance at hospital appointments: a systematic review. *J Telemed Telecare*. 2011; 17(7):358–364
7. Alaeddini A, Yang K, Reeves P, Reddy CK. A hybrid prediction model for no-shows and cancellations of outpatient appointments. *IEEE Transactions on Healthcare Systems Engineering*. 2015; 5(1):14–32
8. AlMuhaideb S, Menai MEB. A new hybrid metaheuristic for medical data classification. *International Journal of Metaheuristics*. 2014; 3(1):59–80
9. Baronti F, Micheli A, Passaro A, Starita A. Machine learning contribution to solve prognostic medical problems. In: Taktak AF, Fisher AC (editors) *Outcome Prediction in Cancer*, Elsevier, Amsterdam, The Netherlands. 2007; pp 261–283
10. Elvira C, Ochoa A, Gonzalez JC, Mocho F. Machine-learning-based no show prediction in out-patient visits. *International Journal of Interactive Multimedia and Artificial Intelligence*. 2018; 4(7):29–34.
11. Topuz K, Uner H, Oztekin A, Yildirim MB. Predicting pediatric clinic no-shows: a decision analytic framework using elastic net and bayesian belief network. *Annals of Operations Research*. 2018; 263(1):479–499
12. Mohamed K, Mustafa A, Tahtamouni S, Taha E, Hassan R. A quality improvement project to reduce the 'no show' rate in a paediatric neurology clinic. *BMJ Open Quality*. 2016; 5(1), DOI 10.1136/bmjquality.2016.0003789.
13. Nelson AH, Herron D, Rees G, Nachev P. Predicting scheduled hospital attendance with artificial intelligence. *npj Digital Medicine*. 2019;2:1–7.
14. Bigby J, Giblin J, Pappius EM, Goldman L. Appointment Reminders to Reduce No-Show Rates: A Stratified Analysis of Their Cost-effectiveness. *JAMA* [Internet]. 1983 Oct 7 [cited 2019 May 12];250(13):1742–5. Available from: <https://jamanetwork.com/journals/jama/fullarticle/388161>
15. Harris SL, May JH, Vargas LG. Predictive analytics model for healthcare planning and scheduling. *European Journal of Operational Research* [Internet]. 2016 Aug 16 [cited 2019 Jul 16];253(1):121–31. Available from: <http://www.sciencedirect.com/science/article/pii/S0377221716300376>
16. Berg B, Murr M, Chermak D, Woodall J, Pignone M, Sandler R, Denton B. Estimating the cost of no-shows and evaluating the effects of mitigation strategies. *Medical Decision Making*. 2013; 33(8):976–985
17. Cohen WW. Fast effective rule induction. In: Prieditis A, Russell SJ (editors) *International Conference on Machine Learning*, Morgan Kaufmann. 1995; pp 115–123
18. Hulten G, Spencer L, Domingos P. Mining time-changing data streams. In: *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, NY, USA, KDD '01. 2001; pp 97–106
19. AlMuhaideb S, Menai M. HColonies: a new hybrid metaheuristic for medical data classification. *Applied Intelligence*. 2014; pp 1–17, URL <http://dx.doi.org/10.1007/s10489-014-0519-z>, doi:10.1007/s10489-014-0519-z
20. Huysmans J, Dejaeger K, Mues C, Vanthienen J, Baesens B. An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. *Decision Support Systems*. 2011; 5(1):141–154
21. Pazzani MJ, Mani S, Shankle WR. Acceptance of rules generated by machine learning among medical experts. *Methods of Information in Medicine*. 2001; 40(5):380–385
22. Lavaric N. Selected techniques for data mining in medicine. *Artificial Intelligence in Medicine*. 1999; 16(1):3–23
23. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*. 1997; 30(1):1145–1159
24. Hosmer JDW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*, 3rd edition. Wiley, New York. 2013
25. Stone M. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*. 1974; 36(2):111–147
26. Hanczara B, Dougherty ER. The reliability of estimated confidence intervals for classification error rates when only a single sample is available. *Pattern Recognition*. 2013; 64(3):1067–1077
27. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *International Conference on Artificial Intelligence*, Morgan Kaufmann. 1995; vol 14, pp 1137–1145, URL <http://dblp.uni-trier.de/db/conf/ijcai/ijcai95.html>
28. Furnkranz J, Widmer G. Incremental reduced error pruning. In: *International Conference on Machine Learning*. 1994. pp 70–77
29. Domingos P, Hulten G. Mining high-speed data streams. In: *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, NY, USA, KDD '00. 2000. pp 71–80
30. Hoeffding W. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*. 1963; pp 13–30
31. Maron O, Moore A. Hoeffding races: Accelerating model selection search for classification and function approximating. In: *Advances in Neural Information Processing Systems*. 1994. pp 59–66
32. Dove HG, Schneider KC. The usefulness of patients' individual characteristics in predicting no-shows in outpatient clinics. *Medical Care*. 1981; 19(7):734–740
33. Daggy JK, Lawley MA, Willis D, Thayer D, Suelzer C, DeLaurentis PC, Turkan A, Chakraborty S, Sands L. Using no-show modeling to improve clinic performance. *Health Informatics Journal*. 2010; 16(4):246–259
34. Lotfi V, Torres E. Improving an outpatient clinic utilization using decision analysis-based patient scheduling. *Socio-Economic Planning*. 2014; 48(2):115–126, DOI <http://dx.doi.org/10.1016/j.seps.2014.01.002>
35. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. Chapman & Hall, New York, NY. 1984
36. Dravenstott R, Kirchner HL, Stromblad C, Boris D, Leader J, Devapriya P. Applying predictive modeling to identify patients at risk to no-show. In: Guan Y, Liao H (editors) *Proceedings of the 2014 Industrial and Systems Engineering Research Conference*, Institute of Industrial Engineers, Inc. (IIE). 2014; pp 1–9
37. Huang Y, Hanauer D. Patient no-show predictive model development using multiple data sources for an effective overbooking approach. *Applied Clinical Informatics*. 2014; 5(3):836–860
38. Kurasawa H, Hayashi K, Fujino A, Takasugi K, Haga T, Waki K, Noguchi T, Ohe K. Machine-learning-based prediction of a missed scheduled clinical appointment by patients with diabetes. *Journal of Diabetes Science and Technology*. 2016; 10(3):730–736.