International Journal of Mathematical, Engineering and Management Sciences
Vol. 7, No. 4, 558-574, 2022
https://doi.org/10.33889/IJMEMS.2022.7.4.036

Machine Learning for Prediction of Clinical Appointment No-Shows

Jeffin Joseph
Department of Management,
Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India.
Corresponding author: jeffinjoseph@hotmail.com

S. Senith
Department of Management,
Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India.
E-mail: senith@karunya.edu

A. Alfred Kirubaraj
Department of Electronics and Communication,
Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India.
E-mail: alfred@karunya.edu

S. R. Jino Ramson
Saveetha School of Engineering, Tamil Nadu, India.
E-mail: jinoramson@gmail.com

(Received on December 16, 2021; Accepted on June 06, 2022)

Abstract
A no-show occurs when patient misses his appointment for visiting doctor in an outpatient clinic. No-shows result in inefficiencies in scheduling, capacity wastage and discontinuity in care. The study aims to develop and compare different models for predicting appointment no-shows in a hospital. The no-show estimation was made using five algorithms including Logistic Regression, Decision Tree Classifier, Random Forest, Linear Support Vector Machine and Gradient Boosting. The performance of each model is measured in terms of accuracy, specificity, precision, recall and F measure. The receiver operating characteristic curve and the precision-recall curve are obtained as further performance indicators. The result shows gradient boosting is more evident in giving consistent performance. The categorical variables used for prediction are gender, mapped age, appointment type, previous no-shows, number of previous no-shows, appointment weekday, waiting interval days, scholarship, hypertension, diabetes, alcoholism, handicap and SMS received.

Keywords- Healthcare, Machine learning, Patient no-shows, Hospital management, Predictive analytics.

1. Introduction

A "no show" occurs in a healthcare facility when a patient misses a scheduled appointment mostly for visiting doctor in an outpatient clinic. It is not easy for the healthcare systems to reschedule another patient to the time slot when a patient misses his appointment. Missed appointments are always a serious challenge to clinics and healthcare facilities all over the world. The impact of these missed appointments is significant to the revenue as well as costs of hospitals (Alaeddini et al., 2011). In the patients' health perspective, no-shows are very severe due to the adverse effects and health problems which it can bring to the patient (Schectman et al., 2008). The effect of such missed appointments is very severe in certain cases where the demand for those services is high and expensive (Daggy et al., 2010).

One of the strategies used to mitigate the impact of no-show is adjusting the service time. Some other strategies used to reduce the effect of missed appointments are introducing cancellation policies, allowing random walk-ins and implementing overbooking. Many healthcare facilities never consider the risk factors associated with no-shows and simply adopt the strategy of overbooking or random walk-ins. When all the scheduled and overbooked patients arrive, the practitioner is forced to do overtime (Huang and Zuniga, 2012). Sending reminders in the form of messages, making phone calls and sending emails are also strategies used by the clinics to reduce the no-shows (Downer et al., 2005; Ritchie et al., 2000). In order to use these strategies, the probability of no-show must be assessed. No-shows are not occurring randomly and the pattern can be predicted. The variables used for such prediction include demographic variables like gender, age, marital status and appointment characteristics like day of the week, month and lead time (Dantas et al., 2018; Peng et al., 2014).

Popular methods used for the no-show prediction described in the literature are Logistic Regression, Neural Networks, Bayesian models, Markov based models and Stacking methods (Carreras-García et al., 2020). Logistic Regression model is a widely explored model for prediction of no-shows whereas methods such as Gradient Boosting, Random Forest, Decision Tree and Linear SVM are very less explored. Thus, the comparative study, comprising of all methods is relevant for existing literature. The current study uses supervised machine learning techniques such as Logistic Regression, Decision Tree, Random Forest, Linear Support Vector Machine and Gradient Boosting to predict the appointment no-shows. The study used variables gender, mapped age, appointment type, previous no-shows, number of previous no-shows, appointment weekday, waiting interval days, scholarship, hypertension, diabetes, alcoholism, handicap and SMS received. All of the variables used in the study are categorical variables, with the majority of them being extracted by the researchers from the raw data in order to improve the current study's performance. All the models are implemented using Python with the Scikit-learn library. Accuracy, Specificity, Precision, Recall, and F measure are obtained as performance measures for each model. Additional performance indicators such as the Receiver Operating Characteristic curve and the Precision-Recall curve are generated. Majority of the related studies have not used Precision-Recall curve to assess the performance of the predictive model.

The research paper is organized as follows: initially, related studies are discussed in section 2, which is followed by the conceptual analysis of the different prediction methods in section 3. Then in section 4, the framework of the study is mentioned in detail. Following that, the variables are analyzed and new variables are extracted for better prediction in section 5. In section 6 each model's performance is assessed and the results are discussed in section 7. The limits are identified in section 8 and the conclusion is presented in section 9.

2. Related Studies

Most of the related studies use Logistic Regression to predict appointment no-shows. Logistic regression is actually solving the limitation of linear regression for categorical variables. It makes use of the maximum likelihood estimation of probability log function and is frequently used for classification problems like show or no-show (Peng et al., 2002). Predictions using Logistic Regression make better results and at the same time are easy to implement. A study conducted to predict appointment no-shows of diabetes patients in University of Tokyo hospital using Logistic Regression (Kurasawa et al., 2015) gives a high level of performance. The performance of the model is analyzed using Receiver Operating Characteristic curve (ROC) and the value of Area Under the Curve (AUC) is obtained as 0.958. The study also measured precision, recall and F measure and is found as 0.757, 0.659 and 0.704 respectively. The study only employed one method, although other approaches may have been used to improve the prediction performance.

In another study Torres et al. (2015), Multiple Logistic Regression analysis model is used to predict missed appointments and considered variables such as age, sex, race, language, insurance, marital status, clinic ZIP code, type of doctor, day of week, time of day and season for the prediction The study used demographic and appointment characteristics but not considered any clinical characteristics for prediction.

Other than the Binomial Regression, Multinomial Regression takes more than two output variables. A study by Hong and Alaeddini used Multinomial Logistic Regression with regularization to predict variables shows, no-shows, tardiness and cancellations (Hong & Alaeddini, 2017). But the study has not considered some of the important appointment characteristics such as lead time and previous no-show. Another similar study used Logistic Regression to predict the appointment no-shows in radiology examination centers (Harvey et al., 2017). The study analyzed the performance of the Logistic Regression model using Receiver Operating Characteristic curve (ROC) and Area under the Curve (AUC). More prediction approaches, such as Random Forest and Gradient Boosting, could have enhanced the study.

In a comparative study Devasahay et al. (2017) used both Logistic Regression model and Decision Tree model to predict appointment no-shows. The performance measures sensitivity and specificity of each model is obtained and is compared. The study used only two methods in prediction and also not measured performance in terms of ROC and Precision-Recall curve. In another study (Pence et al., 2018) three Logistic Regression models are compared, by plotting Receiver Operating Characteristics (ROC) curves. The study separately considered the different predictor variables such as previous missed appointments, demographic variables and clinical & psychosocial characteristics and identified their prediction capacity. In another study Elvira et al. (2018) carried out outpatient no-show prediction by building model based on Gradient Boosting with an Area under the Curve (AUC) of 0.74.

In another study Mohammadi et al. (2018) built three machine learning models to predict clinical appointment no-shows. Logistic Regression, Multilayer Perceptron, and Naïve Bayes classifier models are used with tenfold cross validation. The performance measures are compared and find that the method Naïve Bayes is better than other two methods based on Accuracy and AUC. The study may have used performance measures including precision-recall curve to analyze the performance. Another study Srinivas & Ravindran (2018) proposed a framework with five different machine learning algorithms Logistic Regression, Neural Network, Random Forest, Gradient Boosting and Stacking in the study for optimizing appointment scheduling system. Performance measures such as Accuracy, Precision, F measure and Area Under the Curve (AUC) are used to evaluate the best machine learning algorithm.

Another study AlMuhaideb et al. (2019) used Decision Trees, Hoeffding and JRip for predicting patient no-shows. The Receiver Operating Curve is drawn and the AUC for JRip is obtained as 0.776 and for Hoeffding it is 0. 861. According to a study Ahmadi et al. (2019) predicted appointment no-shows and late cancellations in two stages. In first stage used the prediction techniques Decision Tree, Random Forest and Naïve Bayes and in the second stage applied the Stacking method. In another study Nelson et al. (2019) for predicting attendance applied Logistic Regression, Random Forest, Support Vector Machine and AdaBoost algorithms and the performance is measured using Receiver Operating Characteristic curve (ROC) and Area Under the curve (AUC). The performance of the models is also measured using Precision-Recall curve and Average Precision. The study included data from two hospitals, which might be reduced to one hospital to improve the model's performance. In another research Alshaya (2019) proposed a framework by using different machine learning algorithms such as Logistic Regression, Linear Support Vector Machine, Stochastic Gradient Descent, K-neighbor and Random Forest for prediction. The study used various balancing techniques, and dimensionality reduction methods. It tackles the problem of data imbalance to avoid the bias in the trained models.

In another study Nasir et al. (2020) used Logistic Regression, Support Vector Machine, Random Forest and Artificial Neural Network for predicting no-shows. The performance measures such as Sensitivity, Specificity, Accuracy and AUC are analyzed and compared to find the best model. In a study Daghistani et al. (2020) analyzed the performance of five machine learning algorithms to predict outpatient's appointment no-shows. The machine learning techniques employed in the study are Logistic Regression, Random Forest, Linear Support Vector Machine, Gradient Boosting and Multilayer Perceptron and the best model is selected based on Accuracy, Precision, Recall, AUC and F-measure. Both the studies may also have considered Precision-Recall curve for analyzing the model's performance. According to another research Fan et al. (2021) used Logistic Regression, K-nearest neighbor, Boosting, Decision tree, Bagging and Random Forest to design prediction models for patient no-show in online outpatient appointments. The researchers focused on online appointments and discovered that the bagging method provides more accurate predictions. The Precision-Recall curve and average precision may have been used in the study to assess each model's performance.

In another study Dashtban & Li (2021) analyzed no-show using Naïve Bayes, K-nearest neighbor, Decision tree, Random Forest and Logistic Regression. The study also used deep learning sparse stacked denoising autoencoders to predict no-show and in the study the method logistic regression proves more performing. Thus; the related studies give a glimpse to the theoretical background and provide empirical evidence to use of machine learning in predicting clinical appointment no-show. Most of the previous literature used Logistic Regression and the current study will add more evidence to the use of other machine learning methods. Other than many of the related studies, the current study used only categorical variables to improve the prediction performance. A new categorical appointment variable, 'number of previous no-shows,' have been added to the research. The study also used the Precision-Recall curve as a performance measure, which many previous studies had ignored when evaluating the model's performance.

3. Approaches for Prediction

Machine learning algorithms automatically learn patterns to make accurate predictions from past observations. Supervised machine learning particularly predicts the output based on evidence. It learns a function on sample input-output pairs, which is further used to map an input to an output (Azuaje, 2006). The training phase of the supervised algorithms uses the inputs and the actual outputs to learn and uncover the relationship between them, and the testing phase uses the trained algorithm to make predictions (Sarker, 2021). To predict missed appointments, five approaches employed which make use of five different logics. They are logistic regression, decision tree, random forest, linear support vector machine and gradient boosting. linear regression is a form of regression that's used to predict binary and categorical outcomes like appointment show and no-show in the most efficient way possible (Kheirkhah et al., 2015). One of the simplest and most widely used machine learning approach for forecasting missed appointment is logistic regression. Binomial logistic regression is used to estimate the likelihood of a dependent variable with two alternative outcomes such as show and no-show (Maalouf, 2011).

The Logistic Regression model uses independent variables to predict the dependent categorical variable whereas the Decision Tree models, take a completely different approach to decision logic to classify data objects into a tree-like structure. A decision tree is similar like a normal tree with roots, branches and leaves. It is a decision support tool that uses a tree-like model of decision-making process and its possible consequences (Patel and Prajapati, 2018). It consists of decision nodes that are connected by branches, starting from a node called the root node and terminates at a node which is called the leaf node, with an outcome. A mathematical algorithm is used to select a feature with a corresponding threshold at each decision node in order to separate the observation into two or more subgroups. For classification problems, the Decision tree model can be a powerful machine learning algorithm. (Osisanwo et al., 2017).

Random forest and gradient boosting are two ensemble methods for improving base model prediction. The averaging ensemble method is used in Random Forest, whereas the boosting method is used in Gradient boosting. The Random Forest (Breiman, 2001) generates numerous decision trees at random in order to produce accurate and consistent results. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features (Kundu and Suranjan Das, 2019). Gradient Boosting employs weak learners or decision trees, as well as an additive model that uses an iterative and systematic method for adding trees one at a time.

On the other hand, the Linear SVM method seeks out a hyper plane or border between data classes that maximize the margin between them (Patnaik et al., 2020). The data in a Support Vector Machine is plotted as points in n-dimensional space. Prediction of show and no-show is done by finding the hyper plane that separates the two sets. SVM selects the most extreme cases to contribute in the creation of the hyper plane (Kotsiantis et al., 2006). Support vectors are the extreme instances, and hence the algorithm is termed as Support Vector Machine.

All the five machine learning approaches for prediction are implemented through the Scikit-learn library in python. The Scikit-learn is the Machine Learning library which consists of tools for regression, classification, dimension reduction and clustering. Several other Scikit libraries such as Pandas, NumPy and Matplotlib are also used by the current study in order to acquire and analyze the data.

4. Framework of Study

The framework of the study shown in Figure 1 gives a detailed description of the entire study. The raw dataset is cleaned and pre-processed for the analysis. Data pre-processing is done in order to convert the messy and noisy real data into quality data. In the first stage of pre-processing the data is checked for duplicates. The cleansed data is then used to extract new variables from the existing variables in order to improve prediction accuracy.

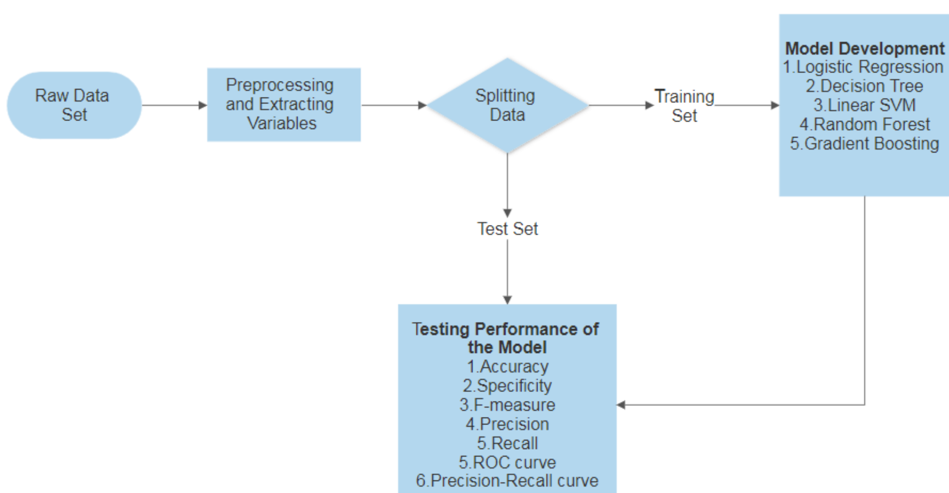

Figure 1. Framework of study.

The pre-processed and extracted data is then divided into training set and test set. In the model generation process, the program learns from an initial set of data which is usually called as the training set and the model is developed from this training set. The prediction is made on another set which is usually called as the test set. Here for the problem presented in the current study, 80% of the total data is used as training set and 20% of data is used as test set (GENÇ & TUNÇ, 2019). There are thirteen input variables considered for the study. The output variable is patient attendance which has two values show and no-show. The machine learning models Logistic Regression, Decision tree, Linear SVM, Random Forest and Gradient Boosting are implemented to determine the missed appointments. The trained models are then tested using the test set. Each model is analyzed and compared for performance using indicators such as Accuracy, Specificity, F-measure, Precision, Recall and Receiver Operating Characteristic Curve. Precision-Recall curve is also drawn and the average precision is determined.

5. Variables for Prediction

The study extracted appointment data from a public hospital in Brazil. The data consisted of 62299 patients and their 110527 appointments which is a large data for implementing machine learning techniques. The raw data obtained from the hospital database consisted of variables such as patient id, appointment id, gender, scheduled day, appointment day, age, neighborhood, scholarship, diabetes, alcoholism, handicap, SMS received and no-show. Out of the 110527 appointments 88208 are shows and 22319 are no-shows. Missed appointments or no-shows account for 20 percent of the total appointments considered for the study. New variables such as waiting interval days, appointment week days, mapped age, appointment type, previous no- shows and number of previous no-shows are extracted from the existing variables for better prediction results. The variable neighborhood is excluded since it decreases the performance of models. All the variables are categorized in order to improve the performance of Models. Most of the prediction models in the related studies used variables such as age, gender, and appointment type, previous no-shows, appointment week day and lead time (Carreras-Garcia et al., 2020). In current study variables 'previous no-shows' and 'number of previous no-shows' are considered separately.

5.1 Appointment Week Days

A new variable extracted from the raw data is the appointment week days. The appointment week days are extracted from the appointment day data. Figure 2 describes the distribution of appointment no-shows based on appointment weekdays. The bar chart represents the number of appointments by patients on each week day and the line chart shows the number of patient no-shows on that day of the week.

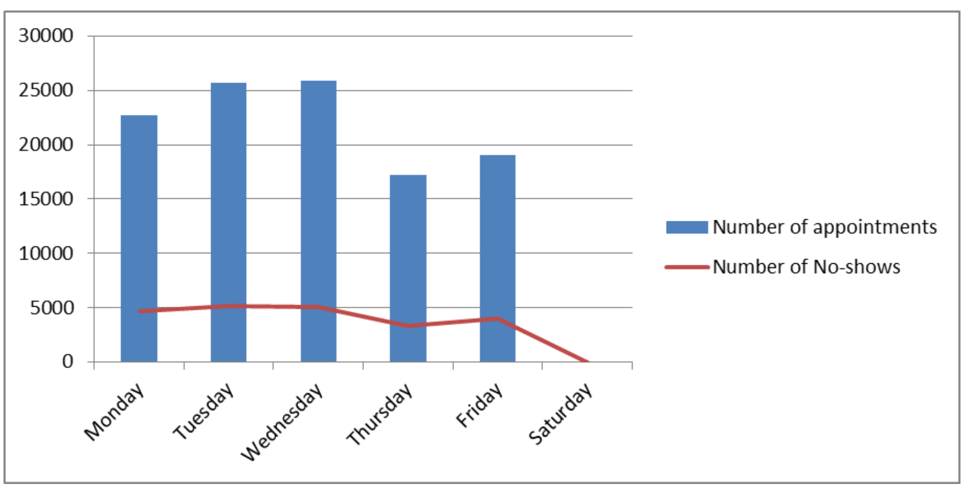

Figure 2. Distribution of no-show based on appointment week days.

5.2 Mapped Age

The variable age is categorized into five child, teenager, and young adult, adult and senior. The data categorization is done in order to improve the performance of the prediction models. The distribution of the no-show based on the mapped age is given below in the chart in Figure 3. Bars represent the number of appointments and lines represent the number of no-shows.

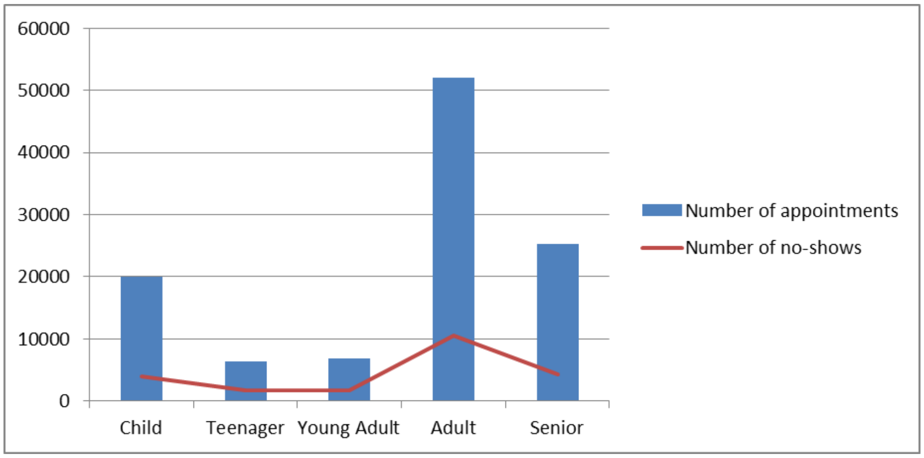

Figure 3. Distribution of no-show based on mapped age.

5.3 Waiting Interval Days

Waiting interval days is one of the new variables extracted from the raw data. The variable waiting interval days is calculated from the variables scheduled day and appointment day. Scheduled day is the day on which the appointment is booked and appointment day is the actual date of appointment on which the patient is supposed to visit the hospital. The waiting interval days are the difference between scheduled day and appointment day. Further the variable waiting interval days is categorized into five including Same day, Within two days, Within a week, Within a month and more than a month. Figure 4 represents the distribution of no-show on waiting interval days. The variable waiting interval days is plotted on the X axis. Number of patient appointments and number of no-shows are represented on the Y axis.

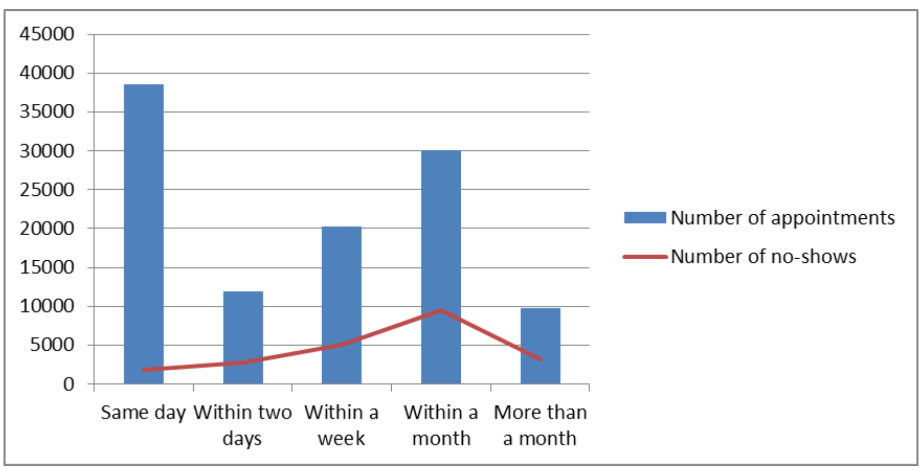

Figure 4. Distribution of no-show based on waiting interval days.

Another variable derived from the appointment data is the appointment type which determines whether the appointment is a fresh appointment or a repeated appointment. Previous no-show describes whether the patient has got any previous no-show history or not. And the number of previous no-shows gives the total number of missed appointments of a particular patient. Both these variables are extracted from the no-show data. Number of previous no-shows is again categorized into no previous no-shows, less than two previous no-shows, between three and ten no-shows, and more than ten no-shows.

The other variables used for prediction are diabetics, hypertension, alcoholism, scholarship and handicap. The variable handicap has got four types. The numbers 1 to 4 indicate number of disabilities of the patient. Thus, the variables used in the study are gender, mapped age, appointment type, previous no-shows, number of previous no-shows, appointment weekday, waiting interval days, SMS received, scholarship, hypertension, alcoholism, diabetics and handicap. The above variables are used with the different classification algorithms to develop the models in predicting patient appointment no-show. Most of the variables used in the study are previously used by many related studies to predict clinical appointment no-shows. A descriptive analysis of the variables towards no-show is given below in the Table 1.

6. Performance Measures

The confusion matrix summarizes the performance of the prediction model. It compares the actual values with the values those predicted by machine learning models. First the case of Decision tree model, as the Figure 5 demonstrates, out of total predicted no-shows 3888 are true positives and are no-shows and 948 are false positives and are show-ups. And out of the total predicted shows in the Decision tree model 16690 are true negatives and 580 are false negatives and are no-shows. Confusion matrix for Logistic Regression model is also shown in the Figure 5 which explains out of the predicted no-shows 3886 are true positives and are no-shows. And the 886 are false positives and end in show-up. And out of the total predicted shows 582 become false negative and are no-shows and 16752 are true negative and are show-ups.

Table 1. Descriptive analysis of variables to show/no-show.

Variables		Total	Show	No show
Gender	Male	38687	30962	7725
	Female	71840	57246	14594
Mapped Age	Child	19945	15910	4035
	Teenager	6343	4653	1690
	Young Adult	6856	5131	1725
	Adult	52086	41475	10611
	Senior	25297	21039	4258
Appointment type	Fresh	62299	50106	12193
	Repeat	48228	38102	10126
Previous no-show	Yes	33697	11378	22319
	No	76830	76830	
Number of previous no-show	No	76830	76830	
	Less than 2 No-shows	31408	10519	20889
	Between 3 &10 No-shows	2245	847	1398
	More than 10 No-shows	44	12	32
Appointment weekday	Monday	22715	18025	4690
	Tuesday	25640	20488	5152
	Wednesday	25867	20774	5093
	Thursday	17247	13909	3338

Table 1. *Continued*

	Friday	19019	14982	4037
	Saturday	39	30	9
Waiting interval days	Same day	38563	36771	1792
	Within two days	11938	9223	2715
	Within a week	20247	15190	5057
	Within a month	30068	20523	9545
	More than a month	9711	6501	3210
SMS received	Yes	35482	25698	9784
	No	75045	62510	12535
Scholarship	Yes	10861	8283	2578
	No	99666	79925	19741
Hypertension	Yes	21801	18029	3772
	No	88726	70179	18547
Alcoholism	Yes	3360	2683	677
	No	107167	85525	21642
Diabetics	Yes	7943	6513	1430
	No	102584	81695	20889
Handicap	0	108286	86374	21912
	1	2042	1676	366
	2	183	146	37
	3	13	10	3
	4	3	2	1

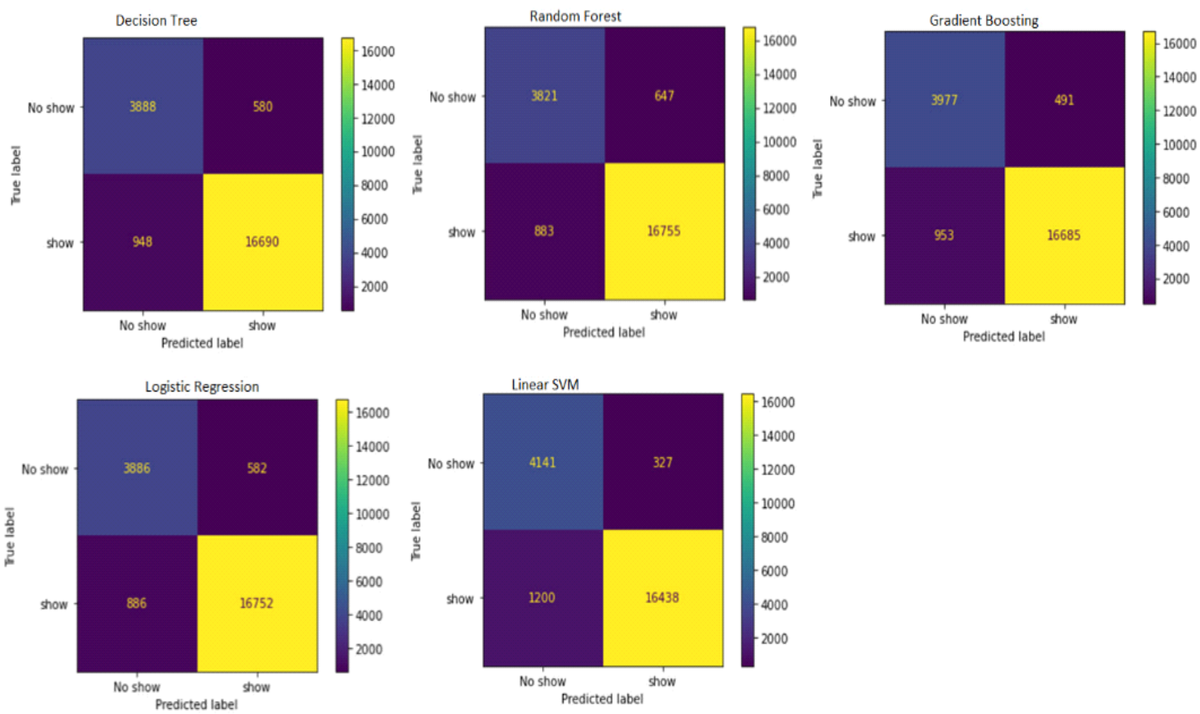

Figure 5. Confusion matrix for different machine learning models.

In the confusion matrix of random forest algorithm among the predicted no-shows 3821 are true positive and are no-shows but 883 are false positives and are shows in the actual values. And out of the predicted shows 16755 are true negative and 647 are false negative and are no-shows.

In the confusion matrix for Linear SVM out of the predicted no-shows 4141 are true positive and are no-shows and 1200 are false positives and are show-ups. Out of the predicted shows in Linear SVM 16438 are true negatives and 327 are false negatives. And in the confusion matrix of Gradient Boosting classifier algorithm, out of the predicted no-shows 3977 are true positives and 953 are false negatives and among the predicted appointment shows 16685 are true negatives and 491 are false negatives and are no-shows.

A few performance measures, including accuracy, specificity, and sensitivity, can be evaluated using the confusion matrix. A test's accuracy is determined by its ability to appropriately distinguish between no-shows and show-ups. We should determine the percent of true positive and true negative in all analyzed cases to measure a test's accuracy (Skaik, 2008). This can be expressed mathematically as:

$$\text{Accuracy}=TP + TN/P + N \qquad (1)$$

where, N is number of show-ups and P is the number of no-shows. TN is the number of patients who were previously predicted as show-ups and TP is the number of patients who were previously predicted as no-shows.

Accuracy is not a reliable measure especially in cases with imbalance problems. When there is a disparity in positive and negative values, it is called an imbalance problem (Chawla, Japkowicz and Kotcz, 2004). In the current study there exist imbalance problems, the number of patient appointment no-shows are very less compared to the patients show-up for appointments. And as a result, the model requires more dependable measures to determine the performance. A test's sensitivity refers to its capacity to appropriately identify patient appointment no-show (Skaik, 2008). In order to determine that the study calculate the fraction of true positive in no- shows. This can be expressed as:

$$\text{Sensitivity} = TP/P \qquad (2)$$

where, P is the number of no-shows and TP is the number of patients correctly classified as no-shows.

The capacity of a test to correctly determine appointment show-up is its specificity (Skaik, 2008). In order to estimate that we calculate the fraction of true negatives in show-up. This can be expressed mathematically as:

$$\text{Specificity} = TN/N \qquad (3)$$

Precision is the fraction of actual no-shows predicted (true positive) to the sum of correctly predicted no-shows and correctly predicted shows (true positive plus true negatives). The measure of precision is mathematically expressed as:

$$\text{Precision} = TP/TP + TN \qquad (4)$$

Another performance measure is the F-measure. F-measure assesses how well precision and recall are harmonized. When F-measure is high, it indicates that both precision and recall are high and a lower F-measure indicates high precision-to-recall imbalance (Skaik, 2008). F-measure score is expressed as:

F measure = 2*Precision*Recall/Precision + Recall (5)

The performance measures accuracy, sensitivity (recall), specificity, precision and recall for each model are evaluated and is are given below in Table 2.

Table 2. Performance measures of the models.

Model	Logistic regression	Decision tree	LSVM	Random forest	Gradient Boosting
Accuracy	0.93	0.93	0.93	0.93	0.93
Sensitivity (Recall)	0.87	0.87	0.93	0.86	0.89
Specificity	0.95	0.95	0.93	0.95	0.95
Precision	0.81	0.80	0.78	0.81	0.81
F measure	0.84	0.84	0.84	0.83	0.85

6.1 Precision-Recall Curve

The Precision-Recall curve is a novel way to assess a predictive model's performance in predicting appointment no-shows (Ozenne et al., 2015). It depicts the trade-off between precision and recall (also known as sensitivity) over all potential threshold values. The Precision-Recall curve focuses on the predictive model's ability to detect no-show, ignoring the correctly anticipated show-up (true negatives), which are the majority. Precision-Recall plot is more useful when evaluating predictive models with imbalanced data (Saito and Rehmsmeier, 2015). The Precision-Recall curve makes it simple determining which operating point the precision and recall are high.

Average precision (AP) summarizes the precision recall curve into a single value. It is the weighted sum of precisions at each threshold with the weight is the increase in recall from the prior threshold. Average precision and the area under the precision-recall curve are common ways to evaluate a Precision-Recall curve.

The Precision-Recall curves for the five models are shown in Figure 6. The average precision obtained for each model is displayed on the chart. The AP values for the model's Logistic Regression, Decision tree, Random Forest, Linear SVM and Gradient Boosting are 0.94, 0.91, 0.93, 0.94, and 0.94 respectively.

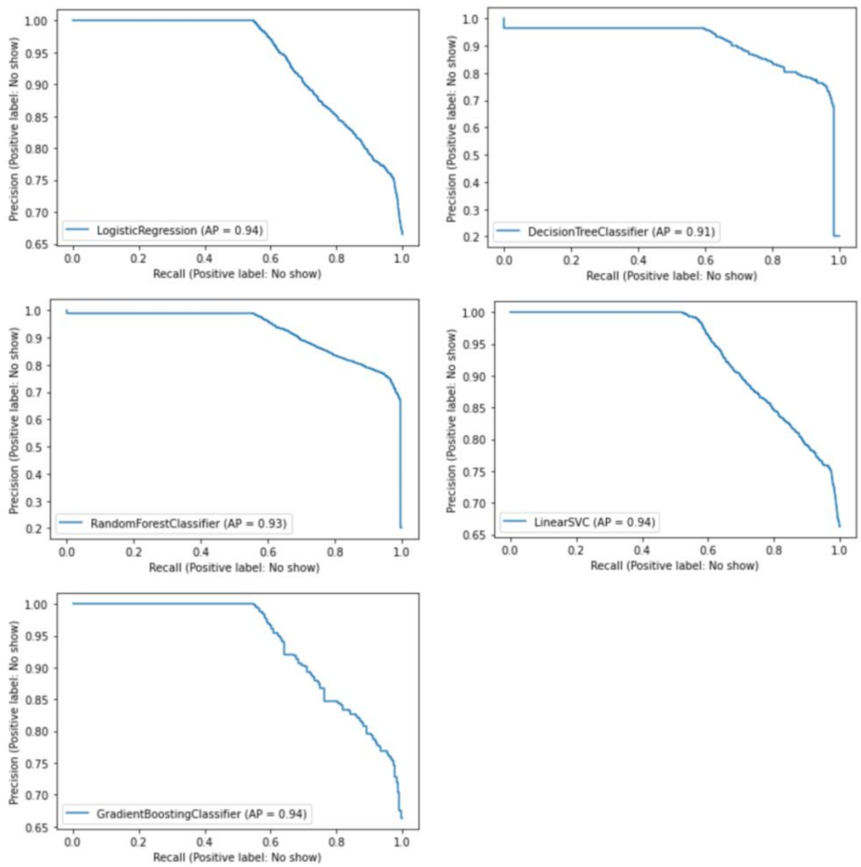

Figure 6. Precision -Recall Curve for different models.

6.2 Receiver Operating Characteristic Curve

The Receiver Operating Characteristic (ROC) curve is a graph that illustrates how well a prediction model performs at various thresholds. The Area under the curve (AUC) indicates to what extend the predictive model distinguishes between show and no-show. Higher AUC shows how better the model differentiates between patients who are showing up and those who don't. Receiver Operating Characteristic curve is formed by plotting FPR (false positive rate) on X-axis and TPR (true positive rate) on Y-axis (Bradley, 1997).

$$TPR = TP/TP + FN \tag{6}$$

$$FPR = FP/TN + FP \tag{7}$$

TPR is sensitivity which shows the correct number of predicted no-show among all actual no- show. FPR is the fraction of incorrect number of predicted show-up to all actual show-up. In short on ROC curve, it is sensitivity on the Y-axis and (1- specificity) on the X-axis.

The AUC value ranges from 0 to 1, with an AUC of value 0.5 indicating no discrimination (i.e., ability to distinguish the difference between show-ups and no-shows), 0.7 and 0.8 are considered acceptable, 0.8 and 0.9 are great, and more than 0.9 is remarkable (Mandrekar, 2010). The receiver operating characteristic curve for the five predictive models' Linear Regression, Decision Tree, Random Forest, LSVM and Gradient Boosting are drawn on a single chart (Figure 7) for the intent of comparison. The AUC values for all the approaches are found remarkable and are 0.98 for Logistic Regression, Random Forest, Linear SVC and Gradient Boosting and is 0.97 for Decision tree model.

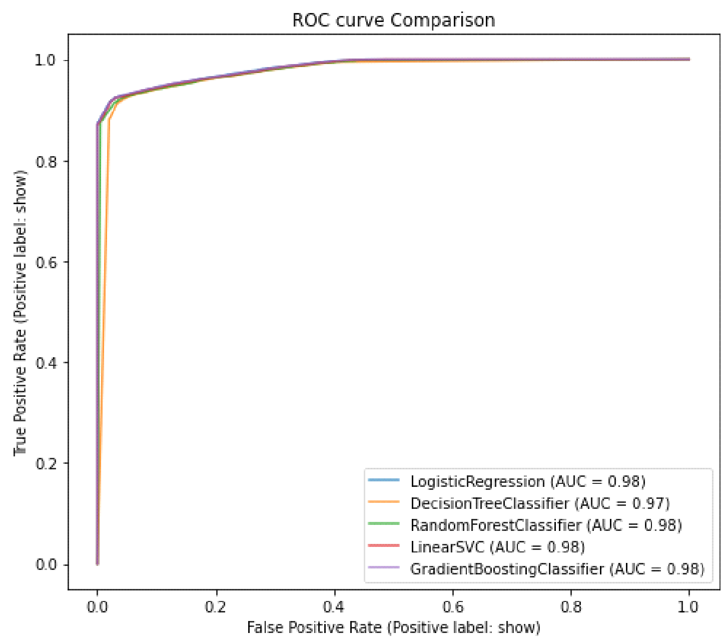

Figure 7. Receiver operating characteristic curve comparison.

7. Discussion of Results

Logistic Regression, Decision Tree model, Random Forest classifier, Linear SVM and Gradient Boosting are the models which are developed and implemented. The five models are compared based on the performance measures accuracy, sensitivity, specificity, precision, and F measure. The performance measure values of the five models are given in table 2. It is identified that all models performed well and the accuracy for all the models are around 0.93. In a comparable study (Alshaya, 2019), the accuracy of all major methods employed to predict appointment no-show, including Logistic Regression, Random Forest, and Support Vector Machine, was found to be 0.68. In comparison to most other models in the literature, the models in this study are more accurate. The precision is measured high 0.81 for Logistic Regression, Random Forest and Gradient Boosting. The recall or sensitivity is high for Linear SVM and is 0.93. Gradient Boosting has a sensitivity of 0.89; Logistic Regression and Decision Tree have a sensitivity of 0.87, while Random Forest has a sensitivity of 0.86. Figure 7 shows the Receiver Operating Characteristic curve (ROC) and the Area under the Curve (AUC) for each of the five models. Linear Regression, Random Forest, Support Vector Machine, and Gradient Boosting all have a high Area under the Curve of 0.98. The AUC for the decision tree model is 0.97. The AUC is high when compared to the AUCs observed for most of the prediction models in the literature. The AUC obtained for a study (Nasir et al., 2020) was 0.8678 for its highest performing model Random Forest. In another study (Daghistani et al., 2020) to predict appointment no-shows, the highest performing model Gradient Boosting has an Area under the Curve of 0.81.

The Precision-Recall curve is constructed for all models, and the average precision is calculated, as illustrated in Figure 6. Linear SVM, Gradient Boosting, and Logistic Regression models have an average precision of 0.94, while Random Forest and Decision Tree have an average precision of 0.93 and 0.91, respectively. From the performance indicators it is clear that the predictions made by different models are accurate, specific and robust. The performance metrics established in this study are higher than the performance measures for no-show predictions found in the literature. Out of the methods employed in the study Gradient Boosting is giving high value for all performance indicators. Demographic variables, appointment features, and clinical characteristics such as handicap and alcoholism were all employed in the study. The study also considered history of no-shows as two variables, previous no-shows and number of previous no-shows. Also, the predictive models' high-performance metric demonstrates the significance of the variables applied for prediction. The Precision-Recall curve and average precision are also used to assess performance in the current study.

8. Limitations

The current study comes with certain limitations. First the proposed prediction models are evaluated with outpatient appointment data from one public hospital in Brazil. The study is confined to only one clinic and hence the findings cannot be generalized. The model needs to be evaluated with other hospitals in other parts of the world with different characteristics and demographics. Secondly the study used only thirteen variables gender, mapped age, appointment type, previous no-shows, number of previous no-shows, appointment weekday, waiting interval days, scholarship, hypertension, diabetes, alcoholism, handicap and SMS received to predict the clinical appointment no-shows. The study can be done by including more demographic and environmental variables like race, ethnicity, marital status and transportation mode. Thirdly the study has assumed missed appointments and cancellations as no-shows and the cancellations are not separately taken into account. The study can be modified if the missed appointments and cancellations are considered separately.

9. Conclusion

The study designed five models that can predict missed appointments of patients in a clinic with high accuracy. Different performance measures such as Accuracy, Specificity, Precision, Recall, F-measure; Receiver Operating Characteristic Curve and Precision-Recall curve are used to evaluate the performance of different models. The performance measures of different models are compared and evaluated. The study used categorized variables gender, mapped age, appointment type, previous no-shows, number of previous no-shows, appointment weekday, waiting interval days, scholarship, hypertension, diabetes, alcoholism, handicap and SMS received. The study used a new categorical appointment variable number of previous no-shows as a predictor in the study. In addition, the study used Precision-Recall curve and average precision for performance analysis. Also, the predictive models' high-performance metric demonstrates the significance of the variables for prediction applied in this research. It is evident from the study that Gradient Boosting is showing high values for all performance measures employed. The practical application of predictive models will undoubtedly aid in the development of a more targeted scheduling system for healthcare institutions, particularly outpatient clinics. The study's findings assist medical practitioners and clinics in predicting appointment no-shows, which is crucial to clinic efficiency. Clinics can adopt strategies like overbooking, adoption of variable service timings and allowing random walk-ins based on the predictions obtained from the models (Molfenter, 2013). Further research can be done to implement methods to reduce the class imbalance problem. Also, cross validations and feature selections can be performed to understand the significance of the variables used for prediction. Again, there is scope for further research if new variables are derived from the existing variables. Next plan of the researchers is to develop an efficient scheduling system for the outpatient clinic of a hospital, incorporating the predictive models.

Conflict of Interest

The authors declare that they have no conflict of interest regarding the publication of this article. All authors have checked and agreed to the submission.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors would like to thank the editor and anonymous reviewers for their comments that help improve the quality of this work.

References

Ahmadi, E., Garcia-Arce, A., Masel, D.T., Reich, E., Puckey, J., & Maff, R. (2019). A metaheuristic-based stacking model for predicting the risk of patient no-show and late cancellation for neurology appointments. *IISE Transactions on Healthcare Systems Engineering*, *9*(3), 272-291. https://doi.org/10.1080/24725579.2019.1649764.

Alaeddini, A., Yang, K., Reddy, C., & Yu, S. (2011). A probabilistic model for predicting the probability of no-show in hospital appointments. *Health Care Management Science*, *14*(2), 146-157. https://doi.org/10.1007/s10729-011-9148-9.

AlMuhaideb, S., Alswailem, O., Alsubaie, N., Ferwana, I., & Alnajem, A. (2019). Prediction of hospital no-show appointments through artificial intelligence algorithms. *Annals of Saudi Medicine*, *39*(6), 373-381. https://doi.org/10.5144/0256-4947.2019.373.

Alshaya, S. (2019, December 10). *Predicting No-show Medical Appointments Using Machine Learning*. SpringerLink. https://link.springer.com/chapter/10.1007/978-3-030-36365-9_18?error=cookies_not_supported&code=650bf76c-afdc-4ad0-850c-2cc486b22a2d.

Azuaje, F. (2006). Witten IH, Frank E: Data mining: practical machine learning tools and techniques 2nd edition. *BioMedical Engineering OnLine*, *5*(1). https://doi.org/10.1186/1475-925x-5-51.

Bradley, A.P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, *30*(7), 1145-1159. https://doi.org/10.1016/s0031-3203 (96)00142-2.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32. https://doi.org/10.1023/ a:1010933404324.

Carreras-García, D., Delgado-Gómez, D., Llorente-Fernández, F., & Arribas-Gil, A. (2020). Patient no-show prediction: a systematic literature review. *Entropy*, *22*(6), 675. https://doi.org/10. 3390/e22060675.

Chawla, N.V., Japkowicz, N., & Kotcz, A. (2004). Editorial. *ACM SIGKDD Explorations Newsletter*, *6*(1), 1-6. https://doi.org/10.1145/1007730.1007733.

Daggy, J., Lawley, M., Willis, D., Thayer, D., Suelzer, C., DeLaurentis, P. C., Turkcan, A., Chakraborty, S., & Sands, L. (2010). Using no-show modeling to improve clinic performance. *Health Informatics Journal*, *16*(4), 246-259. https://doi.org/10.1177/1460458210380521.

Daghistani, T., AlGhamdi, H., Alshammari, R., & AlHazme, R.H. (2020). Predictors of outpatients' no-show: big data analytics using apache spark. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00384-9.

Dantas, L.F., Fleck, J.L., Cyrino Oliveira, F.L., & Hamacher, S. (2018). No-shows in appointment scheduling - a systematic literature review. *Health Policy*, *122*(4), 412-421. https://doi.org/10.1016/ j.healthpol.2018.02.002.

Dashtban, M., & Li, W. (2021). Predicting non-attendance in hospital outpatient appointments using deep learning approach. *Health Systems*, 1-22. https://doi.org/10.1080/20476965.2021.1924085.

Devasahay, S.R., Karpagam, S., & Ma, N.L. (2017). Predicting appointment misses in hospitals using data analytics. *MHealth*, *3*, 12. https://doi.org/10.21037/mhealth.2017.03.03.

Downer, S.R., Meara, J.G., & da Costa, A.C. (2005). Use of SMS text messaging to improve outpatient attendance. *Medical Journal of Australia*, *183*(7), 366-368. https://doi.org/10.5694/j.1326-5377.2005.tb07085.x.

Elvira, C., Ochoa, A., Gonzalvez, J.C., & Mochon, F. (2018). Machine-learning-based no show prediction in outpatient visits. *International Journal of Interactive Multimedia and Artificial Intelligence*, *4*(7), 29. https://doi.org/10.9781/ijimai.2017.03.004.

Fan, G., Deng, Z., Ye, Q., & Wang, B. (2021). Machine learning-based prediction models for patients no-show in online outpatient appointments. *Data Science and Management*, *2*, 45-52. https://doi.org/10.1016/j.dsm.2021.06.002.

Genç, B., & Tunç, H. (2019). Optimal training and test sets design for machine learning. *Turkish Journal of Electrical Engineering & Computer Sciences*, 1534-1545. https://doi.org/10.3906/elk-1807-212.

Harvey, H.B., Liu, C., Ai, J., Jaworsky, C., Guerrier, C.E., Flores, E., & Pianykh, O. (2017). Predicting no-shows in radiology using regression modeling of data available in the electronic medical record. *Journal of the American College of Radiology*, *14*(10), 1303-1309. https://doi.org/10.1016/ j.jacr.2017.05.007.

Hong, S.H., & Alaeddini, A. (2017). A multi-way multi-task learning approach for multinomial logistic regression. *Methods of Information in Medicine*, *56*(04), 294-307. https://doi.org/10.3414/me16-01-0112.

Huang, Y., & Zuniga, P. (2012). Dynamic overbooking scheduling system to improve patient access. *Journal of the Operational Research Society*, *63*(6), 810-820. https://doi.org/10.1057/jors.2011.83.

Kheirkhah, P., Feng, Q., Travis, L.M., Tavakoli-Tabasi, S., & Sharafkhaneh, A. (2015). Prevalence, predictors and economic consequences of no-shows. *BMC Health Services Research*, *16*(1). https://doi.org/10.1186/s12913-015-1243-z.

Kotsiantis, S.B., Zaharakis, I.D., & Pintelas, P.E. (2006). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, *26*(3), 159-190. https://doi.org/10.1007/s10462-007-9052-3.

Kundu, M., & Suranjan Das, M. (2019). Predicting autism spectrum disorder in infants using machine learning. *Journal of Physics: Conference Series*, *1362*, 012018. https://doi.org/10.1088/1742-6596/1362/1/012018.

Kurasawa, H., Hayashi, K., Fujino, A., Takasugi, K., Haga, T., Waki, K., Noguchi, T., & Ohe, K. (2015). Machine-learning-based prediction of a missed scheduled clinical appointment by patients with diabetes. *Journal of Diabetes Science and Technology*, *10*(3), 730-736. https://doi.org/10.1177/1932296815614866.

Maalouf, M. (2011). Logistic regression in data analysis: an overview. *International Journal of Data Analysis Techniques and Strategies*, *3*(3), 281. https://doi.org/10.1504/ijdats.2011.041335.

Mandrekar, J.N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, *5*(9), 1315-1316. https://doi.org/10.1097/jto.0b013e3181ec173d.

Mohammadi, I., Wu, H., Turkcan, A., Toscos, T., & Doebbeling, B. N. (2018). Data analytics and modeling for appointment no-show in community health centers. *Journal of Primary Care & Community Health*, *9*, 215013271881169. https://doi.org/10.1177/2150132718811692.

Molfenter, T. (2013). Reducing appointment no-shows: going from theory to practice. *Substance Use & Misuse*, *48*(9), 743-749. https://doi.org/10.3109/10826084.2013.787098.

Nasir, M., Summerfield, N., Dag, A., & Oztekin, A. (2020). A service analytic approach to studying patient no-shows. *Service Business*, *14*(2), 287-313. https://doi.org/10.1007/s11628-020-00415-8.

Nelson, A., Herron, D., Rees, G., & Nachev, P. (2019). Predicting scheduled hospital attendance with artificial intelligence. *Npj Digital Medicine*, *2*(1). https://doi.org/10.1038/s41746-019-0103-3.

Osisanwo F.Y., Akinsola J.E.T., Awodele O., Hinmikaiye J.O., Olakanmi O. and Akinjobi J. (2017). Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology*, *48*(3), 128-138 https://doi.org/10.14445/22312803/ijctt-v48p126.

Ozenne, B., Subtil, F., & Maucort-Boulch, D. (2015). The precision-recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *Journal of Clinical Epidemiology*, *68*(8), 855-859. https://doi.org/10.1016/j.jclinepi.2015.02.010.

Patel, H.H., & Prajapati, P. (2018). Study and analysis of decision tree based classification algorithms. *International Journal of Computer Sciences and Engineering*, *6*(10), 74-78. https://doi.org/10.26438/ ijcse/v6i10.7478.

Patnaik, S., Yang, X., & Sethi, I.K. (2020). *Advances in machine learning and computational intelligence: proceedings of ICMLCI 2019 (Algorithms for Intelligent Systems)* (1st ed. 2021 Ed.). Springer.

Pence, B.W., Bengtson, A.M., Boswell, S., Christopoulos, K.A., Crane, H.M., Geng, E., Keruly, J.C., Mathews, W.C., & Mugavero, M.J. (2018). Who will show? predicting missed visits among patients in routine hiv primary care in the United States. *AIDS and Behavior, 23*(2), 418-426. https://doi.org/10.1007/s10461-018-2215-1.

Peng, C.Y.J., Lee, K.L., & Ingersoll, G.M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, *96*(1), 3-14. https://doi.org/10.1080/ 00220670209598786.

Peng, Y., Erdem, E., Shi, J., Masek, C., & Woodbridge, P. (2014). Large-scale assessment of missed opportunity risks in a complex hospital setting. *Informatics for Health and Social Care*, 1-16. https://doi.org/10.3109/17538157.2014.965303.

Ritchie, P.D., Jenkins, M., & Cameron, P.A. (2000). A telephone call reminder to improve outpatient attendance in patients referred from the emergency department: a randomised controlled trial. *Australian and New Zealand Journal of Medicine*, *30*(5), 585-592. https://doi.org/10.1111/j.1445-5994.2000.tb00860.x.

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, *10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432.

Sarker, I.H. (2021). Machine learning: algorithms, real-world applications and research directions. *SN Computer Science*, *2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x.

Schectman, J.M., Schorling, J.B., & Voss, J.D. (2008). Appointment adherence and disparities in outcomes among patients with diabetes. *Journal of General Internal Medicine*, *23*(10), 1685-1687. https://doi.org/10.1007/s11606-008-0747-1.

Skaik, Y.E.W. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology*, *56*(4), 341. https://doi.org/10.4103/0301-4738.41424.

Srinivas, S., & Ravindran, A.R. (2018). Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework. *Expert Systems with Applications*, *102*, 245-261. https://doi.org/10.1016/j.eswa.2018.02.022.

Torres, O., Rothberg, M.B., Garb, J., Ogunneye, O., Onyema, J., & Higgins, T. (2015). Risk factor model to predict a missed clinic appointment in an urban, academic, and underserved setting. *Population Health Management*, *18*(2), 131-136. https://doi.org/10.1089/pop.2014.0047.