Prediction of Patients No Show to Their Appointments with Machine Learning Algorithms

Burcu Ağdar
Computer Engeenering
Dokuz Eylul Univeristy
Izmir, Turkey
burcu.agdar@ogr.deu.edu.tr

Emirhan Bilge Bulut
Computer Engeenering
Dokuz Eylul University
Izmir, Turkey
emirhanbilge.bulut @ogr.deu.edu.tr

***Abstract*— Hospital appointment systems are important for the regular functioning of the hospital system. However, not attending the appointments can cause problems. It is possible to predict whether a person will come to an appointment by using the data in hospitals. By using real data from hospitals in Brazil and classification models in machine learning, it was desired to reach the model that made the most optimal prediction. In this study, among the classification models examined, the algorithms with the highest percentage of accuracy were RandomForest and K-Nearest Neighbors algorithms.**

Keywords—machine learning, classification, binary data, accuracy, models

I. INTRODUCTION

The health sector is at the forefront of many sectors. Billions of data are produced every year in the field of health and these data are made meaningful by machine learning algorithms. In addition, these algorithms play an important role in the development of the health sector. In regions where patients are concentrated, there may be congestion and crowding at certain times in hospitals. Many different business areas and the public are affected by the intensity that can be experienced in this sector. It is aimed to prevent patient admissions and uncontrolled crowding that may occur with appointment systems. Especially during the pandemic period, it is difficult to get hospital appointments, but in case of the formation and increase of patients who do not come to the appointment, it prevents the demand from being met and is effective in increasing the patient density faster. The aim of this study is to develop a machine learning algorithm that can predict whether patients who make an appointment from the hospital do or not, and to test its applicability. It aims to observe whether there is a connection or relationship between the patients' attendance to appointments by looking at the information of the patients. With the relationships and estimation models that can be drawn, it can reduce the density that will arise from hospital appointments and prevent patients who do not come with an appointment in the future. In this study, the study was carried out on real hospital records in Brazil. The next sections of the article will proceed as follows . Chapter 2 includes algorithms and methods used by similar studies and studies in the literature. In Chapter 3 , the materials and methods used will be explained . Chapter 4 explains the information about our dataset, the operations performed on it, and includes detailed reviews about the dataset. Chapter 5 includes applied models, relationships and model evaluations-performance comparisons. Chapter 6 describes the applicability and the overall result of our study.

II. RELATED WORK

Tsai and Teng [1], it seeks ways to optimize the patients' getting an appointment or an outpatient examination service. It offers different classification methods in order to reduce the patient density and basically allows a second person to make an appointment at the appointment times of the people who are highly likely to not come after predicting whether a patient will come or not. At the same time, they have put forward a model that will enable the hospital to work in the most optimal and non-intensive way by categorizing the patients, according to their probability of not coming and their time distribution.

Kurasawa et al study [2], it was tried to create a model that predicts the time of not going to the appointments of diabetes patients specifically. It is the first study to predict that diabetic patients will not go to hospital appointments and the aim of the study is to ensure that patients can come to their appointments and receive treatment if it can be predicted early. Although the performance and rates of the algorithm used in the study are stated, the success rate has not been tested in real life, but only on machine learning models.

With the online availability of hospital appointment systems, problems such as forgetting and late cancellation of appointments can occur both for the hospital and for the patients. This type of appointment or not is called a no-show in the literature. Sometimes patients cancel their appointments because of the patient's illness. A missed appointment could negatively impact patient's health condition and even increase the risk of premature death in severe cases.

McQueenie [3] investigated the relationship between the absence of hospital appointments in all-cause mortality in patients with chronic physical and mental health problems over a 3-year period. As a result of her study, she revealed that patients who do not go to their appointments have a higher risk of death. Patients who do not attend an appointment not only harm themselves, but also negatively affect the hospital's resources and effectiveness unnecessarily. Kheirkhah et al. [4] reported a no-show proportion of 18.8% across 10 clinics between the years 2006 and 2008 at a medical center that serves over 76,000 veterans in Texas, with an estimated 14.58 million marginal cost of no-shows in 2008 [5]. The effects of those who did not come to the hospital with an appointment were mentioned in the following studies [6 , 7]. As seen in these sources we have examined , being able to predict whether patients in hospitals will come to their appointments will yield useful results in many respects . In addition to reducing the hospital density , it seems possible to detect the negative effects on the patients .

III. MATERIAL AND METHODS

1.Data Reduction and Transformation

After eliminating the missing and faulty parts of the data, it progressed by applying the backward elimination technique in creating and training the models. First, the model was trained regardless of the correlation values of all features. Optimization measures were examined by playing on the parameters of the results. Afterwards, the features were extracted respectively according to the correlation matrix and the results were compared. By reviewing the data, it was aimed to increase the performance of the model by deriving other features that will be meaningful on ABT .

2.Model Selection and Evaluation

Most of our features consist of categorical data and our target feature is binary value. Continous data are very few in our data and classification algorithms will be applied in model selection. Random Forest, Decision Tree Classifier, K-Nearest - Neighbor, Logical Regression, which are among the classification algorithms, have been extensively tested. In order to increase the performance, methods such as Adaboost were also tried and the results were observed.

- K-Nearest – Neighbor: Although it is one of the most well-known classification algorithms, it actually memorizes data rather than learning it. Such approaches are called lazy learner and try to decide by looking at k predetermined neighbors that are close to the new value according to certain parameters. While applying the algorithm on our dataset, we tried to get better results by changing the size of the dataset.
- Decision Tree Classifier: It is one of the models used in both categorical and continuous data, which is also known as ID3 algorithms and separates decision nodes according to information weight. One of the things we need to pay attention to when using this model is the possibility of slipping into the overfitting state if the outliers in the data are not cleaned, pruning is not performed and if not adjusted correctly. In addition, DTCs may need pruning to increase performance because they contain gready search.
- Random Forest: Although it is based on decision trees, the working mechanism creates trees on the dataset and gives the result with the highest value of the result from these trees. Random forest also reduces the possibility of overfitting with this structure. In addition, it has become one of the fastest working and preferred models by bringing parallel operation with it.
- Logistic Regression: Although the first continuous data comes to mind when regression is mentioned, it is often preferred in problems applied on categorical data and especially in problems where we can distinguish the target feature as binary (0 or 1). It can generally be preferred on data that is not normally distributed or in data sets containing independent samples. It is a model with a high success rate that we predict on our data.
- AdaBoostClassifier : Usually there are decision trees on the basis . It aims to create another tree by changing the data of erroneous and correct results from a tree, thus increasing the success rate. It aims to increase the success rate and become a strong learner by creating trees that are trained with the output of the data of many trees in succession, while the success rate is low in the first tree or model, which is called a weak learner. There are many approaches based on this methodology.

IV. DATASET INFORMATION

1.Initial Data

The data are taken from the dataset named Medical Appointment No Shows on kaggle.com. Data refer to information obtained from hospitals in Brazil. The dataset contains a total of 110527 samples and 14 descriptive features. These features are;

Patient id : The unique number that patients have

AppointmentID: The unique number of each appointment.

Gender: It is the gender of the patients. In this data set, 65% of women and 35% of men are included.

ScheduledDay: It is the date that the patients created the appointment. The dates are between 10 November 2015 and 8 Jun 2016.

AppointmentDay : Date of appointment. Dates are between 29April 2016 and 8 Jun 2016.

Age : It is the age of the patients. When the erroneous data is ignored, the lowest age is 0 and the highest is 115.

Neighborhood : The place where the appointment will take place. There are 81 different locations in the dataset.

Scholarship : Represents the state aid to poor families. It can take values of 1 or 0.

Hypertension : Indicates whether patients have hypertension. It can take the values 1 or 0.

Diabetes : Indicates whether patients have diabetes or not. It can take the values 1 or 0.

Alcoholism : It shows whether the patients use alcohol or not. It can take the values 1 or 0.

Handcap : Indicates whether patients have any disabilities. It can take the values 1 or 0.

SMS_received : Shows whether the patients received sms about the appointment. It can take the values 1 or 0.

No-show: It shows whether the patients come to the appointment or not. It can take the values Yes or No.

2.Data Preparation

There may be incomplete or incorrect information in the data. Although these can be tolerated to a certain extent, when they are large, they should be removed from the table completely. If these missing or faulty data are too few, it is sufficient to remove these samples from the table. There were some wrong data in the age and handcap columns of our data. The age of the patients cannot be less than 0. However, there were -1 values in the data. A small number of rows with this value were removed from the table. On the other hand, handcap values should have only 1 and 0, while 3 and 4 values were also present. These erroneous values were also cleared from the table. The Nan values in the Age column were filled with the mean value of the ages.

Since the time between the appointment time and the dates to go to the appointment can be an important factor, this time has been derived as a new feature. From the dates to go to the appointment, the day of the week was also derived and added as a new feature. Features named Appointmentday and Scheduleday have been deleted from the table.

The values of the features named gender and no-show have been updated as 1 and 0 in order to be used in models. We used the one hot encoding process for digitization, as there are features with more than one categorical value on the date and place of the appointment.

V. EXPERIMENTAL RESULT

1. Creating First Models and Making First Trials

Model	Target	Parameters	Precision	Recall	F1-Score	Support	C Matrix		Accuracy
KNN (Default)	0	{'n_neighbors':5, 'weights': uniform, 'algorithm': auto}	0.81	0.93	0.87	29060	27049	2011	0.78
	1		0.37	0.16	0.22	7334	6164	1170	
KNN (Optimum)	0	{'n_neighbors':2, 'weights': uniform, 'algorithm': auto}	0.81	0.96	0.88	29060	27829	1231	0.79
	1		0.39	0.11	0.17	7334	6557	777	
Random Forest (Default)	0	{ 'n_estimators':100 , 'critarion':'gini' , 'min_samples':2 , 'max_features':'auto' }	0.82	0.92	0.87	29060	26845	2215	0.78
	1		0.40	0.20	0.27	7334	5833	1501	
Random Forest Optimized	0	{'n_estimators': 100 'criterion': 'entropy', 'max_features': 'sqrt', }	0.82	0.92	0.87	29060	26845	2215	0.78
	1		0.41	0.21	0.27	7334	5825	1509	
Decision Tree Classifier (Default)	0	{ 'critarion':'gini' , 'splitter':'best' }	0.83	0.84	0.84	29060	24510	4550	0.74
	1		0.34	0.32	0.33	7334	4965	2369	
Decision Tree Classifier Optimized	0	{'criterion': 'gini', 'max_features': 'sqrt', 'splitter': 'best'}	0.81	0.93	0.87	29060	27049	2011	0.78
	1		0.37	0.16	0.22	7334	6164	1170	
Logistic Regression	0	{ 'penalty':'l2'}	0.80	0.99	0.89	29060	28827	233	0.79
	1		0.31	0.01	0.03	7334	7229	105	
Naive Bayes	0	{'var_smoothing' : 1e-9}	0.83	0.67	0.74	29060	24510	4550	0.62
	1		0.25	0.45	0.32	7334	4965	2369	

We achieved accuracy at different rates, from 62% to 78%, in the models we initially applied on our data. If it is examined in detail, when the KNN algorithm is applied with default values, the estimation of those who do not come to the appointment has very low percentages in precision and recall values. When GridSearch is applied for the KNN algorithm and optimal values are found, an increase of 1% in accuracy is seen. However, the percentage of correct predictions for those who did not attend the appointment is still quite low. When we apply the Random Forest algorithm with the default values and the results obtained after optimizing it with GridSearch, the accuracy values are seen as 78%. Although there is an increase in the estimation of the patients who did not come, it is still not at the desired level. After running the Decision Tree Classifier algorithm with default values and the results obtained from GridSearch, the accuracy values obtained are 74% and 78%, respectively. The Decision Tree Classifier algorithm could not make the desired patient estimation. In Logistic Regression, on the other hand, although the accuracy rate is 79%, which can be considered good, the estimation of the absent patient is very low, 3%. Finally, when we look at Naive Bayes, it is seen that accuracy is much lower than the others. However, the estimation of absent patients is better than other algorithms. However, the rate of 32% is still quite low.

2. Results After Removing Unnecessary Features

Model	Target	Parameters	Precision	Recall	F1-Score	Support	C Matrix		Accuracy
Random Forest	0	{'n_estimators':100 , 'critarion':'gini' , 'min_samples':2 , 'max_features':'auto' }	0.82	0.92	0.87	29060	26785	2275	0.77
	1		0.36	0.18	0.24	7334	6042	1292	
Random Forest(Optimum)	0	{'n_estimators': 2 'criterion': 'entropy', 'max_features': 'auto', }	0.81	0.92	0.87	29060	26873	2187	0.77
	1		0.35	0.16	0.22	7334	6163	1171	
Decision Tree Classifier	0	{ 'critarion':'gini' , 'splitter':'best' }	0.81	0.93	0.87	29060	26963	2097	0.77
	1		0.36	0.16	0.22	7334	6169	1165	
Decision Tree Classifier(Optimum)	0	{'criterion': 'gini', 'max_features': 'sqrt', 'splitter': 'random'}	0.81	0.93	0.87	29060	27021	2039	0.77
	1		0.35	0.15	0.21	7334	6213	1121	

By examining the correlation matrix, it was observed that some features had no effect on the result or were negative. When he removed these featuras and applied Random Forest and Decision Tree Classifier algorithms, 77% accuracy value was obtained in both algorithms. The estimation of incoming patients is seen at a good value of 87% in both algorithms. However, the estimation of the patients who did not come remains very low at 16% and 18%. After applying GridSearch to these two algorithms, there was no change in accuracy.

3. Results After New Feature Derived

Model	Target	Parameter	Precision	Recall	F1-Score	Support	C Matrix		Accuracy
KNN Default Parameter	0	{'n_neighbors':5 , 'weights': uniform, 'algorithm': auto}	0.93	0.93	0.93	29060	27052	2008	0.89
	1		0.73	0.73	0.73	7334	1947	5387	
KNN	0	{'n_neighbors':2 , 'weights': uniform, 'algorithm': auto}	0.89	0.96	0.92	29060	27854	1206	0.87
	1		0.76	0.52	0.62	7334	3533	3801	
Random Forest	0	{ 'n_estimators':100 , 'critarion':'gini' , 'min_samples':2 , 'max_features':'auto' }	0.95	0.93	0.94	29060	27099	1961	0.91
	1		0.75	0.82	0.78	7334	1329	6005	
Decision Tree Classifier	0	{ 'critarion':'gini' , 'splitter':'best' }	0.93	0.94	0.93	29060	27337	1723	0.89
	1		0.75	0.71	0.73	7334	2139	5195	
Logistic Regression with Default Parameter	0	{ 'penalty':'l2'}	0.87	0.96	0.91	29060	27915	1145	0.85
	1		0.73	0.43	0.54	7334	4215	3119	
Naive Bayes	0	{'var_smoothing' : 1e-9	0.86	0.94	0.90	29060	27458	1602	0.83
	1		0.64	0.39	0.48	7334	4475	2859	
Decision Tree Classifier Optimum	0	{'criterion': 'entropy', 'max_features': 'auto', 'splitter': 'best'}	0.95	0.93	0.94	29060	27085	1975	0.91
	1		0.64	0.39	0.48	7334	1295	6039	
Ada Boost	0	{ 'n_estimators':50}	0.95	0.93	0.94	29060	27135	1925	0.91
	1		0.75	0.79	0.77	7334	1512	5822	

A new feature was derived, as the accuracy values and the estimation of patients who did not come had low values. In this feature, patients with multiple appointments were given a score. This score greatly benefited the models in terms of whether patients should come to their appointments. When the KNN algorithm is examined, it is seen that it has 89% accuracy in default values. When the k value is changed to 2, the accuracy becomes 87%. The Random Forest algorithm has an accuracy of 91%. It is the algorithm with the best prediction rate for patients who do not come. Logistic Regression and Naive Bayes algorithms have values of 85% and 83%, respectively. However, recall and precision values are low. Decision Tree Classifier has 89% accuracy at default values and 91% accuracy at optimum values. AdaBoost has the same accuarcy as Random Forest. When the values of all algorithms are looked at, the ones with the best results are AdaBoost and Random Forest algorithms.

VI. CONCLUSION

In summary, it seems possible to predict whether or not people will come to hospital appointments by looking at patient records. It can be used to reduce hospital density by predicting whether patients will come to their appointments with machine learning models. It has been observed how to increase the success rate of the model by making various changes and adjustments to the models we have applied on patient records in Brazil, and which data has the highest information weight (Entropy or gini index). While applying these approaches, these inferences were reached by using the backward elimination technique. This research can be used for estimation by applying it to other patient records.

References

[1] Tsai, P. F. J., & Teng, G. Y. (2014). A stochastic appointment scheduling system on multiple resources with dynamic call-in sequence and patient no-shows for an outpatient clinic. *European Journal of Operational Research*, *239*(2), 427-436.

[2] Kurasawa, H., Hayashi, K., Fujino, A., Takasugi, K., Haga, T., Waki, K., ... & Ohe, K. (2016). Machine-learning-based prediction of a missed scheduled clinical appointment by patients with diabetes. *Journal of diabetes science and technology*, *10*(3), 730-736.

[3] H.M. Choi and J.P. Hobert, The polya-gamma Gibbs sampler for bayesian logistic regression is uniformly ergodic, Electron. J. Stat. 7 (2013), pp. 2054–2064.

[4] P. Kheirkhah, Q. Feng, L.M. Travis, S. Tavakoli-Tabasi, and A. Sharafkhaneh, Prevalence, predictors and economic consequences of no-shows, BMC. Health. Serv. Res. 16 (2016), pp. 13.

[5] Lin, Q., Betancourt, B., Goldstein, B. A., & Steorts, R. C. (2020). Prediction of appointment no-shows using electronic health records. *Journal of Applied Statistics*, *47*(7), 1220-1234.

[6] A.S. Hwang, S.J. Atlas, P. Cronin, J.M Ashburner, S.J. Shah, W. He, and C.S. Hong, Appointment ‘no-shows’ are an independent predictor of subsequent quality of care and resource utilization outcomes, 2015.

[7] L.A. Nuti, M. Lawley, A. Turkcan, Z. Tian, L. Zhang, K. Chang, and D.R. Willis, No-shows to primary care appointments: Subsequent acute care utilization among diabetic patients, 2012.