Predictive Analysis

Liu, D., Shin, W. Y., Sprecher, E., Conroy, K., Santiago, O., Wachtel, G., & Santillana, M. (2022). Machine learning approaches to predicting no-shows in pediatric medical appointments. *Npj Digital Medicine*, *5*(1). https://doi.org/10.1038/s41746-022-00594-w

Introduction

What was the problem being solved?

The study sought to predict no-shows in pediatric medical appointments. No-shows refer to medical appointments that patients schedule with a medical practitioner but fail to show up for on the scheduled date. No-shows result in poor health results for patients, with inefficient utilization of medical resources. There has been limited exploration of the prediction of no-show times, a challenge the study seeks to address in the prediction process. The ability to predict no-shows or highlight possible no-shows is essential in the design and implementation of interventions that could prevent the occurrence, improving patient care and allocation of resources in clinical departments.

Why was this problem important to solve?

Research shows that healthcare providers report a 3-80% no-show rate, which hinders the efficient and timely delivery of preventative, continuous, and check-up services. The result of a no-show is inefficient control of chronic diseases and late presentation for healthcare. No-shows have been associated with environmental factors like traffic, social economic factors, and time conflicts between physicians and patients. No-shows lead to missing data and inefficient care, therefore predicting no-shows was important to prevent chances of missing data occurrences and ensure patients receive e timely medical attendance. It was also important to conduct the study to generate recommendations to the healthcare system on addressing the factors that result in a no-show

How was the data acquired?

The data was acquired from Boston Children's Hospital's primary care pediatric clinic, which had recorded a 20% no-show rate before the study was conducted. The authors collected information from 161,822 hospital appointments made between January 2015 and September 2016 by 19,450 patients. The information collected included medical record number(MRN), patient's age, insurance type, gender, and appointment status.

Methods and Results

What steps were taken to prepare the data?

The recorded patient data was inspected for any missing information, and it was recorded that 77% of the records had missing data. To train the data for better modeling outcomes, data processing methods were then designed to tackle the missingness indicator. The data categories were then converted into binary variables, with 1 representing the presence and 0 absence of the data category, days of the appointment classified between 1 to 7 for Sunday to Monday of each week, and 1 for no show and 0 for show. The study was aimed at children's medical services, therefore any record of a patient above 18 years who received specialized care at the facility was excluded.

How was this problem solved?

The study developed understandable methodologies that predict a patient's risk of a no-show. the results also informed that the history of patients' no-shows and weather were the biggest predictors of future no-shows. The study also recommended the best day to schedule medical appointments. Where appointments were set on days with higher temperatures, humidity, and lower wind speed, there were lower rates of no show. The study, therefore, solved the problem by showing the causal factors and recommending ways forward to address them for better show-up rates.

What modeling techniques were used?

The study employed neural network-based and logistic regression-based binary models. The models were constructed in Keras 2.0 environment using TensorFlow 1.7 in the backend. The first testing in the models determined hyperparameters. This was then followed by validation, where data was split into training sets and testing sets, which took up 75% and 25% each. The datasets were then combined to train the model.

Why did the team choose the methods/models they did?

The absence of studies predicting medical no-shows using neural networks was the main reason the methodology was selected. The neural network approach was used to ensure that the potential non-linear relationships in the different data categories were captured, and the desired outcome was arrived at. The neural models used contained 3 hidden layers to impute missing variables, which eliminated the need for data imputation before modeling and served to improve the performance of the models. The models are also flexible and easily adaptable and have been previously employed in other case studies with missing data in other clinics and hospitals, which made them ideal for the study.

What metrics were used to evaluate the results? Why?

To improve the metrics, a ROC curve was produced by plotting the True Positive Rate(TPR) and False Positive Rate(FPR) in thresholds between 0 to 1. This was carried out to ensure a high-quality prediction, with prediction values closer to 0.5 showing bad quality and values close to one indicating good quality prediction. The study also utilized baseline methods of applying the state of the latest visit to be the predictor of a show or no-show for the next visit, which gave a better prediction than using other previous records.

Conclusion

How were the results of the model implemented?

The results were used to develop a way of predicting the risk of no-shows in pediatric patients. Using the prediction, healthcare practitioners can identify patients at risk of a no-show, therefore employ strategies that will ensure patients show up for their appointments on the set dates. The results were also used to relate the rate of no-shows to causal factors that had been established, to showcase which was the highest causal factor. The results were therefore used to inform on the best day to set medical appointments.

What were the actionable consequences of the case study?

The team recommended analysis of the no-show data for a wider population in different clinical settings for better generalization of trained algorithms and scientific conclusions. Increasing the amount of patient information available on their medical problems improves the performance of the prediction models. The study informed the utilization of SMART-on-FHIR systems in hospitals to ensure accurate record keeping and consideration of patient history when setting up appointments.

What did the team learn from the study?

The team learned that there are plenty of missing data in hospital records that need to be addressed for better prediction of no-shows and patient outcomes in the future. They learned that the facility had a 20% no-show rate that was a result of conflicting times between patient and medical practitioner schedules and the state of local weather on the day of the set appointment. The team also learned that the inclusion of the state of local weather on appointment days would further improve future prediction models and probably be a higher contributing factor to no-shows compared to other factors. Further, the team learned that their prediction methodology could be used in hospitals that employ the SMART-on-FHIR systems for retrieving patient history when setting patient appointments. They also learned that neural network model approaches have higher computational costs compared to logistic models.

How should or would the team approach the problem differently in the future?

With access to more patient information and not limited to what the team had, there would be improved model accuracy. The increased data size for future study will inform a better relationship between the neural networks and regression models which is informed by the size of data space. The inclusion of data like reminder messages for appointments could better inform the no-show rates for the design of better interventions.