Predict No Shows for Medical Appointments

Resource planning is crucial for any business; it's even more so for operations struggling with resource crunch. Government sponsored medical services is a prime example of such an operation. In almost all developing countries , there is severe shortage of medical personnel and at the same time, not ideal living conditions for big chunk of the population further exacerbates need of medical attention.

Governments have tried floating subsidised access to poor families to increase access to healthcare . One big problem which stops better utilisation of this facility is that people make appointments but do not show up without a notice. That time slot could have been given to a needier person if authorities were able to determine who is very likely to not show-up and followup appropriately .

Your task here is to make use of historical records to build a model for predicting a No-Show for an appointement given appoint details, medical history and demographic details of the customer.

Data Files

Medical History = medical_history.csv

Demographic Details = demographic_details.csv

Train Dataset =train.csv

Test Dataset = test_share.csv

Formal Problem Statement

Variable names are self explanatory and there is no formal data dictionary provided by the client .

Your task here is to build a predictive model for predicting No-shows given the appointment details. You need to build your model on the train dataset. **Test dataset does not have a response column; you need to predict those values and submit it in a csv format.**

target column = No-show

Advice : Make use of demographic and medical history data files to get a good model.

Evaluation Criterion Part 1:

You will first attempt Part 1 of this project which is a quiz. You can access it through LMS. This quiz needs to be answered based on exploration of the dataset given and some generic questions about algorithms discussed in the course. Consider only the training dataset for data cleaning and exploration to answer the quiz questions. There will be 10 questions of which you need to get at least 7 correct in order to pass the project.

Part 2:

Here you work on creating the machine learning models and choosing the one which gives the best performance. **You can refer to the Project Process Guides provided in LMS to understand how to approach and work on a project.**

For this project, score will be calculated as:

roc_auc score

This score will be calculated using your predictions for the test file. **You need to score more than 0.66 in order to pass the project submission for this particular project**. You need to submit predicted probabilities [**not the** hard classes] for this project.

Submission:

Please give appropriate names to submission file. Preferable containing your name and attempt number [this is for you to keep track of your different experiment performance] . Submission needs to be a csv file . Any other format like excel , pdf etc will not be graded.

Number of rows in the submission csv should be exactly the same as test data. If this is not taken care of, your submission will not be graded.

You can make as many submissions you want . [We might ask you to submit the script which was used to generate the submission at any time].

In order to clear this project, you are required to clear both, Part 1 as well as Part 2 of this assignment. Wish you all the best!