

Predicting failure in the Brasil's Unified Health System (UHS) to attend patients demands.

A Data Science Methodological Approach.

Abstract:

Here in Brazil we have a public health care services system called Unified Health System (UHS). The institution's main goal is to provide free health care services for citizens on a national scale using Federal Government resources. My question refers to the context of this institution which currently faces various issues related to providing effective and efficient health care services, such as: delay in services, lack of materials and equipment, overcrowding, lack of physicians, unprepared professionals, waiting queue for different specialties and bad management of financial resources. Patients can have all sorts of demands in a hospital. Specialized assistance, proper equipment, medication, food and a bed are just a few examples. And sometimes, for various different reasons, these demands are not met by the service providers, resulting in loss mostly for the patients but also for the institution's and the government's reputations. With this in mind can we predict which demands are not going to be attended, where, how many of them, when and why (causes)? Putting in context, which, where, when, how many times and why, services from the UHS hospitals and facilities have failed, are failing and will fail to attend the different types of demands that patients present? For example, an incoming patient might need an urgent surgical intervention in the chest. If there is no cardiologist surgeon physician available, his demand for a specialized professional won't be met and he might be at death risk. The patient will also have to wait and therefore his demand for urgent assistance won't be met. If there is no proper equipment for this type of surgery, his equipment demand is not met. Another patient might need hospitalization and beds might not be available. The goal of the manager or stakeholder asking this question is therefore to identify, understand and predict failures in the process of providing the health care services (which, where, when and how many) along with what causes them (why), based on previous records (historical data) analysis and continuous training of the model, so actions can be taken in order to solve or mitigate the problems at hand and their consequences.

1. ANALYTICS APPROACH

The approach was determined by the problem at hand, failures in attending daily patients demands, and results expected, reduction of failure episodes and their consequences. As the problem involves yes/no answers (fail/success) for each demand presented in the hospitals, a classification approach would be suitable.

A decision tree classification model might be up for the task. This is because it can be used to identify certain combinations of conditions (causes of failure/success) leading to a specific outcome (failure/success in attending demands).

In the case study, their model also outputted a risk score for the facilities to use, representing the chance of reincidence for each cardiac patient. In a similar fashion, our model would output a risk score representing the chance of failure to attend a specific type of patient demand somewhere, somewhen, etc. This could be extremely helpful for managers of medical facilities and resources to guide decision making in order to reduce the number of occurrences of these failures and their associated impact.

2. DATA REQUIREMENTS

This step's actions are derived from the nature of the problem to be solved and the analytics approach chosen. As I don't know that much about decision tree classification models I'll just point out that:

2.1 In this step the scientist "needs to identify which data are required, how to source and collect them, how to understand or work with them, and how to prepare data to meet the desired outcome"¹ or the approach chosen.

2.2 For our problem we should therefore, identify what are the patients' demands, the possible predictors of failure and success to attend those demands, and what data we have on that. Then we should understand what the data available is about, how we can access it, appropriate formats the data needs to be in so it can be combined with the approach/model and what kind of preparations we will have to make on the data, all before actually collecting it.

3. DATA COLLECTION

Now that you have planned it all in the 'Data Requirements' step, it's time for some action. We then go and actually collect the data available from its sources. In our case, Government and related public institutions open data sources.

We might perform additional surveys to get more data.

4.1 DATA UNDERSTANDING

Now it's time to look back and think about what we have done. The results of the data collection step are data sets. This means we have to check for two things:

4.1.1. How they were constructed, that is, were there any flaws or mistakes in the whole of the construction process?

4.1.2. Is the data that we collected representative of the problem to be solved?

Then, if necessary, you go back to data requirements and data collection steps to make the appropriate changes and check for possible improvements. And if you are all set, you can go forth with the data to the preparation stage - the coolest of all stages.

4.2 DATA PREPARATION

There are a lot of things one could say about data preparation. It often deals with missing values, duplicates and formatting. In general terms the goal is to transform the data into a state it may be easier to work with and that matches with the models to be used. Various techniques are used for the purpose of accelerating these practices, such as feature engineering and automation.

5. MODELING AND EVALUATION

The modeling and evaluation stages are done iteratively. When modeling we'll build models to answer the question. We choose the models based on the question/problem, results expected and the approach. In our case a predictive model should suffice because it "tries to yield yes/no type outcomes" ², as explained before.

We'll play around with different algorithms to ensure that the variables in play are actually required. We'll then train the data (for which the demands met/not-met we already know) to generate the models. And finally, we are ready for evaluation of the model. This step can have two main stages: Diagnostic measures and statistical significance testing.

In the first one, as we are dealing with predictive models, we'll ensure the models are working correctly by using decision trees to evaluate if the answers the models can output are aligned to the initial design. In the second one we'll make sure that the data is being properly handled and interpreted within the models. This avoids unnecessary second guessing when the answers are revealed.

As this is a never ending process we'll then go back to modeling to input the evaluation results in the models to improve their results.

6. References:

1. Video: Data Requirements, week 1
2. Video: Modelling - concepts, week 2