Chapter 13
Predicting Patterns in Hospital Admission Data

Jesús Manuel Puentes Gutiérrez
Universidad de Alcalá, Spain

Miguel-Angel Sicilia
University of Alcalá, Spain

Salvador Sánchez-Alonso
Universidad de Alcalá, Spain

Elena García Barriocanal
Universidad de Alcalá, Spain

ABSTRACT

Predicting patterns to extract knowledge can be a tough task but it is worth. When you want to accomplish that task you have to take your time analysing all the data you have and you have to adapt it to the algorithms and technologies you are going to use after analysing. So you need to know the type of data that you own. When you have finished making the analysis, you also need to know what you want to find out and, therefore, which methodologies you are going to use to accomplish your objectives. At the end of this chapter you can see a real case making all that process. In particular, a Classification problem is shown as an example when using machine learning methodologies to find out if a hospital patient should be admitted or not in Cardiology department.

INTRODUCTION

In this chapter we roughly discuss how to derive conclusions from hospital admission data. We describe a process to identify patterns in the data as well as the description of several concepts needed to carry out that objective. We use different Big Data analytics techniques to achieve our goal. Big Data analytics allowed us to uncover hidden patterns and unknown correlations to start working with available datasets. Then, we are able to improve the operational efficiency and obtain business benefits, in order to follow a system of work. Similar conclusions were reached according to (Powers, Meyer, Roebuck, & Vaziri, 2005), where they use advanced econometric cost modelling techniques to predict healthcare costs using pharmacy data.

Initially, it is advisable to make a study of correlation indexes from the attributes we are going to use. These indexes will give us a better idea about the most appropriate attributes and will allow us to obtain conclusions with the selected dataset.

DOI: 10.4018/978-1-5225-2607-0.ch013

Once we have decided what are the answers we want to know and the type of study we want to accomplish, we need to begin studying the type of data and the type of structure that we have in our dataset. This means that an important part of the available time to develop the study was devoted to prepare our data for the algorithms we would use. For that reason, section 2 details the type of data we can find and how it usually appears.

After preparing the dataset environment, we needed to use the appropriate Machine Learning techniques depending on the type of data we had and on which conclusions we wanted to obtain. In the present day, other studies are using machine learning techniques to predict behaviours in health systems and they select their appropriate techniques to reach them. As an example of this, some researchers would like to know if patients are going to re-enter during the next twelve months as it is done in (Vaithianathan, Jiang, & Ashton, 2012), where they used multivariate logistic regression. Or, perhaps, they would like to predict hospital admissions depending on patient-specific medical history using several types of classification algorithms, according to Wuyang et al. (2015). Since several years ago, those techniques have gradually been introduced in different studies thanks to their effectiveness when making predictions, as we can observe in (Wuyang et al., 2015) too.

At the end of this chapter, we will describe the process followed in a real research with the dataset provided by a Spanish hospital. As we will see, while developing this research, the same procedure described in this chapter was followed to obtain the conclusions. In the following sections we describe everything that is required to carry out such research. Each technique permits to decide which algorithm is appropriated in each case. Specifically, in that research a "yes or no" question type was answered but it is extendable to other types of similar questions. All this also depends on what kind of information we were treating with. In that case, it was generic data from the hospital information system. On that point, we needed to analyse the dataset to know what conclusions we can derive and what kind of answers we expect to obtain. In the mentioned research we had a supervised machine learning problem, because of the type of question we have to answer (which is of the "yes or no" type in this case). We studied a Classification problem because the results we wanted to obtain used labelled data. In that case we already had answers to the questions raised. In other words, we have a dataset where we know what happened previously and then we want to predict what is it going to happen in the next unknown cases. This can be observed in (Valverde, Tejada, & Cuadros, 2015), where the authors use some methodologies comparing several supervised Classification algorithms to predict and analyse feelings. As we have a dataset where we know what happened previously, we can split the dataset in two groups of data, one to train the learning algorithm and another one to use it lately in a testing process, avoiding the overfitting effect. In this way, after the training and testing process, we can compare predicted results with real ones. That means that we can validate the selected algorithm and determine the success rate that it is obtained with its use. Finally, we can observe the best algorithms that can be used with that particular dataset. The rest of this chapter is organized as follows. Section 2 defines the types of data we can find in a dataset and how to classify them. It is important to know where to use it and how. In section 3 we describe machine learning techniques. We have to know when we should use supervised or unsupervised learning methods to achieve the objectives that we had planned. Then, we talk about the different algorithms that can be used to predict our objectives according to the types of data that we have and to the type of technique. Section 4 describes the process followed in a recent research to find out potential assignments in a hospital department, being assisted by the different sections in this chapter. Finally, in sections 5 and 6, we find the conclusions and the references, respectively.