Universidad Torcuato Di Tella

Master in Management+Analytics Thesis Proposal

A Machine Learning Approach for Prediction of Hospital Bed Availability

by

Josefina Dalla Via Monti

Advisor: **Gustavo Vulcano, PhD**
Co-Advisor: **Nicolás Schapchuck, MBA**

December, 2019

1 Background

Bed management is at the core of hospitals' Operations Management. Poor bed management may result in either overcrowded emergency rooms, valuable resources out of use or treatment limitations for critical patients, just to name a few. Patient dissatisfaction is the immediate result, leading to a detrimental effect for the health unit reputation in the long term. In this context, optimization of hospital bed management is key for hospital's success.
Previous work on this topic has been focused on optimizing bed allocation using linear programming and simulation techniques. However, little has been done in terms of improving the quality of the inputs that optimization models rely on: upcoming demand and bed availability forecasts.

2 Research Problem

While endowed with huge amounts of data, hospitals usually make bed management decisions based on physician's hints on bed availability and staff intuitions regarding upcoming demand.This is the case for one of the biggest hospitals in Argentina, located in the city of Buenos Aires, which is the focus of our study.
Even though we do not underestimate the expert knowledge of health professionals and the vast experience of hospital staff, we believe that advanced, predictive analytics can provide a rapid, automated and scalable solution for the prediction of bed demand and supply patterns. Thus, they can help hospitals manage their capacity more efficiently. In this study, we develop a scientific solution to make predictions regarding bed availability at the different hospital departments, using a machine learning approach. Our project aims not to radically change the current way in which the hospital management optimize bed allocation, but rather to provide an improved input for the current bed management processes.

3 Data

In order to address the above stated problem, we have been granted access to the hospital's full database, which comprises information of:

- Admissions
- Patient's clinical history
- Patient's demographic data
- Medical studies and surgeries and its results
- Reallocation of patients among the different hospital departments
- Diagnosis.

The data provided ranges from January 2017 to November 2019.

Furthermore, we are in permanent contact with the hospital's management team which gives us the possibility of collecting qualitative data about the hospital management processes and the main difficulties that they face. This input will be critical in order to deliver an applicable solution that results useful for the hospital's operations. We may also conduct interviews with the hospital staff: physicians, nurses and administrative employees in order to capture relevant information about the daily operations which may give us some insights for modeling the problem.

4 Methods

We will approach the problem using machine learning techniques. More specifically, different supervised learning algorithms will be tested in order to deliver a prediction of the bed availability in the following twenty four hours.

An assessment will be made regarding whether is it more convenient to make this prediction at a patient level (this is, conditional on the patient history up to now, what's the probability that the patient will be discharged in the following hours) using a classification model, or at a department level (given the characteristics of the current inpatients, how many beds will be released in the following hours) using a regression model.

In either cases, we will train, validate and test the models with different sets of data to avoid overfitting as it is a widespread machine learning best practice. The dynamics of the problem application will be considered in the validation framework: models will be trained with data up to a certain point in time, but validation will be on data from a later window time, thus reflecting real use-case conditions. For obvious reasons, patients in the training set will not be present in the validation or testing set.

The performance metric will be defined in consensus with the hospital management team. This may or may not be one of the classic performance metrics for supervised learning algorithms. If necessary, business opportunity costs regarding wrong predictions will be included in this metric.

Special emphasis will be placed on calculating point in time correct features in order to avoid data leakage.

5 Expected Results

By the end of the project, we expect to have an automated solution that, given the occupation characteristics of the different hospital departments at a certain point in time, will let us know how many beds will be vacated in the following twenty four hours.

We expect to provide a solution whose performance is satisfactory in the eyes of the hospital management team. While this satisfaction criteria may seem ambiguous, it results critical not to focus only on algorithm performance, but also to incorporate the management opinion regarding the applicability of the solution.

Once developed the end-to-end model, we plan to run an experiment comparing the

medical staff prediction with the machine learning model prediction in a real-time scenario. The experiment will be run as long as necessary to have significant results on whether the proposed solution is a preferred option or not.

6 Workplan

Defining a linear workplan for a machine learning project seems inappopiate as iteration and continuous improvement is at the core of the methodology. However, we will be tackling, for sure, the following phases:

1. Data collection and exploration:
   Includes getting all the relevant data, detecting needs of data cleansing and performing the corresponding updates. Making sense of the data by discovering relevant trends and relationships. May also include gathering additional qualitative information by scheduling interviews or shadowing sessions with the hospital staff.

2. Modeling:
   Defining the target variable and building an appropriate data representation through an ETL process. Construction of a validation schema by properly splitting the data into training, test and validation sets. Defining the algorithm to implement and coding its implementation.

3. Training and validation:
   Defining the key performance metric. Training the algorithm. Hyperparameter optimization by comparing performance on validation set. May include visualizing the output to assess the intuition behind the results. Presenting output to the hospital management team for feedback.

4. Model Optimization:
   Iterate over phases 2 and 3 trying different algorithms, learning methods and model features until getting satisfactory results.

5. Testing:
   Defining the success metric and a testable hypothesis. Applying the model prediction on a daily basis along with the current "human" prediction and tracking performance of both methods. Comparing results.

Proposed timeline

- *January 15th: Proposal acceptance/rejectment.*

  Exploration and data collection.

- *January 31st: Deliverable: Jupyter Notebook with visualization of main insights discovered during exploration.*

  Modeling and training

- *February 28th: MVP presentation to the hospital management team. Deliverable: Jupyter Notebook with the MVP code. Performance report.*

  Model optimization.

- *March 31st: First check-point with advisor. Deliverable: Jupyter Notebook with end-to-end optmized model.*

  Testing and reporting

- *May 10th: Second check-point with advisor. Deliverable: Finished thesis document with testing results. Jupyter Notebook with final code.*

  Corrections, follow-ups on advisors' comments.

- *May 30st: Final deadline. Deliverable: Finished thesis document reviewed plus code.*

Project progress will be tracked on the following GitHub repository:

`https://github.com/josedallavia/A-Machine-Learning-Approach-for-Prediction-of-Hospital-Bed-Availability`

References

[1] Barry Dumas, M. (1985). Hospital Bed Utilization: An Implemented Simulation Approach to Adjusting and Maintaining Appropriate Levels. *Health Services Research.* **20**

[2] Carrascosa, R. "Leading in ML projects: Tools and tricks". *Ciclo de conferencias semanales de Machine Learning.* Mercado Libre, December 2019.

[3] Kumar, K. and Shastri, L., (2018). *How Advanced Analytics Can Improve Hospital Capacity Management.* India: Infosys.

[4] Nguyen, J.; Six, P.; Antonioli, D.; Glemain, P.; Potel, G.; Lombrail, P. and Le Beux, P., (2005). A simple method to optimize hospital beds capacity. *International Journal of Medical Informatics.* **74**, 39-49.

[5] Perelman, L.; Barret, E. and Paradis, J. (1997). *The Mayfield Electronic Handbook of Technical and Scientific Writing.* USA: Mayfield Publishing Company. Viewed [15 December 2019]. Available from: `http://web.mit.edu/course/21/21.guide/home.htm`

[6] Ramachandran, B. and Brownlee, M. (2018).*Bed Management Optimization.* India: Infosys

[7] Suryati, S., (2018). Optimization Model for Capacity Management and Bed Scheduling for Hospital. *IOP Conference Series: Materials Science and Engineering.* **300**, 12-16.