CAPSTONE PROJECT

PREDICTING HOSPITAL ADMISSION AT EMERGENCY DEPARTMENT TRIAGE

Predicting Hospital Admission at Emergency Department Triage

BY CAROL CALDERON

INTRODUCTION

Emergency departments (ED) are the hospital's areas where it is provided treatment for patients with severe illnesses and life-threatening injuries 24 hours a day, 365 days a year. Emergency departments are places with high-level pressure and are fast-paced. Despite the fact that the majority of patients are discharged from the emergency department, it usually is the primary source of hospital admissions and a major source of healthcare spending [1]. Some advantages of optimizing the process of triage and admission of a patient in the emergency department are [2][3][4]:

- **Patients:** better opportunity, safety and satisfaction. Reduce complications, adverse events, patient morbidity and mortality.
- **Healthcare providers:** support in decision making and not only to emergency professionals, prediction could help primary health services to define if the patient needs to consult the emergency room.
- **Hospitals management:** support decision-making regarding resources management, patient flow, and healthcare spending. Improve indicators like stay length, readmission rates, and support protocol definitions.

With hospitals increasing adoption of electronic health records and data warehousing, it has been easier to apply statistical models [5]. As well as literature about is exponentially increasing the knowledge and experiences using machine learning to improve healthcare decision making.

OBJECTIVE

Using machine learning, how might we predict the patient admission into the hospital in the emergency department to provide better information that helps the different stakeholders make better decisions.

METHODOLOGY

This project followed the data science process described in Illustration N. 1. The dataset used in this project was originally collected and published by Hong WS, Haimovich AD, Taylor RA (2018) in their paper, 'Predicting hospital admission at emergency department triage using machine learning.'PLoS ONE 13(7): e0201016. https://doi.org/10.1371/journal.pone.0201016
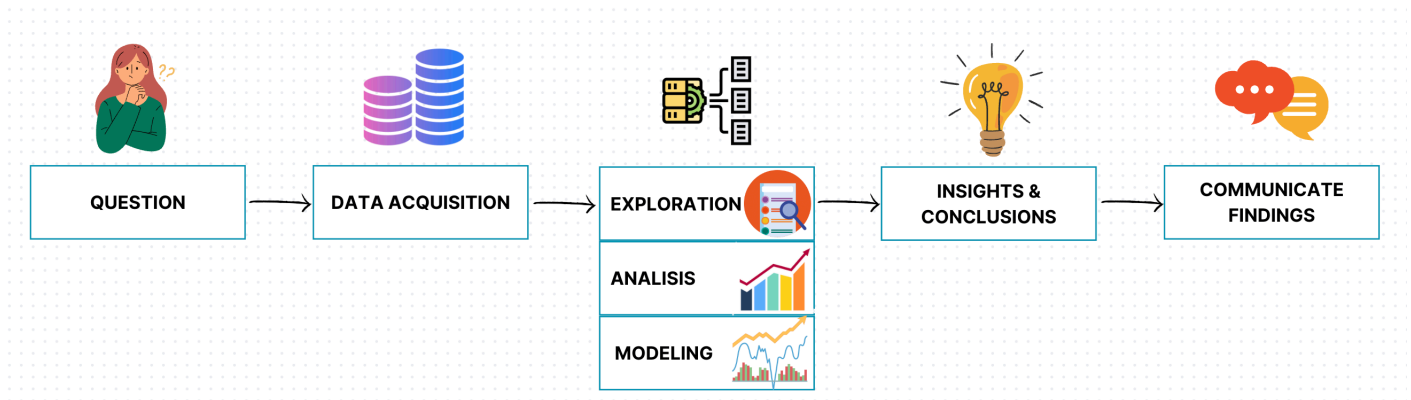
Illustration N. 1 Data Science Process Stages

The data was thoroughly explored, cleaned and preprocessed, before being utilized to build various models. Throughout the modelling stage were applied data balance techniques, feature selection, hyper-parameter optimization and the models were evaluated using different metrics.

The data processed contained more than 400.000 patient records (rows) and roughly 600 features (medical information related with the current and past medical patient situation).

ANALYSIS

The task of predicting whether a patient will be admitted or discharged is a binary classification problem, which involves making one of two possible predictions. There are several machine learning models that can be used to solve binary classification problems like predicting whether a patient will be admitted or discharged. It is usually a good practice to try out multiple models and compare their performance to select the best one for the problem at hand. Considering the size of the dataset that could be computationally expensive to train and test depending the model chose, the following models were explored:

- **Logistic Regression:** simple, interpretable, and computationally efficient, and it can serve as a good starting point for model selection.
- **Random Forest:** known by perform well with large datasets and high-dimensional feature spaces. It can handle a mix of categorical and numerical features, and it's relatively easy to interpret.
- **Gradient Boosting (XGBoost):** known to perform well with large datasets and can handle a mix of categorical and numerical features. It's more complex than Random Forests, but it can achieve better performance with careful parameter tuning.

RESULTS
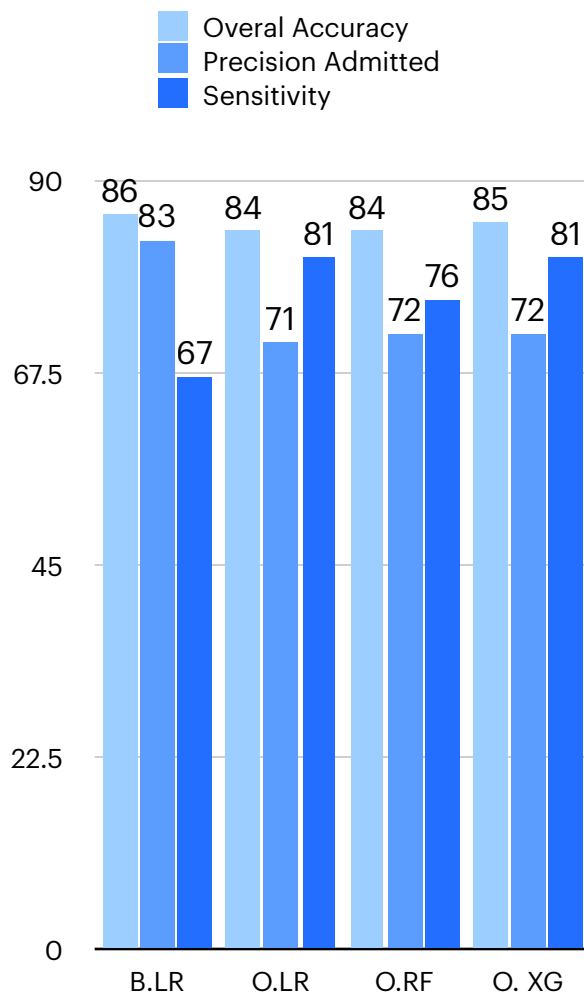
MODELING

About the graph: Opt = Optimized. The values are in percentages.

All three models had almost the same accuracy score of 0.84 (correct predictions over the total number of predictions). However, the precision and sensitivity predicting admitted patients are slightly different for each model.

- **Baseline Logistic Regression (B.LG)** has the lower sensitivity 67% predicting admission. Although, after applied balance data techniques the model sensitivity increased, this means that **Opt. Log. Reg. (O.LR)** could predict 81% of the total admitted patients (18% higher that B.LG).

- **Random Forest (O.RF)** the sensitivity is considerably lower predicting admitted patients (76%) this means the model can be missing important admitted cases, leading to serious consequences for the patients.

- **XGBoost (O.XG)** performs slightly better than logistic regression and random forest, especially regarding sensitivity but the differences are relatively small between logistic and XGBoost.

CONCLUSIONS

MODEL PERFORMANCE

The results suggest that XGBoost performs slightly better than logistic regression and random forest, particularly for class 1 (Admitted patients). Nevertheless, considering interpretability, training time, and computational resources logistic regression is much better, since that XGBoost required more computational resources and longer training time.

FEATURE IMPORTANCE

Predictors with higher coefficients were mainly emergency severity index, chief complains and some medications groups. However, include information about demographics and past medical history improved the model performance, this was demonstrated when reducing the number of variables included in the model reduce the overall model accuracy went down. Additionally, two big groups of data were dropped during preprocessing steps: triage data and past medical history data, for future models include this information have a potential to improve the model performance.

MODEL INTERPRETATION

Overall, logistic regression provides coefficients that can be easily interpreted, while random forest and XGBoost provide importance measures that rank the importance of predictor variables. It could be said that the 3 models have a good degree of interpretability.

GENERALIZABILITY

The fact that the language and clinical practice are standardized offers great advantages when it comes to generalizing these results and expecting similar results using new information.

LIMITATIONS

Hospitals face challenges in integrating disparate data sources, ensuring data quality, and protecting patient privacy when using predictive models. Implementing predictive models can also be disruptive and expensive, requiring infrastructure and technical expertise.

FUTURE DIRECTIONS

The results suggested that increasing the amount of data can improve the predictive performance of a model. However, it is also important to note that a high level of dimensionality in the independent variables can increase model complexity and computational expenses. To address this, future models could consider grouping chief complaints and past medical history features into categories to evaluate model performance and increase the practical applicability of the models in real-world scenarios.

Using machine learning to improve decision-making in hospitals at all levels has the potential to greatly reduce costs related to medical errors and resources distribution, thereby improving not only financial outcomes but also satisfaction among different groups, including patients, healthcare workers, and management agents.

PROJECT URL

https://github.com/CarolMargeth/admission_prediction_project

BIBLIOGRAPHY

[1] Aramide G, Shona K, Keith B, Teresa B. Identify the risk to hospital admission in UK—systematic review of literature. Life (Jaipur). 2016;2(2):20–34.

[2] Rahimian F, Salimi-Khorshidi G, Payberah AH, Tran J, Ayala Solares R, Raimondi F, et al. Predicting the risk of emergency admission with machine learning: Development and validation using linked electronic health records. PLoS Med. 2018;15(11):e1002695. https://doi.org/10.1371/journal.pmed.1002695

[3] Pianykh OS, Guitron S, Parke D, et al. Improving healthcare operations management with machine learning. Nat Mach Intell. 2020;2:266–273. https://doi.org/10.1038/s42256-020-0176-3

[4] Abdulaziz Ahmed, Omar Ashour, Haneen Ali, Mohammad Firouz. An integrated optimization and machine learning approach to predict the admission status of emergency patients. Expert Syst Appl. 2022;202:117314. https://doi.org/10.1016/j.eswa.2022.117314.

[5] De Hond A, Raven W, Schinkelshoek L, Gaakeer M, Ter Avest E, Sir O, Lameijer H, Hessels RA, Reijnen R, De Jonge E, Steyerberg EW, Nickel CH, De Groot B. Machine learning for developing a prediction model of hospital admission of emergency department patients: Hype or hope? Int J Med Inform. 2021;152:104496. https://doi.org/10.1016/j.ijmedinf.2021.104496.