Predicting the Length of Hospital Stay with Machine Learning

Emma Topolovec

Department of Computer Science, Binghamton University

CS 301: Ethical, Social, and Global Issues in Computing

Prof. Weinschenk

December 5, 2022

Abstract

Machine learning models can predict how long a patient stays in a hospital, also known as their length of stay (LOS). Hospitals have used two different algorithms, random forest (RF) and logistic regression (LR), for this application. In multiple cases, including prediction of LOS from heart bypass surgery, cataract surgery, and COVID-19, random forest has exceeded logistic regression in prediction accuracy. The random forest algorithm's ability to accurately process discrete variables allows RF to surpass LR for predicting length of stay. The decision trees of random forest can easily model variables such as a patient's comorbidities, healthcare plan, and sex, while the sigmoid functions of a logistic regression model struggle to accurately model these discrete variables. Hospitals should consider developing random forest models as tools for doctors. Machine learning models can enhance the decision-making skills of medical professionals and hospital administration, leading to better decisions that can improve hospital efficiency and patient health.

Predicting the Length of Hospital Stay with Machine Learning

Knowing how long patients will stay in a hospital can save lives. With information about a patient, machine learning (ML) algorithms such as logistic regression and random forest can calculate the expected length of stay (LOS); however, they differ in effectiveness. Compared to a logistic regression (LR) algorithm, a random forest (RF) algorithm for machine learning has higher accuracy in predicting the length of stay of patients at hospitals, which allows doctors and hospital administration to make informed decisions to improve the health of patients. Throughout multiple comparative studies of ML algorithms, random forest consistently predicts LOS with higher accuracy for different types of surgeries, including cataract and heart bypass surgeries. RF can even predict how long patients will stay in the emergency department. Despite logistic regression's status as a baseline algorithm in many machine learning studies, random forest's methodology allows for better predictions of LOS.

Alternative Technology

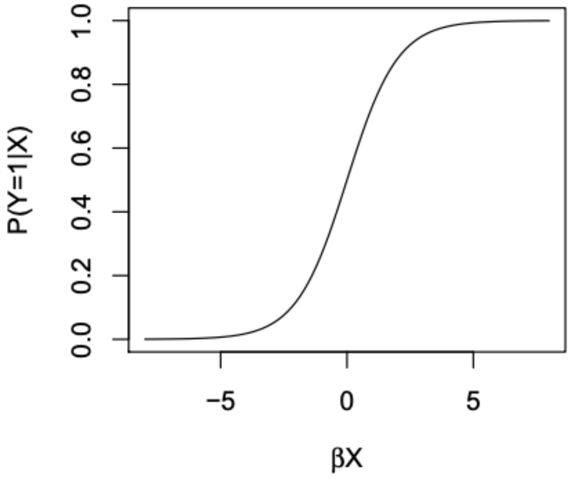
While having distinct methods, machine learning algorithms have the same general structure. They all train with sample data and process input variables to calculate an output. When predicting length of stay, these similarities arise.

To determine LOS, all algorithms use patient data. Hospitals compile past patient information into a training data set. This set consists of a variety of data points about patients, including continuous variables, such as the height, weight, and age of the patient, and discrete variables, such as the sex and healthcare plan of the patient (Profeta, 2021, para. 5). Patient data, along with each patient's length of stay, determine how each ML model will predict future patients' LOS; however, every machine learning algorithm calculates an output differently.

One standard method of ML involves the logistic regression algorithm, which creates and utilizes a specific logistic function called the sigmoid function. As a well-studied algorithm, scientists have used LR as a baseline in ML studies (Etu, 2022, para. 14). For LOS, LR performs by analyzing what lengths of stay occur when inputting training patient data. A logistic curve, seen in Figure 1, models the probability that length of stay exceeds a certain time (Wålinder, 2014, p. 11).

Figure 1

Logistic Function for Logistic Regression

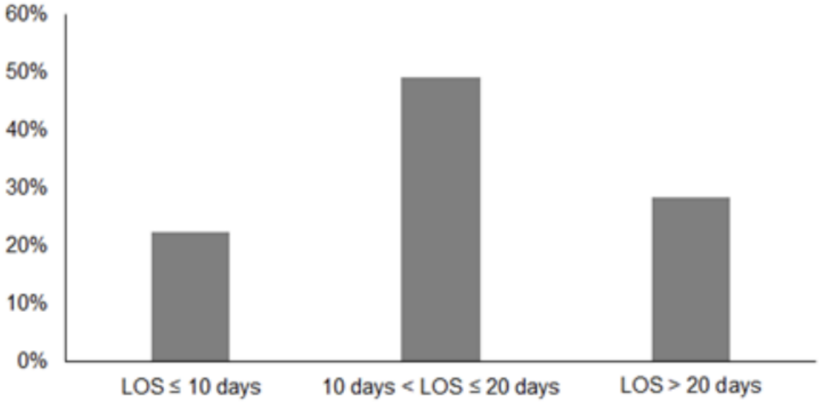


Inputting all variable values from training data into the sigmoid function will create a logistic curve, which can process the data of new patients. Comparing the sigmoid function output of a patient's data to the curve determines that patient's LOS. Suppose a curve models the case if LOS lies above 5 days. An output to the right of the curve would have a LOS that exceeds 5 days, and an output to the left of the curve would have a LOS of 0-5 days. Calculating multiple curves using different output times creates a more detailed model of LOS, allowing the model to predict a more precise time range LOS will fall in. When a new patient enters a hospital for surgery, the different curves calculate the most probable window of length of stay using that patient's information.

Despite having the ability to predict length of stay, logistic regression has flaws. As a logistic function, the sigmoid function uses a strict curve, which can limit possibilities (Wålinder, 2014, p. 15). In some cases, the pattern of LOS contradicts a linear model entirely. Additionally, LR can only calculate discrete, or categorical outputs (p. 15). To predict LOS, a continuous output, a logistic regression model must categorize length of stay into several different set ranges. Figure 2 showcases LOS time ranges of 0-10, 11-20, and 21+ days (Profeta, 2021, p. 3). This reduces the level of detail the algorithm can give. Although these limitations exist, LR still viably predicts length of stay, but it leaves room for improvement.

Figure 2

Distribution of Length of Stay (LOS) of Patients Undergoing Heart Bypass Surgery



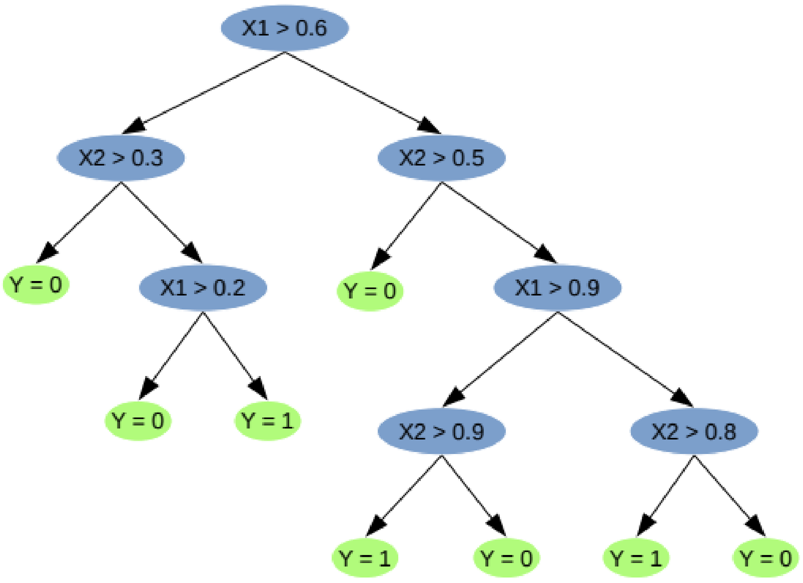
Support

Technical Details

The random forest algorithm takes a completely different approach. The algorithm uses a variety of randomized decision trees to predict future outcomes. Andreas Wålinder (2014), a professional Swedish game developer and former student at Linnaeus University, completed his master's thesis on RF. In the paper, Wålinder describes a decision tree as consisting of several nodes that each split data into two parts based on one of the input variables. Figure 3 shows an example of a basic decision tree (p. 18).

Figure 3

Decision Tree

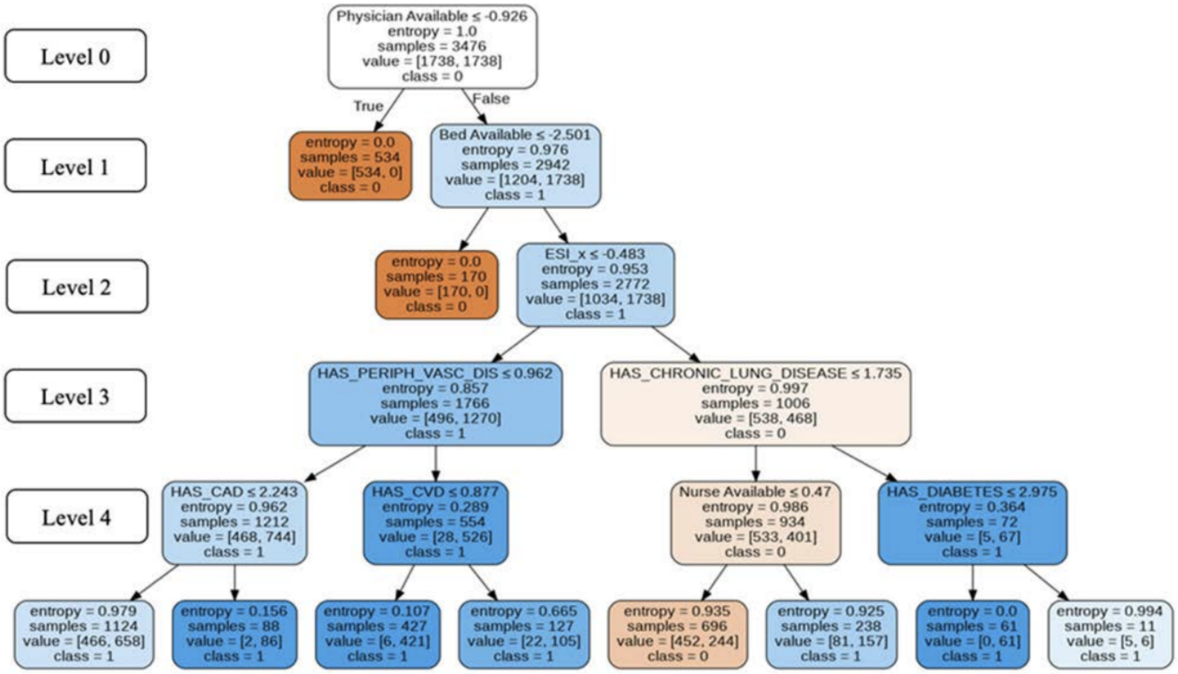


To find the most probable outcome, inputs travel down the tree and narrow down to reach a single decision. In Figure 3, the X1 and X2 variables of the input would be compared to the values in the tree to determine which branches to take until reaching an output value for Y (Wålinder, 2014, p. 18). RF uses multiple of these decision trees, each based on a random set of data points from the training dataset. The trees also use different, random selections of the input variables. When predicting future outcomes, input variables feed into every random decision tree. The most commonly produced decision determines the overall outputted decision (pp. 17-18).

Due to the different decision trees, the random forest algorithm surpasses single decision trees and creates a prediction model that has less variance and higher accuracy. While a single decision tree perfectly models training data, it may overfit the data, leading the tree to develop inaccurate, hyper-specific branches. Several random trees allow for a more general model which tends to predict future outputs with higher accuracy (Wålinder, 2014, p.21).

Figure 4

A Decision Tree for Emergency Department Length of Stay for COVID-19



Like many machine learning algorithms, random forest can model length of stay. The decision trees factor in random sets of discrete and continuous input variables and random subsets of training data from hospitals. Figure 4 shows an example of a decision tree (Etu, 2022, p. 42248). This tree considers the variables of physicians available, beds available, comorbidities, and others, while another tree in the model may only consider the variables of physicians available and patient weight. These two trees would have trained from different, random subsets of the training data. To predict a current patient's LOS, the model inputs the patient's data into every decision tree and outputs the most commonly predicted length of stay.

While random forest has the capacity to create a regression model with a continuous output, it can also produce discrete outputs. For LOS, a random forest model can predict the time range of the LOS. The ability to produce a specific, continuous time may seem desirable; however, separating output into larger time ranges can simplify the model. This also allows for the comparison of RF to algorithms like LR, which can only produce discrete outputs.

RF and LR perform at different levels for different types of applications due to their properties. In LR, every variable in the sigmoid function needs a weight based on how significantly the variable affects the output, but discrete variables rarely allow for the calculation of meaningful weights (Wålinder, 2014, p. 15). This flaw leads to LR falling behind when used with many discrete inputs. LR typically has binary output, so making a prediction between more than two outcomes requires an advanced form of the algorithm called Multinomial logistic regression (p. 15). LR does have benefits, namely simplicity in its methods. The algorithm rarely overfits data; it will make a model that does not rely heavily on training data and will accurately make future predictions (p. 14). Random forest also avoids overfitting through its use of multiple, random decision trees; however, the use of many trees leads to slow performance (p. 21). RF has more versatility, as the algorithm allows for either binary output or the creation of a regression model for continuous outputs (p. 17). For a specific application, a person should choose the algorithm with the most beneficial properties for the data.

When calculating length of stay, a random forest model has more desirable properties. The randomized decision trees of RF surpass the sigmoid functions of LR. While LR may outpace RF in speed, accuracy matters far more than efficiency when finding LOS. LR falls behind in accuracy because LOS involves many discrete inputs, for instance, the type of patient healthcare plan. These inputs need weights for the LR sigmoid function, but inputs like patient comorbidities lack numerical equivalents, limiting the LR model's ability to process these discrete data points. RF can process discrete variables just as well as continuous variables, so the algorithm bypasses this problem. For LOS output, random forest can produce a continuous output or a discrete time range, while logistic regression only outputs LOS in time ranges.

When testing RF's advantages for LOS, researchers have found that RF's accuracy exceeds that of logistic regression. Martina Profeta (2021), a biomedical engineer with a Ph.D. from the Italian Institute of Technology, worked with a team to analyze LOS for heart bypass surgery, concluding that RF has a 96.50% accuracy and LR has an 83.30% accuracy (Profeta, 2021, Table 1). Another study by a team led by Arianna Scala (2021), a biomedical engineer at Ph.D. student at the University of Naples Federico II, analyzed LOS for cataract surgery, placing RF's accuracy at 90.71% and LR's at 86.62% (Scala, 2021, Table 3). Both studies show that a random forest algorithm produces the best results when predicting length of stay. While the researchers only tested two types of surgeries, RF's consistent performance leads to the theory that random forest will outperform logistic regression for most types of surgery. This higher accuracy can make a difference in patient health and hospital management.

Social Impact

The ability to predict how long patients stay at a hospital can increase efficiency and save lives. In 2020, during the beginning of the COVID-19 pandemic, many hospitals became overcrowded. Egbe-Etu Etu (2022), an industrial engineering Ph.D. and researcher at San Jose State University, studied the LOS in hospital emergency departments due to COVID-19. Etu proposes that having a model to predict LOS, especially a random forest model, can help doctors find patients with unusually long lengths of stay, allowing doctors to focus their resources to determine how to reduce LOS for those patients. Millions of people have died from COVID-19, and optimizing hospital management may help prevent future deaths (pp. 42248-42249). With proper knowledge of how long a patient will spend in a hospital bed, doctors and hospital administrators will know how to properly distribute resources such as nurses, medical systems, and medications (p. 42249). By efficiently using a hospital's resources, patients receive treatment faster, which could matter significantly to a patient's health. Reserving enough doctors and nurses on standby can save a patient's life. Analyzing LOS to optimize hospitals allows for an overall increase in patient health.

While many doctors can provide a relatively accurate estimation of LOS, artificial intelligence can surpass that estimation. Doctors may have a mastery of medical issues and treatment, but LOS involves more than just heath. Paperwork, room transfers, and insurance companies can impact the speed at which someone receives treatment. Machine learning algorithms such as random forest factor in many aspects that determine LOS and make computations beyond human capabilities.

Though ML may seem capable of solving every problem, human ability often surpasses algorithms. If a random forest model predicts that a patient will need to stay five days after surgery, doctors should use reasoning before forcing a patient to stay for exactly five days. In reality, a patient may need three days or an entire week to recover. Just like humans, AI lacks perfection. A model may overlook some possible factors for LOS, such as weather or the personality of a patient. The programmer who created the model may miss something that appears obvious to a doctor.

Instead of trusting an algorithm with potentially life-changing decisions, doctors and hospital staff should work in tandem with AI. This process can take an algorithm like random forest to the next level by allowing it to thrive in practical applications. In his TED talk on cooperation between AI and humans, the Chief Operating Officer of Palantir Technologies, Shyam Sankar (2012), explains computer scientist and psychologist J.C.R. Licklider's theory of intelligence augmentation (01:45). Instead of developing technology that replicates humans, Licklider argues that new technology should enhance existing human functions. According to this theory, a random forest model should not replace a doctor. Instead, ML models should act as a tool that enhances the decision-making skills of doctors.

Humans must master the idea of working in harmony with algorithms to unlock the true potential of both man and machine. In a 2012 magazine article about mixing human reason with machine learning, Michelle Vaccaro, an MIT Ph.D. student, and Jim Waldo, the Chief Technology Officer of Harvard University, state "Even if algorithms do not officially make decisions, they anchor human decisions in serious ways" (p. 105). If doctors let a LOS prediction algorithm completely dictate decisions regarding patients, the flaws of the algorithm will appear. Voccaro and Waldo point out that ML can harm people if treated as a magic solution. Hospitals should integrate AI into their systems without removing the oversight of doctors. If doctors can learn to use a random forest model as a tool, the combination of humans and machines will surpass the ability of the individual, leading to the betterment of the healthcare system.

Conclusion

The random forest algorithm would clearly benefit hospitals more than logistic regression. Through multiple studies, RF has predicted length of stay with higher accuracy. The algorithm can process discrete variables, which appear often in calculating LOS, without encountering the same problems as an LR model. This heightened accuracy matters crucially for length of stay, allowing medical professionals and hospital administration to make the most informed decisions to maximize the health of patients.

Medical and computing professionals need to develop and use machine learning models for the betterment of medicine. Random forest models for LOS, when used in tandem with doctors and hospital staff, can save lives. Any medical professional could easily use a well-developed ML model as a tool to enhance their abilities and improve the health of all. The incorporation of artificial intelligence into the medical field needs to increase; doctors and programmers should work together to develop these technologies to better the world.

References

Etu E., Monplaisir L. Arslanturk S., Masoud S., Aguwa C., Markevych I., & Miller, J. (2022, April 18). Prediction of length of stay in the emergency department for COVID-19 patients: A machine learning approach. *IEEE Access, 10*, 42243-42251. 10.1109/ACCESS.2022.3168045

Profeta, M., Ponsiglione, A. M., Ponsiglione, C., Ferrucci, G., Giglio, C., & Borrelli, A. (2021). *Comparison of machine learning algorithms to predict length of hospital stay in patients undergoing heart bypass surgery*. Paper presented at BECB 2021: 2021 International Symposium on Biomedical Engineering and Computational Biology, Nanchang, China. 10.1145/3502060.3503625

Sankar, S. (2012). *The rise of human-computer cooperation*. [Video]. TED. https://www.ted.com/talks/shyam_sankar_the_rise_of_human_computer_cooperation

Scala, A., Trunfio, T. A., Lombardi, A., Giglio, C., Borrelli, A., & Triassi, M. (2021). *A comparison of different machine learning algorithms for predicting the length of hospital stay for patients undergoing cataract surgery*. Paper presented at BECB 2021: 2021 International Symposium on Biomedical Engineering and Computational Biology, Nanchang, China. 10.1145/3502060.3503647

Vaccaro, M., & Waldo, J. (2019, November). The effects of mixing machine learning and human judgement. *Communications of the ACM*, *62*(11), 104-110. https://cacm.acm.org/

Wålinder, A. (2014). *Evaluation of logistic regression and random forest classification based on prediction accuracy and metadata analysis*. [Master's Thesis, Linnaeus University]. DiVA. https://www.diva-portal.org/smash/get/diva2:724982/FULLTEXT01.pdf