Milliman **MedInsight**

Machine Learning for High Risk Patient Identification

Jason Altieri

The healthcare landscape has been shifting toward value-based care since the passage of the Patient Protection and Affordable Care Act (ACA) in 2010. In addition to the growth in Medicare accountable care organizations (ACOs), more and more commercial payers and some Medicaid programs have started using value-based contracting to pay providers. For example, UnitedHealth has reached 60% of its payments going toward value-based programs.[1] This shift has resulted in a need for better ways to identify high-risk patients in order to prevent potentially unnecessary costly services rendered in the emergency room and inpatient settings.

Traditional Methods for Identifying High Risk Patients

Traditionally, identification of high-risk patients has been based on their prior-year costs, risk scores, or highest-risk condition. The prior year's cost and identification of conditions are both naive methods, while many risk scores are developed using classic linear regression techniques. Note that, from this point forward, "linear models" refers to the classic linear regression. Take, for example, the Hierarchical Condition Category risk score models of the Centers for Medicare and Medicaid Services (CMS-HCC) or the U.S. Department of Health and Human Services (HHS-HCC), which are most commonly used in public programs in the United States. These scores are designed to predict future costs based on a patient's demographic factors and clinical conditions, computed using linear modeling.

The strengths of linear modeling include:

- **Interpretability:** Model coefficients provide a clear way to identify the drivers behind the model output.
- **Longevity:** The use of regression models has been a common and successful practice for a long time in a variety of industries.
- **Familiarity:** Linear models are commonly taught in school and are relatively ubiquitous. This familiarity leads to a greater understanding of how and why these models work.

These features make it easier for different stakeholders to reach consensus in situations such as capitation rate setting, business analysis, or policy setting. People are also more likely to trust the results of a model that they feel comfortable explaining.

In comparison to machine learning models, the weaknesses of linear modeling include:

- **Unknown Relationships:** Linear models rely on explicitly identifying relationships and complex relationships in the data can be missed. Most modern machine learning models can identify relationships in the data that are unknown at the time the model is specified.
- **Handling Nonlinearity:** In most real-world applications the relationships between variables are not strictly linear. Classic linear models do not have the inherent structure to address nonlinearity without additional work. For instance, the CMS-HCC risk scoring model includes two-way and three-way interaction terms in addition to the binary condition category indicators. Most modern machine learning models can identify and account for these nonlinearities automatically.

1 https://www.forbes.com/sites/brucejapsen/2018/03/05/unitedhealth-group-hits-64-billion-in-value-based-care-spending/#5505b8c43581.

Figure 1: Linear Fit on Nonlinear Data

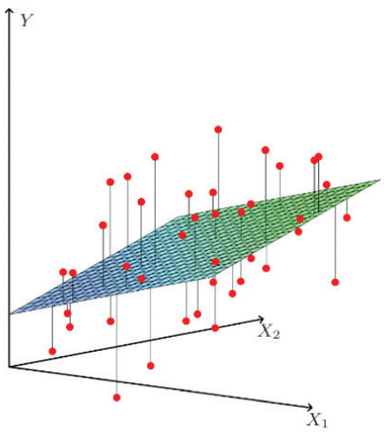
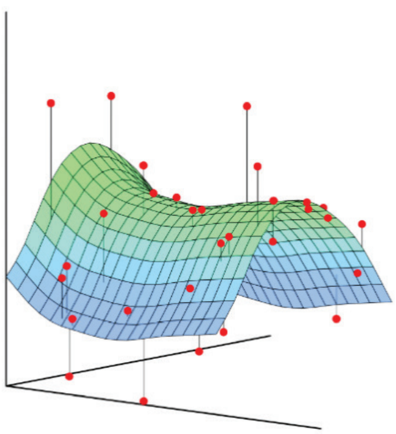


Figure 2: Improved Fit by Accounting for Nonlinearity



- **Multicollinearity:** Classic linear models can easily break down and generate fragile coefficients that don't generalize well in the presence of multicollinearity. Most modern machine learning algorithms can responsibly learn from two features that are highly correlated with each other.
- **Curse of Dimensionality:** A classic linear model usually requires at least 10 times more training observations than features to approach a healthy fit. Most modern machine learning algorithms have a way to smoothly constrain the complexity of their fits. This allows them to produce useful results even when there are far more features than training observations (e.g., DNA markers).

Healthcare has some characteristics that can highlight these weaknesses. First, the availability of data can vary substantially among sources. In some cases, rich claims history along with demographic and electronic health record data may be available. In other cases, there may only be sparse claims data available. The number of potential different combinations of conditions, enrollment, lab values, socioeconomic information, and other data points makes it challenging to identify all the useful relationships among the data points. The inability of classic linear models to identify relationships among the multitude of data points means that their predictive power is limited. The large number of potentially available variables can also make feature selection and pattern recognition a time-consuming process in order to avoid issues with multicollinearity and the curse of dimensionality.

Risk scores in capitation rate setting are intentionally limited to only using medical condition and demographic information. Factors such as the prior year's cost and utilization are not used in risk scoring due to policy concerns. As a result, risk scores are less effective at identifying risk *at the individual level*. Risk scores may fail to accurately capture nonlinearity and have prediction biases *at the individual level*, though they are nonetheless effective in aggregate at the population or subpopulation level. If, however, we try to determine the best *individual* patients to intervene with in a value-based care setting, the ability to better predict single patient outcomes becomes more important. It should be noted that many commercial risk adjusters are currently researching and incorporating non-linear machine algorithms into their models.

There are methods that can be applied to classic linear models (penalization, splines, principal component analysis) to help reduce the issues raised above; however, applying these methods significantly increases the complexity of the model. This reduces the benefits (interpretability, simplicity) that make linear models appealing to begin with. In addition, once many of these techniques are included, the model complexity reaches the point where other types of models should also be considered.

Benefits of Machine Learning Models

Machine learning methods offer an alternative set of methodologies compared to classic linear models. "Machine learning" refers to models such as gradient boosting machines, random forests, neural networks, and other models with less explicit definition of the relationship between input factors and outcomes.

Machine learning models excel in addressing some of the weaknesses of classic linear models. They are good at handling nonlinearity and implicit relationships among input variables. Additionally, many of these models automatically handle feature selection. Feature selection is the machine learning analog of variable selection in classic linear models. Feature selection is the process in machine learning that determines what variables and relationships should be included in the model, just as variable selection is the process in linear models to choose the variables and interactions to include. Many machine learning models have built-in feature selection algorithms that allow the developer to provide as many input factors as possible without needing to explicitly define relationships or carefully curate the variables that are included. This is helpful in healthcare analytics where there are a variety of different variables and complex relationships present in the data.

These benefits come with the trade-off of added algorithmic complexity and harder-to-interpret results. However, many of the models do have algorithms to estimate feature importance, which can aid in understanding how the model arrived at its results.

Using Machine Learning to Identify High-Risk Patients

Machine learning techniques offer the ability to use a variety of features (inputs) to predict outcomes of interest in the identification of high-risk patients:

FEATURES	OUTCOMES
Claims Data	Patient Costs
Enrollment Data	Preventable Admissions
Socioeconomic Data	Medical Record Gaps
Electronic Medical Record Data	Readmission Risk
Admit, Discharge, Transfer Data	

One example of applying machine learning in healthcare is the Milliman PRM Analytics® potentially avoidable cost predictions.[2] These predictions use a gradient boosting machine to identify the patients who are most likely to incur costs in the next six months that could potentially be avoided through care management. Below are a couple of examples from this model that highlight the advantages of machine learning in patient-level predictions.

First, in many populations, the model identifies older patients who have disengaged with the healthcare system as of significantly higher risk than patients who are the same age but have remained engaged. Patients are disengaged when they have not seen a physician in over a year despite being elderly or having been diagnosed with chronic conditions in the past. These patients may or may not have continued to fill prescriptions; however, they are not actively engaged with a physician to monitor their health status. The model picked up that these patients often see a spike in potentially avoidable services following a period of disengagement. The initial reaction from care coordinators to a high potentially avoidable cost prediction for these patients is often "Why is this person projected to be expensive? They incurred no costs last year!" After examining the clinical profile and behavioral patterns of the patients identified by the model, such as the aforementioned spike-disengagement pattern, the risk becomes apparent. Milliman's machine learning models identified hidden relationships that may be overlooked in traditional linear models, unless the linear model developer had prior knowledge about the hidden relationships.

In another case, the model identified an end-stage renal disease (ESRD) patient with a large amount of potentially avoidable costs. Typically, ESRD patients have little potentially avoidable cost due to the fact that, while treatment is expensive, it is medically necessary. In this case, the patient had been delaying treatment until an emergency and receiving dialysis in the emergency room, rather than going to a scheduled outpatient facility appointment. After reaching out, the provider organization determined that the patient did not like their current dialysis provider and was able to save thousands of dollars a month by shifting the patient to a new facility. The ability of the machine learning algorithm to absorb a large number of features and identify latent relationships allowed the model to find a meaningful outlier patient.

2 https://www.prmanalytics.com/assets/white_papers/White%20Paper%20-%20Milliman%20PRM%20Technical%20Process.pdf.

Conclusion

Machine learning models have several advantages over classic linear models when it comes to patient-level predictions that are important in a value-based care framework. Provider organizations can use machine learning models to help identify the members who would benefit most from active care management in an effort to improve patient outcomes and reduce costs under value-based contracting.

CONTACT

Jason Altieri
jason.altieri@milliman.com