panalgo

Machine Learning: Why It's an Essential Tool in the RWE Analytics Toolbox

The investment in real-world data (RWD) shows no signs of slowing, further intensifying the demand to crunch larger datasets and generate novel insights through its analysis. Augmenting traditional statistical techniques with advanced analytic techniques, like machine learning, has become more crucial as the data grows and competition increases. In a recent survey of life sciences executives, 95 percent of respondents said they expect to use machine learning in the coming years to glean real-world evidence (RWE) from the growing volume of data.[1]

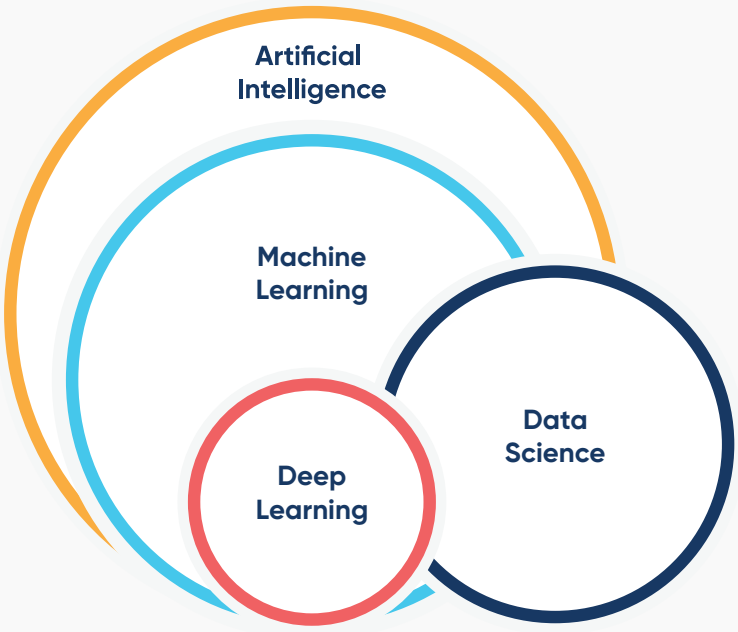
A number of life sciences organizations have started to implement machine learning and invest in advanced analytics to generate RWE and drive competitive advantage. And yet, despite its growing use, there's still uncertainty around machine learning's full potential and the best ways to employ it.

This white paper will examine the impact of machine learning for life sciences Health Economics and Outcomes Research (HEOR) and RWE teams and why it is becoming an essential tool in the analytics toolbox.

Understanding Machine Learning

Machine learning is a subfield of artificial intelligence that continuously learns from patterns in massive amounts of data and uses those patterns to make predictions as new data is introduced.[2] The ability to learn and continually improve through experience is key to the power of machine learning, which can now be programmed to handle specific tasks, analyze large amounts of data and provide results that are crucial for effective decision-making.

Data science is the inter-disciplinary field that uses scientific methods, processes, and algorithms to understand and analyze phenomena in data. In other words, it is used to extract knowledge and insights from data. Deep learning is a subset of machine learning that uses neural networks to compute predictions.



Machine learning is a powerful tool in the analytics toolbox. Many statistics and machine-learning methods can be used for both prediction and inference. Statistical methods have a long-standing focus on inference, achieved through creating and fitting a project-specific probability model. Alternatively, machine learning concentrates on prediction, by using general purpose learning algorithms to find patterns in often rich and unwieldy data with minimal assumptions.

Key Machine Learning Algorithms

Machine learning algorithms are typically broken out into two main categories.[3]

Supervised Learning

the most prevalent form of machine learning - the data contains labels that tell the machine exactly what patterns to identify.[4] The goal is to find and approximate these patterns so that when new data is introduced, output variables can be predicted. Examples include gradient-boosted trees, random forest, k-nearest neighbors, support vector machines and regularized regression methods.[5]

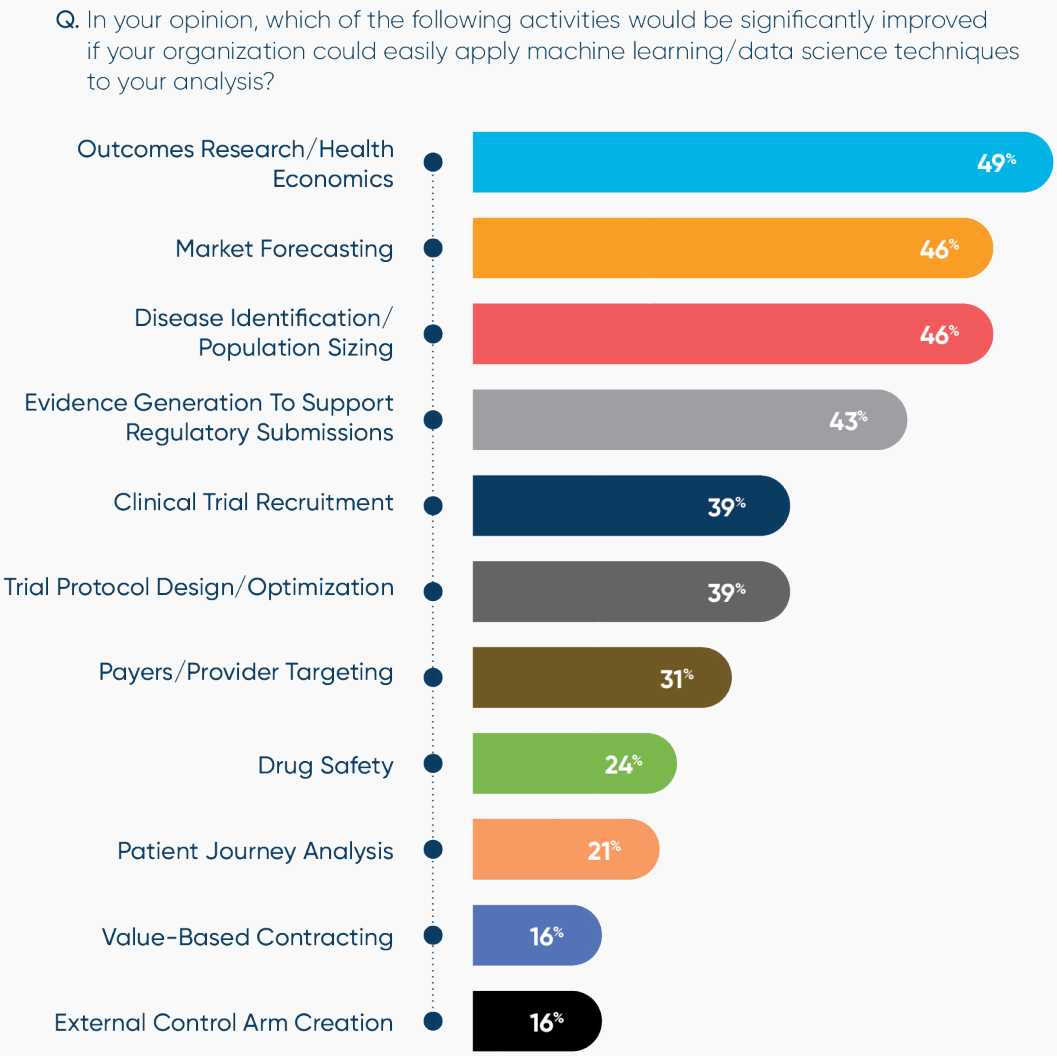
Unsupervised Learning

In unsupervised learning, the data has no labels and the machine is told to look for any pattern it can find.[6] Underlying or hidden structure in the data is modeled to learn more about the data. Examples include principal component analysis[7], hierarchical or k-means clustering.[8]

Current View of Machine Learning

In Panalgo's 2021 Benchmark Report, we asked 100 analytics leaders across global life sciences to identify areas they thought would be improved with the use of machine learning. The report identified eleven areas where machine learning techniques would make a significant impact, including health economics and outcomes research as the number one area that would see significant improvement from machine learning (49%). Respondents also cited the areas of market forecasting (46%), disease identification/population sizing (46%), evidence generation for regulatory submissions (43%), clinical trial recruitment (39%) and trial protocol design/optimization (39%).[9]

Figure 16. Activities that would be significantly improved if machine learning could be applied



Source: *2021 Panalgo Benchmarking Report*

Another industry report estimates that big data and machine learning in pharma and medicine could generate up to $100 billion annually in new revenue and cost savings thanks to better decision making, optimized innovation, improved efficiency of research, enhanced clinical trials and the development of new tools.[10]

The Impact on RWE and HEOR Teams

Predicting high-cost patients for certain diseases, identifying risk factors for hospitalization or ICU utilization, and predicting health outcomes are all essential research goals used to lessen the burden on the healthcare system and lower costs. Traditionally, researchers have used linear regression models to analyze these trends, but with today's sheer volume of data, the high number of predictors in any given research project may not be suited to traditional analytical techniques.

Machine learning can be used to:

- Generate predictions and novel insights from large, unwieldly datasets
- Manage massive amounts of secondary data from dissimilar sources
- Predict patients with undiagnosed disease
- Identify patients likely to develop a disease
- Uncover possible adverse events and predict clinical events such as readmissions or relapses
- Predict patient milestones (e.g. A1C reductions) and disease activity scores

CASE STUDY

Predicting Inpatient Relapse in Multiple Sclerosis Patients Using First-Line Disease Modifying Therapies

Challenge

Nearly a million people are currently living with multiple sclerosis (MS) in the United States.[11] The annual cost to treat this disease ranges from $57,000 to $62,000. Relapse among MS patients is associated with disability progression and worsening outcomes. Understanding the drivers of relapse can inform improved management of the disease. Gaining a more complete understanding of these drivers requires creating an evidence-based, data-driven decision rule to discriminate between patients with and without an inpatient relapse.

Approach

Machine learning, with its capacity to analyze huge amounts of data quickly while optimizing predictive performance, is uniquely positioned to discover novel insights that can augment MS care management. Using claims data and Panalgo's machine learning module, IHD Data Science, the study examined patient characteristics to identify predictors of inpatient MS relapse. The IHD Data Science module provided the ability to rapidly and efficiently test different models enabling researchers to measure performance across six different model types: XGBoost, Random Forest, Neural Network, Regularized Logistic Regression, Support Vector Machine, and Logistic Regression. Features included demographics, comorbidities, concomitant medications, healthcare resource utilization, disease modifying therapies (DMTs), route of administration and proportion of days covered for DMTs.

Results

The XGBoost ML model had the strongest ability to predict inpatient MS relapse compared to all models tested. Identified predictors of relapse included previous inpatient or emergency room visit with an MS diagnosis, the number of MS-related encounters, use of home care services and durable medical equipment, epilepsy / convulsions, paralysis, urinary tract infections, potential medication side effects (nausea and vomiting) and use of muscle relaxants, anticonvulsants and antidepressants. The values for the predictors that best identified MS inpatient relapse from the model were used to derive an actionable decision rule. Specifically, the model demonstrated that MS patients are more likely to have a relapse if they 1) have 30 or more unique comorbidities, or 2) have a previous emergency room visit with an MS diagnosis and 10 or more previous MS related encounters or 3) have 20 or more previous MS related encounters.

Impact

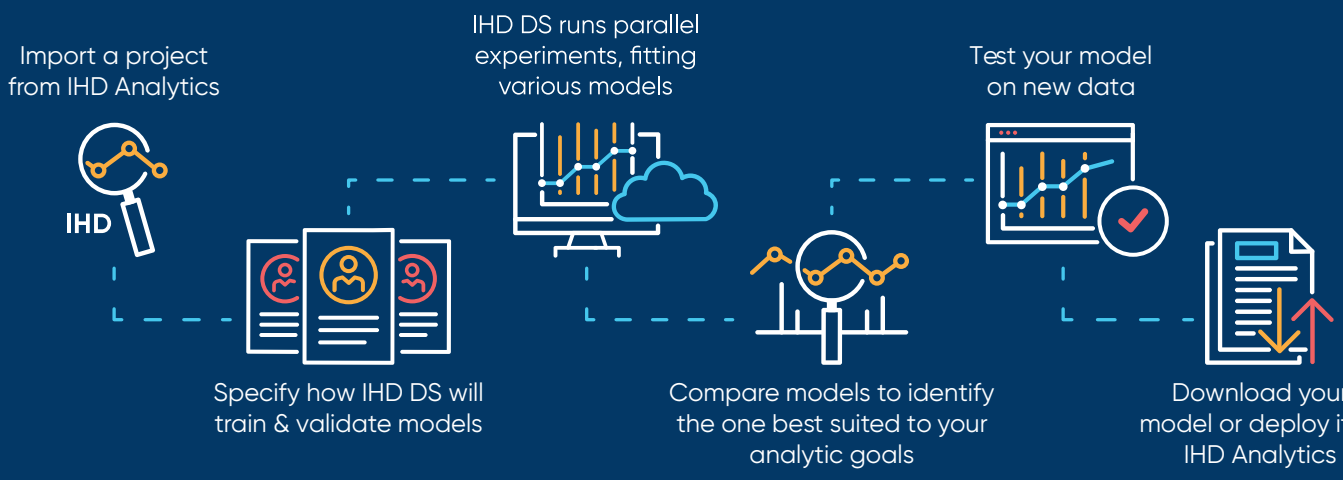
The IHD Data Science module allowed researchers to assess the patient's full diagnosis history and evaluate a myriad of factors that may be related to MS relapse. Understanding the drivers of MS relapse beyond well-known factors allows life sciences to more accurately target patients for therapy and gives providers the ability to identify novel subgroups of patients who are at high risk for relapse and improve disease management.

Conclusion

As the volume of RWD continues to grow at a rapid pace, tools like machine learning are becoming increasingly crucial to learn from the data and provide predictive insights.

Panalgo's Data Science module brings the power of machine learning to life sciences researchers, allowing them to leverage the full potential of data and work effectively with a large number of covariates, identify predictors and perform advanced analysis using a variety of machine learning techniques all without the need for custom programming.

How IHD Data Science (DS) Works



The module is built on Panalgo's IHD platform and provides a single environment to easily train, validate and test models against multiple datasets, as well as allow users to seamlessly expand analytics projects to generate new findings and drive product success.

References

1 *Mission Critical, Biopharma companies are accelerating real-world evidence adoption, investment, and applications.* Deloitte Insights, 2018

2 https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/

3 *Commonly used Machine Learning Algorithms (with Python and R Codes),* Sunil Ray, *Analytics Vidhya,* September 9, 2017

4 https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/

5 *Top 10 Machine Learning Algorithms List (2021 Updated),* MIT-ADT University, MIT Centre for Future Skills Excellence

6 https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/

7 *A One-Stop Shop for Principal Component Analysis,* by Matt Brems, Towards Data Science, April 17, 2017

8 *Unsupervised Learning: K-means vs Hierarchical Clustering,* by Valentina Alto, Towards Data Science, July 8, 2019

9 *Panalgo State of Healthcare Analytics Benchmarking Report: Data Analytics and Machine Learning in Life Sciences*

10 *How big data can revolutionize pharmaceutical R&D,* Jamie Cattell, Sastry Chilukuri, and Michael Levy, McKinsey & Company, *Pharmaceutical & Medical Products,* April 2013

11 National MS Society, https://www.nationalmssociety.org/What-is-MS, Accessed May 7, 2021

Learn how IHD streamlines healthcare data analytics and empowers you to generate and share trustworthy results faster at:

panalgo.com/contact

panalgo

Panalgo provides software that streamlines healthcare data analytics by removing complex programming from the equation. Our Instant Health Data (IHD) software empowers teams to generate and share trustworthy results faster, enabling more impactful decisions. To learn more visit us at www.panalgo.com