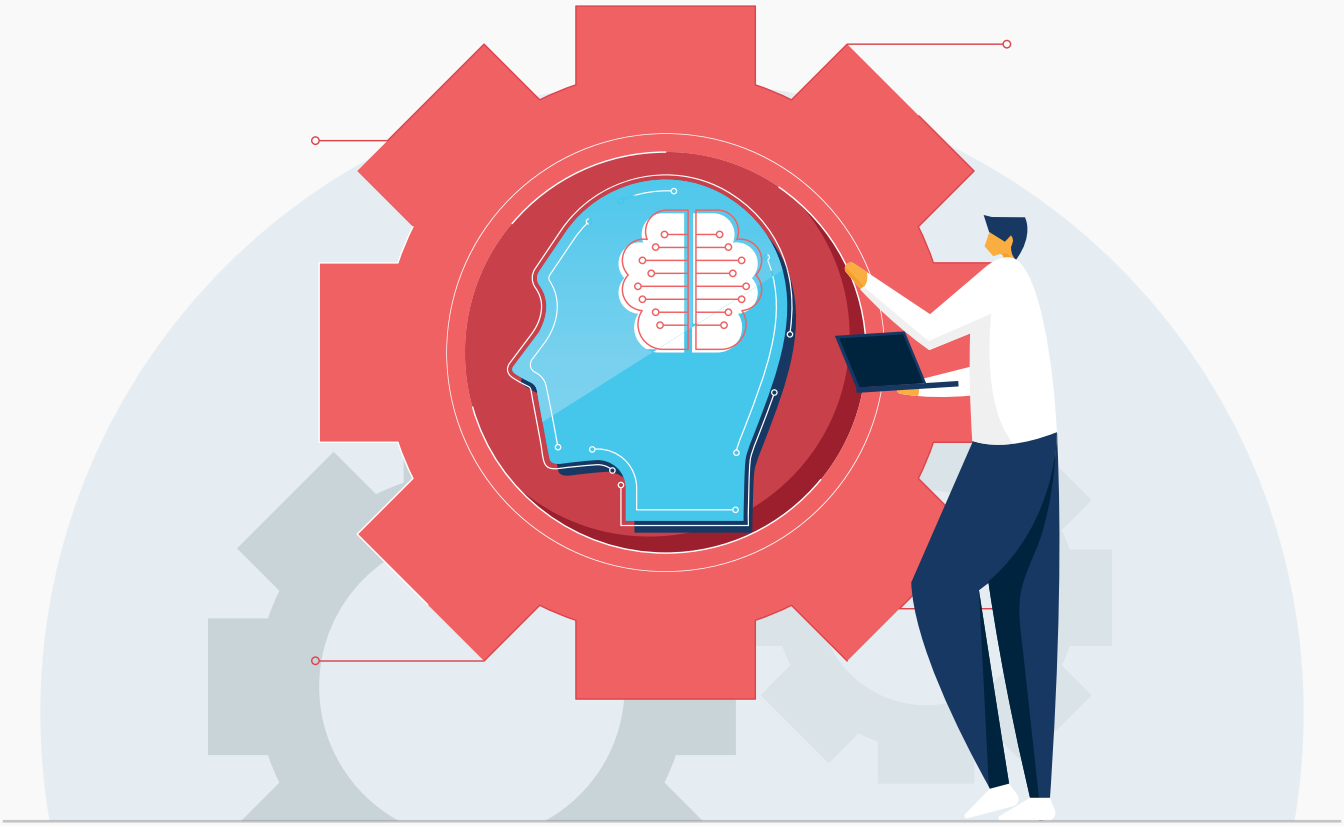
panalgo

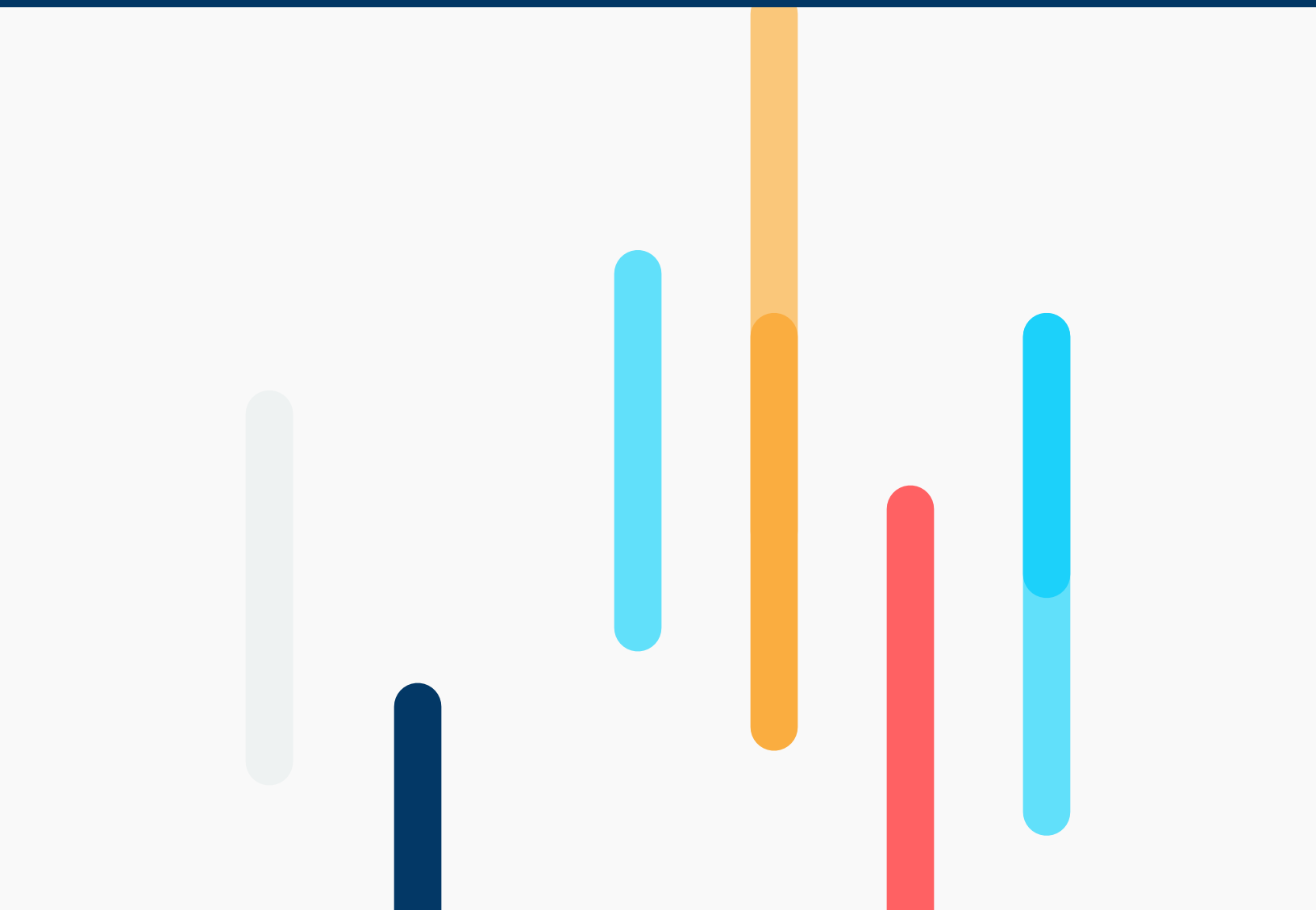


VOLUME #2

How Machine Learning is Being Leveraged to Analyze Real-World Data

A Collection of Recent Research and Use Cases

Table of Contents



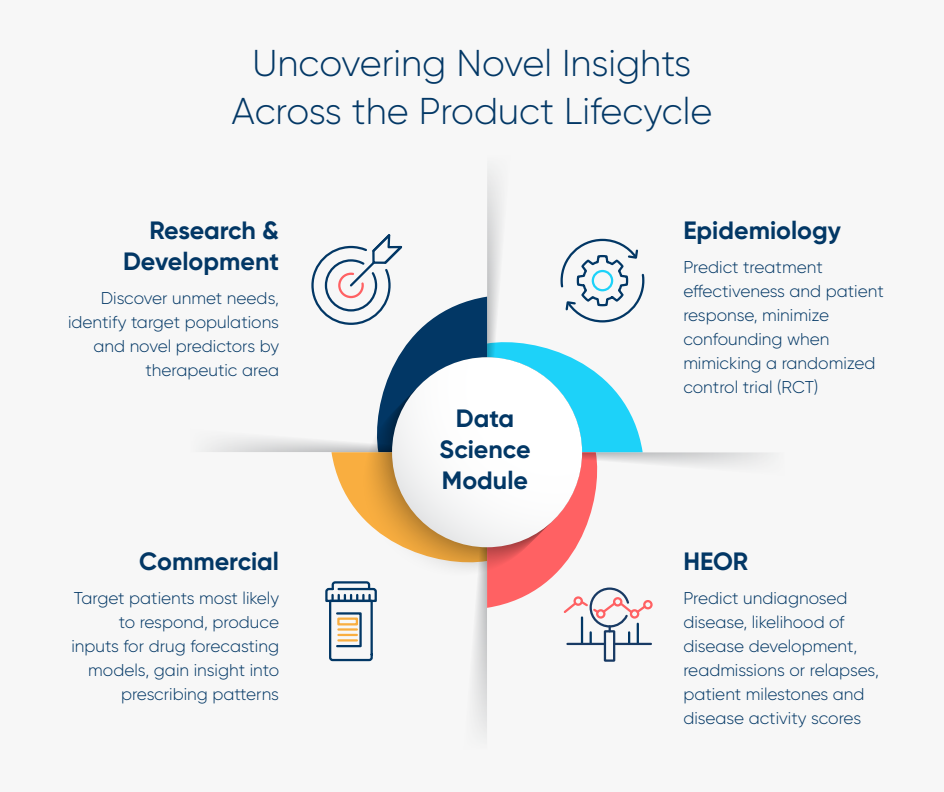
INTRODUCTION

In our first edition of *How Machine Learning is Being Leveraged to Analyze Real-World Data*, we discussed how machine learning is rapidly becoming an essential tool in the analytics toolbox by showcasing 12 recent use cases where machine learning was employed to analyze real world data (RWD).

In this edition we look at an additional 12 use cases that further demonstrate the power and growing acceptance of machine learning. As healthcare data continues its exponential growth, machine learning offers an opportunity to leverage the full potential of data, work with a large number of covariates, identify predictors and perform advanced analysis.

Machine Learning Across the Entire Product Life Cycle

Machine learning can be employed across the product life cycle to predict everything from patient response to treatment effectiveness to undiagnosed disease.



Current View of Machine Learning

In a recent survey of life sciences executives, 95% 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. Additionally, respondents in the Panalgo 2021 Benchmarking Report reported a positive outlook on the benefits of machine learning and data science. Two-thirds (66%) of respondents indicated that outcomes research/health economics would significantly improve with ML, while more than half of respondents reported significant improvements to trial protocol design/optimization (59%), market forecasting (57%), clinical trial recruitment (57%), disease identification/population sizing (56%), and evidence generation to support regulatory submissions (52%) with advanced analytics like ML.

To better understand the impact of machine learning, we have compiled the following 12 use case studies across healthcare, including predicting outcomes, analyzing highly skewed data, identifying drug-drug interactions, and more.

PREDICTING TREATMENT OUTCOMES

USE CASE #1

Development and evaluation of a predictive algorithm for unsatisfactory response among patients with pulmonary arterial hypertension using health insurance claims data

Janssen Scientific Affairs, LLC.

Published: *Current Medical Research and Opinion*

Pulmonary arterial hypertension (PAH) is a form of pulmonary hypertension in which blood vessels in the lungs are narrowed, blocked, or destroyed. The condition affects 15–50 persons per million in the U.S., with a higher prevalence among women and a mean age at diagnosis of about 50 years. Researchers at Janssen Scientific Affairs designed a retrospective study using RWE and ML to develop a predictive algorithm for unsatisfactory treatment response among patients with PAH who had just started their first PAH therapy.

Approach

Researchers identified adult patients with PAH who were newly initiated on their first PAH therapy using health claims from Optum's de-identified Clinformatics Data Mart Database. They developed an algorithm using a tree-based machine learning strategy called random survival forests. This type of algorithm can be used to build predictive models based on right-censored survival data. From a group of 500 survival trees, the algorithm predicted the probability that a patient would not have an unacceptable response at a given point after that month. Researchers then averaged the predictions of the 500 survival trees to create a prediction for the random survival forest.

Results

A total of 4,781 adult patients with PAH were included in the study sample. The random survival forest algorithm identified the 20 most important risk factors including PAH-related outpatient visits, pulmonologist visits, and days since PAH diagnosis. The algorithm demonstrated a good ability to predict unsatisfactory response to initial PAH therapy (C-statistic: 0.732); the algorithm had 82% sensitivity and 55% specificity at 12 months.

Conclusion

The machine learning algorithm reliably identified those patients at higher risk of unsatisfactory response following the initiation of PAH-specific therapy. Using the best predictors, researchers developed a simplified risk score containing seven variables as a generalizable tool for discovering high-risk patients who might be candidates for combination therapy. The full algorithm and simplified risk score could have many uses in practice, particularly in providing earlier combination therapy for higher-risk patients or helping payors identify patients at risk before their treatment escalates.

USE CASE #2

Machine Learning Prediction of Treatment Outcome in Late-Life Depression

UCLA

Published: *Psychiatry*

Late-life depression (LLD) is a common disorder among the elderly and is often associated with poor quality of life, increased risk of cognitive decline, and increased risk of suicide. Recent evidence suggests that including data that spans different types and contexts (imaging, texts, or genetics) can enable machine learning algorithms to better predict the outcome of treatment for LLD. In this study, researchers at UCLA compared the predictive performance of three machine learning models using differing combinations of sociodemographic characteristics, baseline clinical self-reports, cognitive tests, and structural magnetic resonance imaging (MRI) features to predict the outcomes of treatment in patients with LLD.

Approach

Researchers combined data from two clinical trials conducted with adults aged 60 and older who suffered from depression, including response to escitalopram and Tai Chi. They defined remission as a score of 6 or less on the 24-item Hamilton Rating Scale for Depression (HAMD) at the end of 24 weeks of treatment. They constructed subsets of features using baseline sociodemographic and clinical features and / or gray matter volumes (GMVs). The team then compared three machine learning classification algorithms: (1) Support Vector Machine-Radial Bias Function (SVMRBF), (2) Random Forest (RF), and (3) Logistic Regression (LR).

Results

Of the three algorithms studied, the combined feature set determined by the RF and SVMRBF algorithms performed better than the clinical and GMV feature sets with a final cross-validated AUC of 0.83 ± 0.11 and 0.80 ± 0.11, respectively, and both passed the permutation analysis. The LR was the best performing algorithm using GMV features alone (AUC 0.79 ± 0.14) but did not pass permutation analysis using any feature set. The three classifiers performed significantly different for all three features sets. The researchers also identified anterior and posterior cingulate volumes in the brain, depression characteristics and self-reported health-related quality scores as important predictive features of treatment response.

Conclusion

The results of this study – one of the first to use ML and multi-modal data to uncover predictors of treatment response in LLD – shows that including clinical and structural MRI features dramatically increases predictive capability. The features identified in the studies are among those that have been found previously in geriatric depression, meaning further studies in the area are warranted. Furthermore, the results suggest that machine learning coupled with multi-modal data may aid in the development of a non-invasive, precision approach to the management of LLD.

USE CASE #3

Prediction of Non-Response to First-Line Methotrexate Treatment in Rheumatoid Arthritis: A Real-World Data Analysis Using Machine Learning

Panalgo & OM1

Published: *Science Direct*

Methotrexate (MTX) continues to be the anti-rheumatic drug first prescribed for the treatment of rheumatoid arthritis (RA), but response to the drug varies. Predicting which patients will not respond to MTX would greatly benefit clinicians who could then seek alternative or additional medication at an earlier stage of the disease and provide a more effective treatment plan for patients. Researchers at Panalgo and OM1 conducted a study to help identify predictors of non-response to methotrexate (MTX), using a longitudinal clinical cohort of RA patients and machine learning methods.

Approach

Researchers partitioned the data using a 75%/25% split to train, validate and test the predictive model with 3-fold cross validation. They considered a wide range of models including traditional and regularized logistic regression, XGBoost, support vector machine, random forest, and feed-forward neural network models. The best model was selected using the area under the ROC curve (AUC) and average precision, Brier score, accuracy, recall, precision, F1 score, negative predictive value and specificity were assessed.

Results

The study of 6,648 patients revealed the top predictors of MTX non-response included increasing baseline clinical disease activity index (CDAI), use of GABA analogs, analgesics, oral steroids, glucocorticoids, anxiolytics or sedatives, history of mood disorders, younger age and being female. Being from the Northeast appeared protective of MTX non-response. The inclusion of CDAI, a score not commonly available in claims or general purpose EHR datasets, contributed significantly to the model performance.

Conclusion

The study identified predictors of MTX non-response with strong predictive accuracy using machine learning. Further research is needed to better understand the potential role of comorbidities, like mood disorders, and their management on treatment success.

PREDICTION OF DISCHARGE STATUS AND READMISSIONS

USE CASE #4

Prediction of Discharge Status and Readmissions after Resection of Intradural Spinal Tumors

Stanford University School of Medicine

Published: *Neurospine*

Intradural spinal tumors are uncommon types of tumors, comprising about 8% of all central nervous system tumors. Surgeries to resect these tumors are complicated and nationwide studies suggest the complication rates could be as high as 18%, ranging broadly from wound hematoma and hemorrhage to urinary and pulmonary issues. Most research studying the connection between clinical characteristics and surgical outcomes have been limited both in size and diversity, minimizing interpretability across large, diverse healthcare systems and geographical regions. Researchers at Stanford University School of Medicine decided to leverage machine learning tools to conduct a study to better predict postresection outcomes for patients with spinal tumors.

Approach

Using the IBM MarketScan Claims Database, researchers identified adult patients receiving surgery for intradural tumors between 2007 and 2016. Key data points included nonhome discharge, 90-day post-discharge readmissions, hospitalization duration, and postoperative complications. They developed risk modeling using a regularized logistic regression framework (LASSO: least absolute shrinkage, and selection operator) and validated the machine learning model in a subset of data. They used multiple regression and multivariable logistic regression to assess study objectives.

Results

The study identified 5,060 patients who received resection of intradural spinal tumors. The study's machine learning models performed better than models using only tumor-specific or patient-specific features in discrimination of the key factors: AUC for nonhome discharge (0.786); 90-day readmissions (0.693); Brier score for nonhome discharge (0.155); 90-day readmissions (0.093). Patients predicted to be highest risk for nonhome discharge required continued care 16.3 times more frequently (64.5% vs. 3.9%) and patients predicted to be at highest risk for post-discharge readmissions were readmitted 7.3 times as often as those predicted to be at lowest risk (32.6% vs. 4.4%).

Conclusion

Using a varied set of clinical characteristics, researchers created and validated two machine learning risk models for predicting nonhome discharge and post-discharge readmissions. These models performed significantly better than approaches using only tumor- and patient-level characteristics and may be able to help improve precision care in this area.

DEVELOPING PERSONALIZED TREATMENT PLANS

USE CASE #5

Producing personalized statin treatment plans to optimize clinical outcomes using big data and machine learning

Premera Blue Cross, OptumLabs® and the University of Minnesota

Published: *Journal of Biomedical Informatics*

Almost half of Americans 65 years of age and older take statins, an extremely effective treatment for lowering low-density lipoprotein cholesterol, preventing atherosclerotic cardiovascular disease (ASCVD), and lessening overall mortality from all causes. Unfortunately, more than half of these individuals stop taking statins within a year, preventing them from obtaining the drugs' critical benefits. Currently, physicians usually make their decision to prescribe statins based on just a few patient data elements, resulting in a reactive strategy of managing symptoms associated with statins (SAS) after they present. Researchers at Premera Blue Cross, OptumLabs® and the University of Minnesota sought a more forward-looking approach and conducted a proof-of-concept study using a machine learning personalized statin treatment plan (PSTP) platform to find the ideal treatment plan to prevent and minimize the dangers of patients discontinuing their statin treatment.

Approach

To effectively create the PSTP platform, researchers used de-identified administrative claims data from the OptumLabs® Data Warehouse, consisting of medical and pharmacy claims, laboratory results, and enrollment records for more than 130 million commercial and Medicare Advantage (MA) members. The study used a decision plot to characterize the proactive strategy and real-world data to solve the clinical problem of statin discontinuation.

Results

Researchers came up with three results. First, when compared with standard practice, the PSTP platform prescription recommendations show much lower risk of SAS and discontinuation. Second, machine learning's ability to take into account more aspects of data enabled the proactive prescription strategy to perform markedly better, especially the artificial neural network approach. Third, the study highlighted a way of including optimization constraints for personalized medicine and physician/patient shared decision making.

Conclusion

While more research is needed to best determine the clinical impact of the PSTP platform, these encouraging results show that leveraging large data sources using machine learning can help develop personalized healthcare treatment plans and precision-health initiatives.

IDENTIFYING RISK FACTORS FOLLOWING TREATMENT DISCONTINUATION

USE CASE #6

Risk factors associated with skeletal-related events following discontinuation of denosumab treatment among patients with bone metastases from solid tumors: A real-world machine learning approach

Amgen Inc. & University of Pennsylvania

Published: *Journal of Bone Oncology*

Clinical practice guidelines recommend using agents that target the bones for preventing skeletal-related events (SREs) among patients who experience bone metastases from solid tumors. However, there is a lack of real-world data on how individual risk factors can contribute to SREs, specifically for patients who stop using denosumab, a drug used to prevent or treat certain bone problems. Researchers from Amgen Inc. and University of Pennsylvania wanted to pinpoint risk factors associated with the occurrence of SREs after a patient has discontinued denosumab. They used machine learning to help identify patients who had a higher risk of developing SREs when they stopped their denosumab treatment.

Approach

Researchers evaluated patients diagnosed with incident bone metastases from primary solid tumors between January 1, 2007 and September 1, 2019 from the Optum PanTher Electronic Health Record database. They used extreme gradient boosting to develop an SRE risk prediction model evaluated on a test dataset. They also used Shapley Additive Explanations (SHAP) to examine multiple variables as potential factors for SRE risk including patient demographics, comorbidities, laboratory values, treatments, and denosumab exposures. The team also conducted univariate analyses on risk factors with the highest importance from pooled and tumor-specific models.

Results

Of a total of 1,414 eligible adult cancer patients, 80% were assigned to model training and 20% to model evaluation. Researchers evaluated the meaningful model performance by an area under the receiver operating curve score of 77% and an F1 score of 62%. Model precision was 60%, with 63% sensitivity and 78% specificity. SHAP uncovered several key risk factors of SREs following denosumab discontinuation for the tumor-agnostic and tumor-specific models including prior SREs, shorter denosumab treatment, number of clinic visits per month, hospitalizations, and more.

Conclusion

The machine learning approach used to identify SRE risk factors reinforces treatment guidance of continuous use of denosumab and could support clinicians' ability to better evaluate patients' need to continue denosumab and ultimately, improve patient outcomes.

ANALYZING HIGHLY SKEWED DATA

USE CASE #7

Predicting in-hospital length of stay: a two-stage modeling approach to account for highly skewed data

Duke University

Published: *BMC Medical Informatics and Decision Making*

In the early stages of the COVID-19 pandemic, researchers at Duke University were asked to identify which elective surgeries would need the use of additional resources so that they could be delayed to free up resources for pandemic patients. They created and put into operation a clinical decision support (CDS) tool to predict expected length of stay (LOS), intensive care unit needs, ventilator use, and whether or not a patient would have to be discharged to a skilled nursing facility. While this model was beneficial due to the immediate need during the pandemic and did provide helpful information about resource utilization at the time, researchers were interested in gaining a deeper understanding of predicting LOS as a continuous outcome, something the original model couldn't do. In this study, they present their findings after evaluating various models for predicting in-hospital LOS after an elective surgery.

Approach

Using electronic health record data on length of stay from 42,209 elective surgeries at Duke University Health System (DUHS), the researchers compared different loss-functions (mean squared error, mean absolute error, mean relative error) such as algorithms (LASSO, Random Forests, multilayer perceptron) and data transformations (log and truncation). Finally, they evaluated the performance of a two-stage hybrid classification-regression method. First, they generated a classifier to determine whether a patient would have a short or long LOS (stage 1) and then used a random forest regressor and log transformations to determine the precise length of stay among those with a short LOS (stage 2).

Results

The results highlighted both the challenges and considerations necessary when applying machine learning methods to skewed outcomes. The final mean absolute error (MAE) suggests that the model predictions are off by less than 1 day (~ 16 h) for LOS < 4 days and less than 2 days for LOS between 4 and 7 days.

Conclusion

Researchers found that the two-stage hybrid model was the best approach to predicting LOS and that this type of multi-stage machine learning tool could be helpful to clinical support when making scheduling decisions during times of constrained hospital resources. However, while it was able to predict short term LOS with good accuracy, it was not able to predict long term LOS. It should also be noted that this approach may not be effective in other settings. Despite its limitations, findings increase trust in and support use of this CDS tool in DUHS and highlight the considerations when creating CDS tools in other settings.

PREDICTING CLINICAL OUTCOMES

USE CASE #8

Using Machine Learning Applied to Real-World Healthcare Data for Predictive Analytics: An Applied Example in Bariatric Surgery

Johnson & Johnson

Published: *Decision-Analytic Modeling: Past, Present, And Future*

While laparoscopic metabolic surgery (MxS) can result in the remission of type 2 diabetes (T2D), treatment response to the surgery can vary and not all patients experience remission. This uncertainty contributes to many eligible patients deciding against the surgery. Many patient-level prediction (PLP) models have been developed to predict T2D remission, but few have been externally validated and most were conducted with relatively small samples. Researchers at Johnson and Johnson developed an open-source predictive analytics platform applying machine learning techniques to a common data model. They developed and validated a predictive model of antihyperglycemic medication termination in patients who underwent MxS and experience remission from T2D.

Approach

Researchers selected the Truven MarketScan Commercial Claims and Encounters [CCAE] database to identify a large target population of patients undergoing metabolic surgery who had a baseline diagnosis of T2D and antihyperglycemic medication treatment. They then trained a LASSO logistic regression model, a type of machine learning, against the CCAE database using one repetition of 10-fold cross-validation. They evaluated model discrimination using the AUC and model calibration by inspecting a calibration plot and externally validated the trained PLP model by applying it to a separate database (Optum Clinformatics Database [Optum]) and evaluating its model discrimination.

Results

The target population meeting the inclusion criteria included 13,050 patients from the CCAE database and 3,477 from the Optum database. Antihyperglycemic medication cessation rates between 1 and 2 years post-surgery were 72.9% in CCAE and 70.8% in Optum. Researchers observed that the three predictors contributing the most to medication cessation were use of noninsulin glucose-lowering drugs, having undergone gastric bypass surgery, and younger age. The model showed good internal discriminative accuracy (area under the curve [AUC] = 0.778 [95% CI = 0.761-0.795] in CCAE test set N = 3527) and external validation showed good transportability (external AUC = 0.759 [95% CI = 0.741-0.777] in Optum N = 3477).

Conclusion

In this study, researchers developed a well-performing machine learning model to predict antihyperglycemic medication cessation after metabolic surgery. Results suggest that machine learning models based on readily available real-world data can improve healthcare decision-making and generate insights for improving clinical practices and optimizing outcomes. In the future, establishing prerequisite technological infrastructure will be necessary to implement such models for real-world decision making.

USE CASE #9

Improving the Prediction of Clinical Success Using Machine Learning

FasterCures (Milken Institute) and IMT Institute for Advanced Studies

Published: *MedRxIV*

When conducting pharmaceutical research, evaluating the odds of a new drug's success as it moves through the development process often relies on historical data and rudimentary analytical methods. In recent years however, the growing power of machine learning provides a new tool for these evaluations and can aid in uncovering the more promising drugs. To evaluate the usefulness of machine learning in this case, researchers from FasterCures (Milken Institute) and IMT Institute for Advanced studies trained and validated a variety of machine learning algorithms on a large database of projects.

Approach

Researchers trained various algorithms on a database of drug development projects to forecast whether the clinical research phases they were conducting would succeed or fail. Each project reflected a combination of input and output data. The input data summarized the attributes of each project like the features of the molecule, intended market, and company, while the output data showed the status of its most advanced clinical research phase (success, failure, or on-going).

Results

The team was able to predict the clinical success and failure of the projects with average balanced accuracy of 83% to 89% - much higher than the 56% to 70% of the method conducted using historical data. They also discovered the key variables which are most likely to contribute to trial success. They were also able to apply the algorithm to products currently in the drug pipeline, expecting to predict their success or failure with better accuracy.

Conclusion

Findings suggest that pharmaceutical companies can use machine learning models like these to improve the quantity and quality of their new drugs. Furthermore, the adoption of machine learning analytics across the industry could reduce the industry's risk profile with important effects on structure, investment in R&D, and innovation costs.

PREDICTING OBESITY IN ADULTS

USE CASE #10

Predicting Obesity in Adults Using Machine Learning Techniques: An Analysis of Indonesian Basic Health Research 2018

The Ministry of Research, Technology/National Research, and Innovation Agency of Indonesia

Published: *Frontiers in Nutrition*

Obesity has been closely tied to increased risk of chronic disease morbidity and mortality. Understanding the association between risk factors and the occurrence of obesity is challenging and the typical regression approach restricts the analysis to only a few predictors and imposes assumptions of independence and linearity. A study funded by The Ministry of Research, Technology/National Research, and Innovation Agency of Indonesia decided to use several machine learning methods such as Logistic Regression, Classification and Regression Trees (CART), and Naïve Bayes to discover the presence of obesity using publicly available health data.

Approach

The study used the data of 618,898 respondents from the 2013 national-level survey of Indonesian Basic Health Research (RISKESDAS) who met the study criteria. Researchers then selected a subset of variables that were significant causes of obesity and applied them to the selected cohort to improve the overall predictive performance of the classification. They addressed data imbalance – one or more classes dominating the whole data as major classes - using Synthetic Minority Oversampling Technique (SMOTE) to predict obesity based on risk factors available in the dataset. The team then evaluated and compared various machine learning methods including Logistic Regression, CART, and Naïve Bayesian.

Results

Out of 618,898 respondents, 134,709 (21.77%) were identified as obese and 484,189 (78.23%) as non-obese. The result revealed that the Logistic Regression method showed a better accuracy compared to the other methods with AUC = 0.798 while the CART method showed better sensitivity. The study identified 21 selected variables which play a prominent role in increasing the risk of obesity in adults, including consumption of alcoholic or sugary beverages, mental / emotional disorders, and fruit / vegetable consumption.

Conclusion

Results from this study suggest that incidence of obesity is underestimated in the Indonesian population and provide evidence for updates to public health policy and clinician care. Furthermore, this study demonstrates that applying machine learning methods on publicly available health data is a promising strategy to improve our understanding of and solve for complex and pressing public health problems.

PREDICTING DRUG-DRUG INTERACTIONS

USE CASE #11

Machine learning-based quantitative prediction of drug exposure in drug-drug interactions using drug label information

Ministry of Food and Drug Safety, Republic of Korea

Published: *Digital Medicine*

A drug-drug interaction (DDI) occurs when the effects of one drug are changed by another drug taken previously or in combination, and a DDI can lead to a drug's withdrawal from the market. The FDA has historically recommended that DDIs be evaluated in clinical trials, which are costly and time consuming, resulting in many unidentified DDIs. Recently, machine learning techniques have provided a simpler way of predicting drug-drug interactions (DDIs) but there are still limitations. For example, there is currently no systematically constructed database with pharmacokinetic (PK) DDI information, nor is there a machine learning model that predicts PK fold change (FC) with that database. To fill this gap and better identify DDIs, researchers at Seoul National University and Jeonbuk National University in the Republic of Korea proposed a PK DDI prediction (PK-DDIP) model for DDI prediction with high accuracy, while constructing a reliable PK-DDI database.

Approach

Researchers selected 3,627 PK DDIs from 3,587 drugs using 38,711 Food and Drug Administration (FDA) drug labels. Using this data, they proposed a PK DDI prediction model (PK-DDIP model) that would quantitatively predict the fold change of drug PK parameters in DDIs. Also, they distributed a standalone application providing predicted fold changes and reported fold changes of PK parameters, anatomical therapeutic chemical (ATC) code-based alternative drug choices, and single nucleotide polymorphism (SNP) action information.

Results

The PK-DDIP machine learning model successfully predicted PK parameter fold changes when two drugs were administered at the same time. The in-sample mean-squared error was 0.2494 and the root-mean-squared error of 5-fold cross-validation was 0.5959. The model predicted the fold change of the area under the time-concentration curve (AUC) within ± 0.5959 and the prediction proportions within 0.8–1.25-fold, 0.67–1.5- fold, and 0.5–2-fold of the AUC were 75.77, 86.68, and 94.76%, respectively. After externally validating their results, researchers found that this model had high prediction performance for newly updated FDA drug labels.

Conclusion

Traditionally, a drug-drug interaction was determined through clinical trials or a physiologically based pharmacokinetic (PBPK) model – a computer modeling approach that incorporates blood flow and tissue composition of organs to define the pharmacokinetics (PK) of drugs. This makes it impossible to conduct drug interaction tests for all possible drug pairs. This machine learning approach to predict PK parameter fold change is a much more flexible model since it is not limited to a specific gene list. It is expected that this machine learning model will enable the identification of DDIs before performing human-based trials, which will significantly save time, cost, and improve treatment outcomes.

USE CASE #12

A Machine Learning Method for Drug Combination Prediction

National Key R&D Project, the Special Projects for Technological Innovation in Hubei, and the Fundamental Research Funds for the Central Universities, China

Published: *Genet*

In recent years, drug combination has become an important research topic in the pharmaceutical industry, but experiment-based methodologies are time-consuming and costly. While computational methods have been proposed to address these problems by starting from existing drug combinations, these methods usually only include molecular structure information with a limited set of characteristics, making it difficult to efficiently evaluate drug combinations. Researchers from Hubei Key Huazhong Agricultural University in Wuhan, China aimed to improve prediction accuracy by using the neighbor recommender method combined with ensemble learning algorithms on similarity-based multifeatured drug data.

Approach

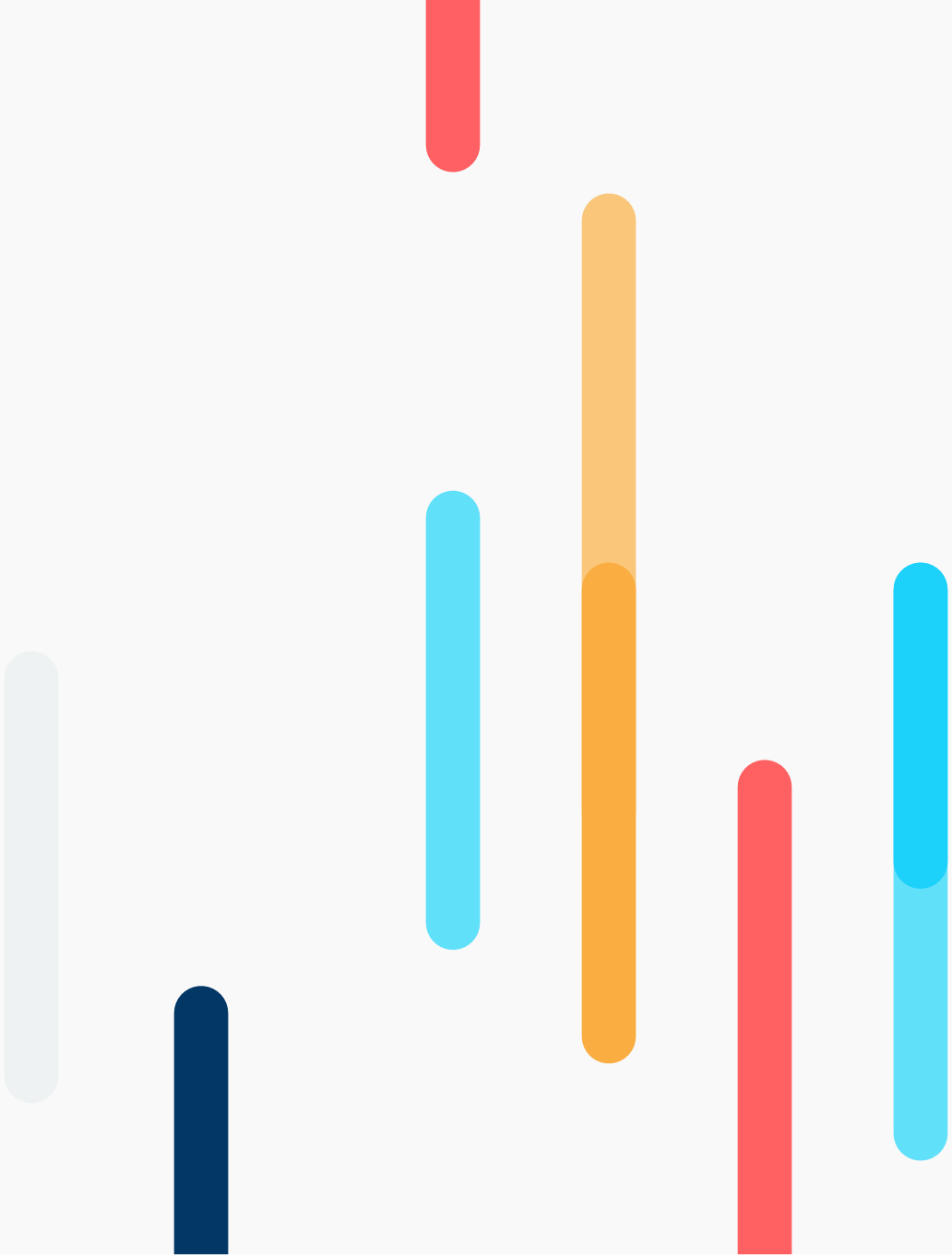
The primary method took a three-step approach: use the drug feature profiles to build the similarity feature-based model; select the useful features and construct the ensemble model for drug combination prediction; and deploy the model for drug combination predictions and conduct experimental validation. Since they used the neighbor recommender method (NRM) to generate models based on five features, they adopted the two most commonly used ensemble rules: the weighted average ensemble and classifier ensemble to obtain better model performances. They adopted generalized linear model (GLM) classifier rules to finalize the output from the base predictors.

Results

Upon comparing the methods, researchers found that their ensemble models outperformed traditional machine learning algorithms such as support vector machine (SVM), Naïve Bayes (NB), and Logistic Regression. In addition, they were able to identify seven candidate drug combinations for the anticancer drug, paclitaxel, and verify that the two of the combinations have promising effects.

Conclusion

Predicting drug combinations with machine learning methods can reduce costly experiments and efficiently uncover potential drug combinations. The biological experimental results for a predicted drug combination (paclitaxel and monobenzone) validated the ensemble model prediction. The researchers believe their methods are a promising approach to find potential drug combinations.



Summary

As evidenced by these and our previous booklet of use cases, stakeholders across healthcare are increasingly leveraging the power of machine learning to examine real world data.

Unlike statistical methods, machine learning concentrates on prediction, by using general purpose learning algorithms to find patterns in often rich and unwieldy data with minimal assumptions. What does this mean for you? Machine learning offers the ability to improve analyses for initiatives like product development and launch, understanding patient populations, determining unmet medical needs, predicting patient outcomes or disease recurrence, and scrutinizing real-world drug performance.

About 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. Our Data Science Module combines the speed of IHD with the power of machine learning.

Click here to learn more or set up a demo of the IHD Data Science Module.