DataRobot: Optimizing Biotechnology

Computational Biologists now have Automated Machine Learning to generate and test hypotheses.

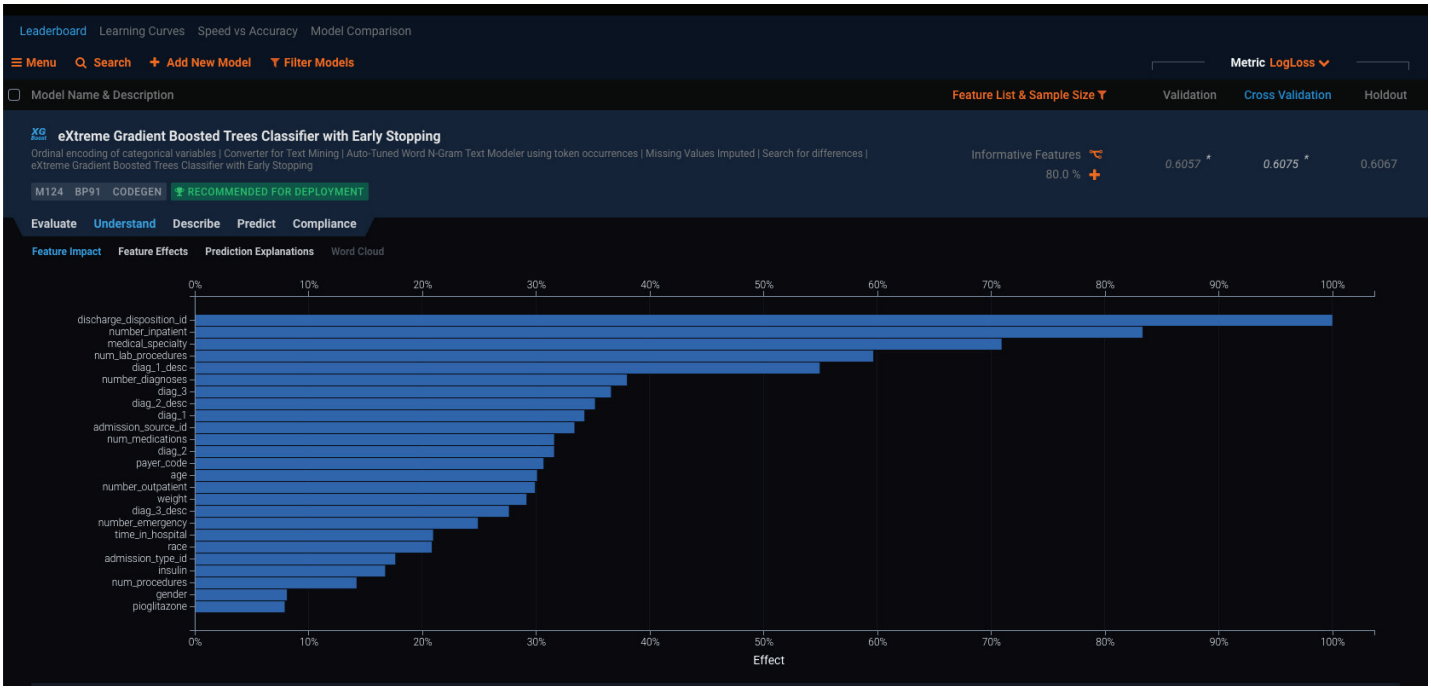
With a library of hundreds of the most powerful open source machine learning algorithms, DataRobot automates the end-to-end model building and deployment process. The platform encapsulates every best practice and safeguard to help organizations accelerate and scale their data science capabilities.

Discovering meaningful insights is fundamental to biological and medical data analyses and is considered as important as prediction accuracy. DataRobot is unique among available data science platforms in that it automatically provides intelligent data visualization and insights that may be used to identify potential biomarkers. For clinical projects, where a large amount of the data is unstructured text, DataRobot has built-in text mining capabilities that integrates this text and uses it as part of the modeling process.

The versatile nature of DataRobot ensures that it supports all avenues of healthcare delivery, including forecasting medical outcomes, finding features associated with high-cost patients, helping to detect medical insurance fraud and optimizing drug manufacturing and shipment.

DataRobot and Target Validation

Identifying novel drug targets for target validation and therapeutic development are essential activities in modern data-driven drug discovery. One of the greatest challenges to drug development is the analysis and integration of large amounts of data generated by sequencing technologies, clinicians and high-throughput compound screening. A central philosophy to drug discovery is fail fast – if your new drug is going to fail, it's better to predict that failure as early into the drug discovery process as possible. DataRobot can support scientists in this major analysis task and help them to both fail fast and succeed fast.



Predictive analytics plays an important part in assessing whether a gene is druggable based on its 3D structure, sequence and interaction information. For example, at a leading academic organization, the Institute of Cancer Research UK, their existing target modeling process used to take three months to complete. DataRobot not only shortened this process to five days but also gave them an average accuracy increase of 5%. The insights provided by DataRobot were also invaluable in determining which of the predicted potential targets were more "interesting," such as from previously undrugged families.

DataRobot has also been used to prioritize new candidate targets in the areas of epigenetics, immunotherapy and DNA Damage Repair. In all cases, DataRobot optimized the current models in both time-scales and accuracy.

DataRobot and Candidate Compound Selection

After a potential drug target has been validated, the next step is to identify some candidate compounds that may be investigated further. Feature representation of chemical compounds can end up very wide and, therefore, noise and sparsity can be a problem for some models. In one instance, DataRobot was used to predict the IC50 of one million compounds. There was an additional constraint of using the smallest number of features possible as the customer wanted insights on chemical substructures for scaffold hopping. The speed of DataRobot, together with its intelligent model and feature selection, allowed the customer to have a prioritized list of candidate compounds in a fraction of the time it took for manual modeling and feature selection. The insights provided by DataRobot helped the customer validate these candidates further and select compounds for a confirmatory bioassay.

DataRobot has also been used for compound ADMET and mutagenicity prediction.

DataRobot and Epidemiology

One of the caveats of predictive analytics for patient data is the low number of samples that may be present. In these cases, it is essential to run repeated experiments with different testing data to ensure the model is stable and true. For example at the Clinical Practice Research Datalink (CPRD), and the goal was not only to predict exacerbation events in asthma and chronic obstructive pulmonary disease patients but also to identify laboratory result biomarkers. The customer's current models were overfitting and not holding true for future patients. DataRobot's speed and advanced training capabilities allowed repeated experiments to be run using different training protocols, for example. Training DataRobot on the CPRD patient data from 2011-2014 achieved 93% accuracy when predicting exacerbation events in patients from 2015.

Summary

In each case DataRobot produced quicker, more accurate models than in-house benchmarks. DataRobot added valuable biological insights to replace in-house that supported black box predictive analytics, and the speed of DataRobot allowed for several different feature lists to be implemented. DataRobot's built-in guardrails allow for building stable models and can be confidently used by interdisciplinary scientists.

Who Uses DataRobot?

No matter your industry, automated machine learning delivers a competitive advantage. DataRobot is transforming businesses across industries, including:

- Healthcare
- Medical Research/Biotech
- Pharma
- Banking/Fintech
- Public sector
- Insurance
- Marketing
- Sports analytics
- Hospitality and travel
- Manufacturing
- Energy
- Retail

DataRobot

DataRobot offers an enterprise machine learning platform that empowers users of all skill levels to develop and deploy machine learning and AI faster. Incorporating a library of hundreds of the most powerful open source machine learning algorithms, the DataRobot platform automates, trains, and evaluates models in parallel, delivering AI applications at scale. DataRobot provides the fastest path to AI success for organizations of all sizes. For more information, visit: **www.datarobot.com**

Contact Us

DataRobot
One International Place, 5th Floor
Boston, MA 02110, USA

www.datarobot.com
info@datarobot.com

powered by aws