nutonian

Performance Genetics Sees into a Thoroughbred's Future, With the Help of Eureqa™

Shopping for a racehorse? Performance Genetics LLC can put you on the right track.

Since 2011, they've been helping clients in North America, Europe, Australia, and Japan to spot horses with the highest potential to become elite racehorses. To this forecasting challenge Performance Genetics brings expertise in horsemanship and horse physiology, but what sets them apart is an intensely computational approach driven by data from pedigrees, analysis of genetic variation, and biomechanical measurements — measurements such as stride length and the time each hoof spends on the ground, all captured at a high speed gallop.

Collecting this kind of biomechanical data, and getting maximum value from it, are high priorities for the company. Collection happens largely at "two-year-old breeze shows," where horses catalogued for sale are run in front of dozens of purchasing agents, racehorse trainers, and buyers. "We'll set up our high-speed cameras at 7:00 a.m. and film every horse, straight through to 5:00," says Performance Genetics cofounder and CEO Byron Rogers. "It's not uncommon for us to capture 300 plus horses in a day." The resulting videos are then manually marked up, and a custom app converts the markup into numeric data, ready for analysis.

Enter Eureqa

"My partner, Alan Porter, saw Eureqa featured in a New Scientist article ("Move over, Einstein: Machines will take it from here") and suggested I take a look," Rogers says.

Finding the graphical interface engaging and intuitive, he soon had Eureqa working on a historical data set. "Based on other ways of looking at the data, we already knew the answers we wanted," he says, "and Eureqa quickly kicked up an algorithm that not only gave us those answers but was immediately applicable in the field as well."

Since then Eureqa has been Rogers' tool of choice for the development of models relating biomechanical data to future performance. Here's how it works: Horses that have already had substantial careers, and for which Performance Genetics has pre-career video data, are given performance scores based on career earnings. Eureqa then builds models (i.e., algorithms) that connect the

pre-career data to those performance scores. Those models are then used to predict future performance for the current crop of not-yet-raced two-year-olds.

Rogers now has measurements for close to 2,500 horses, and with each new batch of performance data Eureqa makes refinements. "The models seem to be working extraordinarily well," he says. "We're seeing horses that the algorithm scored highly go on to become top-class racehorses, and in many cases the market wasn't picking up on that potential. So instead of selling in the range of three to four hundred thousand dollars, which is common for a top-class horse that looks great to the naked eye at the sale, these horses sold for twenty or thirty thousand. The data we can gather from the high speed video and the algorithm that Eureqa has created from that data is exposing a market inefficiency that our clients can take advantage of."

Filtering Out the Noise

Eureqa isn't the only arrow in Performance Genetics' analytical quiver. Regression modeling, grammatical evolution, and neural networks have all yielded useful results. Eureqa, though, has shown an unequaled ability to produce algorithms that filter out noise and marginal factors. "Originally, we were gathering 30 to 40 data points for each horse," Rogers says. "Eureqa showed us we could get that down to less than 10, with no significant loss of predictive accuracy." With a streamlined algorithm and a reduced video-markup burden, Rogers can now film those 300 plus horses as they "breeze," mark up the videos, generate numeric data,

run it through the algorithm, and get performance score predictions, all as the event proceeds. This on-site near-real-time intelligence gives an edge to Performance Genetics' clients as they decide which horses merit a second look. "There are only a couple of days between the time that they breeze and the time they sell, and you have to sort through a lot of horses," Rogers added. "The fact that Eureqa significantly reduced the measurement time has resulted in us being able to spend more time in growing our client base at the sale, which was an unexpected benefit."

Discovering New Signals

Eureqa has done more than just help filter out noise, it has helped bring in new signals. The conceptual clarity of its output has spurred insights that have subtly redirected data collection. The team had assumed, for example, that stride length would be of primary importance — horses with longer strides than others would naturally cover more ground — but Eureqa found stride frequency to be a far more powerful determinant. This led Rogers to reposition his cameras so as to capture stride frequency with more nuance and greater accuracy. The resulting data, in turn, has led to further increases in accuracy of the model.

The unusual transparency of Eureqa's process provides additional fuel for insight. Rogers and Porter often find value in looking back through the steps Eureqa took en route to an optimal algorithm. "We'll see certain data points popping up repeatedly," he says, "and that gets us thinking: Why was it looking at this? And why not that? And should we be including other measurements that limit or underlie those factors? It's led to some interesting new avenues of research."

Looking Forward: One Model to Bind Them All

Analysis of genetic variation already plays a role in Performance Genetics' horse-evaluation process — they regularly test for particular genetic variants known to be associated with speed and endurance in thoroughbred racehorses. Additionally, cardiovascular measurements that look at the capacity of the cardiovascular system, a key determinant of athletic potential, are taken by ultrasound. With genetic data for almost 600 horses, and cardiovascular data for over 1,000 (and both numbers growing rapidly), Rogers believes he will soon be in a position to discover a good number of additional data points that point towards elite performance.

To bring it all together, Rogers is working on a larger project that will examine over 10,000 records, each containing about 120 data points covering three generations of racehorses. As he prepares to include these additional domains of data in the modeling process, he foresees an important role for Eureqa. "As with the biomechanical data, I expect that Eureqa's ability to pick out the important data points in these other areas, along with its ability to make use of prior solutions, will be integral in the development of a unified model," he says. "It's going to be very interesting."

About Eureqa

Eureqa is Nutonian's next-generation technology for extracting patterns, insight, and value from data. For more information or to get started with a free trial visit www.nutonian.com.