Health System Applies Linguamatics NLP Platform To Improve Care And Population Insights

A health system in the Midwest needed to support the extraction of heart failure factors that were trapped in patient notes. Using the Linguamatics natural language processing (NLP) platform, it has delivered real world insights across 100,000 patients in just three months.

By providing access to mountains of data previously locked up in clinicians' freeform notes, Linguamatics NLP has contributed to millions in savings.

QUICK FACTS

Situation: A large medical device manufacturer commissioned our client, a health system in the Midwest, to help understand the outcomes for heart failure patients, especially those fitted with cardiac resynchronization therapy (CRT) devices. A manual chart review approach would have taken approximately 55 years to complete. They needed an NLP tool to enable them to extract these insights from patient notes rapidly and effectively.

Solution: The Linguamatics NLP platform was used by the customer's data science team to index 34 million patient notes and extract heart failure factors such as left ventricle ejection fraction, cardiac related symptoms and medical device outcomes.

Success: Results were delivered in just three months by two analysts, and have been presented to the Food and Drug Administration (FDA) as a source of real world evidence (RWE). Accuracy levels are an impressive 95–99%, and demonstrate how the NLP platform can support population-scale analysis of patient outcomes.

Situation

Our customer, a health system in the Midwest, provides contract research services for medical device and pharmaceutical clients, to support the use of RWE in FDA submissions. As a large health system with a mature and consolidated Epic electronic health record (EHR) system, it has a significant data set of patient treatments and outcomes. In addition, it has a precise supply chain management system that tracks every device and treatment from manufacturer to patient.

"Our combination of a single mature EHR and ability to track exactly who received what device is a key differentiator when we are working with life science companies," said a V.P. at the organization.

A recent study, sponsored by a medical device company, required analysis of 100,000 congestive heart failure (CHF) patients to assess the impact of fitting CRT devices, and capture patient outcomes.

While identifying and pulling records for this population is relatively trivial for this health system, a number of the attributes needed by the study were trapped in clinical notes. These attributes indicate the severity of the disease and the impact of the device, for example:

- ejection fraction—a measure of how well the heart is pumping out blood;

- New York Heart Association classification—a four-level indication of severity;

- symptoms including dyspnea (shortness of breath), fatigue and dizziness; and

- the CRT model and whether it was explanted (removed), and why.

Manual chart review of this data is impractical at best; even at only one hour per chart, it would take over 55 full-time equivalent (FTE) years to complete. Our client had tight deadlines to deliver the project, and knew it would need to use some innovative approaches to meet its target.

The organization had been evaluating the Linguamatics NLP platform in another part of the business, and a V.P. attended an internal demo showing how ejection fraction could be extracted from clinical notes. Once that connection had been made, he and his team began working with Linguamatics to use NLP to deliver the project. "I could see that Linguamatics could help us with our project, but we hadn't used NLP techniques before so we weren't out of the woods quite yet. My team of data scientists needed to get up to speed with the tool and integrate it into our workflow for us to be successful."

"Our results show that NLP is a reliable and accurate method to extract relevant data from clinical notes in a CHF population."

— Director, Data Science

Solution

Following initial training, our customer was able to jump-start the project using NLP resources built by Linguamatics. A major advantage for Linguamatics is having access to 30 million medical transcripts—this supports the NLP platform's data driven methods to explore large data sets for specific concepts, and design extraction algorithms based on frequency and linguistic patterns.

"It's a great opportunity to look at the impact of a high-end technology on a very sick heart failure population—without NLP we couldn't do this."

— A cardiologist at the health system

To identify such a large patient population, the organization went back to 2011 to gather patients with a heart failure diagnosis. Once identified, clinical records were extracted going back three years from the date of diagnosis, and forward to the present day—resulting in approximately 34 million clinical notes. This comprised eight years of data that was loaded into the Linguamatics platform. The platform pre-builds large-scale indexes that incorporate clinical ontologies, and numeric and linguistic patterns, which enables analysts to quickly and iteratively develop and modify their NLP algorithms. Some issues with the EHR export formatting required the use of pre-processing. This is a common issue that many groups run into when pulling document sets using SQL queries. However, due to the complexity of the data warehouse tables, this is often the only way to extract a focused set of documents for analysis.

Learning the NLP platform was a rapid process for the data scientists, but the delivery of results for the project was still a major undertaking. The initial algorithms needed to be tuned to the local data set, a typical practice for NLP to tailor the algorithms for the local writing style.

"We looked at all the attributes across each year to assess how well the algorithms were performing, and iterated until we were happy with the results," said the Director of Data Science. "By sampling sets of random records, we could see how the balance of sensitivity and specificity changed with our modifications. We knew the data would be presented to the FDA, so we wanted to triple check everything. My team manually reviewed 100 records for each year to get our final accuracy scores. We especially wanted to check for false positives, so we reviewed echo reports done by cardiologists that had no NLP results; these were the most likely errors, but fortunately we didn't see any issues there."

Success

The health system team was able to complete the project on time and at an impressive 95–99% accuracy level. Most importantly, the whole analysis was completed in three months by two analysts with no previous experience in NLP; an improbable manual task that would have taken over 55 FTE years to complete. "Linguamatics was really invested in helping us meet our deadlines and deliver the accuracy level we wanted," said the Director of Data Science.

Our customer subsequently presented the results to the FDA as RWE for pacemaker devices and outcomes. Reaction to the project has been one of amazement that such large-scale analysis is possible—let alone completed. Such use of large data sets and NLP techniques has the ability to transform clinical trials and post-market surveillance, cutting time to market and demonstrating true insights into patient outcomes. The team is moving on to its next contract project with another medical device company; it will continue to use the Linguamatics NLP platform to support this work, but with more automation to support the processing of the large document sets.

Looking to the future and broader application of NLP, this health system is exploring how the Linguamatics platform can support patient safety and care quality. Another key area is reducing the amount of clinical documentation required by physicians, which continues to cause concern in the healthcare community. "We were able to show the life cycle of a heart failure patient to see risk factors for heart failure, medications, labs performed, date of heart device implantation and outcomes such as ejection fraction results. The project showed [the NLP platform] has the potential to ease the EHR burden on physicians," said the Director of Data Science.

"As we look to the future of healthcare, [our organization] is well placed to support its patients and provide the outcomes analysis that will lead to better understanding of treatment options and improved health," said a V.P. at the health system.

"The results have been amazing to me; it's restored my faith in NLP's ability to get us out of this data-capturing conundrum."

— A cardiologist at the health system

≡IQVIA