Using Machine Learning to Identify Sub-Groups of Patients that Benefit from Treatment

Executive Summary

Primary Question

How can machine learning (ML) methods help uncover populations in which the intervention is most effective?

Key Results

- Cytel's machine learning capabilities helped the client identify a subset of patients that could potentially benefit from their treatment.
- Machine learning methods prevented discarding of a drug as ineffective in this disease area.
- The efficiency and improvement in results were found to be comparable when employing robust pre-processing and data cleaning/wrangling, as opposed to utilizing more advanced ML techniques.

Overview

In a large pharmaceutical client's double-blind study in severe infectious disease, the patients failed to meet the primary endpoint. A subset of the patients showed benefit with the treatment, but a variety of simple sub-group analyses were unable to identify which patients could benefit from active treatment versus standard of care. Cytel used machine learning (ML) to help the client identify a subset of patients that could potentially benefit from their treatment.

Methods and Result

Cytel collected a large dataset from the client that comprised over 50 biomarkers and every patient's clinical data. Our team explored various supervised machine learning techniques to assess nonlinear associations potentially present in the data. After a thorough assessment, Random Forests for the Binary Endpoints and LASSO for the continuous endpoint (secondary) were chosen for the final modeling, based on their robust operating characteristics. To estimate the PITE and CATE accurately, thereby enabling identification of critical predictive variables, the X-learner approach was used (Duan et al. 2019). These identified variables were pruned based on clinical relevance and the known mechanism of action (MoA) of the drug. A further improved analysis was carried out using knockoff filters (Candes et al. 2015 JRSSB) to control the false positive chances of finding a predictive variable.

Additionally, our team worked with the client's clinicians and the data-driven models were adjusted for clinical meaningfulness to yield potential sub-populations that benefit differentially from active treatment.

Finally, Cytel was able to help the client identify a subset of patients that could potentially benefit from their treatment. The patient groups identified were consistent with the drug's presumed MoA, which added confidence to the findings. The client is now evaluating whether to pivot the development of this therapeutic to a more specific population for whom the treatment effect is going to be very high.

ROI Summary

Time to Market

18 Months

Potential Additional Risk-Adjusted NPV

$10M+