DIY Analytics for HMOs: How You Can and Why You Should

ref: 26X-D4P6B6P3A6

Your submission is limited to a maximum of 350 words (combined).

# Background

There is a robust market for predictive analytic products in both healthcare and health insurance. Some provide locally-installed software (e.g., Johns Hopkins’ ACG System) while others offer services which give predictions for uploaded data (e.g., JVion). Both are expensive, but provide significant value.

The increased availability and quality of open-source software for data science re-opens the 'build or buy?' question. Many of our organizations straddle both market segments (providers and insurers) affording more data than is usable by most vended software. We hypothesize that this combination of data may yield greater insights than either would alone. We will tell the story of an effort at 'building' a local model—similar to the risk of hospitalization component in ACG—to predict who is likely headed into the hospital for prioritizing Case Management efforts.

## Methods

We pulled diagnoses, procedures, pharmacy fills, BMI measures, blood pressures, and lab results for a cohort of 70k chronically ill patients observed during a 3-month period in 2016. We then identified inpatient admissions among those patients in the subsequent 12 months. Using a data-science approach, we trained several classifiers from Python's scikit-learn package to discriminate between patients who were and were not hospitalized on the basis of their claims and clinical data together. We used the remainder of the data to evaluate those trained classifiers.

# Results

The Support Vector and Random Forest classifiers performed best, producing samples of more than 50% hospitalized people among their top 100 most-likely patients, comparable to the predictor of inpatient risk that ACG produces (SVC:63%, RF:55%, ACG:46%), though ACG relies on 12 months of data prior to the observation period rather than our use of only 3 months.

# Conclusion

With modest efforts, VDW data, free software, commodity hardware, and almost no clinical expertise, we constructed a predictor of inpatient risk that worked as well for our purposes as ACG, despite using fewer months of input data. Our results demonstrate that 'building' analytics tools is a viable choice for organizations desiring to explore analytics, especially if they have access to more data sources than is expected by vended solutions.