# 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 the health care and health insurance segments. Some products are vended software you install locally and run data through (e.g., Johns Hopkins' Adjusted Clinical Group (ACG) software) and some are services to which you upload your data and get predictions back (e.g., JVion). Both can be expensive, and both can provide significant value.

At the same time these services and products were being developed, open-source software for doing data science has gotten faster and easier to use. So the 'build or buy?' question has never been more relevant. Many of our organizations straddle both market segments--are providers and insurers. It is intuitive that the combination of data from both segments may yield greater insights than either would alone. The proposed presentation will tell the story of an effort at 'building' a local model to predict who is likely headed into the hospital in order to prioritize Case Management efforts.

## Methods

We pulled all 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 looked for inpatient admissions among those patients in the subsequent 12 months. Using a data-science approach, we trained several classifiers from Python's Scikit-Learn library to discriminate between patients who were and were not hospitalized on the basis of their claims and clinical data. We then used the remainder of the data to evaluate those trained classifiers.

## Results

The Support Vector and Random Forest classifiers were best, producing samples of more than %50 hospitalized people among their top 300 most-likely patients. Both of these compare favorably to the predictor of inpatient risk that ACG produces (using 12 months of data).

## Conclusion

Using only modest efforts, VDW data, free software and commodity hardware--and almost no clinical expertise--we were able to produce a predictor of inpatient risk that worked as well for our purposes as ACG. This, despite using fewer months of input data (3 months vs. 12). Our results establish that 'building' analytics tools is absolutely a viable choice for organizations desiring to get into analytics.