

How Expensive Are We? Cloud Based Cost Prediction Service for Medical Claims Data

Shanu Sushmita, Aftab Hasan,
Garima Khulbe, Stacey Newman,
Martine De Cock, Ankur Teredesai,
Senjuti Basu Roy, David Hazel
Center for Data Science
Institute of Technology
University of Washington, Tacoma
[sshanu, aftabh,
garima1, newmsc,
mdecock, ankurt
senjutib, dhazel]@uw.edu

Virendra Prasad
Edifecs, Bellevue
Bellevue-WA
VirendraP@edifecs.com

ABSTRACT

In this demonstration proposal, we describe an interactive predictive modeling framework for the exploration of health-care costs. It allows for fast analysis and estimation of costs, as well as ease of predictive model deployment. End users (typically managers at Accountable Care Organizations – ACOs, or benefit managers at insurance companies) can upload basic demographics, recent diagnoses, and comorbidities of a population for the past year, and browse through population characteristics, estimates of the healthcare costs associated with the next year and the factors that contribute most to high costs. The proposed framework can be used for cost predictions and visualizations for three scenarios: (1) General population, (2) Sub-population or health specific cohorts, e.g., Congestive heart failure (CHF), Diabetes, etc. and (3) Individual level view with “what-if” analyses. For the given population, it can identify top the 100 “most expensive” beneficiaries¹ in the coming year. An additional feature is that it allows development and deployment of predictive models for scoring. The back-end system contains cloud based prediction services hosted on the Microsoft Azure infrastructure that allow the easy deployment of models encoded in Predictive Model Markup Language (PMML)² and trained using either Spark MLlib or various non-distributed environments.

¹A beneficiary, in the context of this proposal, is an individual (patient) who receives the benefits of a health insurance contract.

²Predictive Model Markup Language (PMML) provides an open standard for representing data mining models. In this way, models can easily be shared between different applications avoiding proprietary issues and incompatibilities.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.