covermymeds®

Team Red Eye: Predicting Patient Pay

When a patient purchases medication at a pharmacy, the out-of-pocket cost is determined by a complex set of negotiations between **pharmacies**, **insurance companies**, and **drug manufacturers**. There are a wide array of different kinds of coverage plans for prescription drugs, which generally separate drugs into different **multi-tiered formularies**. For each plan and within each tier, coverage is determined by differing **deductibles**, **coinsurance**, and **copays**. Moreover, the effect of third-party discounts, such as those from the drug manufacturers or other agents, further complicates the calculation of patient pay. Due to these complicated factors, **actual out-of-pocket expenses are difficult to predict**.

While information about insurance formularies and drug discount programs are in principle publicly available, this information is spread out and hard to aggregate in one place. In contrast, **actual pharmacy transaction data is available in large abundance**. For example, CoverMyMeds provided us with a data set of just under 14 million pharmacy transactions. Although this data is simulated, it presents similar opportunities and challenges as the real-world data that companies like CoverMyMeds have access to. **Understanding** and being able to **accurately model** large-scale pharmacy transaction data is relevant to a variety of stakeholders, including **healthcare providers**, **pharmacies**, **insurance companies**, **drug manufacturers**, and most of all **patients**, for whom out-of-pocket medication costs may seem surprising or even unfeasible, even when equivalent medication may be available at a lower cost.

In our project we have built models to **predict patient pay accurately** with an average 3.87% MAPE error for our random forest model, and 1.41% MAPE error for our histogram gradient boosted model. While the gradient boosted model has a lower error, the random forest model is capable of predicting expenses for insurance plans that may not have been seen in the training phase. We have also built an app that allows doctors to easily interface with the predictions and to better inform patients about medical costs. Since there is a wide variance in costs associated with branded drugs, we suggest building separate models for generic and branded drugs. In future work, these modeling approaches could be combined into a useful product for patients. Such a product will aid patients in choosing between different insurance plans, and in choosing among alternative medications when prices are unnecessarily high.