How much will your prescription drugs cost?

Predicting copayments with machine learning

Will Hardt, Karan Srivastava, Christine Sun, Funing Tian
Team Walnut (Fall 2022)

When a patient is prescribed medication from healthcare providers, their net copayment at the pharmacy is determined by a complex system involving many factors, such as the specific drug treatments and the patient's insurance. Currently, patients and doctors do not have a method of checking expected copayment costs before prescribing medication. Machine learning presents considerable opportunities to improve patient-facing drug recommendations. In this project, we survey many regressors for predicting copayment costs based on patient insurance plan, prescribed drug, patient diagnoses, and other factors. With this, we hope to build the foundations for future systems that will inform doctors and patients about potential costs of medication before prescription to help patients work with doctors to find affordable treatments for their conditions.

For this project, our group has chosen to work on a simulated dataset created by CoverMyMeds, a healthcare technology company that aims to improve medication access by reducing administrative waste and prescription abandonment. The dataset is composed of simulated transactions from different pharmacies that were taken across a single year. The identifiers in the dataset include the transaction date, the pharmacy, the diagnosis of the patient, the drug prescribed, the patient's insurance plan, and the co-payment required from the patient.

Our key performance indicators were Root Mean Squared Error (RMSE) and Root Mean Squared Log Error (RMSLE). RMSE is a commonly used metrics that measure the typical deviation, in dollars, between our predicted copay and the actual copay. RMSLE penalizes underestimates more than overestimates, which is important since it is more harmful to underestimate a copayment than to overestimate it. We surveyed various regression models and found our best performing model was the Random Forest, with a RMSE of 15.6 and RMSLE of 0.343. Over 80% of our predictions were within $6 of the actual copayments; this $12 range can be compared to the $55 spread between the 10th and 90th percentiles of copays in the test set.

Future work required to successfully implement our work in hospitals and pharmacies,would involve training existing models, designing recommendation systems, and creating interactive user interfaces. First, in order to make our models more robust, we need to train our models with additional factors, such as geographic location, prior diagnosis, and the drug formulary from different insurance companies. Secondly, we would use a clustering system to group together medications with similar expected costs and formulary statuses depending on patient insurance. Finally, we would need to develop an interactable and intuitive interface and feedback loop for healthcare providers and insurance companies to create and review recommendations based on patient information and predicted costs.