MEMOTEXT

Using Private Insurance Claims to Predict the Onset of T2DM

Background

Type 2 Diabetes Mellitus is a complex disease that is difficult to understand from claims data alone. It is currently difficult to identify plan member risk of developing T2DM. The sooner the risk for an individual for developing T2DM can be identified, the more proactive and preventative measures can be taken.

Goals

1) Use ‘prediabetes’ insurance claims data to engineer new risk features and to predict T2DM
2) Compare results to the manual method of risk identification using traditional rule-based tagging.

Methodology

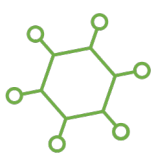
- Claims data dating from 2010-2013 was pre-processed and mined
- Built features to predict the onset of T2DM in the 3-year period ahead
- Applied 3 different algorithms to compare: logistics regression, random forest , and XGBoost for prediction

Feature Engineering

Demographics (Sex, Age)

Total Conditions

Hypertension Meds



Cholesterol Med Use

Unique Drug Prescriptions

Adherence (PDC statins)

MEMOTEXT

Results

- XGBoost was used to achieve the best performing model
- Comparing model results to manual rule tagging, it is noted that the model had great improvement on metrics such as accuracy, precision, false positive, and specificity

	Manual Rules	VS.	Machine Learning
Accuracy	66.4%		83.0%
Recall	55.1%		41.5%
Precision	8.6%		14.0%
False Positive Rate	31.1%		13.8%
Specificity	67.1%		85.4%

Identifying T2 risk using ML is a great improvement over rule-based tagging. However, a balance between ML & human intervention is required.

Discussion

Private insurance drug-level claims data can be used to predict which plan members may be at risk for developing diabetes. The use of machine learning leads to higher precision and a decreased false positive rate, which can allow digital health interventions and higher touch resources to be proactively targeted at the right people.