Junjie Luo[2], Abhimanyu Kumbara[1], MS, MBA, Anand K. Iyer, PhD, MBA[1], Mansur E. Shomali, MD, CM[1], Guodong (Gordon) Gao[2], PhD

[1]WellDoc, Inc. and [2]Johns Hopkins Carey Business School, Center for Digital Health, and Artificial Intelligence (CDHAI)

Using an Automated, Real-time Data-Enabled Feature Engineering Process to Predict Future Weight Outcomes

Research Introduction

Value Proposition

- Collect different types of patient-generated health data
- Use AI algorithms to extract insights from the data
- Provide personalized digital health coaching for people with cardiometabolic chronic conditions and their healthcare providers

Optimal, long-term chronic condition management requires in-the-moment coaching that utilizes a user's "360-degree data", including continuous glucose monitoring (CGM) senser data, connected device data (e.g., weight scales, BP cuffs, etc.) and other self-management data such as food, activity, labs, and medications.

Being able to predict outcomes associated with these data requires a robust engineering framework that extracts meaningful features at scale from real-time/near real-time data sources, which can then be used to make accurate outcome predictions.

We propose one such framework in this study to extract meaningful features from a variety of data sources, which are then used to predict weight outcomes.

Methods

Reviewed real-world data from 2443 patients with type 1 and type 2 diabetes who were enrolled in digital health coaching platform.

Data set included CGM data, along with Medication, Education, Diet, Activity, and Labs (MEDAL) engagement data.

Model training features were extracted for every 'case' point, which is defined as a MEDAL entry at a given timestamp for a given patient.

Outcome variable was categorized as achieving weight loss (>=3% weight loss) or not achieving weight loss (<3% weight loss).

Dataset was split into a training set and testing set with a ratio of 7:3.

Ensemble of machine learning models was implemented to choose the best model for weight loss prediction.

A logistic regression model was also implemented to understand the specific impact of predictor variables on achieving weight loss outcomes in the future.

welldoc

Results

01

The LightGBM model was highly accurate in predicting the binary “achieved or did not achieve” weight loss outcome variable **2 months into the future, with an a**ccuracy rate of 0.93.

02

This model had an **AUC of 0.96.**

03

Engaging with education content or connecting a device 1 week or even 1 day before the prediction case point is positively correlated with a good weight loss outcome in the future.

welldoc

Conclusions

1. The comprehensive feature extraction framework allows us to build dynamic feature profiles of patients, with temporal features that can be used to make accurate predictions or learn individual-level factors which can impact various health outcomes.
2. We have shown how our feature engineering framework allows us to make accurate weight loss predictions using CGM and MEDAL data to learn important factors that influence weight loss in the future.
3. Robust feature engineering frameworks can enable and power the next generation of highly scalable and automated personalized digital health solutions to manage chronic conditions.