Capstone III Proposal

Problem Identification

Problem Statement

How can individuals with type 1 diabetes (and perhaps the providers of insulin pumps) use data from glucose monitors and fitness trackers to help them improve glucose levels by incorporating glucose forecasts for the next few hours to better gauge required insulin amounts?

Context

Type I diabetics use insulin injections to keep their blood glucose levels near a normal range. Other factors, such as intake of carbohydrates and physical activity, also influence blood glucose. Patients can already buy insulin pumps that integrate continuous glucose monitors (CGM) to semi-automate insulin dosing based on current blood glucose values and trend. With an extended forecast of blood glucose based on recent physical activity, meal carbohydrates and insulin dose, blood glucose levels could be managed more proactively and thus help avoid spikes in glucose that require reactive injecting of insulin.

Criteria for success

The goal is to predict blood glucose for the next four hours with an error of no more than 15%. Achieving this accuracy is required in order to be a useful guide for people with type 1 diabetes to incorporate these forecasts when deciding on the optimal dosage for their next insulin injection.

Scope of solution space

The scope is the prediction of glucose based on past values of glucose, insulin, and activity. Other biometric variables, such as weight or age will not be considered (not available). No suggestions for insulin amounts will be generated. An individual with type 1 diabetes will still have to use good judgment to make decisions about insulin dose.

Constraints

Blood glucose management requires an understanding of the mechanisms and dynamics of factors that influence blood glucose. This means that even with a blood glucose forecast, correct action can only be taken with a good amount of knowledge and experience. Also, the forecast relies on the accuracy of the manual entries of grams of carbohydrates and units of insulin injected. Some diabetics may find it onerous to track this kind of data.

Stakeholders

The only direct stakeholder for this project is the person with type 1 diabetes who provided the data. However, the project could also serve as a proof of concept that may interest other diabetics and makers of insulin pumps.

Data source

This Kaggle data set https://www.kaggle.com/datasets/lacofloris/type-1-diabetes-blood-glucose-prediction contains three types of records from one person with type 1 diabetes collected over a few months. It includes Fitbit data, data from a continuous glucose monitor (CGM), and manually entered insulin and carbohydrate amounts.

Outline of Problem Solution

Modeling

The goal is to create a time series forecast for blood glucose. Since there are exogenous variables, such as carbohydrate intake and insulin doses, a model that can accommodate these will be used. SARIMAX is a well known regression based model for time analysis. In addition, at least one other model that is based on newer, neural network architectures, will be explored, such as Darts Transformer or TCN.

Data Wrangling

The data for some of the variables are spread across several files that need to be combined and any overlap removed. Glucose measurements by the CGM are taken in small intervals (every few minutes), but injections of insulin or meals are less frequent, so resampling is needed to create regular intervals for all time series.

EDA

- Check for outliers and gaps in the data
- Find a good resample rate
- Check time series for stationarity
- Use decomposition to find any seasonality
- Use ACP and PACP to determine order of SARIMAX model

Data Preprocessing and Transformation

- No transformation, since values are not vastly spread
- The last four hours worth of data are split off and reserved for the forecast (‘test’), the remainder will be used to train the models

Deliverables

- Jupyter notebook for all steps of the project
- Project report
- Slide deck