Diabetic Patients Monitoring

CSULA Advisor: Dr. Mohammad Pourhomayoun
Medtronic Liaison: Alex Zhong
Real Time Data: Andrew Padilla, Koenrad MacBride
Prediction Engine: Mohammadreza Vahedi, Wun Woo
Insight Engine: Christopher Fong, Yosep Kim

Table of Contents

Introduction

Background

Diabetes is a disease that causes a person's blood glucose (BG) levels to be too high. Insulin, a hormone created by the pancreas, helps regulate the level of glucose in the blood. A person is said to have type 1 diabetes when their body can not produce insulin at all. Diabetic patients need to closely monitor their BG level and maintain it within a specific range. Hyperglycemia (high blood glucose) and Hypoglycemia (low blood glucose) are two conditions that may happen for diabetes patients and they can have destructive effects on their body organs.

Medtronic has provided data collected from type 1 diabetes patients. This data was collected via a CGM (Continuous Glucose Monitoring) system as well as from fitbit. The CGM sensor is a patch, typically applied to the abdomen area, that has a small needle entering but not penetrating the skin. The sensor needs to be calibrated 3-5 times per day with the use of the standard fingerstick method. These sensors are very accurate and can help patients monitor their blood glucose in real time. In an attempt to use this data to benefit type 2 patients, Medtronic has combined CGM data and FitBit activity metrics with the goal of predicting future BG levels in real time. If this prediction can be made accurately, we can apply this method to type 2 patients who do not use a CGM system solely using fitbit and fingerstick measurements.

Project Goal

To develop a backend infrastructure to generate real time insights and predictions to be applied to type 2 patients.

Real Time Data

Overview

Real Time Purpose : To provide Real Time data streaming capability to our insight engine and prediction engine
Real Time Goal : To fully connect a real time simulator using a microservices data infrastructure with Apache Kafka, Cassandra and Storm for insights and prediction.

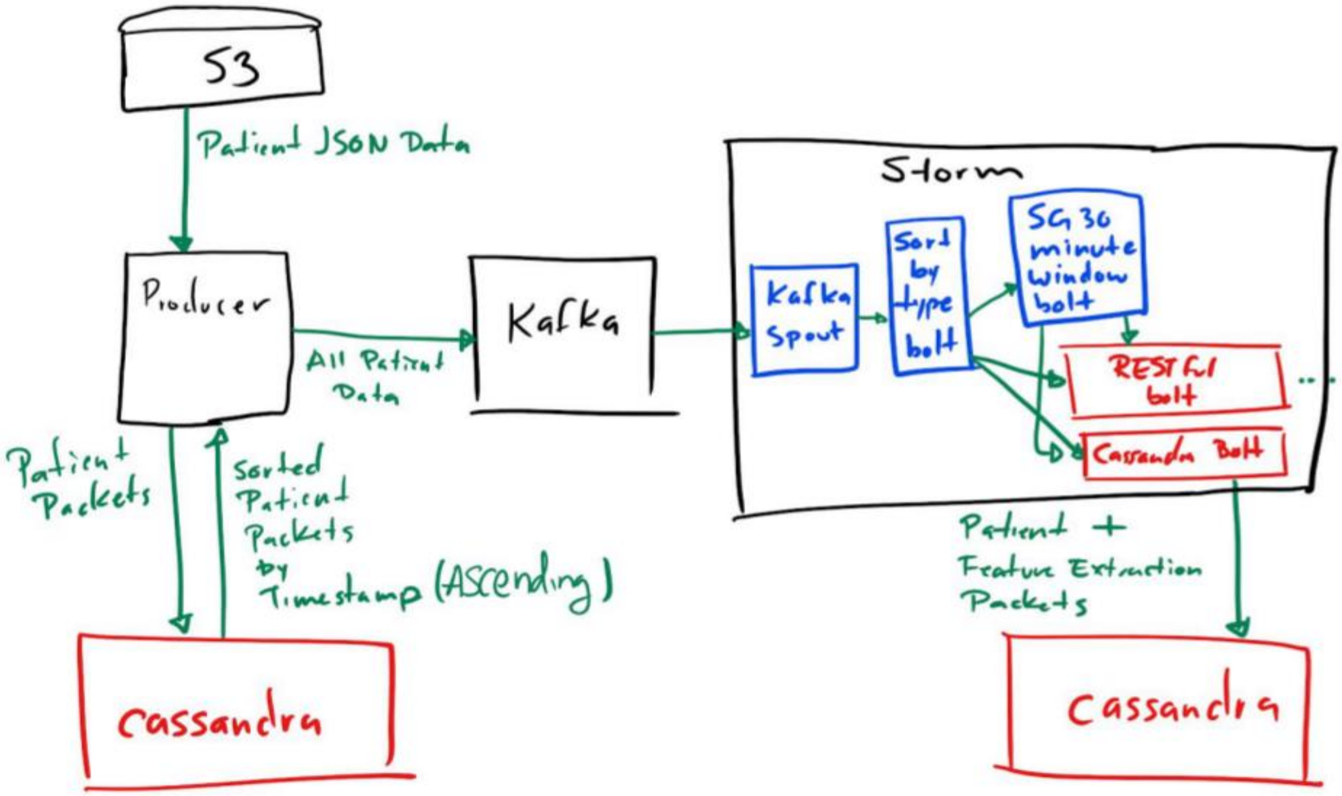
Progress

Kafka

- Kafka Broker, Zookeeper running on local and virtual machine
- Kafka Producer R1 - Amazon S3 as a data store & Connection/packet generation to Kafka
- Kafka Producer R2 - Created and Serialized data packets as JSON objects. Cast date to (date + 12:59:59) for packets without time. SG and CaloriesOut data handled.
- Kafka Producer R3 - Connected to Cassandra database for mass storage and sorting of data packets. ActivityTimeSeries, CGMData, HRTimeSeries patient data handled.
- Kafka Producer R4 - All patient data is handled, including log data and nutrino data.

Storm

- Storm server and zookeeper running on local and virtual machine
- Storm Kafka Spout
  - Reads stream off kafka topic
  - Implements GSON encapsulation for easy processing
- Storm Bolts
  - 1. Sort By Type
    - emits data-packets to individual streams according to the data type of the packet
  - 2. PostBolt
    - posts data packet to a specified URL
    - 17 streams are currently being POSTed to insight engine
  - 3. Custom Window Bolt (Windowed Feature Extraction)
    - implemented to calculate the average of sg for 30 minutes and output a packet
    - Extracting statistical features can be easily implemented now



Insight Engine API

- Jupyter notebook, python, flask, gentelella installed on virtual machine
- Python-Flask
  - Implemented routes for incoming data from the POST/RESTful Storm bolt.
  - Asynchronous task handling using a RabbitMQ server and Celery encapsulation.
  - Cassandra database insertion and dataframe aggregation
  - Data passing to Gentelella
- Gentelella - A free dashboard interface template
  - Real time graphs generated and visible to user
  - Insights displayed as well

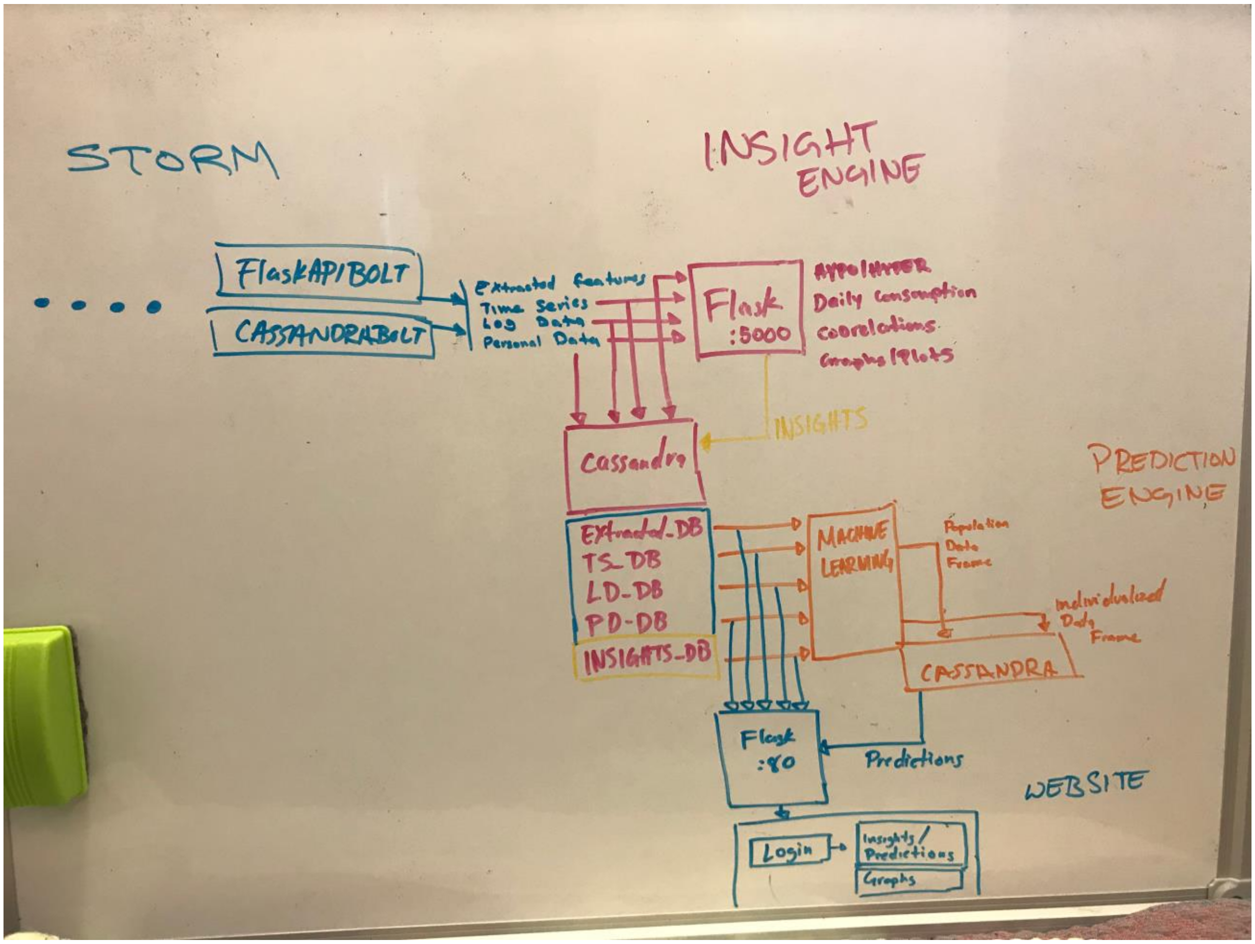
Infrastructure

5 Virtual Machines running Ubuntu Linux. Each machine running a microservice

- Kafka
- Cassandra
- Storm
- Insight
- Prediction

Database Layout

Cassandra was installed on a virtual machine. The database schema can be found in section the System Architecture section.



STORM
INSIGHT ENGINE
FlaskAPIBOLT
CASSANDRABOLT
Extracted features
Time Series
Log Data
Personal Data
Flask :5000
HYPO/HYPER
Daily consumption
coorelations
Graphs/Plots
INSIGHTS
Cassandra
Extracted_DB
TS_DB
LD_DB
PD_DB
INSIGHTS_DB
PREDICTION ENGINE
MACHINE LEARNING
Population Data Frame
Individualized Data Frame
CASSANDRA
Flask :80
Predictions
WEBSITE
Login
Insights/Predictions
Graphs

Prediction Engine

Prediction Engine includes the following steps:

Step 1: Data Preprocessing

The main task in this phase was converting raw data to clean data which is feasible for analysis.

- **Data Cleaning**
  We checked the validity of all patients' data. The validated data were converted from JSON format to CSV, which is feasible for analysis. We set up and used a virtual machine to process the raw data from all patients:
  Number of patients processed so far: 93

- **Missing Value Imputation**
  In the process of data cleaning, we encountered some missing values that could affect the accuracy of our prediction. We handled the missing values in two steps:

  - Step1 (personalized missing value imputation): Missing values of each patient replaced by the mean of the same feature for that patient.

  - Step2 (overall missing value imputation): After combining the data of all patients together, again we replaced the remaining missing values by the mean of the same feature, but this time for all patients. This way we could handle all the missing values of all patients.

- **Normalization**
  Since features have different ranges and units, we used Scikit Learn normalization routines to standardize the data for next steps.

Step 2: Feature Extraction

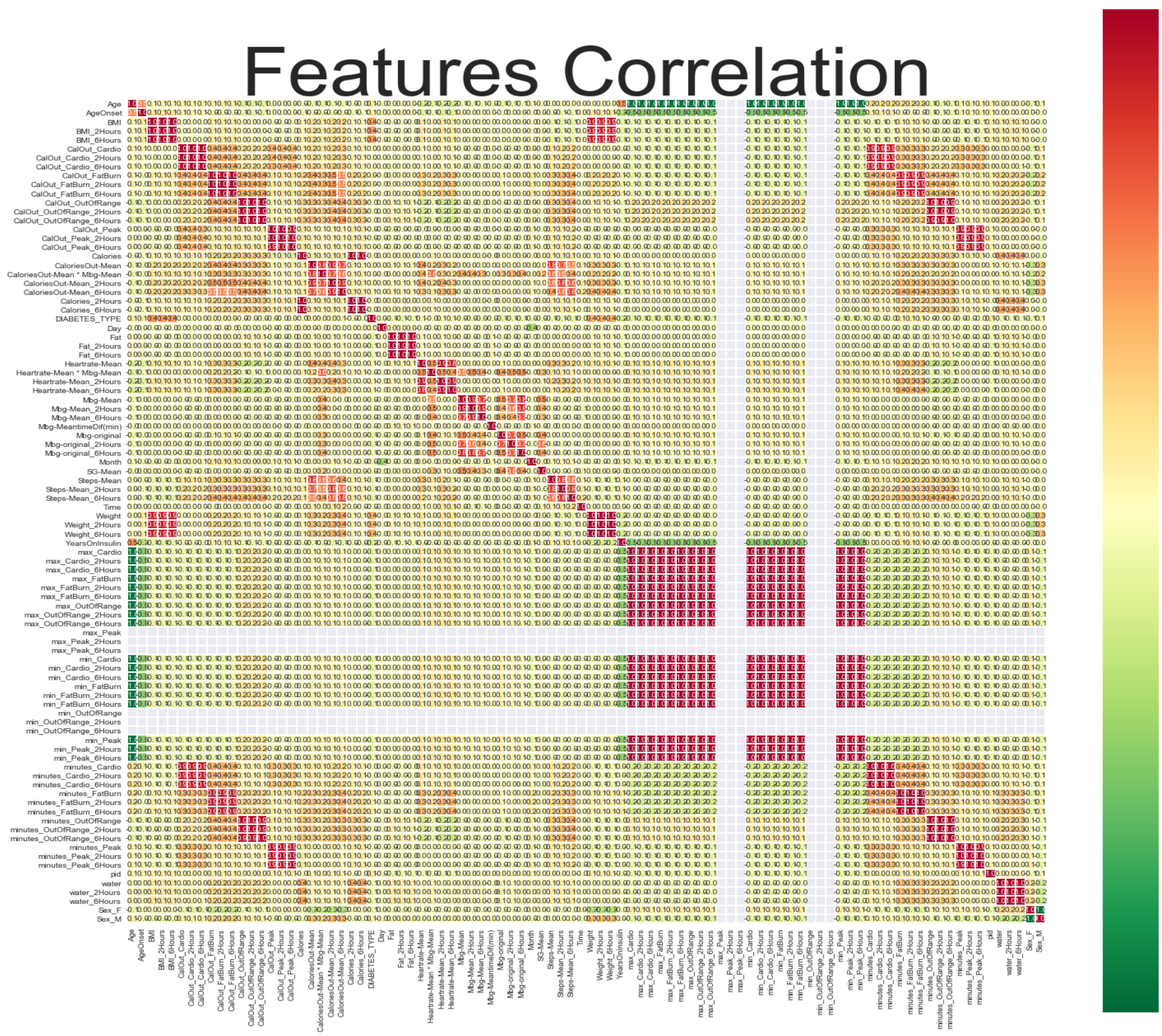
Average of SG within every 30 minute is considered as a data sample. Three different time windows have been used to extract features for each data sample from each data component: mean of past 30-minute, past 2-hour, and past 6-hour for every data component.

Overall, 92 features extracted for EVERY 30 minutes of every patient data.
194,891 data samples generated from 93 patients.
So, the size of feature table is 194,891 by 92.

Step 3: Feature Selection

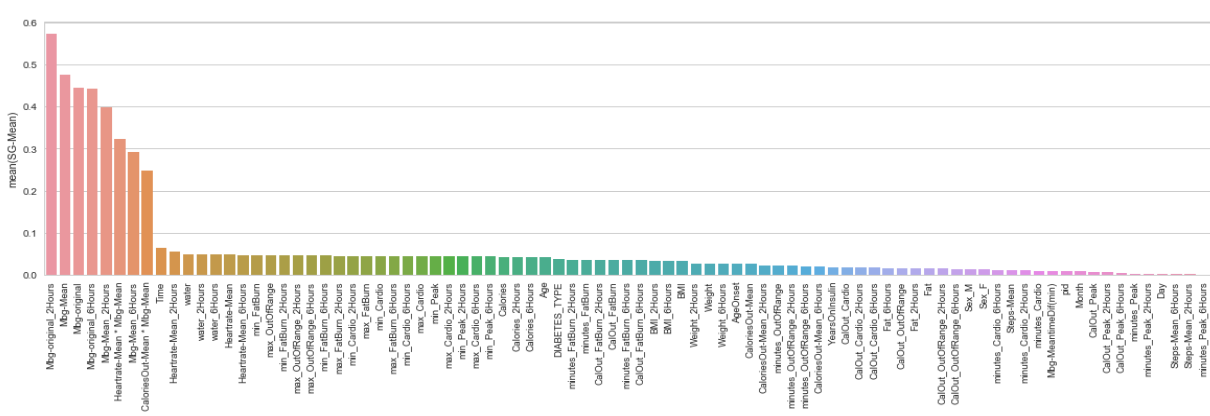
Feature selection methods have been used to identify and remove unneeded, irrelevant and redundant attributes from data that have destructive impact on accuracy of the model. Apart from removing irrelevant features, we tried to find the best combination, which can lead to the higher accuracy. Having this in mind, we calculated the correlation between each feature and the target, and then sort the features based on feature correlation with the target. Then, we made our feature combinations by choosing features from sorted list in order and according to their correlation.

The figure below shows all feature correlations with each other:



Correlation coefficient between all features and SG

SG-Mean	1.000000	minutes_FatBurn_6Hours	0.034271
Mbg-original_2Hours	0.573757	CalOut_FatBurn_6Hours	0.034159
Mbg-Mean	0.477063	BMI_2Hours	0.032177
Mbg-original	0.444299	BMI_6Hours	0.032170
Mbg-original_6Hours	0.443066	BMI	0.032168
Mbg-Mean_2Hours	0.397805	Weight_2Hours	0.026356
Heartrate-Mean * Mbg-Mean	0.323618	Weight	0.026350
Mbg-Mean_6Hours	0.292017	Weight_6Hours	0.026348
CaloriesOut-Mean * Mbg-Mean	0.247940	AgeOnset	0.025644
Time	0.063062	CaloriesOut-Mean	0.025568
Heartrate-Mean_2Hours	0.053866	CaloriesOut-Mean_2Hours	0.022198
water	0.048077	minutes_OutOfRange	0.021927
water_2Hours	0.047904	minutes_OutOfRange_2Hours	0.021384
water_6Hours	0.047399	minutes_OutOfRange_6Hours	0.020235
Heartrate-Mean	0.047297	CaloriesOut-Mean_6Hours	0.018981
Heartrate-Mean_6Hours	0.045926	YearsOnInsulin	0.018008
min_FatBurn	0.045072	CalOut_Cardio	0.017875
max_OutOfRange	0.045072	CalOut_Cardio_2Hours	0.017693
min_FatBurn_2Hours	0.045070	CalOut_Cardio_6Hours	0.017427
max_OutOfRange_2Hours	0.045070	Fat_6Hours	0.015488
max_OutOfRange_6Hours	0.045065	CalOut_OutOfRange	0.015004
min_FatBurn_6Hours	0.045065	Fat_2Hours	0.014952
max_FatBurn_2Hours	0.044508	Fat	0.014731
min_Cardio_2Hours	0.044508	CalOut_OutOfRange_2Hours	0.014551
max_FatBurn	0.044508	CalOut_OutOfRange_6Hours	0.013630
min_Cardio	0.044508	Sex_M	0.012672
max_FatBurn_6Hours	0.044507	Sex_F	0.012672
min_Cardio_6Hours	0.044507	minutes_Cardio_6Hours	0.010344
max_Cardio	0.044347	Steps-Mean	0.010215
min_Peak	0.044347	minutes_Cardio_2Hours	0.009721
max_Cardio_2Hours	0.044347	minutes_Cardio	0.009357
min_Peak_2Hours	0.044347	Mbg-MeantimeDif(min)	0.009191
max_Cardio_6Hours	0.044345	pid	0.008358
min_Peak_6Hours	0.044345	Month	0.007902
Calories	0.042081	CalOut_Peak	0.005684
Calories_2Hours	0.041780	CalOut_Peak_2Hours	0.005352
Calories_6Hours	0.041701	CalOut_Peak_6Hours	0.003953
Age	0.041599	minutes_Peak	0.002105
DIABETES_TYPE	0.036832	minutes_Peak_2Hours	0.001729
minutes_FatBurn_2Hours	0.034903	Day	0.001630
minutes_FatBurn	0.034762	Steps-Mean_6Hours	0.000827
CalOut_FatBurn_2Hours	0.034500	Steps-Mean_2Hours	0.000820
CalOut_FatBurn	0.034346	minutes_Peak_6Hours	0.000472



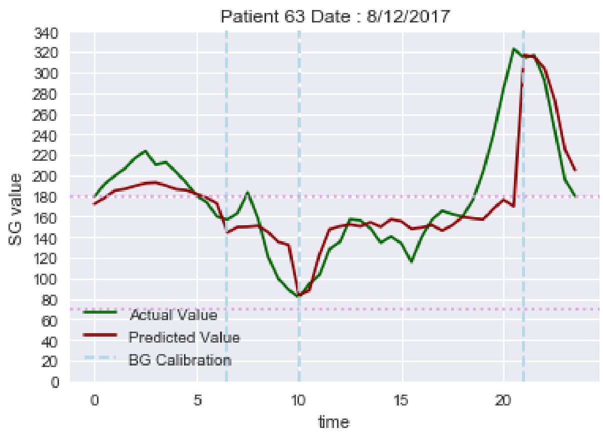
Step 4: Dimensionality Reduction

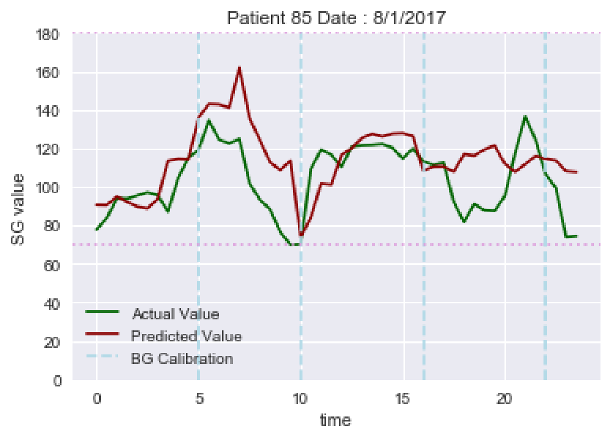
Using PCA, we tried to check how the dimensionality reduction can reduce our error and improve our accuracy. Again, for dimensionality reduction, we calculate the optimum number of features that could lead to minimum error and highest accuracy.

Step 5: Prediction using Machine Learning Algorithms

The total number of data samples after combining all 93 patients' data is 194,891. Random Forest and Neural Network algorithms were used for training/prediction.

Algorithm	**MAPE Error**
Random Forest Regressor	27.9
MLPRegressor	29.6





Insight Engine

Overview

Automatically create insights based on each individual patient. And display graphs and helpful tips to the user in real time.

Step 1: Data Preprocessing

Refer to Prediction Engine steps for details.

Step 2: Feature Extraction

Refer to Prediction Engine steps for details.

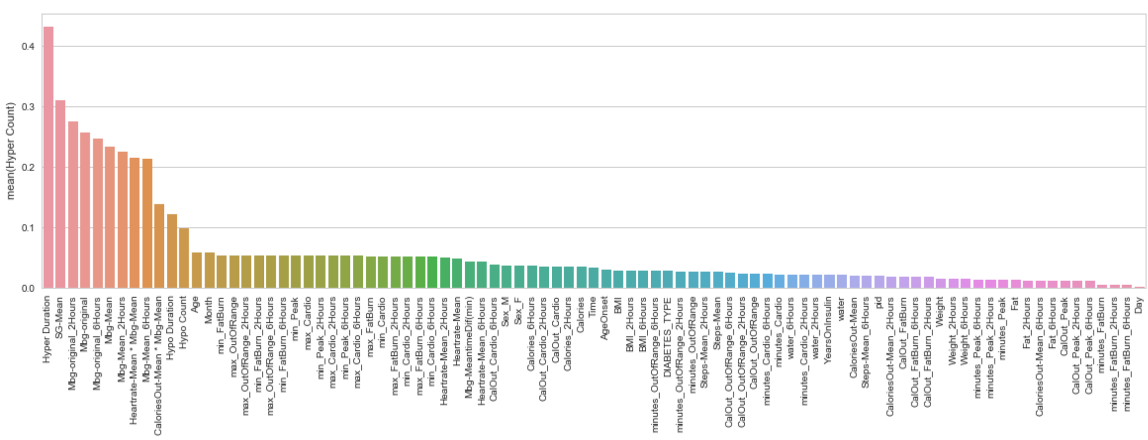
Step 3: Hyper and Hypo events and its Durations

- The team had worked with graphs to check if the data was viable. We had looked at Sensor Glucose and Hyper and Hypo events. Next used these events and created durations of these Hyper and Hypo events. The start of an event was after 15 minutes over or under the allotted hyper and hypo conditions. The end of an event is marked when they are within the normal SG range for at least 10 minutes.

- Medtronic team conveyed the need to have a correlation with the outcomes and to develop insights based on rankings. The purpose is to look for insights instead of having static values do it. The prediction engine team helped come up with top-ranked features for the insights.

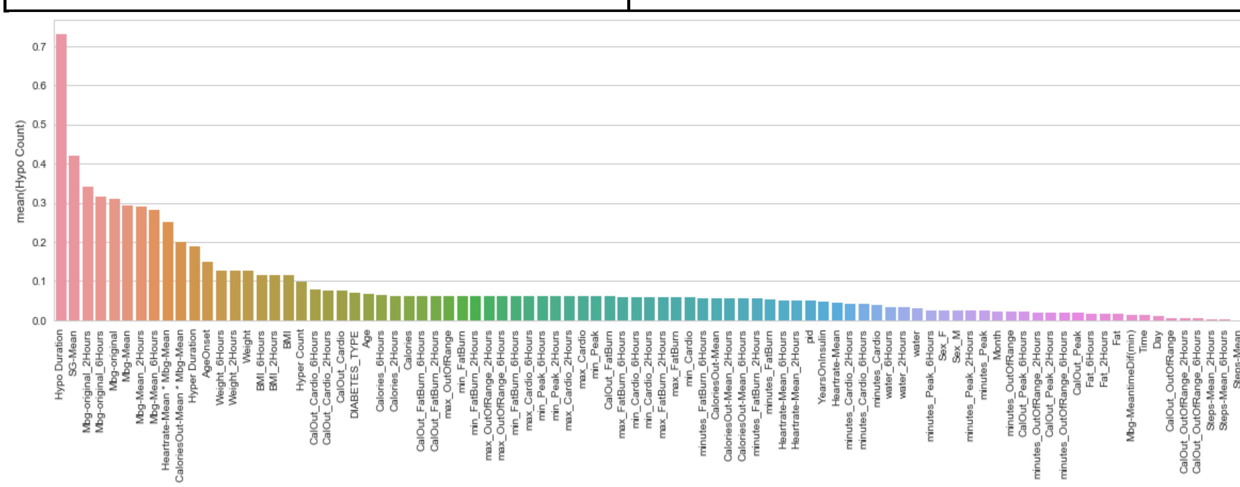
Correlation between all features versus Hyper Count

Hyper Count 1.000000
Hyper Duration 0.431635
SG-Mean 0.310531
Mbg-original_2Hours 0.275118
Mbg-original 0.256415
Mbg-original_6Hours 0.246400
Mbg-Mean 0.233512
Mbg-Mean_2Hours 0.225431
Heartrate-Mean * Mbg-Mean 0.215299
Mbg-Mean_6Hours 0.213062
CaloriesOut-Mean * Mbg-Mean 0.138313
Hypo Duration 0.121694
Hypo Count 0.098516
Age 0.057546
Month 0.057424
min_FatBurn 0.053351
max_OutOfRange 0.053351
max_OutOfRange_2Hours 0.053347
min_FatBurn_2Hours 0.053347
max_OutOfRange_6Hours 0.053336
min_FatBurn_6Hours 0.053336
min_Peak 0.052268
max_Cardio 0.052268
min_Peak_2Hours 0.052264
max_Cardio_2Hours 0.052264
min_Peak_6Hours 0.052254
max_Cardio_6Hours 0.052254
max_FatBurn 0.051720
min_Cardio 0.051720
max_FatBurn_2Hours 0.051717
min_Cardio_2Hours 0.051717
max_FatBurn_6Hours 0.051707
min_Cardio_6Hours 0.051707
Heartrate-Mean_2Hours 0.049406
Heartrate-Mean 0.048752
Mbg-MeantimeDif(min) 0.043732
Heartrate-Mean_6Hours 0.042514
CalOut_Cardio_6Hours 0.037525
Sex_M 0.036586
Sex_F 0.036586
Calories_6Hours 0.035799
CalOut_Cardio_2Hours 0.035361
CalOut_Cardio 0.034803
Calories_2Hours 0.034656
Calories 0.034111
Time 0.033559
AgeOnset 0.028995
BMI 0.028681
BMI_2Hours 0.028667
BMI_6Hours 0.028634
minutes_OutOfRange_6Hours 0.028103
DIABETES_TYPE 0.027847
minutes_OutOfRange_2Hours 0.026742
minutes_OutOfRange 0.026030
Steps-Mean_2Hours 0.025725
Steps-Mean 0.025655
CalOut_OutOfRange_6Hours 0.024711
CalOut_OutOfRange_2Hours 0.023800
CalOut_OutOfRange 0.023253
minutes_Cardio_6Hours 0.022912
minutes_Cardio 0.022019
water_6Hours 0.021795
minutes_Cardio_2Hours 0.021674
water_2Hours 0.021277
YearsOnInsulin 0.021087
water 0.021011
CaloriesOut-Mean 0.020178
Steps-Mean_6Hours 0.020009
pid 0.019381
CaloriesOut-Mean_2Hours 0.017865
CalOut_FatBurn 0.017303
CalOut_FatBurn_6Hours 0.017292
CalOut_FatBurn_2Hours 0.017282
Weight 0.014860
Weight_2Hours 0.014846
Weight_6Hours 0.014811
minutes_Peak_6Hours 0.013256
minutes_Peak_2Hours 0.013133
minutes_Peak 0.013079
Fat 0.012316
Fat_2Hours 0.012006
CaloriesOut-Mean_6Hours 0.011286
Fat_6Hours 0.011177
CalOut_Peak 0.010754
CalOut_Peak_2Hours 0.010721
CalOut_Peak_6Hours 0.010670
minutes_FatBurn 0.004471
minutes_FatBurn_2Hours 0.004396
minutes_FatBurn_6Hours 0.004360
Day 0.002044



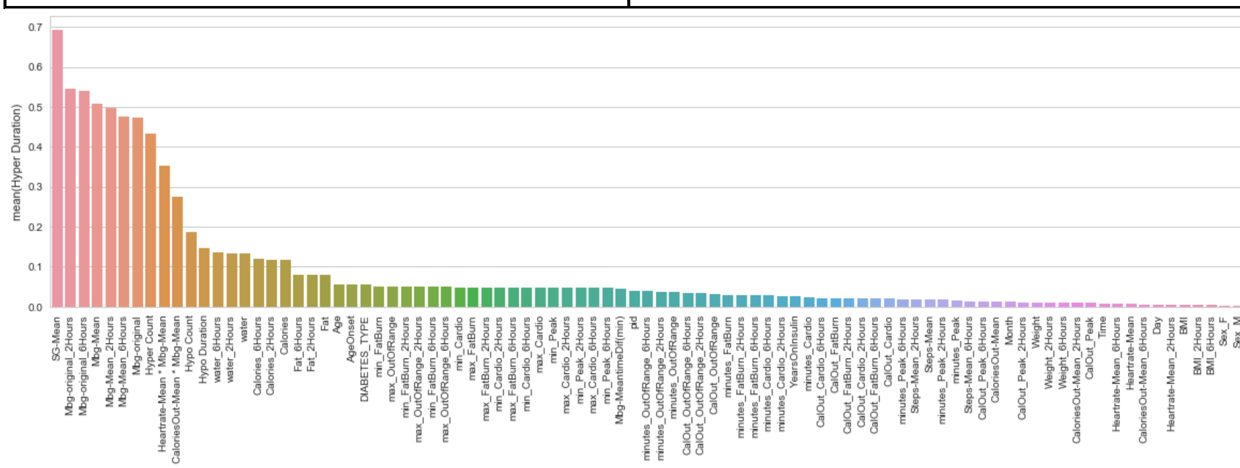
Correlation between all features versus Hypo Count

Hypo Count 1.000000
Hypo Duration 0.731197
SG-Mean 0.421450
Mbg-original_2Hours 0.341226
Mbg-original_6Hours 0.317096
Mbg-original 0.311218
Mbg-Mean 0.293565
Mbg-Mean_2Hours 0.289629
Mbg-Mean_6Hours 0.281933
Heartrate-Mean * Mbg-Mean 0.250806
CaloriesOut-Mean * Mbg-Mean 0.198811
Hyper Duration 0.187984
AgeOnset 0.147907
Weight_6Hours 0.127137
Weight_2Hours 0.127105
Weight 0.127099
BMI_6Hours 0.116156
BMI_2Hours 0.116113
BMI 0.116105
Hyper Count 0.098516
CalOut_Cardio_6Hours 0.077406
CalOut_Cardio_2Hours 0.076880
CalOut_Cardio 0.075997
DIABETES_TYPE 0.070882
Age 0.068596
Calories_6Hours 0.063326
Calories_2Hours 0.062981
Calories 0.062773
CalOut_FatBurn_6Hours 0.062291
CalOut_FatBurn_2Hours 0.060936
max_OutOfRange 0.060906
min_FatBurn 0.060906
min_FatBurn_2Hours 0.060904
max_OutOfRange_2Hours 0.060904
max_OutOfRange_6Hours 0.060899
min_FatBurn_6Hours 0.060899
max_Cardio_6Hours 0.060340
min_Peak_6Hours 0.060340
min_Peak_2Hours 0.060338
max_Cardio_2Hours 0.060338
max_Cardio 0.060337
min_Peak 0.060337
CalOut_FatBurn 0.060321
max_FatBurn_6Hours 0.059479
min_Cardio_6Hours 0.059479
min_Cardio_2Hours 0.059476
max_FatBurn_2Hours 0.059476
max_FatBurn 0.059474
min_Cardio 0.059474
minutes_FatBurn_6Hours 0.055804
CaloriesOut-Mean 0.055359
CaloriesOut-Mean_2Hours 0.055166
CaloriesOut-Mean_6Hours 0.055156
minutes_FatBurn_2Hours 0.054778
minutes_FatBurn 0.054273
Heartrate-Mean_6Hours 0.050197
Heartrate-Mean_2Hours 0.049653
pid 0.048975
YearsOnInsulin 0.046731
Heartrate-Mean 0.045480
minutes_Cardio_2Hours 0.040657
minutes_Cardio_6Hours 0.040449
minutes_Cardio 0.039456
water_6Hours 0.033378
water_2Hours 0.032130
water 0.031573
minutes_Peak_6Hours 0.025838
Sex_F 0.025417
Sex_M 0.025417
minutes_Peak_2Hours 0.025126
minutes_Peak 0.024746
Month 0.021438
minutes_OutOfRange 0.020827
CalOut_Peak_6Hours 0.020706
minutes_OutOfRange_2Hours 0.020434
CalOut_Peak_2Hours 0.019852
minutes_OutOfRange_6Hours 0.019613
CalOut_Peak 0.019429
Fat_6Hours 0.016383
Fat_2Hours 0.015997
Fat 0.015847
Mbg-MeantimeDif(min) 0.014794
Time 0.012412
Day 0.009421
CalOut_OutOfRange 0.005153
CalOut_OutOfRange_2Hours 0.004787
CalOut_OutOfRange_6Hours 0.004017
Steps-Mean_2Hours 0.002433
Steps-Mean_6Hours 0.001965
Steps-Mean 0.000785



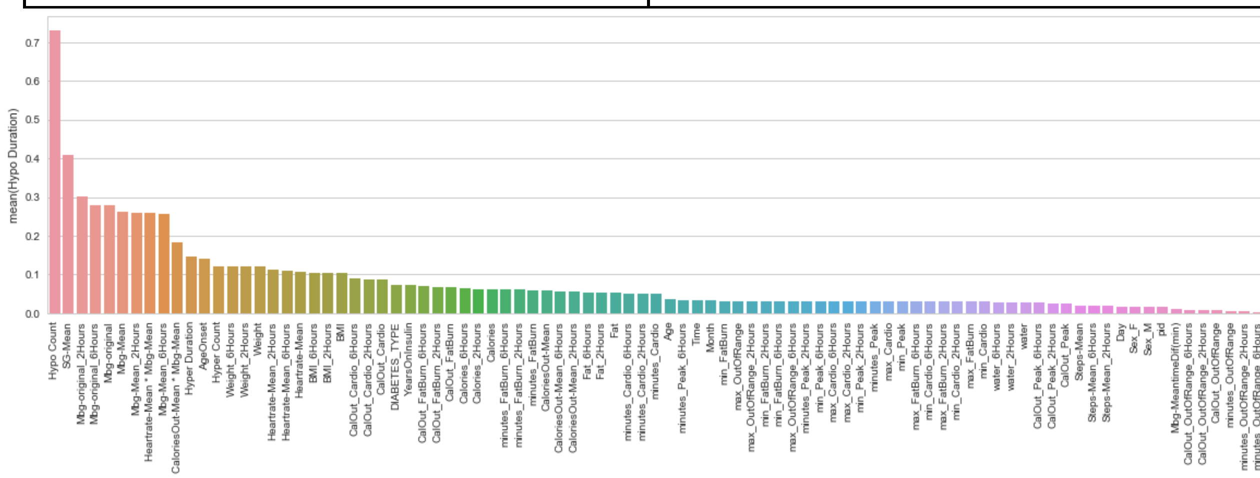
Correlation between all features versus Hyper Duration

Hyper Duration 1.000000
SG-Mean 0.692502
Mbg-original_2Hours 0.544804
Mbg-original_6Hours 0.538946
Mbg-Mean 0.506610
Mbg-Mean_2Hours 0.496731
Mbg-Mean_6Hours 0.476087
Mbg-original 0.472885
Hyper Count 0.431635
Heartrate-Mean * Mbg-Mean 0.353218
CaloriesOut-Mean * Mbg-Mean 0.274696
Hypo Count 0.187984
Hypo Duration 0.145295
water_6Hours 0.135587
water_2Hours 0.134185
water 0.133537
Calories_6Hours 0.120432
Calories_2Hours 0.117385
Calories 0.116116
Fat_6Hours 0.080596
Fat_2Hours 0.080294
Fat 0.080171
Age 0.056223
AgeOnset 0.056146
DIABETES_TYPE 0.055858
min_FatBurn 0.050613
max_OutOfRange 0.050613
min_FatBurn_2Hours 0.050608
max_OutOfRange_2Hours 0.050608
min_FatBurn_6Hours 0.050595
max_OutOfRange_6Hours 0.050595
min_Cardio 0.046609
max_FatBurn 0.046609
max_FatBurn_2Hours 0.046607
min_Cardio_2Hours 0.046607
max_FatBurn_6Hours 0.046601
min_Cardio_6Hours 0.046601
max_Cardio 0.046491
min_Peak 0.046491
max_Cardio_2Hours 0.046489
min_Peak_2Hours 0.046489
max_Cardio_6Hours 0.046482
min_Peak_6Hours 0.046482
Mbg-MeantimeDif(min) 0.044810
pid 0.039475
minutes_OutOfRange_6Hours 0.038470
minutes_OutOfRange_2Hours 0.037353
minutes_OutOfRange 0.036744
CalOut_OutOfRange_6Hours 0.033769
CalOut_OutOfRange_2Hours 0.032954
CalOut_OutOfRange 0.032469
minutes_FatBurn 0.029872
minutes_FatBurn_2Hours 0.029626
minutes_FatBurn_6Hours 0.028762
minutes_Cardio_6Hours 0.027699
minutes_Cardio_2Hours 0.026140
YearsOnInsulin 0.025135
minutes_Cardio 0.024605
CalOut_Cardio_6Hours 0.021700
CalOut_FatBurn 0.021149
CalOut_FatBurn_2Hours 0.021038
CalOut_Cardio_2Hours 0.020832
CalOut_FatBurn_6Hours 0.020643
CalOut_Cardio 0.020610
minutes_Peak_6Hours 0.018607
Steps-Mean_2Hours 0.018088
Steps-Mean 0.017730
minutes_Peak_2Hours 0.017122
minutes_Peak 0.016613
Steps-Mean_6Hours 0.013075
CalOut_Peak_6Hours 0.012212
CaloriesOut-Mean 0.012098
Month 0.011733
CalOut_Peak_2Hours 0.010963
Weight 0.010809
Weight_2Hours 0.010809
Weight_6Hours 0.010801
CaloriesOut-Mean_2Hours 0.010692
CalOut_Peak 0.010555
Time 0.008511
Heartrate-Mean_6Hours 0.007365
Heartrate-Mean 0.007042
CaloriesOut-Mean_6Hours 0.005716
Day 0.005536
Heartrate-Mean_2Hours 0.004728
BMI 0.003520
BMI_2Hours 0.003515
BMI_6Hours 0.003512
Sex_F 0.002467
Sex_M 0.002467



Correlation between all features versus Hypo Duration

Hypo Duration 1.000000	minutes_Cardio 0.049625
Hypo Count 0.731197	Age 0.036157
SG-Mean 0.408548	minutes_Peak_6Hours 0.032716
Mbg-original_2Hours 0.301977	Time 0.032566
Mbg-original_6Hours 0.279004	Month 0.032027
Mbg-original 0.278180	min_FatBurn 0.031883
Mbg-Mean 0.261168	max_OutOfRange 0.031883
Mbg-Mean_2Hours 0.259692	max_OutOfRange_2Hours 0.031880
Heartrate-Mean * Mbg-Mean 0.259385	min_FatBurn_2Hours 0.031880
Mbg-Mean_6Hours 0.255682	min_FatBurn_6Hours 0.031870
CaloriesOut-Mean * Mbg-Mean 0.182910	max_OutOfRange_6Hours 0.031870
Hyper Duration 0.145295	minutes_Peak_2Hours 0.031474
AgeOnset 0.140298	min_Peak_6Hours 0.030957
Hyper Count 0.121694	max_Cardio_6Hours 0.030957
Weight_6Hours 0.120359	max_Cardio_2Hours 0.030946
Weight_2Hours 0.120323	min_Peak_2Hours 0.030946
Weight 0.120314	minutes_Peak 0.030943
Heartrate-Mean_2Hours 0.111968	max_Cardio 0.030941
Heartrate-Mean_6Hours 0.110892	min_Peak 0.030941
Heartrate-Mean 0.107818	max_FatBurn_6Hours 0.029741
BMI_6Hours 0.103070	min_Cardio_6Hours 0.029741
BMI_2Hours 0.103022	max_FatBurn_2Hours 0.029726
BMI 0.103010	min_Cardio_2Hours 0.029726
CalOut_Cardio_6Hours 0.088860	max_FatBurn 0.029720
CalOut_Cardio_2Hours 0.087657	min_Cardio 0.029720
CalOut_Cardio 0.086518	water_6Hours 0.028651
DIABETES_TYPE 0.074203	water_2Hours 0.028355
YearsOnInsulin 0.072232	water 0.028185
CalOut_FatBurn_6Hours 0.068676	CalOut_Peak_6Hours 0.026589
CalOut_FatBurn_2Hours 0.067250	CalOut_Peak_2Hours 0.025184
CalOut_FatBurn 0.066573	CalOut_Peak 0.024607
Calories_6Hours 0.063848	Steps-Mean 0.020370
Calories_2Hours 0.062819	Steps-Mean_6Hours 0.018771
Calories 0.062352	Steps-Mean_2Hours 0.017819
minutes_FatBurn_6Hours 0.061382	Day 0.016473
minutes_FatBurn_2Hours 0.060271	Sex_F 0.016092
minutes_FatBurn 0.059732	Sex_M 0.016092
CaloriesOut-Mean 0.057464	pid 0.015868
CaloriesOut-Mean_6Hours 0.056987	Mbg-MeantimeDif(min) 0.011338
CaloriesOut-Mean_2Hours 0.056901	CalOut_OutOfRange_6Hours 0.008168
Fat_6Hours 0.052408	CalOut_OutOfRange_2Hours 0.007101
Fat_2Hours 0.052045	CalOut_OutOfRange 0.006633
Fat 0.051901	minutes_OutOfRange 0.004345
minutes_Cardio_6Hours 0.051148	minutes_OutOfRange_2Hours 0.003815
minutes_Cardio_2Hours 0.050951	minutes_OutOfRange_6Hours 0.002597



Step 4: Automating Insights

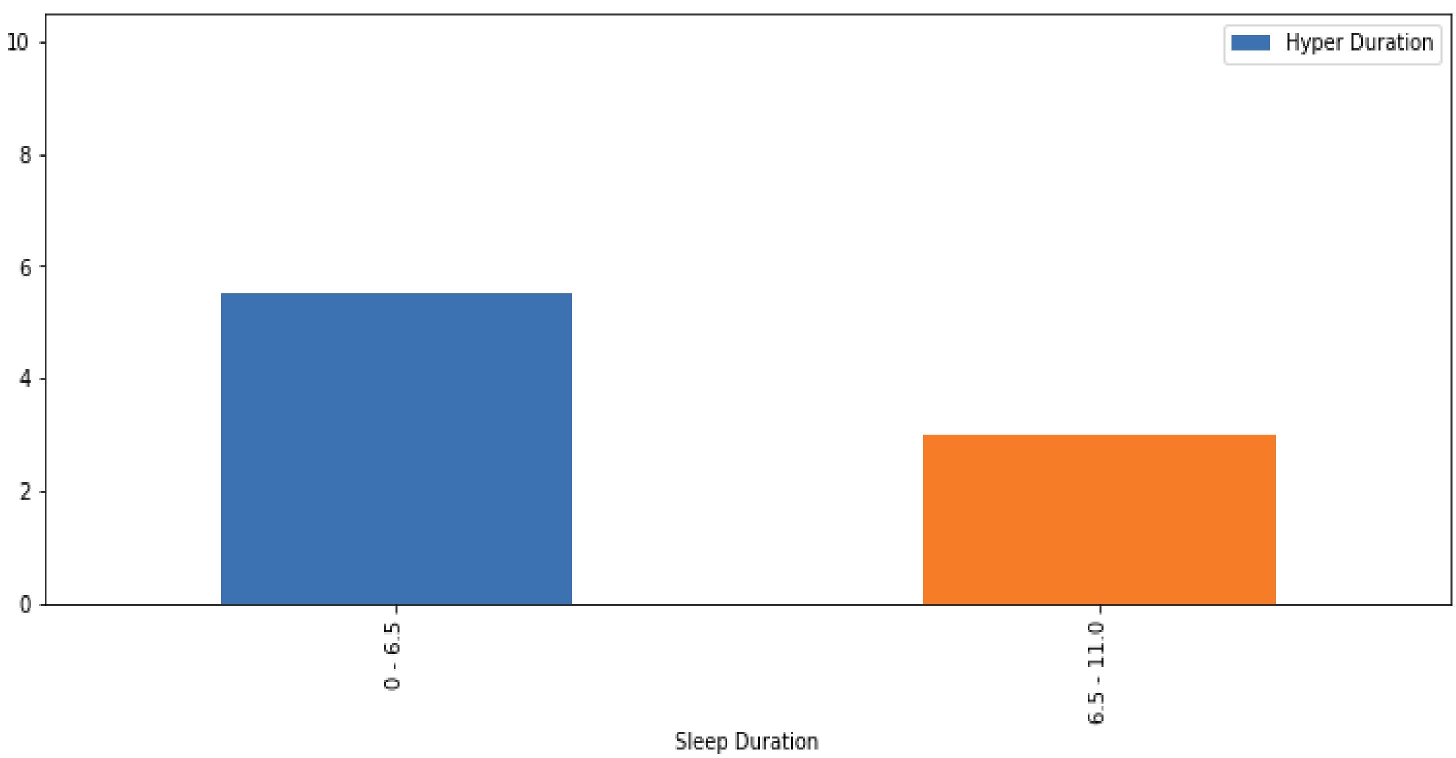
- Medtronic Data Science team has specified and communicated what type of automation they wanted. We discussed graphs on hyper/hypo events and their durations. We then created insights using these features. These features include hyper/hypo durations with sleep duration, step count and calories out.

.

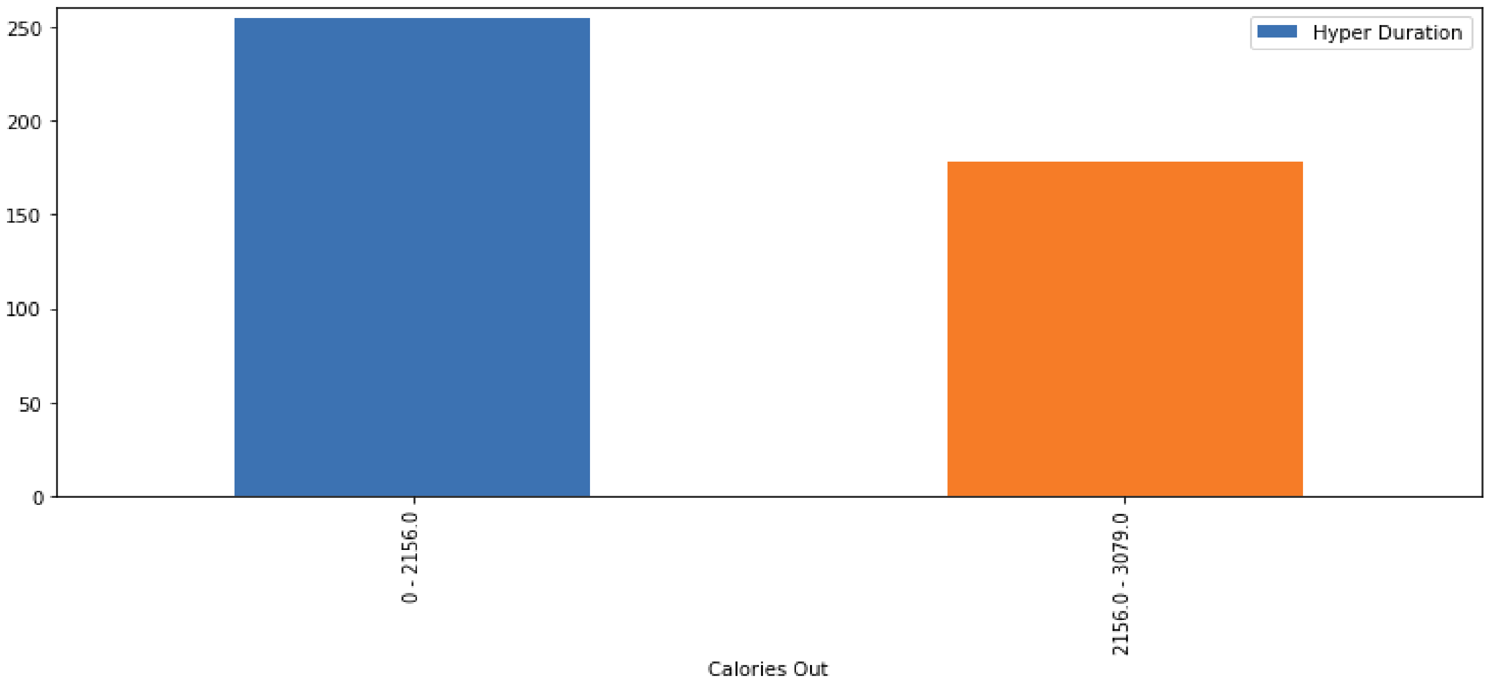
- We have created an algorithm for automating our insights. Each column on the graphs is automatically generated based on the data of each individual patient.

- As you can see from the sample graphs and insights automatically generated below, the value of each bar is based on either the number of steps a day, amount of sleep overnight, and calories outputted daily of the patient. The outcome is that the insights suggest the patient take certain steps in order to lower their duration spent in Hyperglycemia per day.

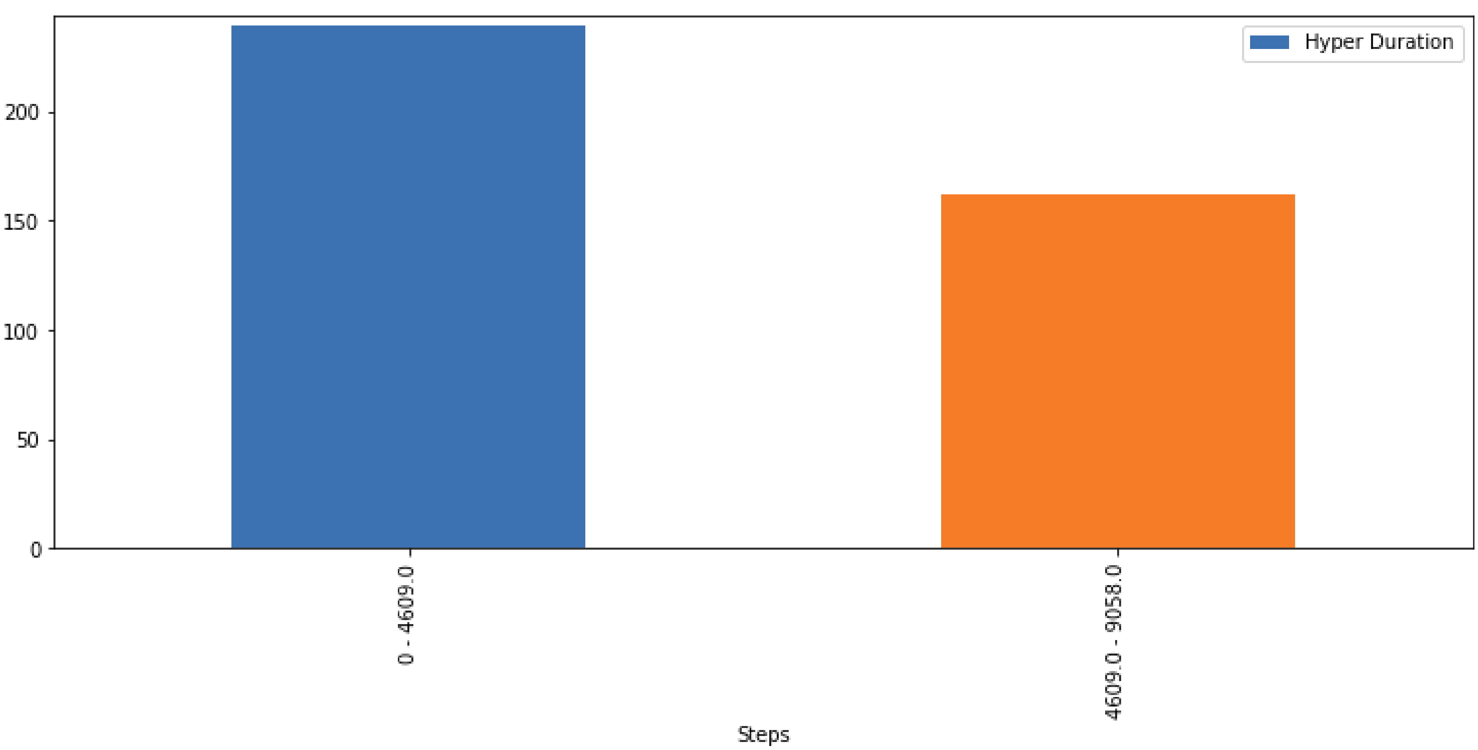
Hi, when you have more than 6.5 Hours of Sleep overnight, a day, you will get 30% reduction in Hyper Duration!



Hi, when you have more than 2156.0 Calories Out a day, a day, you will get 30% reduction in Hyper Duration!



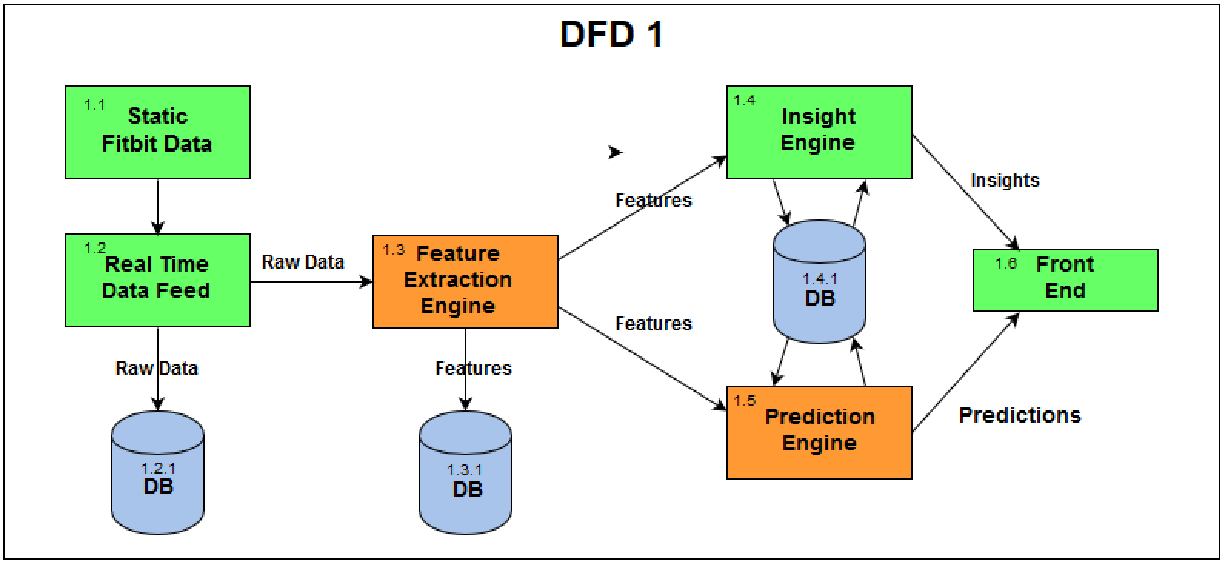
Hi, when you have more than 4609.0 Steps a day, a day, you will get 30% reduction in Hyper Duration!

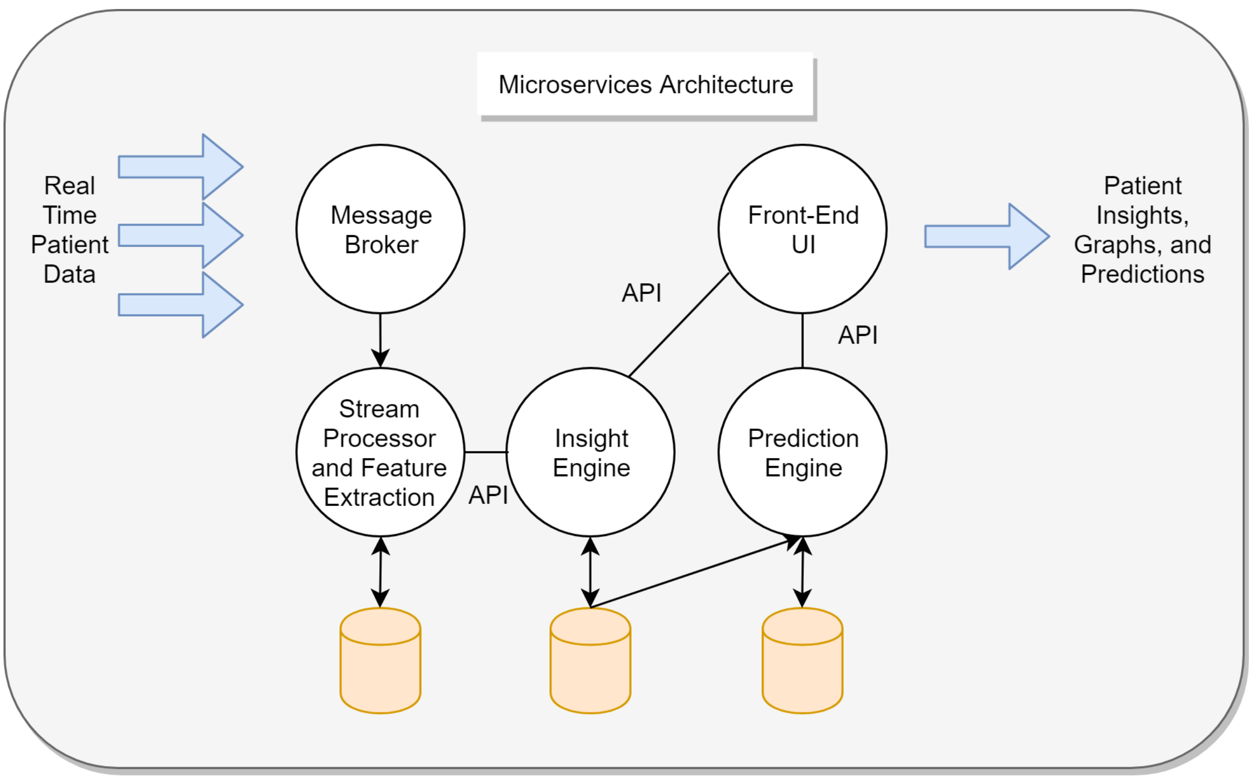


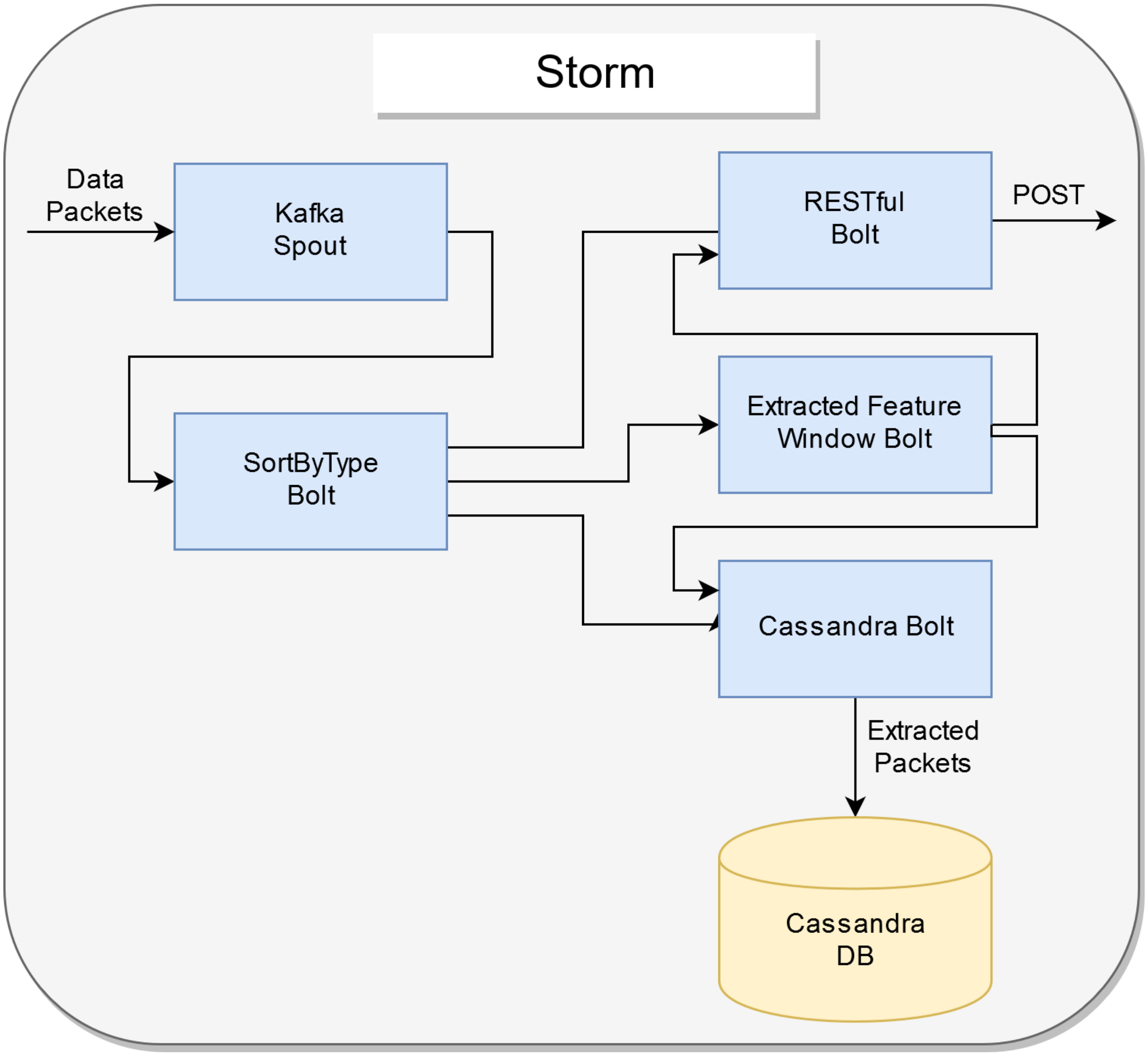
Technologies Used

System Architecture

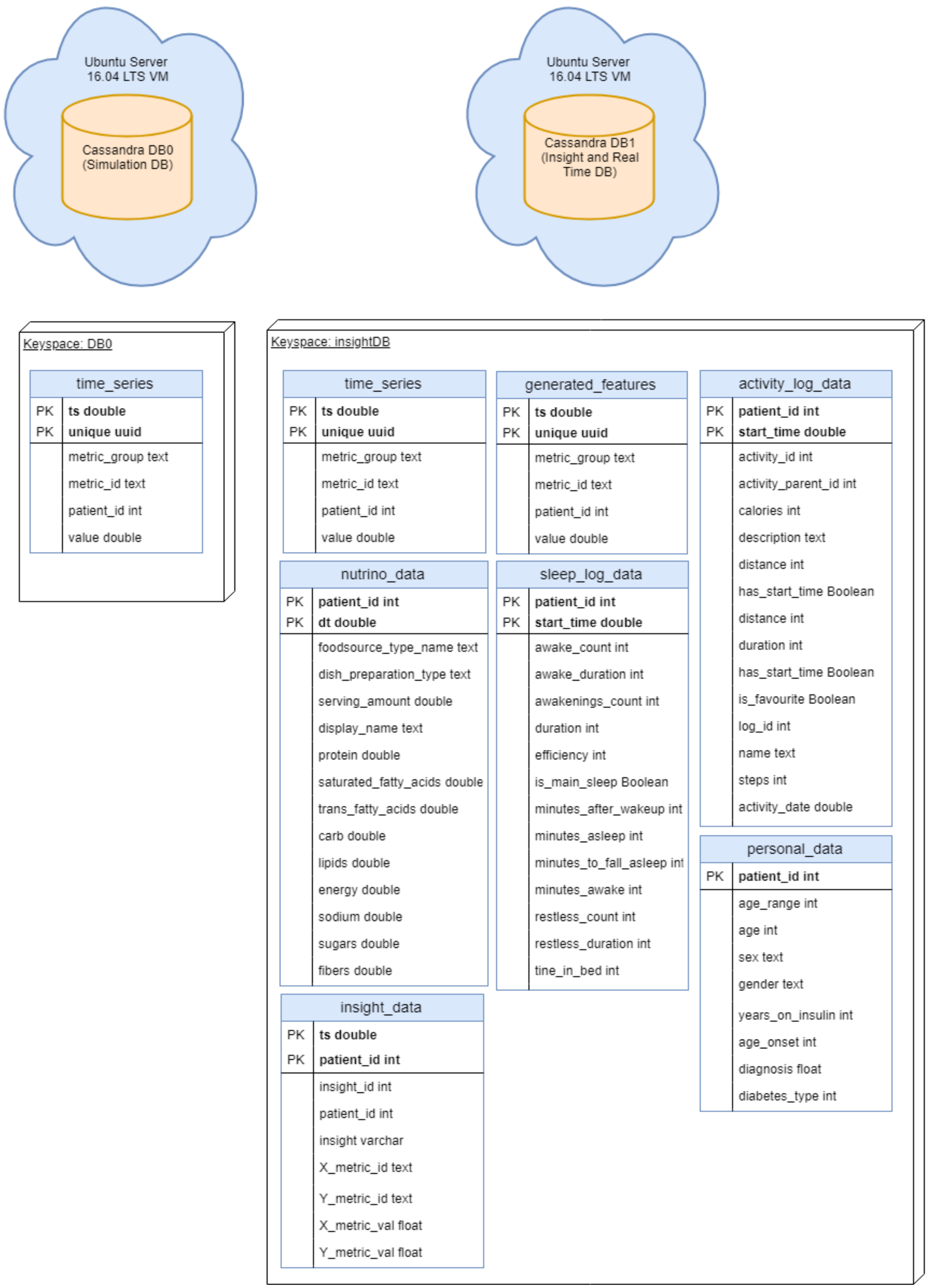
Data Flow






Storm
Data Packets
Kafka Spout
RESTful Bolt
POST
Extracted Feature Window Bolt
SortByType Bolt
Cassandra Bolt
Extracted Packets
Cassandra DB

Database Schema



Conclusion and Results

After combining the real time engine with both the Prediction Engine and the Insight Engine, we have completed a Gentelella dashboard: