ST 495 Final Project Report

Group E

Executive Summary

This report contains information on device data in the MAUDE database which focuses on Dexcom glucose monitors. Exploratory data analysis was conducted on three models (G4, G5, G6) to observe trends in the reports within devices. The time series analysis decomposed the data to observe trends and seasonality, as well as forecasting frequencies of reports in years to come. This approach will allow for prediction on how many devices may fail in the future from the current data. Text analyses were performed on the subsetted data to extract topics for time series analysis. The text analysis extracted topics from the device reports to address current problems and find the most common reason for device reports.

Problem Setting

We will be trying to answer the following questions:

Does the frequency of newer reports on Dexcom Continuous Glucose Monitor devices follow the trend of frequency of previous reports?

What sort of problems need to be addressed in the future?

We chose to investigate these research problems because continuous blood glucose monitor systems were the second most reported device in the MAUDE database, and out of the glucose monitor manufacturers, Dexcom is the most popular. This analysis will be useful in determining if new models are actually better than old models for Dexcom Glucose Monitor devices.

Data Description

We first extracted device data from the MAUDE database, focusing on reports on Dexcom Continuous Glucose Monitor devices including models G4, G5, and G6, which spans from 2016 to 2021. We also extracted narrative data from the MAUDE database specifically on reports on the Dexcom Continuous Glucose Monitor G6 device from 2018 to 2020.

In regards to the data for topic modeling, we will be working with the original case report which is found in (B.5) of the case report form located here: https://www.fda.gov/media/76299/download. Approximately one-half of the dataset involves these original reports. The other half involves updates or descriptor information.

Methods

Once the device reports were submitted to include only relevant information such as the number of reports and the date reported, the variables were formatted so that they could be graphed. The data visualization included area graphs for each report as well as a stacked area graph to compare the different models. A proportional area graph was created as well to better compare the three devices.

To better understand the patterns observed in the area graphs from EDA, we apply time series analysis to investigate the change in density of the device reports by month of all 3 generations of Dexcom Continuous Glucose Monitor systems: G4, G5, and G6. First, we convert our data to time series data. Then, for each device's time-series data, we use STL decomposition

(Seasonal and Trend decomposition using LOESS) to convert the data into underlying components. These components include:

1. Trend, which represents the overall progression of the data;
2. Seasonality, which represents the variation in the data that occurs in a yearly pattern; and
3. The remainder, which represents the random noise that is not accounted for in the other two components.

STL decomposition is an additive decomposition method, which means that the components add up to the original data. The LOESS (Locally-Weighted Smoothing) approach is used as it is a non-parametric smoothing technique that does not require prior knowledge of the distribution of the data.

Following the decomposition, we move on to perform time series forecasting. From our research we learned that both G4 and G5 models are discontinued from Dexcom by the end of 2020, so we focus on forecasting future G6 report count.

We fit a SARIMA (Seasonal Autoregressive Integrated Moving Average) model over the G6 data from May 2018 to December 2020. There are 2 approaches to fitting the model, using the built-in R function to select the best ARIMA model for our G6 data or finding the tuning parameters manually then fitting an ARIMA model using those tuning parameters. We opt for the first approach because there are likely inconsistencies in finding the tuning parameters manually and for SARIMA models, slight differences in certain parameters such as the AR terms and MA terms do not lead to drastic changes in the forecast. After finding the best SARIMA model, we run the residual analysis with the autocorrelation function to confirm the fit of the model.

Then we can forecast the future values for device report count by month using our SARIMA model. We then compare the predictions with the actual device report count data collected in 2021 from the MAUDE database and produce an error plot. Now that we have insight into projected reports, we can begin to look deeper into what these reports are about.

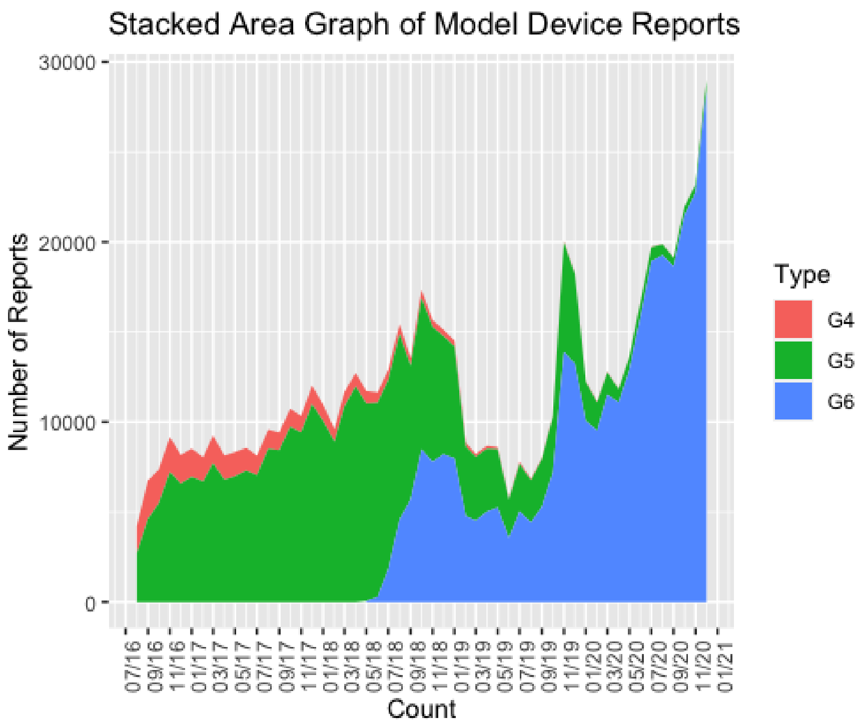
We can explore past report problems and observe how these issues changed in frequency throughout the lifespan of the G6 model. With the aid of natural language processing, we can extract topics from each report. These topics can then be tracked over time to gain insight into what issues might arise in the future. A large part of our natural language processing task involves preprocessing. We are working with the subsetted data mentioned in the 'Data Description' section. First, we will remove regular expressions from the text as well as change all words to lowercase. As a result, the text is more uniform. Next, we will remove stop words. One important aspect to note in this step is the word 'not'. 'Not' can be problematic as it technically is a stop word but in our case 'not' provides useful information for certain topics so this word will not be added to stop words. We have also appended certain stop words after the first iteration of our model building. Next, we will use lemmatization to reduce words down to base forms. We chose lemmatization over stemming to aid in the interpretability of our models' output. Because our task is unsupervised, the model is evaluated based on human judgment. With the clean text, we will use TF-IDF to convert the documents (reports) into a document term matrix. This method is needed to gain weighted frequencies of words found in each report. This converts our text data into numeric form.

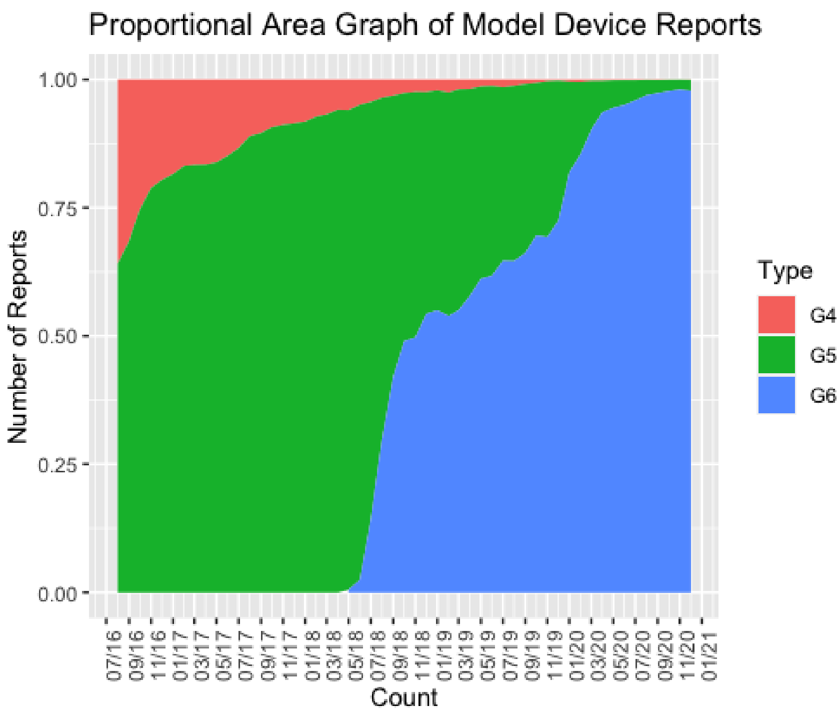
Now that our data is preprocessed, we will explore models. We are going to consider the following three models: Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA), and Non-Negative Matrix Factorization (NMF). LSA will allow for an effective way to convert our document term matrix into two matrices. The output will be a document-topic matrix and a topic-term matrix. LDA will take a different approach and output a probability for the number of topics specified for model building. Due to the iterative nature of LDA, this model will not be explored to the full extent due to limited computational resources. Lastly, NMF will produce results similar to LSA but the output is strictly positive. NMF is a linear algebra-based, dimensionality reduction technique that will split the document term matrix into a document-topic matrix and a topic-term matrix. In all cases, the models will be evaluated based on the interpretability of the topics. We will explore a different number of topics for each model and compare words in each topic. Once the best model has been chosen, we will label all reports with a topic. These reports will then be mapped back to the date so we can analyze how topics have changed over time. After topics are extracted from each report, we can conduct a time series analysis to look at various components of each series.

Results

EDA

When EDA was conducted on the data, it was observed that the devices contained seasonal peaks which may be associated with factors such as the seasonality of diabetes or manufacturing. The G4 and G5 devices were discontinued at the beginning of 2020 which contributed to the decline of reports. When compared to the number of reports for G5 and G6 devices, the G4 reports made up a smaller percentage of total device reports. The increase in reports for G5 and G6 devices may be due to the increase in the number of people with diabetes who need these devices. With an increase in production, there is also an increase in reports so, in order to predict the number of future reports, a time series analysis was conducted. We also observe a sharp spike in total device reports around the end of 2019, when the coronavirus pandemic began. This is likely because the coronavirus's impact is more severe on diabetic patients compared to patients without diabetes, so more and more people become cautious about diabetes and use the Dexcom glucose monitors, which potentially led to the sharp increase in the report about device failures around that time.





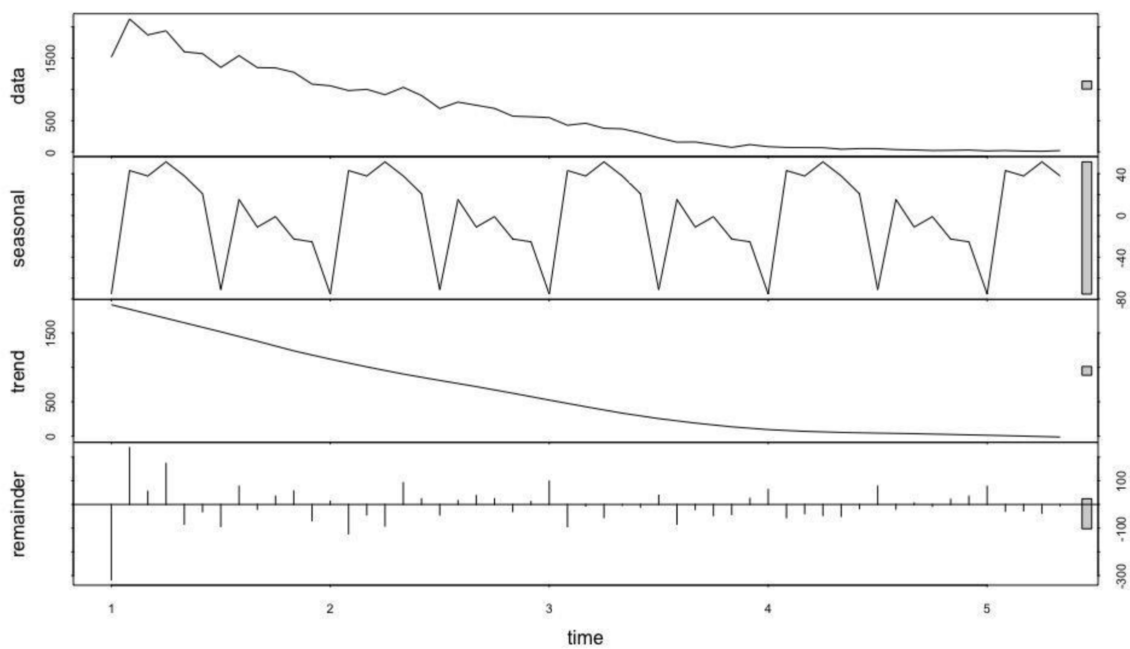
STL Decompositions

The STL decomposition of the time series data produces 4 graphs. From top to bottom, they are:

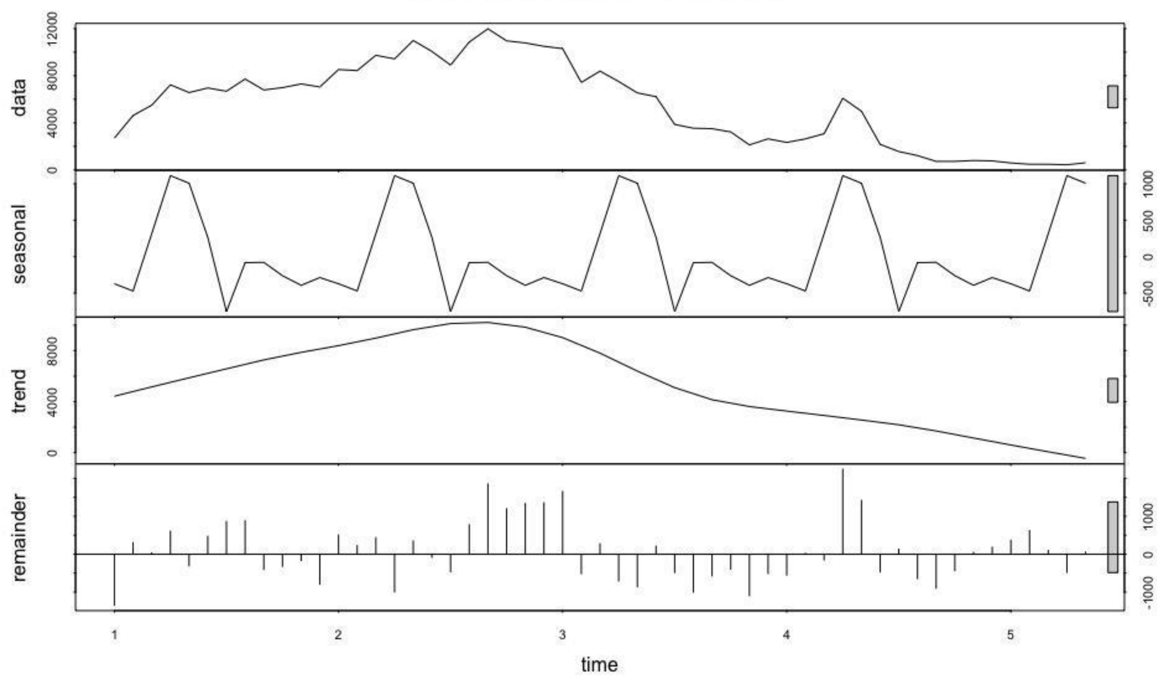
1. Data vs. time: A line graph of the report count by month over time.
2. Seasonal vs. time: Seasonality component of the data over time.
3. Trend vs. time: Trend component of the data over time
4. Remainder vs. time: Random component of the data over time.

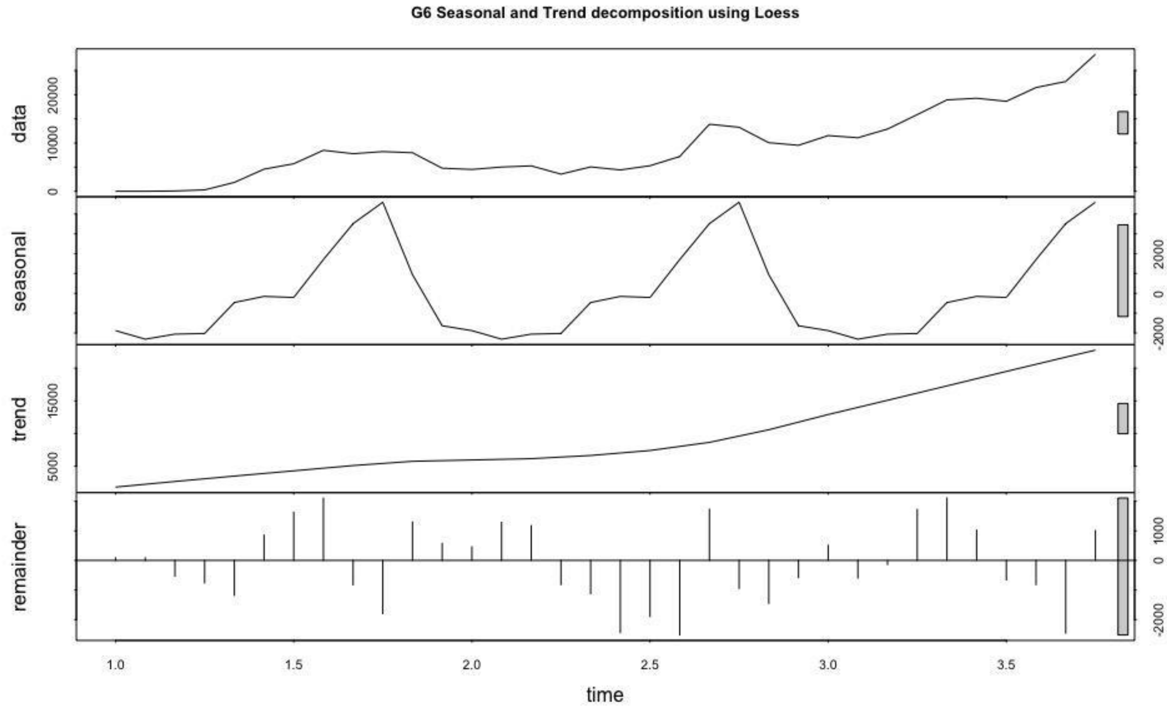
The time axis is in the unit of year, but our data is measured by months, so there are 12 observations in the span of 1 unit of time, evident in the remainder graph.

G4 Seasonal and Trend decomposition using Loess



G5 Seasonal and Trend decomposition using Loess

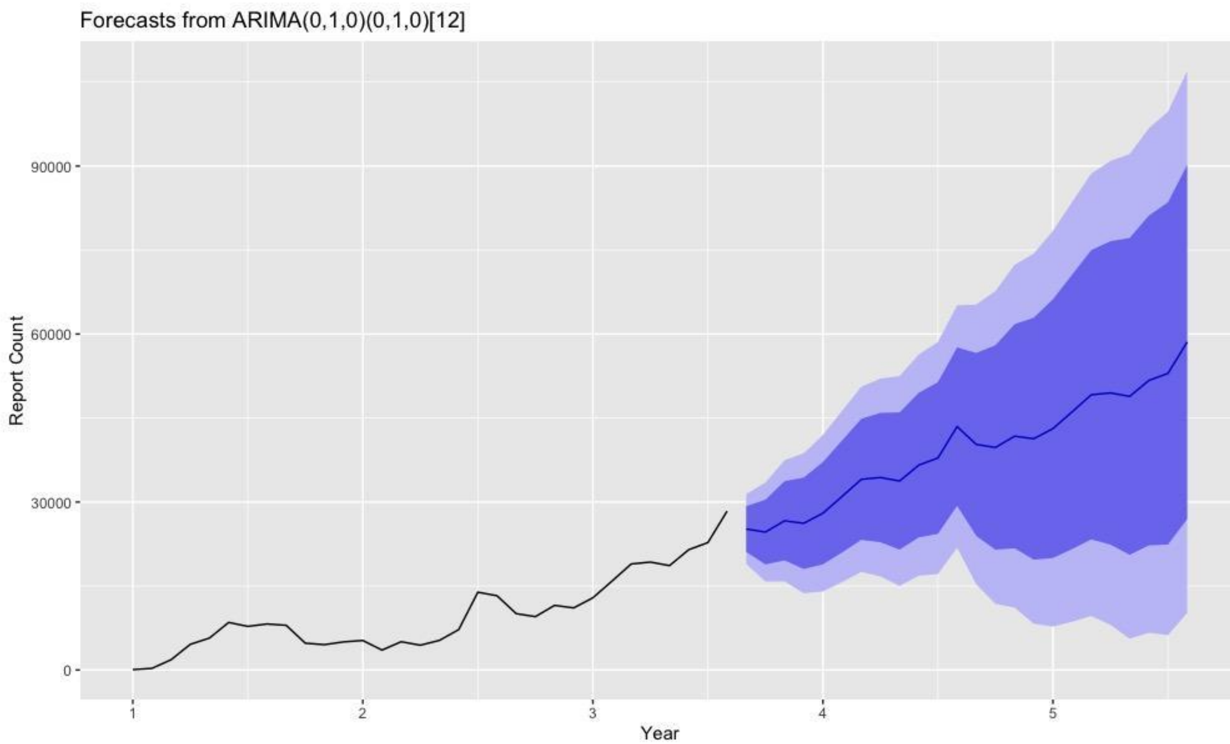




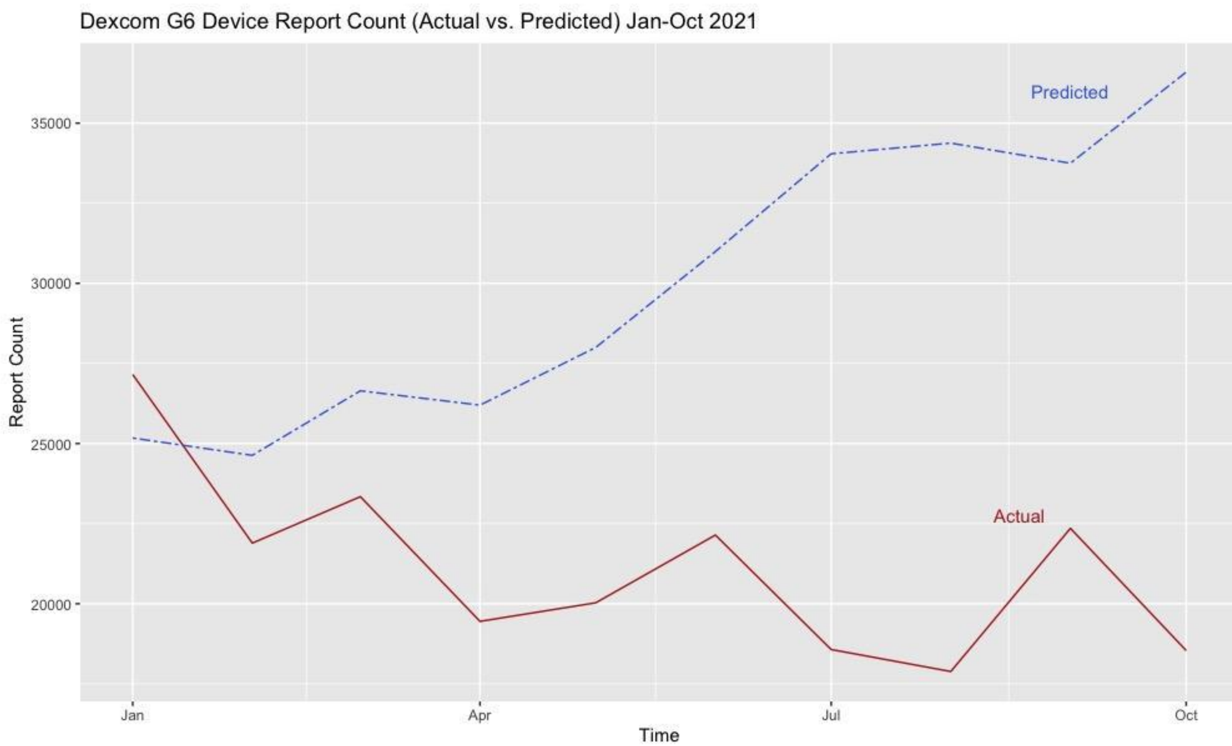
The G4 data starts in August 2016, represented by "1" on the x-axis. From the Seasonal graph, we observe a low report count around August and February and a high report count around December. The data without seasonality shows a strictly decreasing trend, converging to 0.

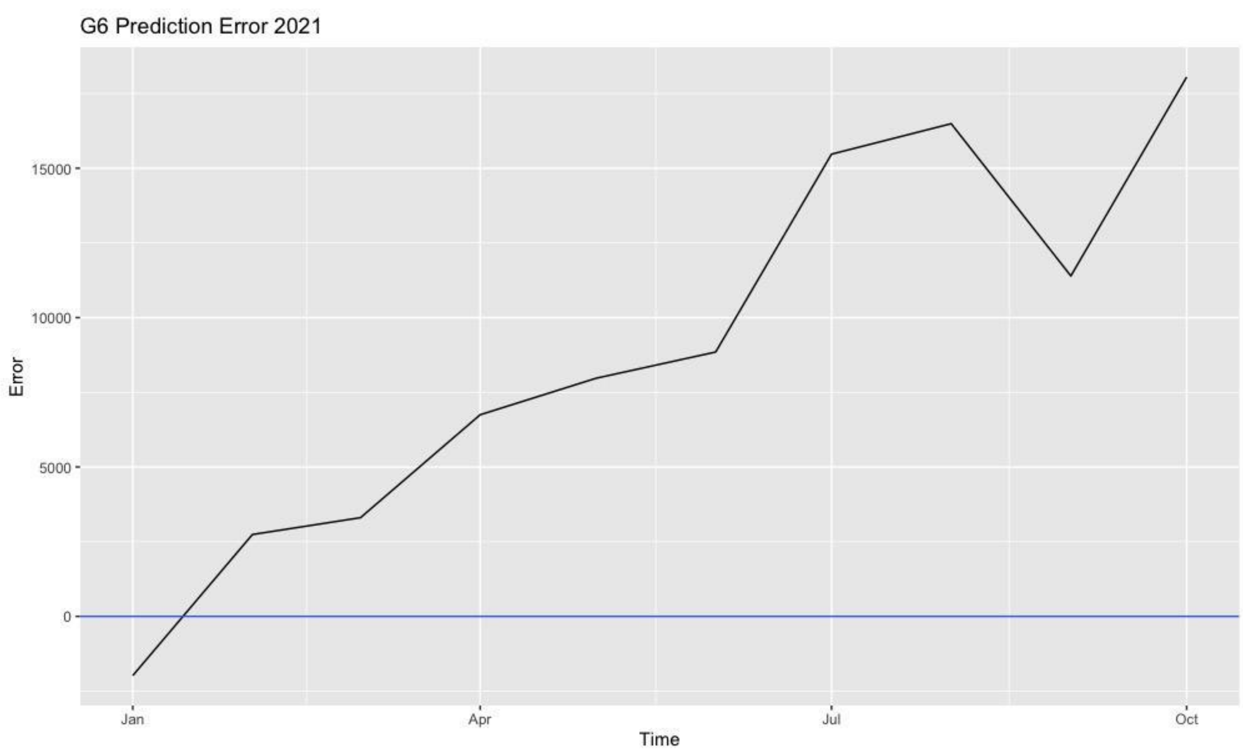
The G5 data also starts in August 2016, represented by "1" on the x-axis. From the Seasonal graph, we observe a similar pattern as the G4 data: low report counts around the end of summer and February and high report counts around the beginning of winter. The overall trend of data without seasonality shows a steady increase from August 2016 to February 2017 then a constant decrease from February 2017 to the end of 2020, converging to 0. In the remainder graph, we observe unusual upward spikes around the end of 2019. This is not picked up by the seasonality component because the spike could be the result of the coronavirus pandemic that has just started around that time, which is not a regular yearly event.

The G6 data starts in May 2018, as it was debuted around that time. May 2018 is represented by “1.0” on the x-axis, May 2019 is represented by “2.0”, etc. From the Seasonal graph, we observe low report counts around the summer, high report counts around February. The overall trend of data without seasonality shows a steady increase ever since G6’s debut. G4 and G5 data converging to 0 are likely due to both G4 and G5 being discontinued and replaced by the new generation model G6.



The best ARIMA model chosen by R is ARIMA(0,1,0)(0,1,0)[12], which indicates that this is a seasonal ARIMA model with 1 differencing to stationarity and 0 non-seasonal AR term, 0 non-seasonal MA term, 0 seasonal AR term, 0 seasonal MA term. The forecast of the G6 data is conducted based on the best ARIMA model, displayed above. The blue line represents the mean of the predictions, the darker blue region represents the 80% confidence interval of the prediction, the lighter blue region represents the 95% confidence interval of the prediction. The forecast follows a general increasing trend and there is a seasonal pattern.





Comparing the predicted frequency and the actual frequency of G6 report data in 2021 from January to October, we see that prediction is growing further apart from the actual data as time goes by. The actual frequency of the G6 reports seems to decrease beginning 2021, while we have predicted the frequency to continue growing from our ARIMA model. This does not necessarily prove that our model is ineffective because we have a limited test data size, and the actual data still falls in the 95% confidence interval of our prediction.

Latent Semantic Analysis (LSA)

When training the LSA model the main parameter we focused on changing was the number of topics. We fit the model using all numbers in the range of two to ten to identify the most interpretable topic results. We discovered the optimal number of topics for LSA was 3. When we ran the model with more than three topics we began to notice overlap. Overall the results of the model provided good results for three topics. The top ten words for each topic are found in the table below. Note, each topic contained 6,010 words.

Topic Number	**Top 10 Words for Each Topic**
Topic 1	'reported', 'transmitter', 'not', 'data', 'allegation', 'injury', 'intervention', 'medical', 'failed', 'occurred'
Topic 2	'loss', 'signal', 'performed', 'one', 'hour', 'within', 'window', 'share', 'log', 'investigation'
Topic 3	'reading', 'glucose', 'bg', 'meter', 'event', 'patient', 'additional', 'information', 'available', 'customer'

When looking at each topic, we can infer what the overall idea of the topic is. After model building, we can classify each report with a value for each topic. Sample output is provided below for the first report (document) in our dataset.

Document/Report Case 1	
dexcom made aware inaccuracy continuous glucose monitor blood glucose meter sensor inserted hip data provided evaluation complaint confirmation probable cause could not determined labeling indicates patient use belly abdomen patient year old choose upper buttock look place belly upper buttock padding sensor not tested approved site talk hcp best site additional event patient information available	
Topic 1	0.13050372978763036
Topic 2	-0.07553725138049512
Topic 3	0.3293313997958946

One important aspect of our output is negative values. These values are measuring the similarity each vector has with each topic. We will compare these results to NMF, which restricts output to positive values.

Latent Dirichlet Allocation (LDA)

As previously mentioned in the 'Methods' section, our lack of computational power hindered our ability to explore more than 3 topics for this model. We also had to reduce the number of iterations to 10. We took a small sample of our data and re-trained the LDA model. The words of each topic on this subset were more interpretable as compared to LSA and NMF. Although this was the case for a subsample, due to scalability issues we did not use this model to extract topics.

Non-Negative Matrix Factorization (NMF)

As with our LSA model, we will fit the model and evaluate topics for a specified number of topics between two and ten. The output for NMF was nearly identical to LSA. We determined the best number of topics was three. In the table below we have listed the top ten words for the three topics. Note, as with LSA, each topic contained 6,010 words.

Topic Number	Top 10 Words for Each Topic
Topic 1	'reported', 'transmitter', 'failed', 'error', 'data', 'injury', 'intervention', 'medical', 'allegation', 'occurred'
Topic 2	'loss', 'signal', 'performed', 'one', 'hour', 'within', 'window', 'share', 'investigation', 'log'
Topic 3	'reading', 'glucose', 'bg', 'meter', 'event', 'customer', 'patient', 'additional', 'information', 'available'

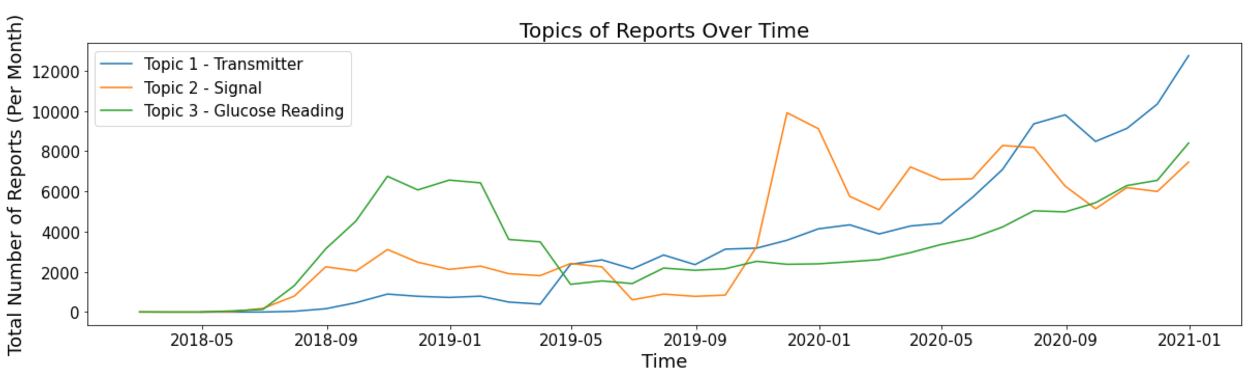
We can see how NMF classified the first report below.

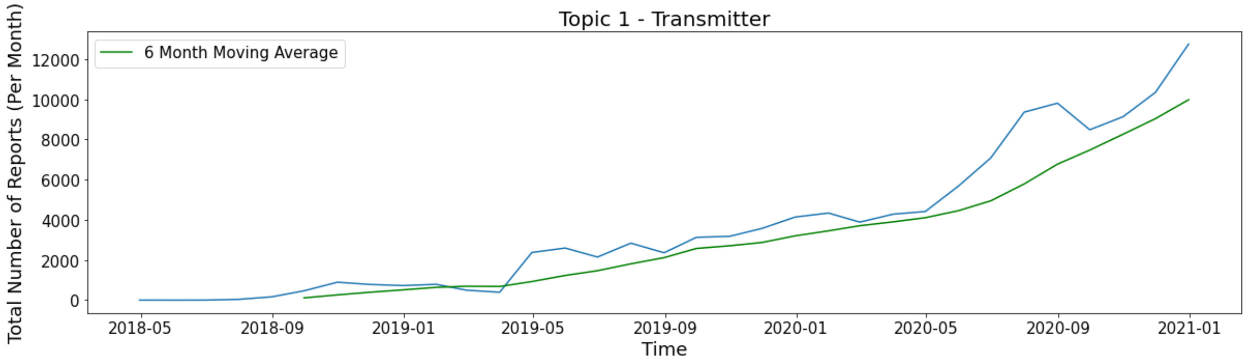
Document/Report Case 1	
dexcom made aware inaccuracy continuous glucose monitor blood glucose meter sensor inserted hip data provided evaluation complaint confirmation probable cause could not determined labeling indicates patient use belly abdomen patient year old choose upper buttock look place belly upper buttock padding sensor not tested approved site talk hcp best site additional event patient information available	
Topic 1	0.00172729
Topic 2	0.0
Topic 3	0.02750821

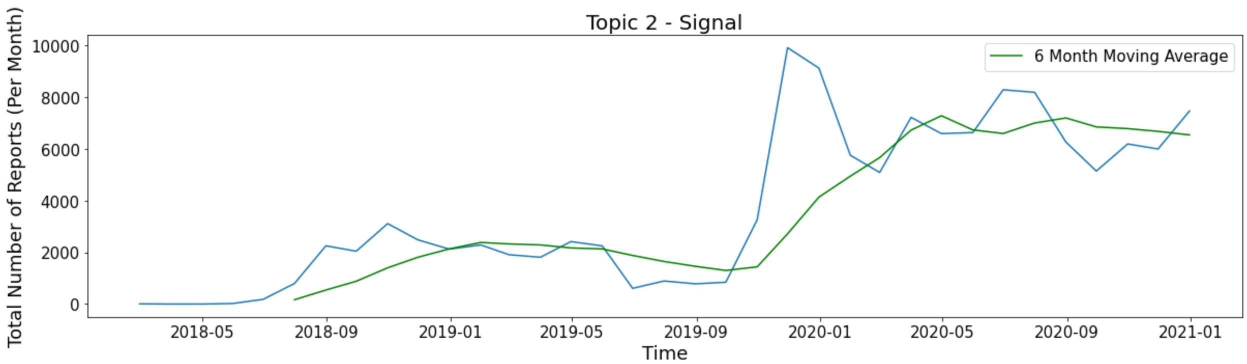
From the results both techniques, LSA and NMF, produce similar results. The notable difference with NMF is the non-negative output. Because the results are very similar, working with strictly positive output will avoid any discrepancies when choosing the best topic for a report. For that reason we will choose to use NMF to label each report case. From our NMF results, we will classify each topic as follows.

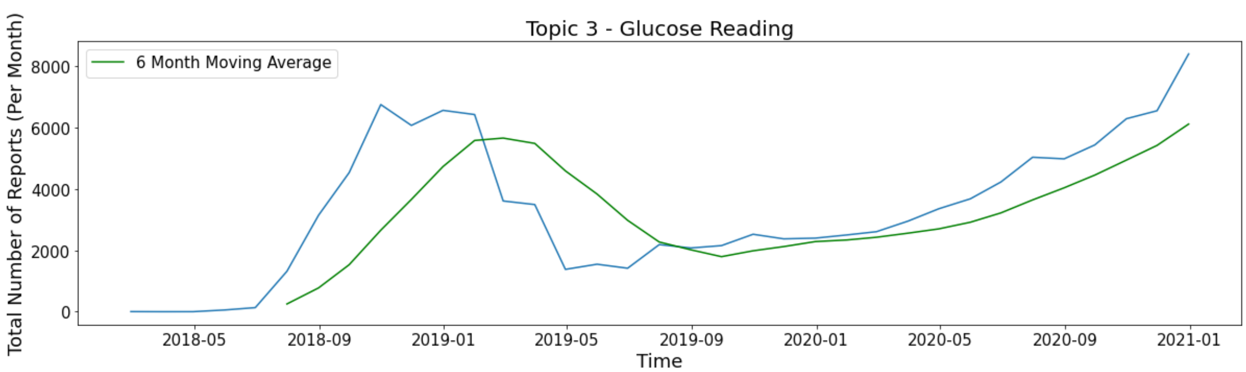
Topic Number	**Top 10 Words for Each Topic**	**Topic Category**
Topic 1	'reported', 'transmitter', 'failed', 'error', 'data', 'injury', 'intervention', 'medical', 'allegation', 'occurred'	Reports relating to the devices transmitter
Topic 2	'loss', 'signal', 'performed', 'one', 'hour', 'within', 'window', 'share', 'investigation', 'log'	Reports relating to the devices signal (loss)
Topic 3	'reading', 'glucose', 'bg', 'meter', 'event', 'customer', 'patient', 'additional', 'information', 'available'	Reports relating to the devices glucose readings

Post Topic Analysis









Discussion

From time series forecasting, we are able to answer the question "Does the frequency of newer reports on Dexcom Continuous Glucose Monitor devices follows the trend of frequency of previous reports?". The prediction indicates an increasing trend for the G6 reports in 2021, and the actual data falls in the 95% confidence intervals, however, it shows a slightly decreasing trend.

From the topic plots, we are able to answer the question “What sort of problems need to be addressed in the future?”. When considering reports involving glucose readings, we see a large increase in the first year the G6 device was released. This increase dropped back to the moving average and followed a stable uptrend. This could provide evidence that early in the G6’s lifetime, there were issues with readings but this issue was solved. When considering reports involving issues with signals, we see a large increase after about 18 months of release. After this increase, the trend seems to remain flat. This explanation for this change in means is unknown and could be explored in future studies. When considering reports involving transmitters, we see a constant increase in trend since G6’s release. Around July 2019, transmitters overtake signals to become the most occurring issue. Since this date, the difference between transmitter issues and other issues has been increasing. Because of this, issues with transmitters need to be addressed in the future.

In the future, more sales data or production/ manufacturing reports could be investigated in order to get more information on the reason for the increase in reports. This could be due to more awareness of this database that is causing more people to report on devices or other factors relating to actual device failures. As mentioned in ‘Results’, when performing LDA we were limited by our computational resources. Our subsample results outperformed both NMF and LSA, so in the future, we could gather more resources and possibly fine-tune our current topics.

References

- Tsujimoto T, Yamamoto-Honda R, Kajio H, Kishimoto M, Noto H, Hachiya R, Kimura A, Kakei M, Noda M. Seasonal variations of severe hypoglycemia in patients with type 1 diabetes mellitus, type 2 diabetes mellitus, and non-diabetes Mellitus: clinical analysis of 578 hypoglycemia cases. Medicine (Baltimore). 2014 Nov; https://pubmed.ncbi.nlm.nih.gov/25415670/
- Gopisetty, Divya. "Dexcom will discontinue G4 and G5 CGM systems United States" https://diatribe.org/dexcom-will-discontinue-g4-and-g5-cgm-systems-united-states
- Crocker, Bryant. "How to Predict a Time Series" https://towardsdatascience.com/significance-of-acf-and-pacf-plots-in-time-series-analysis-2fa11a5d10a8
- Salvi, Jayesh. "Significance of ACF and PACF Plots In Time Series Analysis" https://towardsdatascience.com/significance-of-acf-and-pacf-plots-in-time-series-analysis-2fa11a5d10a8
- https://online.stat.psu.edu/stat510/lesson/4/4.2
- Hyndman, Rob. Athanasopoulos, George. *Forecasting: Principle and Practice (2nd ed)* https://otexts.com/fpp2/stl.html
- Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011. https://radimrehurek.com/gensim/auto_examples/index.html
- Data Source: https://www.fda.gov/medical-devices/mandatory-reporting-requirements-manufacturers-importers-and-device-user-facilities/manufacturer-and-user-facility-device-experience-database-maude