Continuous Glucose Monitoring Analysis Final Report

Sai Rajuladevi, Jonathan Wang, Nehal Chaturvedi

95451 – Making Products Count: Data Science for Product Managers

David Steier

October 13th, 2022

Executive Summary

The purpose of the Data Science for Diabetes Management Group (DSDMG) is to analyze and research the status quo around current Diabetes technologies to fast-track solutions to existing problems. In 2019, diabetes was the direct cause of 1.5 million deaths. Thus, it is imperative that solutions are effective and quickly implemented. In this report we analyze current CGM sentiment, compare Dexcom and FreeStyle Libre, and make recommendations for existing companies to market their strengths and bolster the weaknesses, or for disruptors to enter the market and capitalize on the unmet needs of the consumer.

Current General Sentiment

The goal of our generalized CGM sentiment and product analysis was to roughly understand consumer sentiment before extensive model building and giving recommendations. Our social media analysis found that CGMs are an effective, life-changing, and convenient alternative to the previous status quo (finger-pricking), but can be expensive and not covered by insurance. This causes CGMs to be inaccessible for a substantial portion of the Type-1 population.

Current Players

The main two CGMs dominating the market are the Dexcom G6 and FreeStyle Libre 2. Other related products include, but are not limited to the Omnipod system (integrates with Dexcom), Tandem's TSLIM (insulin pump integrates with Dexcom G6), Ambrosia Blucon sensor that integrates with Libre, and a few others.

Recommendations

As a diabetic patient, if you can afford Dexcom or get it covered by insurance, it is your best option. However, if you cannot, Libre is a wonderful, extremely affordable alternative that is much better than the status quo of finger-sticking multiple times a day.

Introduction

The purpose of The Data Science for Diabetes Management Group (DSDMG) is to analyze and research the status quo around current Diabetes technologies to fast-track solutions to existing problems. Our goal in this report is to identify the current strengths and weaknesses of CGMs (specifically Dexcom and Libre) in the current market to spur innovation.

We analyzed the dataset “Diabetes Continuous Glucose Monitoring – Data Export.xlsx.” provided by Anupam Singh and the team at 113 Industries. This set contains information about continuous glucose monitoring (CGM) which analyzes data from a sensor inserted under the skin for diabetic patients.

Diabetes is a metabolic disease that causes high blood sugar when the body cannot produce enough insulin (Type 1 – 10%) or cannot use the insulin it does make (Type 2 – 90%). In 2019, diabetes was the direct cause of 1.5 million deaths, and with the advent of the COVID-19 pandemic, is a significant contributor to mortality from COVID-19. Diabetes cannot be cured but can be managed. The goal of diabetes management is to reduce A1C levels, which measure the amount of hemoglobin (a protein in red blood cells) that has glucose attached to it. Glucose levels are usually measured with finger-stick blood glucose tests, but an emerging alternative is the use of continuous glucose monitoring (CGM) which analyzes data from a sensor inserted under the skin. Since CGM is always on, glucose levels can be tracked in real-time to see how glucose levels change throughout the day. CGMs are usually used for type 1 diabetes.

For better understanding, we can draw a comparison of CGMs to tire pressure monitoring systems in cars. In the past, tire pressure gauges were used mostly to measure tire pressure. Now, tire pressure sensors are built into car systems’ which scan continuous real-time data. Finger-stick blood glucose tests and automated tire-pressure sensors utilize very similar principles in that real-time data must be closely monitored.

General CGM Analysis

To get a better understanding of the CGM context, we manually read through the posts by sampling ~100 posts in the dataset. Additionally, we selected the column 'SourceType' = 'Original' to filter out comments for each post. We then categorized these manual readings for the two CGM products by filtering on 'SourceType' = 'r/Dexcom' and 'r/FreestyleLibre', which were different subreddits for each product.

We then found the following general trends using random sampling (Refer to Appendix A for some examples of post snippets):

1. Customers expressed a mostly positive experience when compared to traditional finger-pricking solutions. They conveyed that this technology has been life-saving and much more convenient.
2. Customers were struggling with the tech support for CGM, and there was an evident knowledge gap in setting up and initialization for each device.
3. Affordability was the main concern for many customers, as it is commonplace for doctors to not understand the need and thus not recommend CGMs. It is also common for customers' insurance to reject covering CGMs under their existing plans.
4. There were knowledge gaps about Insurance coverage, the geographical coverage of different CGM products, and the precise accuracy of the monitors.
5. There were many threads about general advice and recommendations over CGM experiences.

Following this initial analysis, basic patient expectations were understood. Generally, patients are seeking a way to better manage glucose levels every day, have fewer low blood glucose emergencies, and increase affordability.

The unmet needs of patients were also understood. To summarize, individual doctors may not understand the importance of CGMs which can act as a barrier to entry, and insurance can refuse to cover the costs.

Data Preparation

The first step before analyzing the data was to import and clean the Diabetes Continuous Glucose Monitoring data export. The main steps were to select 'Post ID', 'Sound Bite Text', 'Title', 'Source Type', 'Source Name', 'Sentiment', 'Positive Objects', and 'Negative Objects' as the main data columns, and create an 'entire_text' column appending the 'Sound Bite Text', 'Title', 'Positive Objects', and 'Negative Objects'. This 'entire_text' column would then be used for the creation of the train/test set. The target variable we chose was sentiment. The sentiment targets were mapped to '0' for 'mixed, '1' for 'negative', '2' for 'neutral', and '3' for 'positive'.

Afterward, a second pre-processing step was performed to tokenize the ‘entire_text’ column and filter out stop words and commonly occurring CGM topics to create a ‘lemmatized’ and ‘cleaned entire text’ column. The outputs were then passed to convert the lemmatized list into a training set. TF-IDF vectorization parameters were added, and a trigram approach was used to convert the lemmatized list into phrases. The trigram approach helped with analyzing the context behind the ‘Sound Bite Text’ and the opinions people have regarding diabetes monitoring. All the steps were then encapsulated into functions to make capturing the data simpler. For the overall sentiment analysis, we utilized a 20 document minimum and a 90% maximum frequency as parameters for the TF-IDF.

```
clean_and_lemmatize(df, column, dexcom=False, libre=False):
create_training_set(df,column, min_df, max_df, dexcom=False, libre=False)
```

Finally, three datasets were constructed, one using the entire data frame, and the two using filtering criteria for ‘Dexcom’ and ‘Libre’ in the ‘entire_text’ column of the data frame. This allowed us to split the analysis into 3 stages for modeling: an overall sentiment analysis of opinions from the entire data frame, a Dexcom-specified product sentiment analysis, and a Libre-specified product sentiment analysis.

Exploratory Data Analysis

Data Summary

Overview

Overview | Alerts 124 | Reproduction

Dataset statistics

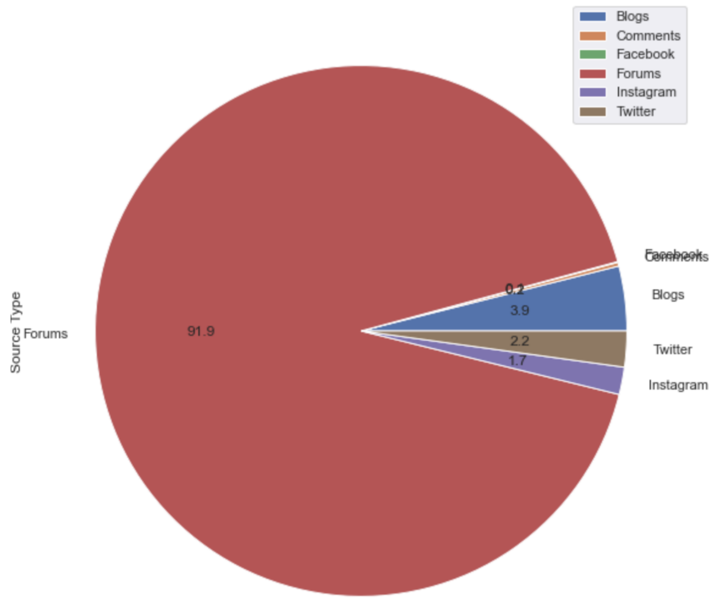
Number of variables	63
Number of observations	37844
Missing cells	1662490
Missing cells (%)	69.7%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	18.2 MiB
Average record size in memory	504.0 B

Variable types

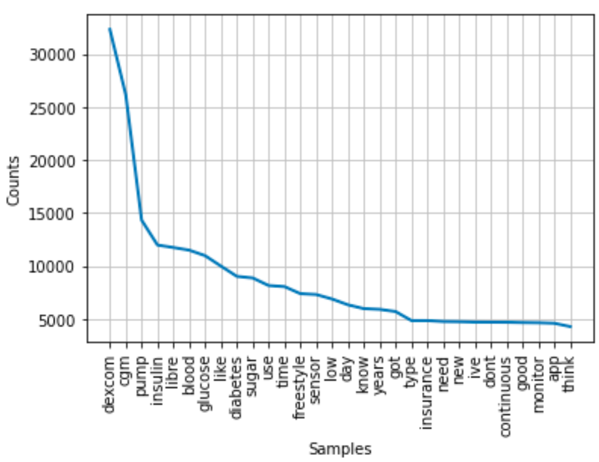
Categorical	30
Unsupported	26
Boolean	1
Numeric	6

	Post ID	Sound Bite Text	Title	Source Type	Sentiment	Positive Objects	Negative Objects	Source Name
0	BRDRDT2-t1_imq98sr	My numbers are great now. Estimated a1c of 7%i...	Have you been denied a second/third pump? Feel...	Forums	Neutrals	number	NaN	r/diabetes_t1
1	BRDRDT2-t1_impbcf4	I tried it for a little while. No side effects...	Metformin	Forums	Positives	NaN	NaN	r/diabetes_t1
2	1565738759353602048	i ran out of characters. youtu.be/RWgl2PDhQiM ...	NaN	Twitter	Positives	dexcom g6, omnipod system	NaN	NaN
3	17944607459251789	MY lunch! Ate at 10:30am \n1 unit NovoLog insu...	NaN	Instagram	Neutrals	NaN	NaN	NaN
4	BRDRDT2-t1_imq8h9m	This is also because like a soak in a hot tub ...	No bath salts, bath oils, soaks?	Forums	Neutrals	NaN	NaN	r/diabetes

The Pandas Profiling Report shows 63 variables, which we then filtered to 8 variables that can filter data out on source (Source Type and Source Name), Sentiments (Sentiment, Positive Objects, Negative Objects), and text (Source Bite Text and Title).

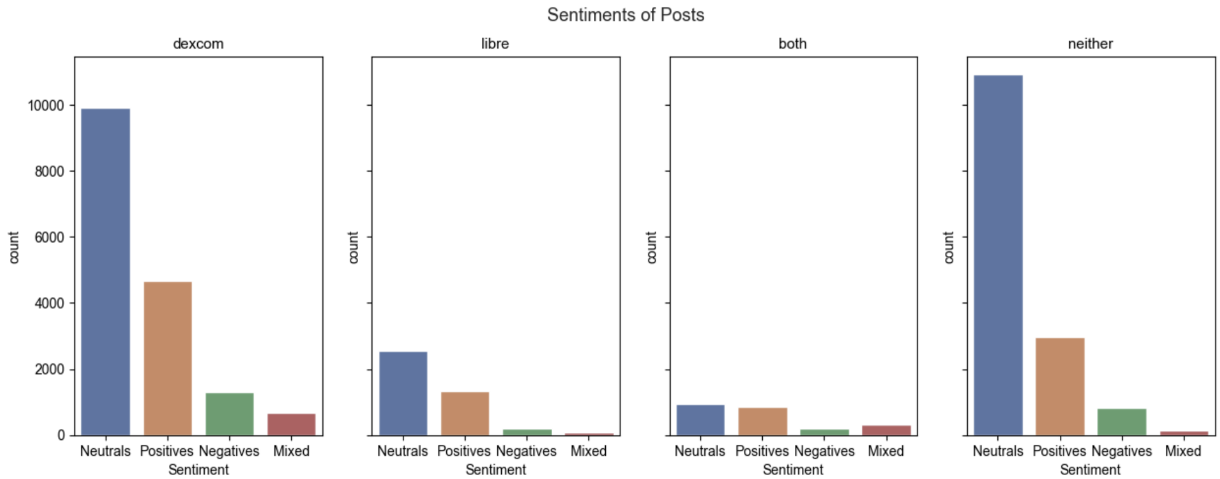


The above pie chart shows the distribution of posts grouped by source type. A vast majority comes from forums (dominated by Reddit). Thus, this overall analysis arguably can only be applied to Reddit users, and cannot be generalized to the overall population. However, due to the convenient structure (posting, commenting, replying) and popularity of the website built for discussions, compared to alternatives such as Instagram or Blogs, users likely come to Reddit to learn and post about Diabetes and CGMs. Thus, we assume this dataset is representative of the general diabetic populace. We recommend future work to investigate whether or not Reddit has a representative sample (or at least the related subreddits).



Above we plotted the most used words (other than common words) across all posts.

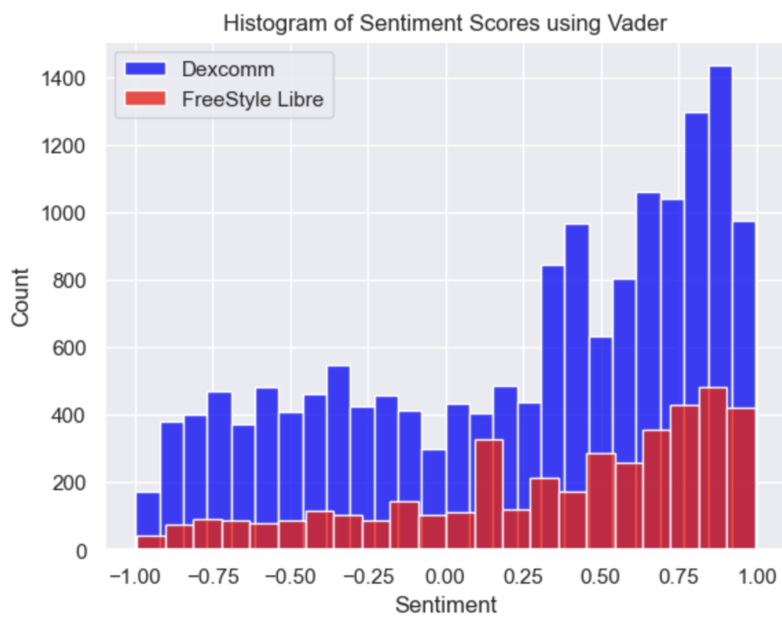
Sentiment Plots



We bucketed data based on the use of tokens ‘Dexcom’ and ‘Libre’. After plotting the data of sentiments (provided with the dataset), we found a large number of ‘neutrals’, seeing this trend, we decided to do a VADER Analysis for each text.

Vader Sentiment Analysis

Vader from the python ‘nltk’ package polarizes a sentence for positive or negative sentiment. We cleaned the data for stop words and other filtering criteria and ran Vader. There were a large number of ‘Neutrals’ present in the underlying data, so we excluded these sentiments (0s) and plotted the data.



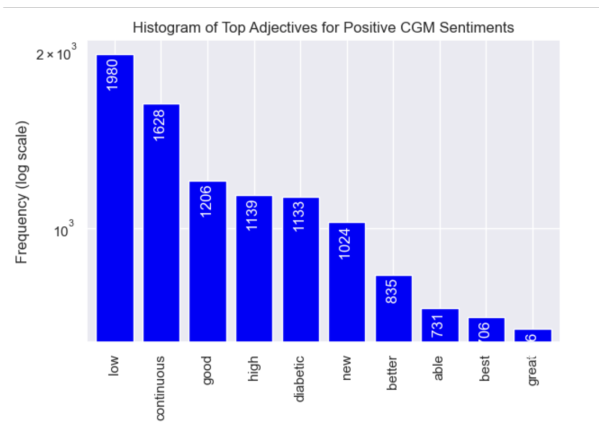
One obvious getaway from this is the paucity of FreeStyle Libre data. There was also equal distribution of positive and negative sentiments for both CGM products according to the histogram. We calculated the average sentiment for each.

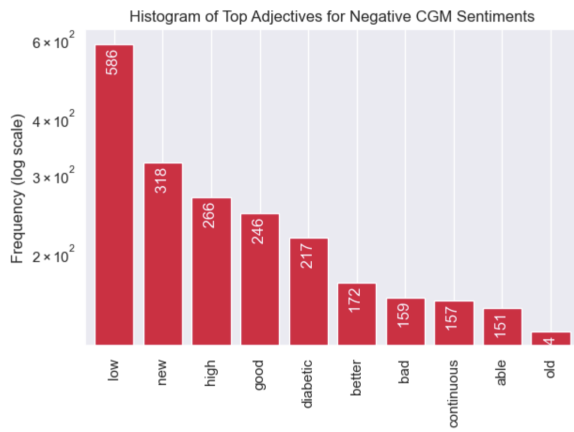
```
In [13]: pd.Series(scores_dexcom).describe()

Out[13]: count    16107.000000
         mean         0.245483
         std          0.570538
         min         -0.996500
         25%         -0.250000
         50%          0.401900
         75%          0.750600
         max          0.997600
         dtype: float64
```

```
In [14]: pd.Series(scores_libre).describe()

Out[14]: count    4185.000000
         mean        0.350072
         std         0.521803
         min        -0.991300
         25%         0.027400
         50%         0.493900
         75%         0.784500
         max         0.995600
         dtype: float64
```

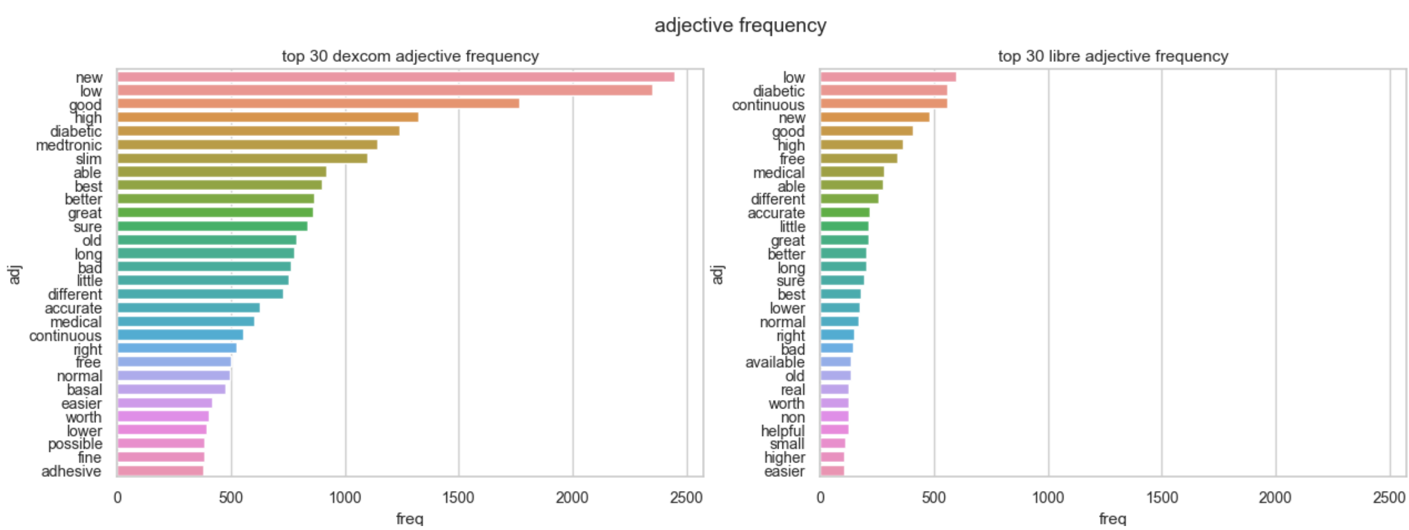
As seen above, Libre has a higher average sentiment, but we have to account for a smaller dataset size (4,185 elements vs. 16,107 elements) to make a sound judgment. To extract further information, we performed POS tagging to get the most common adjectives, nouns, and verbs for each CGM product.

Top Adjectives for Positive/Negative CGM Sentiments.





Top 30 Adjectives for Dexcom/Libre.



Modeling Approach

We utilized different 3 stages for modeling: an overall sentiment analysis of opinions from the entire data frame, a Dexcom-specified product sentiment analysis, and a Libre-specified product sentiment analysis.

For each one of these stages, we utilized 3 machine-learning classification models from the Scikit-learn package distribution: Multinomial Naive Bayes, Random Forest, and XGBoost. This resulted in a total of 9 models. The purpose of utilizing a larger number of classification models was to isolate key findings between each of the 3 datasets. By starting with general takeaways on the topic of CGM, we then focused our efforts on differentiating sentiment between Dexcom and Libre products.

We finally enabled hyperparameter tuning as an optimization technique for each one of these models. Using GridSearchCV, we performed an exhaustive search over specified parameter values for each estimator. These approaches were nested into the following function to make modeling simpler.

```
run_classification_model(model, parameters, features, labels, hyperparam=True)
```

Results

TF-IDF vectorization was used on the cleaned and lemmatized data to create the three different train and test sets (‘df’, ‘df_dexcom’, and ‘df_libre’). We utilized an 80% to 20% train/test split, enabled hyperparameters, and ran an accuracy, F1 score, and classification report on each model.

Entire Dataset Multinomial Naive Bayes

```
Accuracy: 0.6473774606949398
F1 Score: 0.7820223181571322
```

```
top 5 negative words:
 ['time leave medtronic' 'covered pharmacy benefit' 'insurance tried deny'
 'check time day' 'device attached body']
top 5 positive words:
 ['doesn look like' 'time leave medtronic' 'medtronic wondered experience'
 'switch medtronic tandem' 'lot issue signal']
```

Dexcom Dataset Multinomial Naive Bayes

```
Accuracy: 0.6152060359837492
F1 Score: 0.7502736507232448
```

```
top 5 negative words:
 ['weekly vent thread' 'hybrid closed loop' 'daily meme day'
 'hour new sensor' 'build app update']
top 5 positive words:
 ['diagnosed year ago' 'medtronic guardian sensor' 'use tandem slim'
 'isn approved arm' 'highly recommend getting']
```

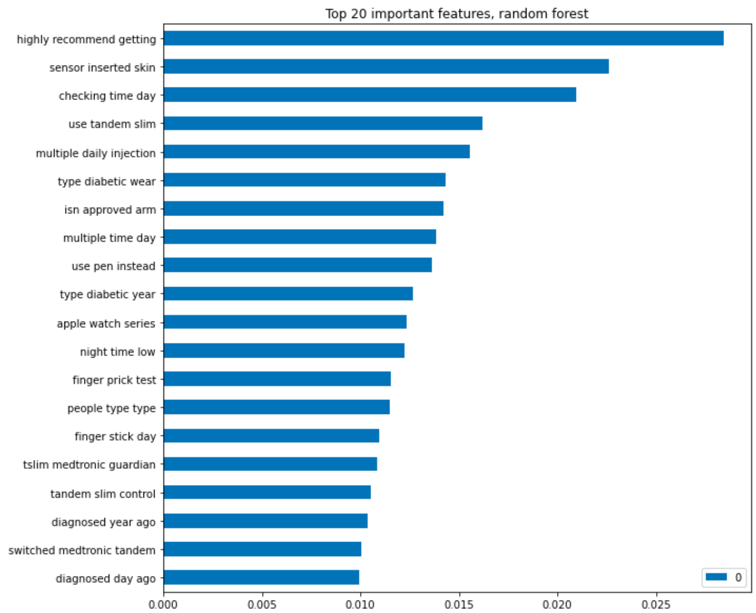
Libre Dataset Multinomial Naive Bayes

```
Accuracy: 0.6193771626297578
F1 Score: 0.7366690948825351
```

```
top 5 negative words:
 ['insurance wont cover' 'closest thing cure' 'diabetic year battling'
 'mark medtronic wondered' 'paying tslim supply']
top 5 positive words:
 ['medical supply company' 'diagnosed year ago' 'device attached body'
 'noticing app having' 'month way high']
```

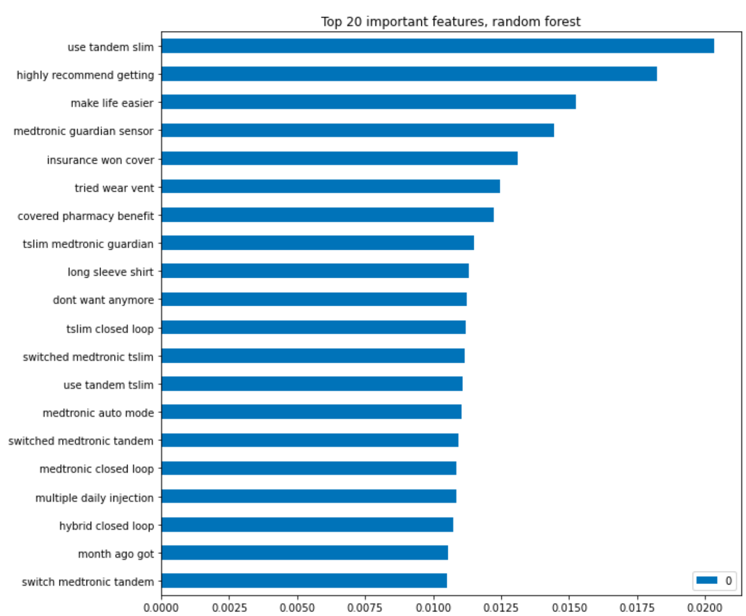
Entire Dataset Random Forest

```
Accuracy: 0.6455278108072401
F1 Score: 0.7707403322939178
```



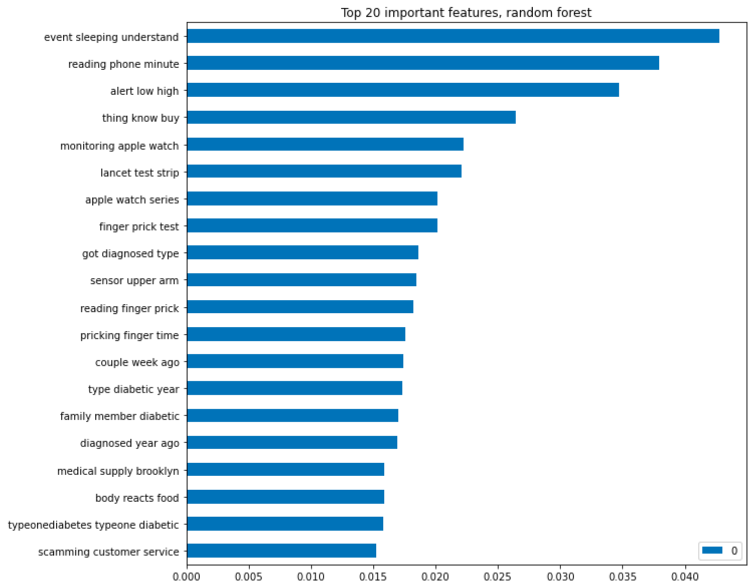
Dexcom Set Random Forest

```
Accuracy: 0.6114335461404526
F1 Score: 0.7328153999273096
```



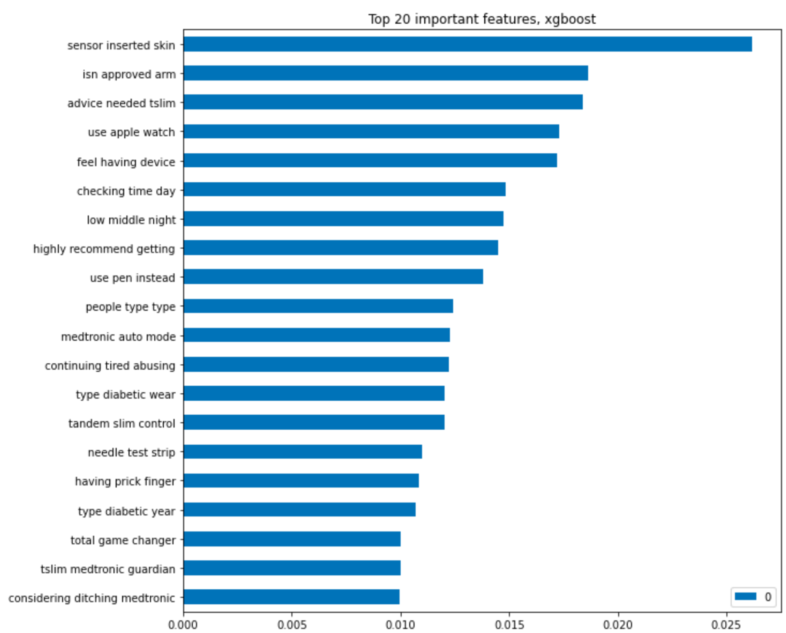
Libre Set Random Forest

```
Accuracy: 0.6101499423298731
F1 Score: 0.7296081312306629
```



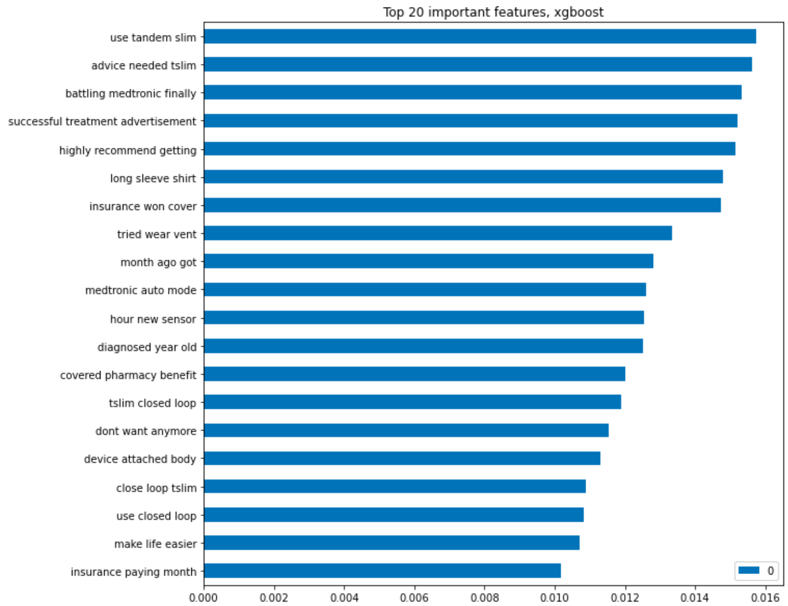
Entire Dataset XGBoost

```
Accuracy: 0.6471132249966971
F1 Score: 0.7766002006998747
```



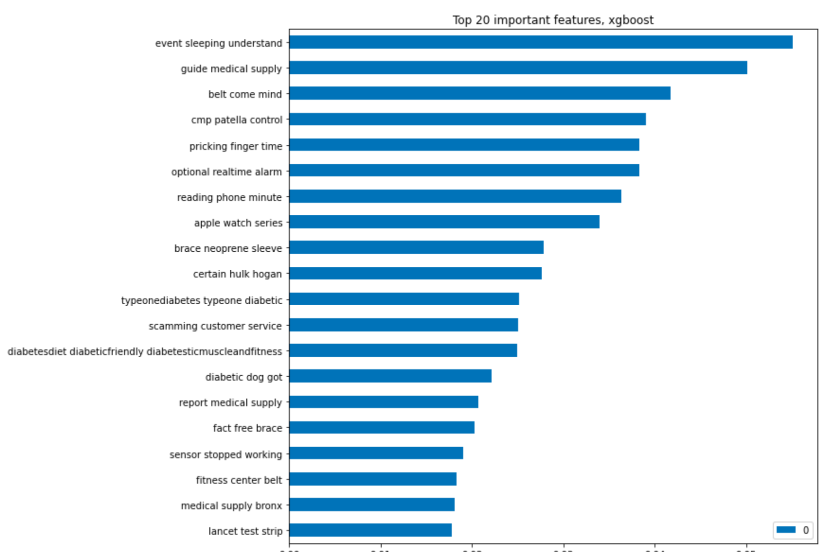
Dexcom Set XGBoost

```
Accuracy: 0.6146256529309344
F1 Score: 0.7431466481568667
```



Libre Set XGBoost

```
Accuracy: 0.6101499423298731
F1 Score: 0.7337886108929081
```



Please see the attached Jupyter notebook for unsuppressed outputs with confusion matrices and precision-recall curves.

Result Analysis

A few key insights can be gleaned from the abbreviated results. First, in terms of performance, regardless of the model, we ended up with similar accuracies of around 0.65 and F1 values of around 0.75. Upon filtering the dataset by company, these values dropped down slightly but otherwise remained fairly static. An important point uncovered from the confusion matrices (see attached Jupyter notebook), is that we can see a clear negative trend associated with the Multinomial Naive Bayes model in predicting the ‘mixed’ and ‘negative’ (0 and 1) target classes. For instance, in the confusion matrices, there are no observations at all for these sentiments, and ‘neutral’ and ‘positive’ (2 and 3) are overcompensated. The Random Forest Classifier and XGBoost models on the other hand were able to capture more of the overall distribution among the target classes, especially for ‘positive’ sentiment.

A common theme between all models is that each strongly over-classifies the ‘neutral’ (2) target, regardless of the dataset. In all the confusion matrices, there are over 4000 observations shown for the true and predicted ‘neutral’ label, compared to less than 500 for all other combined classes. As shown in the exploratory data analysis, we strongly believe this is related directly to the influx of ‘neutral’ sentiment observations in the dataset, which is causing our trained models to overcompensate for the ‘neutral’ target.

Second, in terms of sentiment for the general dataset, we were able to uncover very interesting details about general product opinions in the market. For instance, the most negative phrase uncovered from the Multinomial Naive Bayes model was “time leave Medtronic”. Medtronic, which is a company which is known for its insulin pump technology, may want to analyze why the general population is making the shift toward other products in the diabetes market. Additionally, for the Random Forest and XGBoost classifiers, one of the most important features was “highly recommend getting”, which is a generally positive phrase helping with sentiment prediction. Although this phrase is not yet associated with a specific product, we were satisfied with the model being able to understand that praise should be weighed equally with criticism.

Upon further analysis, we filtered the entire dataset into Dexcom and Libre splits. In terms of sentiment for the Dexcom filtered dataset, the Multinomial Naive Bayes model uncovered the phrase “weekly vent thread " as one of the most negative features in the model. From the Random Forest and XGBoost models, two of the most important words were “highly recommend getting” and “make life easier.” In terms of sentiment for the Libre filtered dataset, the Multinomial Naive Bayes model uncovered the phrase “insurance won't cover" as one of the most negative features in the model. From the Random Forest and XGBoost models, two of the most important phrases were “scamming customer service” and “event sleeping understand.” Based on these results, we can see clearly that there are notable differences between the products that are captured by the machine learning models.

Conclusion

Overall, we see Dexcom products being highly praised as compared to Libre products. While it is arguably the best product, upon an external research analysis we believe that there are still major downsides. First of all, you must get it through insurance. Unlike Libre, which is very easy to acquire, Dexcom is locked behind insurance coverage which can cause major problems for many. Moreover, it is extremely expensive ($700+) which drives many customers away. Lastly, based on our analysis sentiment points to their customer service as lackluster (compared to Libre’s excellent customer service). On the flip side, Libre can be bought “over the counter” and also has options for insurance coverage (but typically gets rejected). Considering Libre can boast about their “1/3rd the cost of other CGM systems,” we find it a reasonable assumption that this price point is a worthwhile cost for the benefits of CGMs. Given all these benefits, however, there are a few aspects of the actual product that fail to meet customer expectations. For instance, Libre does not track well during sleep.

To summarize, while Dexcom has high praise for its effectiveness, and is typically seen in a better light relative to Libre, they do have “Weekly Vent Threads,” where users will vent about the product every week in a dedicated discussion board. Therefore, Dexcom is not without its downsides even on the product quality side. With that being said, we still recommend customers utilize Dexcom products if they can get access to them and aim for Libre products over no CGM tools at all.

Because of certain issues highlighted above, we see a trend of patients discussing shifting to new products like Medtronic. To curb this shift, Dexcom and Libre should invest in cheaper insurance coverages and better customer services respectively.

For better data analysis, we believe the current sentiment column is limited in function; two numerical columns (one positive sentiment, one negative sentiment) with ranges from -5 to 5 will help in better modeling and accuracy. Furthermore, we discovered a substantial class imbalance for Dexcom and Libre. More balanced data will help improve the analysis further.

To better understand the problem, we want to venture out of the constraints of only social media analysis. We believe that survey data will be a great benefit to analysis and that internal analytics data such as customer lifetime value, churn from products, and revenue incrementality will help differentiate between the business motives of Dexcom and Libre and allow for more useful predictive modeling on top of current sentiment analysis techniques.

Appendix

a. …"I had diabetes for 23 years before I finally started using a CGM and pump. Up until that point, I was adamant that I would never use them. Long story short, the benefits vastly outweighed the discomfort and I got used to wearing both and would never go back to MDI [multiple daily doses of insulin] and finger pricks…"
b. "...I always thought they were A Big Deal to insert and maintain. ... The needles end up creating biohazard waste that you have to dispose of carefully. And, perhaps most importantly for someone who types as much as I do, sticking your fingertips with needles hurts. A continuous glucose monitor solves those three problems…"
c. "And I was trying, but it was bare minimum. I'd check, I'd carb count, I'd give insulin, but then my blood sugars would still be so high. I wasn't coming back down like I should've been, but since I didn't have a CGM I couldn't see what was happening…"
d. Must be affordable: "This morning I found out insurance won't cover my order of dexcom sensor and transmitter until deductible is paid. So I need to pay 1,335 dollars for the order. That's more than my paycheck."
e. Some patients are not getting CGMs recommended through their doctors, as they don't understand the need for it. Essentially not getting ease of access.
f. "After jumping through several eight inch hoops lit on fire and spinning around at a speed of 93mph, I've finally gotten my endo to send the prescription I needed for dexcom sensors/transmitter. I'm picking up through CVS pharmacy and the total is $700 which is an immediate NO. I'm supposed to start on the tslim with Control IQ but there is no way in hell I'm coughing up $700 It's just not possible for me."
g. "Sounds like he was not knowledgeable about Diabetes or is just darn lazy! Although my primary is a sweetheart and know some about the D, I did have pressure her to get me

the CGM Libre. First she said no I didn't need it, then when I explained why I wanted it, she gave in. Even within the same practice, doctors had different opinions. It's nuts.”

h. “might have something to do with the way your body is positioned while you're sleeping, tossing and turning has given me lows that make my mom call me in the middle of the night just to make sure I'm ok lol also might just be the cgm itself.. hard to tell with these things sometimes.”