ANALYTICS, FACEBOOK, AND FAMILY HEALTH: AN APPLICATION OF TEXT MINING ON PARENTING IN PEEL

August 2018

Table of Contents

List of Figures and Tables

Executive Summary

At the Region of Peel, the Multichannel Contact Centre (MCC) within the Family Health division uses communication technology to provide information and support to families with children ages 0 to 6 (Region of Peel, 2018). One of these services is the Parenting in Peel Facebook page. Previously, there was no systematic approach to analyzing the content of the Facebook page, and therefore a strategy to increase the effectiveness of the page did not exist.

This report introduces text mining and discusses how it was used to analyze the Parenting in Peel Facebook posts to increase the effectiveness of the page. More specifically, text mining was used to discover which topics have the highest interaction rates and to determine if re-posting affects the number of interactions (likes, comments, and shares) that posts get.

Based on the post-matching analysis, the report confirms that re-posting is ineffective and is associated with a decrease in the number of interactions; therefore, the recommendation is to post original content as often as possible.

Moreover, the results of the topic identification and clustering reveal that both the distribution of topics and the number of interactions per topic are not uniform. Certain topics such as newborns, infant sleep, and immunization have much higher like, comment, or share rates than other topics. This helps in identifying customer demands by understanding what motivates someone to interact with a post.

The report concludes that text analytics provides an effective methodology of analyzing data from the Parenting in Peel Facebook page, and contributes to the strategy of increasing user engagement. The methods used provide a good basis for the Health Analytics division to build upon for future analysis of Parenting in Peel.

1.0 Introduction

Parenting in Peel is a Facebook page run by registered nurses in the Family Health division at the Region of Peel that supports parents of children ages 0 to 6 by posting tips and information, and interacting with page followers.

While it is possible to track the page's 'reach' (number of page views, likes, messages, etc.), no clear strategy of how to increase user engagement exists. Using text mining to develop a methodology to analyze the content of the posts will help understand what topics and types of posts receive the most likes, comments, and shares, thereby improving the effectiveness of the Facebook page.

This report analyzes how text mining techniques such as topic identification, clustering and visualization contribute to the strategy of increasing engagement of the Parenting in Peel Facebook page.

2.0 Analysis

2.1 Background

The goal of the Parenting in Peel text mining project is to develop a methodology to analyze and understand data from the Parenting in Peel Facebook page, and create a strategy to increase the effectiveness of the page (Region of Peel, 2018). Applying text mining techniques such as topic modeling, clustering, and visualization contribute to the strategy by determining why certain topics receive the highest interaction rates, and whether re-posting content affects the number of likes, comments, and shares.

2.2 Post Matching

The purpose of the post matching analysis is to determine if re-posting Facebook content on the Parenting in Peel page affects the number of likes, comments, and shares that posts get when they are re-posted. To do this, posts from two non-overlapping time periods (July 1, 2016 to February 15, 2018 and February 15, 2018 to May 22, 2018) were compared using a matching algorithm. The algorithm identified posts from the second time period that were identical or nearly identical to posts from the first time period.

2.2.1 Text Pre-Processing

Pre-processing is an important technique used in text mining to reduce the size of textual data and transform it so that it is easier to work with. The key text pre-processing steps used in this analysis are the following:

- Convert to lower case
- Remove numbers
- Remove punctuation
- Remove special characters
- Remove stopwords
- Remove extra whitespace
- Reduce each word to its root (stemming)

Since the purpose of the matching analysis is to match the content of the posts, removing unnecessary words, symbols, numbers, and suffixes will ensure that the algorithm is accurate. Also, since the Parenting in Peel posts are written by different people, the writing styles are not always the same. Using text pre-processing helps account for these differences.

2.2.2 Matching Algorithm

To identify re-posts, every post from the first time period was compared to every post from the second time period using a matching rate: the number of matching words between two posts divided by the number of words in the longer of the two posts. If two posts have at least 70% of the same words, the algorithm considers them to be equivalent. The following table shows three examples of matching posts, and the dates on which they were posted.

Old Post	New Post	Old Date	New Date
Yes or No? Some days, being a parent is hard work!	Yes or No? Some days, being a #parent is hard work!	2017-11-28	2018-04-03
Why take a chance? Better safe than sorry!	Why take a chance with #alcohol? Better safe than sorry!	2017-11-14	2018-04-10
Did you know? Peel Public Health offers breastfeeding instructional videos in 7 different languages! Check out this link for videos about latching, skin to skin with baby and other topics	Did you know? Peel Public Health offers breastfeeding instructional videos in 7 different languages! Check out this link for videos about latching, skin to skin with baby and other topics #breastfeeding #skintoskin	2017-05-30	2018-03-07

Table 1: Post Matching Examples

The next section presents the results of the matching algorithm and compares the interactions of the re-posts to the interactions of the original posts to determine if there is an increase, decrease, or no change.

2.2.3 Results and Discussion

Based on the matching algorithm, there were 39 posts from the second time period that were re-posts. In total, the re-posts had 72 interactions, or 1.85 interactions per post. This dropped from 459 interactions for the original posts, or 11.77 interactions per post. This is a decrease of 84%, which indicates that re-posting is ineffective, and that original content should be published as often as possible.

Time Period	Total Interactions	Average Interactions	Total Likes	Avg. Likes	Total Comments	Avg. Comments	Total Shares	Avg. Shares
01-07-2016 to 15-02-2018	459	11.77	386	9.90	33	0.85	40	1.03
15-02-2018 to 22-05-2018	72	1.85	63	1.62	7	0.18	2	0.05

Table 2: Post Matching Comparison Results

2.3 Topic Identification

In addition to matching text to identify re-posts, text mining was used to accurately assign topics to Parenting in Peel posts published over a two-year period (April 1, 2016 to April 30, 2018). Grouping posts into topics makes it possible to calculate interaction rates to determine which topics are the most popular. In addition, creating visualizations such as wordclouds of the posts in each topic gives insight into why certain topics receive higher like, comment, and share rates than others.

2.3.1 Algorithm

Like the post-matching analysis, pre-processing was used on the text of the posts prior to applying the topic identification algorithm. Essentially, for the topic identification process, the words of each post are compared to a list of keywords classified under different topics. Some of the topics include: breastfeeding, newborns, infant sleep, and immunization. Depending on the frequency of keywords in a post, the topic that has the highest keyword count for that post is assigned its overall topic. The diagram below illustrates a simplification of the topic identification process.

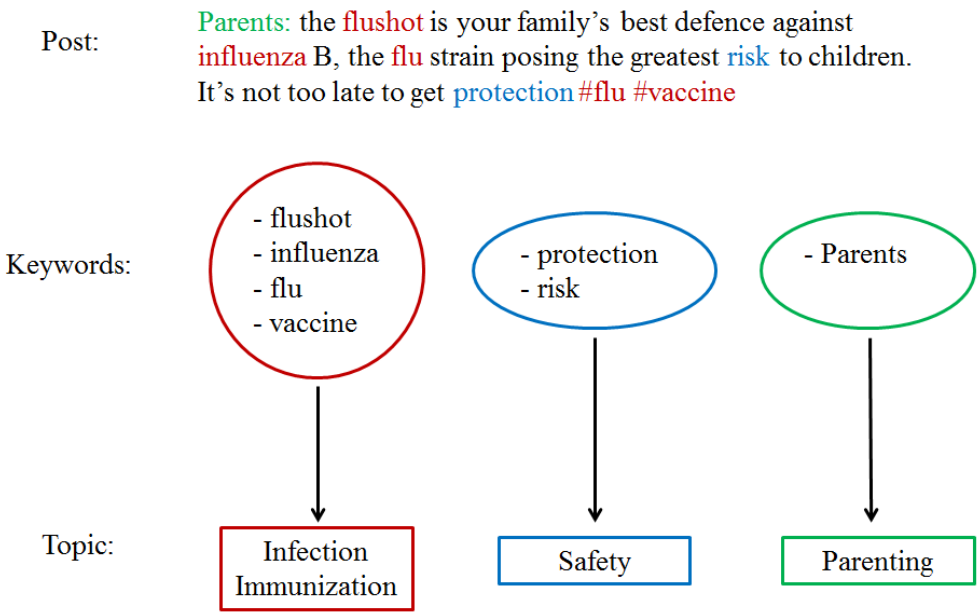

Figure 1: Topic Identification Process

In this case, this post would be classified under the infection immunization topic. The next section presents the results of the topic identification algorithm and discusses the distribution of posts per topic.

2.3.2 Results: Distribution of Topics

From April 1, 2016 to April 30, 2018, the most frequent topics on the Parenting in Peel page (excluding greetings) were safety, nutrition, and breastfeeding. The topics that were the least common were preconception, infant sleep, and families. The results show that the distribution of posts per topic is not uniform; however, this does not mean that the most frequent topics receive the highest interaction rates, as the next section reveals.

Rank	Topic	Total Posts
1	Greetings	205
2	Safety	80
3	Nutrition	76
4	Breastfeeding	72
5	Pregnancy	40
6	Early Learning	37
7	Attachment	31
8	Newborns	29
9	Parenting	28
10	Intro to Solids	26
11	Oral Health	22

Rank	Topic	Total Posts
12	Child Care	22
13	Infection Immunization	17
14	PMD	17
15	Physical Health	17
16	Programs & Services	13
17	Physical Literacy	11
18	Healthy Development	10
19	Social Connectedness	9
20	Families	6
21	Infant Sleep	5
22	Preconception	1

Table 3: Topic Identification Results

2.3.3 Interaction Rates by Topic

With the posts classified by topic, each of the like, comment, and share rates as well as the overall interactions rates can be calculated to determine which topics people are most interested in. This helps to increase the effectiveness of the page by understanding which topics people interact with the most.

2.3.3.1 Results: Overall Interaction Rates

Like the distribution of posts per topic, the number of interactions per topic is not uniform. The topics that had the top five overall interaction rates were: newborns, infant sleep, oral health, early learning, and infection immunization. This is not consistent with the previous results for the number of posts per topic; even though there were only five posts about infant sleep in two years, this topic had the second highest overall interaction rate (19.60 interactions per post).

Examining each of the like, comment, and share rates for each topic will provide even further insight into which topics are the most popular.

Rank	Topic	Interaction Rate
1	Newborns	26.52
2	Infant Sleep	19.60
3	Oral Health	19.00
4	Early Learning	18.86
5	Infection Immunization	17.06
6	Programs & Services	16.77
7	PMD	15.41
8	Parenting	12.79
9	Families	11.83
10	Nutrition	11.59
11	Breastfeeding	10.68
12	Pregnancy	10.15
13	Social Connectedness	8.67
14	Safety	7.45
15	Healthy Development	7.3
16	Attachment	6.19
17	Physical Literacy	6.18
18	Intro to Solids	5.5
19	Physical Health	5.18
20	Greetings	4.7
21	Child Care	4.23

Table 4: Overall Interaction Rates

2.3.3.2 Like, Comment, and Share Rates

In addition to overall interaction rates, the like, comment, and share rates for each topic can be analyzed.

The topics that had the most likes per post were: newborns, early learning, oral health, and programs and services. Liking a post indicates that the user connects with it in some way or is looking for more information on a topic; for example, some Parenting in Peel posts encourage people to 'like' the post for more information.

Rank	Topic	Like Rate
1	Newborns	19.90
2	Early Learning	16.00
3	Oral Health	14.18
4	Programs & Services	14.15

Table 5: Top Like Rates

When someone has a strong opinion on a topic, they will likely comment on the post. The topic that had the highest comment rate was infant sleep with 12.80 comments per post. One theme under the infant sleep topic that many people commented on is bed-sharing, which is a subject that people have conflicting views on.

Rank	Topic	Comment Rate
1	Infant Sleep	12.80
2	Newborns	5.45
3	Pregnancy	4.44

Table 6: Top Comment Rates

Similarly, sharing a post suggests that the person feels a social responsibility to tell others about the topic. For example, the topic that had the most shares per post was immunization. Like bed-sharing, this is a controversial subject that people feel strongly about.

Rank	Topic	Share Rate
1	Infection Immunization	9.29
2	Parenting	3.18
3	PMD	2.59
4	Oral Health	2.32

Table 7: Top Share Rates

2.3.3.3 Visualization: Wordclouds

Another important text mining technique that was utilized in the Parenting in Peel analysis is visualization, particularly wordclouds. Generating wordclouds helps to tell a story, and gives insight into what motivates someone to like, comment, or share a post.

The figure below is a wordcloud of the most frequently used words in posts about infant sleep, the topic with the highest comment rate. For these posts, people might have commented their opinions on bed-sharing, co-sleeping, and safe sleeping habits for babies.

Figure 2: Worcloud: Infant Sleep

Additionally, examining wordclouds for posts about newborns (highest like rate), and immunization (highest share rate) gives further understanding into why people interacted with these topics.

Figure 3: Wordcloud: Newborns

Based on the wordcloud for newborns, people may have liked these posts looking for information to keep their newborn baby healthy, or for tips on how to manage changes in sleep or crying.

Figure 4: Wordcloud: Immunization

Words that stand out in the immunization wordcloud include: infection, spread, children, and risk; people likely shared these posts to warn others of the infections that can spread when children are not vaccinated.

3.0 Conclusion

Text mining provides an effective methodology to analyze posts from the Parenting in Peel Facebook page. The results from tasks such as topic identification and post matching contribute to the understanding of customer demands by helping to understand what topics people who follow the page are most interested in, and if re-posting affects the number of interactions that posts get.

From the topic identification and analysis of interaction rates, the results indicate that both the distribution of topics and number of interactions per topic are not uniform. Furthermore, the post-matching analysis reveals that re-posting negatively affects the number of likes, comments and shares that posts get.

The text mining methods used in this analysis provide a good foundation for the Health Analytics division to expand on. Additional analyses will help increase the effectiveness of the page even further.

References

Region of Peel. (2018). MultiChannel Contact Center – Facebook Analytics.

Region of Peel. (2018). Parenting in Peel. Retrieved from http://www.peelregion.ca/health/family-health/index.htm