Project final report towards partial fulfillment of the course PRED498

Consumer Needs Analysis and Opinion Mining

PREDICT 498, FALL2016

Team Member1: *Ulla McGee, (Research/Team Communication)*

Team Member2: *Damian Bailey (Senior Analyst, Document Prep)*

Team Member3: *Nicolas Bolanos Parra (Senior Analyst, Programmer, Document Prep)*

Team Member4: *Dmitriy Rudenko (Senior Analyst, Document Prep)*

Faculty Advisor: Nethra Sambamoorthi, PhD

Team Members	**Role**	**Responsibility**
Ulla McGee	Team Lead, Communicator, Writer	Ensure the timely execution of the project tasks, coordinate meetings, keep advisor communications and validate results. Contribute project documentation
Damian Bailey	Senior Analyst, Writer, Communicator	Analyze data, validate results. Create visualizations communicating results and findings. Contribute project documentation.
Nicolas Bolanos Parra	Senior Analyst, Writer, R Programmer, Python Programmer	Create models, collect data, analyze data, and validate results. Contribute project documentation.
Dmitriy Rudenko	Senior Analyst, Writer, Communicator	Coordinate project documentation, editing and submission. Contribute project research and references.

Abstract

This paper presents the term project proposed by the Digital Advertising and Marketing Analytics team for the Predict 498 class, during the fall quarter of 2016. The project is based on a startup company called Ample Foods, founded in San Francisco, California. Ample Foods launched in October 2016, selling meal replacement products. We have developed a solution that focuses on finding the most important needs, attributes and opinions about meal replacement products in a systematic and automated way. Based on our solution, which we applied to a dataset of 2,500 reviews on Amazon.com containing at least 7500 opinions about meal replacement products, we have found that the greatest concern of consumers in this market is the taste of the product, the protein content, and how natural its ingredients are, but not necessarily the price, health benefits, weight loss support, or convenience of the products, as we initially thought. In particular, we found out that taste was the most important aspect, especially for customers who did not think much of the product overall. Although there seems to be a number of competitors with an advantage on the taste appeal of their products, there is still plenty of opportunity for Ample to meet this need and succeed in the market. Other needs, such as the need of non-artificial sweeteners and non-synthetic products were found to be detrimental to successful penetration of meal replacement products. The recommendation is to focus on improving product taste and on combatting the perception that the nutritional elements of the product are “synthetic”.

Introduction

Consumer reviews are a way in which individuals express their opinions, frustrations and recommendations about a variety of products or services; from dinners at restaurants, to consumer electronics. Amplified by the Internet, consumer reviews play a dual role of helping other consumers make better-informed purchasing decisions, while allowing marketers and business leaders make product development, pricing, promotion and placing decisions. However, with the ever-expanding body of reviews being generated each day, any attempt of listening to the "voice of the consumer" without any means to analyze and summarize text content can be an overwhelming task.

This paper focuses on bridging the gap between consumer feedback through online reviews and new ideas for product marketing and development. Through a combined approach of rule-based text analytics, traditional association rule mining, and machine learning techniques for clustering, we propose a method to systematically analyze and summarize text content from consumer reviews. This approach builds on the methods proposed in Lee (2007), introducing alternative methods to perform text analytics, association rule mining and clustering.

In this paper, we begin by introducing the context of our project and the company for which we develop our solution. We first indicate the issues faced by the company and how our solution can help overcome those issues; then, we provide an overview of the data sources and data collection methods we used, followed by the methodology, which details the data processing algorithms, model implementation and model evaluation. In the final sections we discuss our findings, challenges, opportunities for future research and provide additional recommendations.

Background

This project is launched by the Digital Advertising and Marketing Analytics team in the Predict 498 class, fall 2016. Our project is on behalf of Ample – a new startup located in San Francisco. The vision for Ample is to provide a healthy and balanced meal in a bottle that anyone on the go can consume and feel good about. Ample was inspired by athletes who felt that most meal powders contained ingredients that could not support a comprehensive nutrition while optimizing the body's biochemical processes (Ample, 2016). Ample will launch in January 2017, with two products: Ample 400 calories and Ample 600 calories.

Ample was the most funded food project on crowd sourcing site Indiegogo, but in order to tailor its marketing strategy, meet the needs of the consumer, and eventually penetrate the meal replacement market, it needs a mechanism to systematically listen to the needs of consumers. By analyzing what features appeal and don't appeal to customers of its competitor products, Ample will be able to hone in on popular features in its marketing message. Our team will be mining the online user-generated content from social media and ecommerce sites in order to make recommendations as to which marketing and product development strategies the company should pursue.

Problem Statement

Problem Description

One of the greatest challenges faced by Ample as a startup company is the challenge of market penetration into the meal replacement products market. Although market penetration is a broad concept, and it can be interpreted in a variety of ways. The Cambridge English dictionary defines market penetration as a process in which a product or brand becomes bought, used, or known by more and more people. In an effort to maintain the objectivity of our analysis, we have identified our problem statement using an issue tree (Figure 1), where we decompose our problem statement into its main constituents. From the issue tree we extract a number of structural hypotheses that both describe the problem and point to possible solutions:

- There is a multiplicity of consumer segments emerging with different interest that converge on meal replacement products, causing an increase in demand and a corresponding increase in competition, and making it difficult for startup companies like Ample to penetrate the market.
- There are a number of concerns from the customer population related to the bad taste of meal replacement products which are culprits for the decrease in market demand for meal replacement products, and which make it difficult for startup companies like Ample to penetrate the market.
- There are a number of concerns from the customer population related to the health implications of meal replacement products, questioning the quality and quantity of their ingredients, the belief that processed foods are unhealthy and the lack of evidence to support their health benefits, contributing to a bad reputation in the customer population and the overall decrease in market demand for meal replacement products, making it difficult for startup companies like Ample to penetrate the market.
- There are a number of concerns from the customer population related to the high price of meal replacement products, which are to blame, in part, for the decrease in market demand for meal replacement products, making it difficult for startup companies like Ample to penetrate the market.

- There are a number of concerns from the customer population related to the inability of meal replacement products to satisfy hunger, which are responsible, in part, for the decrease in market demand for meal replacement products, making it difficult for startup companies like Ample to penetrate the market.

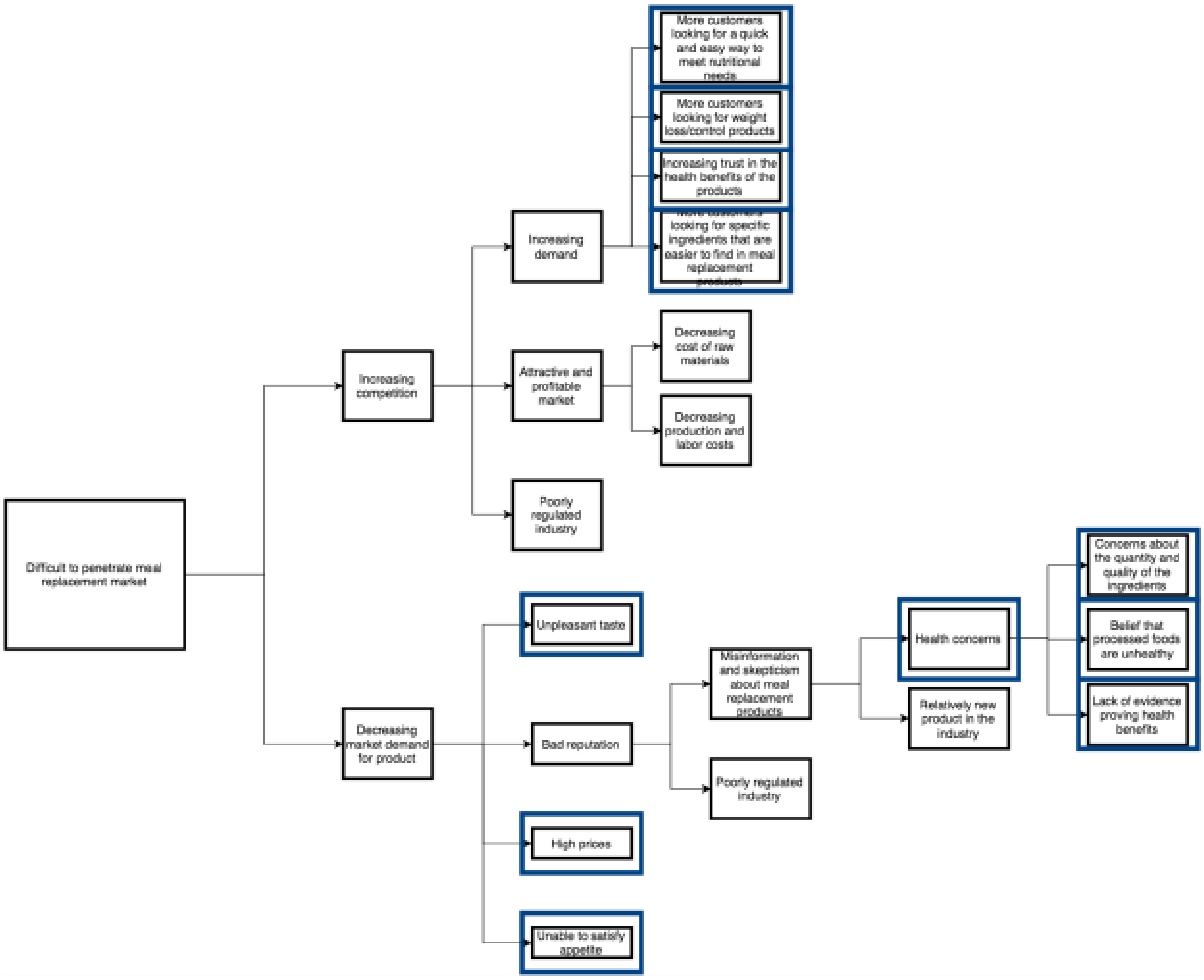

Figure 1 - Issue tree

Project Objective

The objective of our project is to help Ample achieve successful market penetration by completing the following goals:

- Develop a method of systematically collecting, summarizing and analyzing the needs expressed by customers through social media feeds and online reviews.
- Find opportunities for Ample to pursue in their marketing and product development strategy based on consumer's unmet needs.
- Identify the main product aspects driving consumer demand and the consumer opinions associated with those aspects.

Project Justification

The project we present here will not only help Ample gain important insights about the needs of the consumers, it can also be completed and implemented without incurring staggering costs. Other than the development time and expertise required to develop it, the only additional cost originates from the licensing of Tableau, the software we used to develop the dashboard. The data used in the project is publicly available and easy to find, and the technologies used for data pre-processing and modeling are open source, including the applications used for text analytics, rule mining and clustering.

Considering that the project provides a solution to market penetration, one of the most critical problems faced by this company as a startup, while keeping project costs to a minimum, there is enough reason to believe this project is justified.

Review of Literature

Sentiment analysis and opinion mining is one of the most active areas of research in natural language processing in which people's opinions, sentiments, evaluations, attitudes and emotions from written language are analyzed and applied to a plethora of domains, ranging from computer science to management sciences and social sciences (Liu, 2012). Opinions are central to almost every human activity and a vehicle through which researchers (and sponsoring organizations) can understand or enrich their view of business and society.

Although there is a large and growing body of research on the subject, some have voiced concerns about unidimensional measures of sentiment (Miller, 2015). Traditionally, sentiment analysis focuses on calculating a single score based on the assignment of numbers to attributes per rules built from a reference corpus of positive or negative words. However, recent efforts in sentiment prediction point to multidimensional measures. Applications such as SentiStrength attempt to process positive and negative sentiment of formal and informal text in parallel, producing two scores instead of one. Research from psychology has revealed that we process positive and negative sentiment in parallel – hence mixed emotions (Thelwall, Buckley, & Paltoglou, 2012). Other approaches rely on POS (Part of Speech) tagging to associate sentiment terms with a specific subject, thus eliminating irrelevant sentiment about uninteresting subjects (Nasakwa & Yi, 2003). We have adopted a mix of these ideas, using POS tagging to associate opinion terms with specific product aspects while calculating sentiment scores (produced by SentiStrength) to gain insights about the context in which the opinions were expressed, not necessarily to link the sentiment calculation to a specific aspect.

Leveraging a two-stage Apriori rule association mining, Lee decomposed text reviews into association rules from product aspects and the associated opinions (akin to transaction-based market basket analysis, treating precedent basket items as product aspects and posterior items as opinions), and re-applied rule association mining after clustering the aspect-opinion pairs, seeking associations between these pairs. Our analysis adopts a similar approach, although instead of seeking associations between aspect-opinion pairs, we seek associations between clusters of aspects and clusters of opinions, thus finding aspects with similar needs.

In traditional rule mining, the predictive quality of the association rules is measured using a combination of confidence, support and lift. In Borgelt's implementation of the Apriori association rule mining program (which we have used for our analysis), the confidence of an association rule is the proportion of all associated item sets where the association is correct, relative to the total number of cases in which the association rule might be applicable. The support is the proportion of all associated item sets where the rule is applicable, whether it is correct or not. Finally, the lift of an association rule is the ratio of the confidence of the association rule to the prior confidence of an association rule (Borgelt, 2016). In other words, it compares the confidence of making predictions before and after applying the association rule. A lift greater than 1 indicates stronger associations.

Much of the work relating aspect-opinion pair extraction is documented in Liu (2012), where aspect-opinion observations of common opinionated sentence structure are used to extract aspect and/or opinions from text. These observations were used as building block for our analysis, although we have enriched them with additional rules to fit our context and to accommodate for other sentence structures we also deemed important.

Methodology

Once the data has been extracted and put into a semi-structured format we have followed similar methodologies as described in Liu (2012) and Lee (2007). Our analysis is divided into four stages: (1) Generation of product aspects to be investigated (2) aspect, opinion extraction (3) sentiment analysis, and (4) aspect-opinion association rules clustering.

Generation of Product Features

This stage is necessary to ensure the proposed opinion mining project is focused on analyzing aspects and opinions that are being actively discussed in social media feeds and reviews. To generate a list of relevant product features we start with a list of aspects related to meal replacement products provided by the client. Terms such as “price”, “taste”, “convenience”, “ingredients” could be included in the initial list of product features.

For the next step we use the set of rules described in Liu (2012) to uncover additional aspects from the corpus of reviews. Given an aspect and a review containing an instance of that aspect, if the sentence follows the sentence composition pattern stated in the rules, we are able to infer additional aspects that may also be related to the product under study (meal replacements, in this case). To account for plurals, inflections and other word variations we first stem the initial list of aspects and use these stems to identify sentences where new aspects can be generated. The rules we used (shown in Table 1) are based on heuristics and observations of common ways to mention aspects about a product within a review. Since these rules are based on the syntactic relationships of each word in a sentence, it will be necessary to tokenize the reviews by sentence, obtain the Part-of-Speech (POS) for each word, and find the syntactic relationships between the words. For tokenizing and tagging the POS of words, we use the python implementation of the Stanford CoreNLP v3 package (Stanford CoreNLP). Once these data are available, it will be possible to apply these rules to extract additional aspect words. The details of how these rules work are shown in Table 1.

	Observations	Output	Example
Rule 1	$A_{i(j)}$->$A_{i(j)}$-Dep->$A_{j(i)}$ subject to:	Aspect	You can find this product at a fair “price” and <u>value</u>

	$A_{j(i)} \in \{A\}$, $A_{i(j)}$-Dep $\in \{CONJ\}$, POS $A_{i(j)} \in \{NN\}$ Where A = aspect and NN = noun word. In general $A_{i(j)}$-Dep indicates a dependency word of A_i on A_j, in this case, it is any word within the set of conjunctions (denoted by {CONJ})		Input: price Output: value The shake has a good "taste" and texture Input: taste Output: texture
Rule 2	A_i-> A_i-Dep->H<-A_j-Dep<-A_j Subject to: $A_i \in \{A\}$, A_i-Dep=A_j-Dep OR (A_i-Dep = subject AND A_j-Dep = object), POS(A_j) $\in \{NN\}$ Where H = arbitrary word(s) in the sentence such that both A_i and A_j are syntactically related to.	Aspect	"Soylent" has a lot of sugar Input: Soylent Output: sugar

Table 1 – Aspect Extraction Rules

Details of how these rules are implemented are shown in Figure 2 (TextRazor, 2016). After tokenizing the reviews in the dataset, each sentence is modeled as a dependency tree where we look for the patterns defined by the rules in Table 1. Assuming the word "price" is the aspect input, we can infer the word "value" using Rule 1, which states that the input word is related to the output through a conjunction dependency (labeled *conj* in the dependency tree).

This process continues until either we are satisfied with the list of aspects to be explored, or no more aspects can be extracted from the dataset. However, it should be noted that extracting too many aspects may result in too many irrelevant aspects and will therefore add unnecessary noise. Also, for new datasets, it will be good practice to generate a separate list using this methodology to guarantee more relevant and accurate results in the next stages of the analysis.

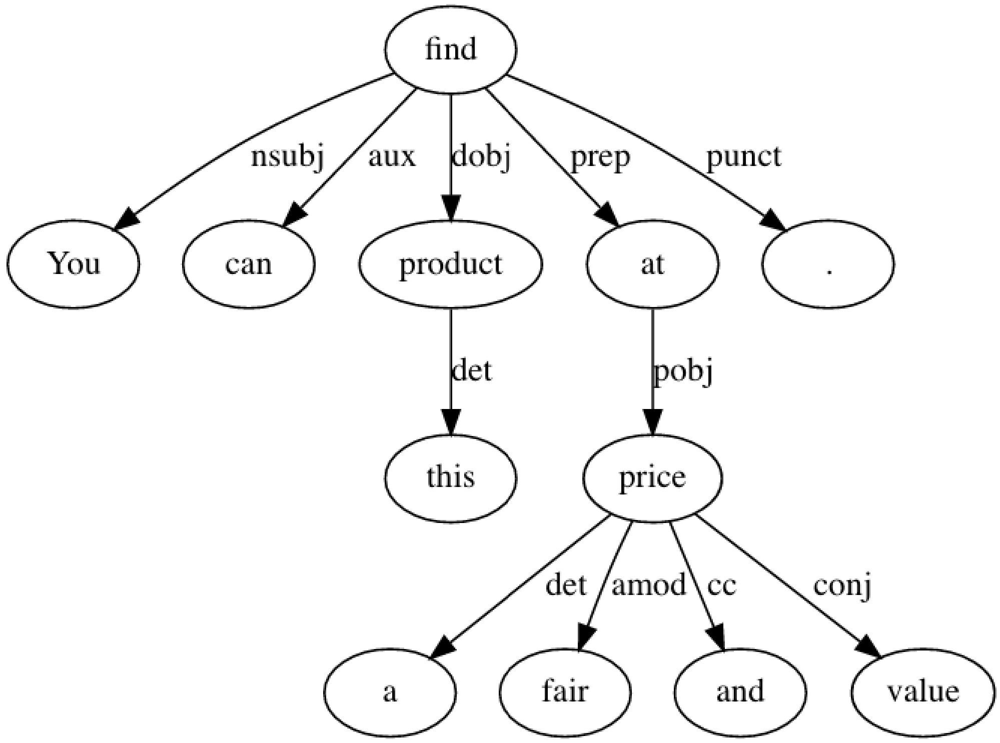

Figure 2 – Sentence Dependency Tree 1

Aspect, Opinion Extraction

Having collected as many relevant aspects about meal replacement products we are ready to extract opinions from the text data. There are many ways to express opinions and the focus of existing work is primarily on adjective, adverb, and noun expressions (Li, Mukherjee, Si, Liu, 2015). Verb expressions are less documented but not less important. Likewise, our project will focus on adjectives, adverbs and nouns to extract opinions (e.g. “The taste is good”) but some verbal expressions will also be considered (e.g. “It makes me feel good”).

Similar rules to those shown in Table 1 are provided in Liu (2012) to extract opinions from aspect words (Table 2). However, these only cover cases in which opinions are expressed as adjectives and aspects as nouns. To enhance our ability to identify aspects, we created additional rules based on sentence structure patterns of other syntactic cases. These are rules 5 through 7 in Table 2. These rules cover cases in which aspects are extracted from verbs (e.g. “It tastes good”, “taste” being the aspect) and cases in which opinions are negated (e.g. “The taste is not bad”, or “It doesn’t taste good”). Although we cannot guaranty these opinions are in fact, opinions about a given aspect, the syntactic relationship is a strong indicator to allow us to make such assumptions.

	Observations	Examples
Rule 3	O->O-Dep->A Subject to: O∈{O}, O-Dep∈{MR}, POS(A)∈{NN} Where O = opinion and MR are specific syntactic relationships (modifiers, direct objects)	The product has good nutritional “value” Input: value Output: nutritional
Rule 4	O->O-Dep->H<-A-Dep<-A Subject to: O∈{O}, O/A-Dep∈{MR} POS(A)∈{NN} Where MR includes amod, nmod, nsubj, dobj, iobj as defined in Marneffe and Manning (2008)	“Soylent” is the best meal replacement shake Input: Soylent Output: best H: shake A: Soylent O: best
Rule 5	VB->XCOMP-Dep->O Subject to: VB∈{VB}, XCOMP-Dep∈{xcomp}, POS(O) ∈{JJ} Where VB is a verb (including all verb variations), xcomp is an open clausal complement as defined in Marneffe et al. (2008) and A is an aspect word consisting of a noun (including all noun variations)	The product makes me “feel” good Input: feel Output: good The product is easy to find at Walmart. Input: find Output: easy

Rule 6	NEG-Dep->H->COMP-Dep->O<-A Subject to: NEG∈{neg}, COMP-Dep∈{ccomp, xcomp, acomp, advcl, dobj}, POS(A)∈{NN} Where NEG represents negative modifiers (ie, not), COMP represents the relationships ccomp, xcomp, acomp, advcl and dobj as defined in Marneffe et al. (2008), and H represents a verb or verb phrase (where verbs are related to each other through COMP relationships).	The shake does not have good nutritional "value" Input: value Output: not nutritional Neg-Dep: not VB: have COMP-Dep: have->dojb<-value O:nutritional A: value The shake doesn't have good "taste" Input: taste Output: not good NEG-Dep: have -> neg <- n't VB: have COMP-Dep: have->dobj<-taste O: good A: taste
Rule 7	NEG-Dep->H->COMP-Dep->VB->O Subject to: NEG∈{neg}, COMP-Dep∈{ccomp, xcomp, acomp, advcl, dobj} Where NEG represents negative modifiers (ie, not), COMP represents the relationships ccomp, xcomp, acomp, advcl and dobj as defined in Marneffe et al. (2008), and H represents a verb or verb phrase (where verbs are related to each other through COMP relationships).	The shake does not make me feel good Input: feel Output: not good Neg-Dep: not H: make->ccomp<-feel COMP-Dep: make me feel->xcomp<-good O:good A: feel

Rule 8	Same as Rule 4 with a negative modifier on H: O->O-Dep-> (NEG-Dep->H)<-A-Dep<-A	"Soylent" is not the best meal replacement shake Input: Soylent Output: not best O: best O-Dep: shake->amod->best NEG-Dep: not H: shake A: Soylent A-Dep: shake->nsubj->Soylent
Rule 9	$O_{i(j)}$->$O_{i(j)}$-Dep->$O_{j(i)}$ Subject to: $O_{i(j)} \in \{O\}$, $O_{i(j)}$-Dep$\in \{CONJ\}$, POS($O_{i(j)}$)$\in \{JJ\}$	The ingredients are organic and nutritious Input: nutritious Ouptut: organic $O_{i(j)}$-Dep: organic->and->nutritious

Table 2 – Aspect-Opinion Extraction Rules

Details of how these rules are applied are shown in Figure 3 (TextRazor, 2016). The decomposition of the sentence "The shake has good nutritional value but it does not make me feel good", provides the opportunity to extract opinions from the words "feel" and "value". We can apply Rule 3 to infer that the word "nutritional" is an opinion about the aspect "value", because it is an adjective and because it modifies (*amod* label in the tree) the word "value". Similarly, we can apply Rule 7 to handle the negative opinion statement "but it does not make me feel good", by noting that the negative dependency (*neg* label in the tree) is acting on the word "make" (notated as H on the rule), and from there searching down the tree for opinionated statements. Since the parent element is negated, we assume that the opinion words, directly or indirectly dependent on the parent, are negated as well. The rule calls for a complement dependency (*ccomp* in the tree) on a verb, which is found in the word "feel", which in turn is related to an opinion, in this case the word “good”, through another complement dependency (*acomp* in the tree). To account for the negation of the parent element, we prepend the word “not” to the opinion, resulting the output “not good”.

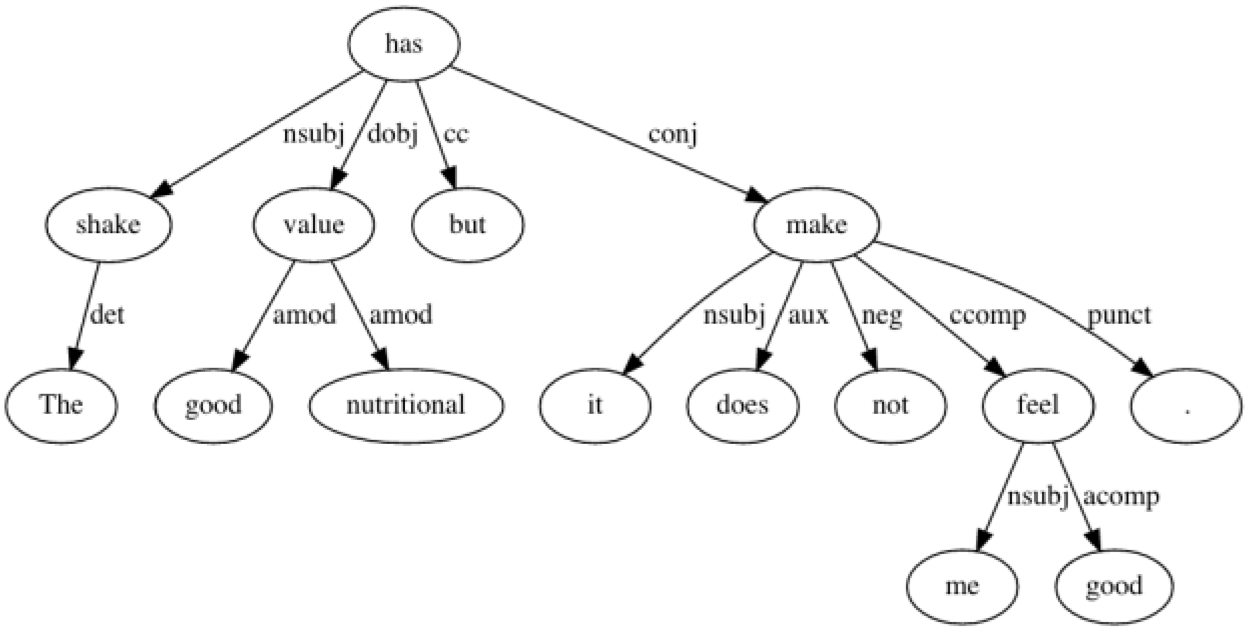

Figure 3 – Sentence Dependency Tree 2

It is worth noting that this methodology suits the extraction of multiple opinions from a single sentence. By applying all the rules in Table 2 separately, we can extract opinions from a rich variety of sentence structures, although occasionally we may find overlapping or even contradicting opinions about an aspect, depending on the sentence. Overlapping might occur if the same opinion is expressed more than once in the same sentence and contradiction generally occurs because some rules do not account for negative dependencies while others do. In any case, overlapping aspect-opinion pairs are counted once and contradicting pairs are always resolved by retaining the negated opinion.

Outside of the scope of our analysis are sentences that express opinions in more abstract ways such as in similes (e.g. “It tastes like cardboard” or “It is as smooth as silk”) or idioms (e.g. “Great bang for your buck”).

In preparation of the aspect-opinion association rules and clustering stage of this analysis it will be necessary to normalize both the opinion and aspect words. This is because having different variants of each word will make it difficult to find clusters. By lemmatizing the opinion and aspect pairings we can make sure that extracted sets like “feeling-greater” and “feels-greatest” are all interpret as “feel-great”. Although less common, yet equally important, some aspect words interpreted as verbs (and used in rules 5 and 7) first required to be transformed into their noun counterparts, to maintain semantic consistency with other noun-based aspects. For example, in the sentence “The product is cheap to buy online” the word “buy” is interpreted as an aspect and transformed into its noun form “purchase”.

Other approaches such as finding synonyms, homonyms, hypernyms and hyponyms could have been applied to further normalize the data but were not implemented in this analysis.

Sentiment Analysis

Understanding the overall sentiment of the context in which opinions were expressed can also provide insights regarding the market’s needs. For example, from the sentence “The taste is great but it does not make up for the terrible side effects.” we might extract the aspect-opinion pair “taste-great” not knowing the context in which the need was mentioned. Thus, understanding the sentiment at the sentence level can provide useful context - perhaps more useful than star ratings, or number of likes, which measure overall sentiment at the review level.

SentiStrength is a Java-based sentiment analysis application designed to estimate positive and negative sentiment in short texts for formal and informal language (Thelwall et al., 2012). It provides scores for positive and negative sentiments ranging from -5 (negative) to 5 (positive). The absolute value of these scores will be interpreted as the overall sentiment score. Table 3 shows examples of SentiStrength in action. It is worth noting that emojis and exclamation points are taken into account when calculating the sentiment score.

Sentence	SentiStrength score	Overall sentiment score
Very good and satisfying.	{4,-1}	3
Very good :) and satisfying.	{5,-1}	4
Very good and satisfying!	{5,-1}	4
The taste is horrible	{1,-4}	-3
The taste is great but it does not make up for the terrible side effects.	{3,-4}	-1

Table 3 – SentiStrength in Action

Aspect-Opinion Association Rules and Clustering

Once the aspects and opinions have been extracted from the reviews, and the sentiment scores have been calculated, the next step is to find aspect-opinion association rules and aspect and opinion clusters.

Our analysis assumes that product needs can be represented as an aspect-opinion pair where the aspect is a feature of the product that is directly related to an opinion. When taken in aggregate, it becomes difficult to interpret these associations because a specific aspect can relate to multiple opinions and vice-versa. The goal of this stage is to find the most common association rules that relate aspects to opinions. We use the Borgelt implementation of the

Apriori association rule mining program, which is implemented in the Arules package in R to obtain the aspect-opinion association rules. To dampen the noise inherent in the data (caused by irrelevant aspect-opinion pairings, for example) we introduce a layer of abstraction by clustering aspect and opinions and then applying the Apriori algorithm a second time where the items are not individual aspects and opinions, but the clusters themselves.

Model Description

We can visualize the aspect-opinion pairs as a set of vertices, representing aspects or opinions, and edges, representing relationships as shown in the following diagram:

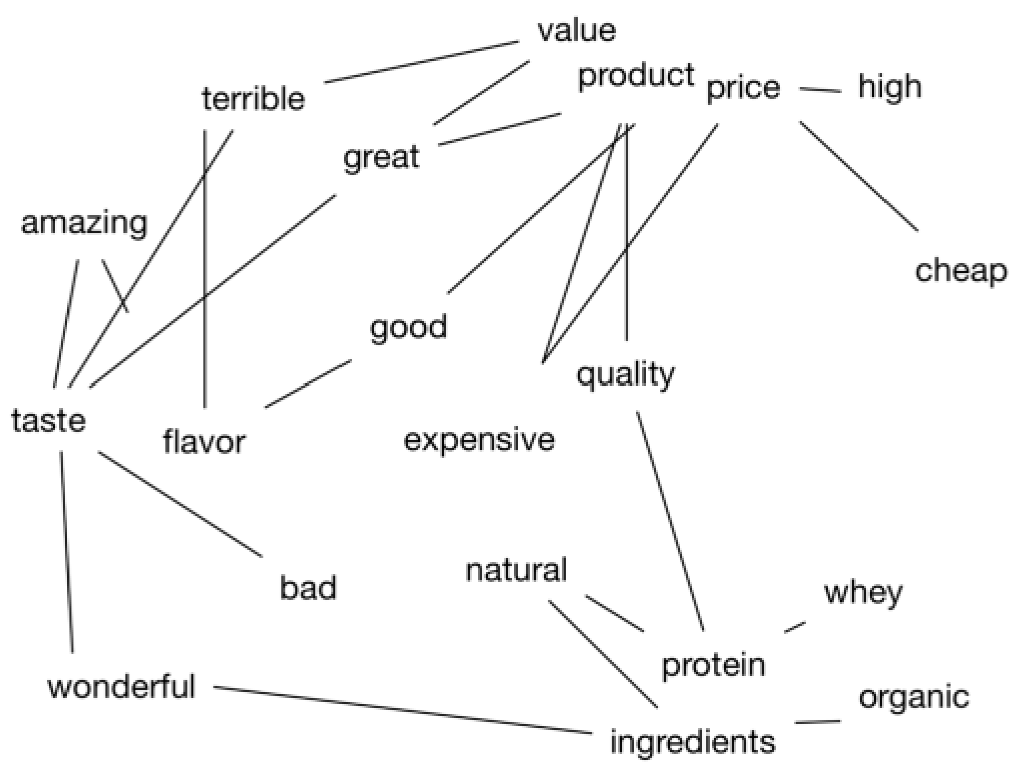

Figure 4 – Aspect-Opinion pairs diagram

Using the dissimilarity function of the Arules package it is possible to find a similarity matrix showing how close each of these relationships are to each other. The distances between relationships, which are also called item sets (i.e., taste ⇔ good) and which describe needs, is calculated using the Jaccard definition of distance. This metric is the measurement of similarity between item sets and is defined as the size of the intersection between every pair of item sets divided by their total size (Batagelj, Ferligoj, & Žiberna, 2006). Once the Jaccard distances are calculated, they are used as input to perform hierarchical clustering, resulting in the clustering of item sets.

However, we are also interested in the clustering of the items themselves (both aspects and opinions) because we want to run the Apriori algorithm using these clusters as inputs, not the words themselves. We want to find the support, confidence and lift of association rules such as {cluster_number_feature=1⇔ cluster_number_feature=2} instead of just {feature=taste⇔opinion=good} for the reasons described above. Visually, with the clustering of item sets we have accomplished this:

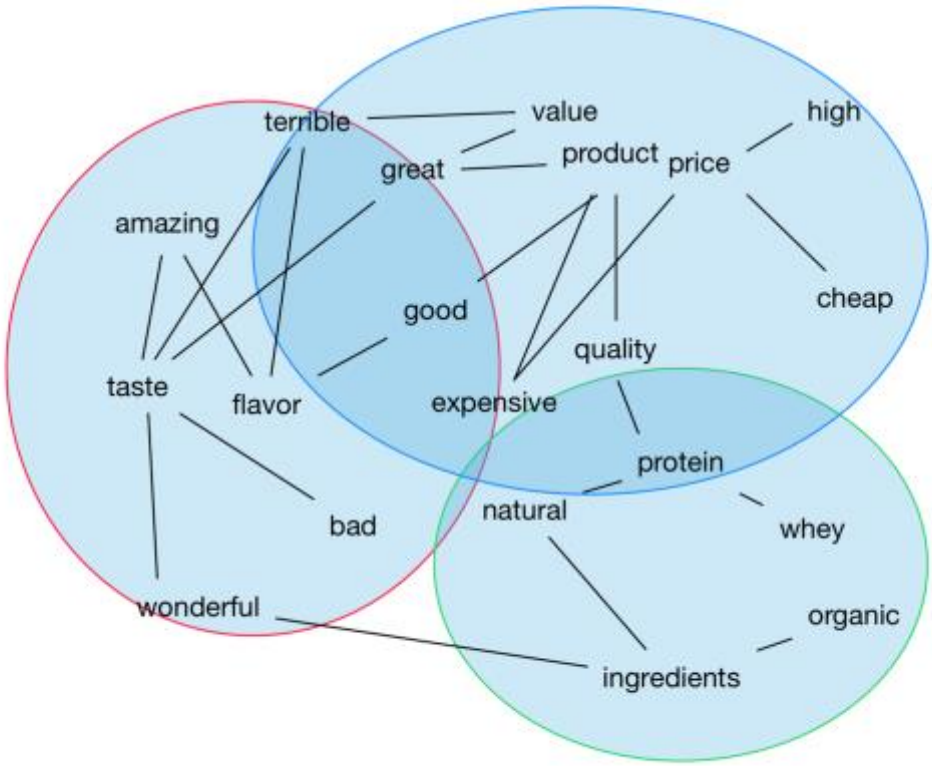

Figure 5 – Item Set Clusters Diagram

For any specific aspect, it is possible to find out which cluster (or clusters) the item is associated with. It is not always possible to have clean clusters that do not overlap, and some aspects or opinions may belong to multiple clusters. For example, the aspect "protein" in the diagram belongs to the green cluster and the blue cluster. So, the question is: how can we separate out the aspects/opinions into their own clusters based on this item set clustering? One way is to assign each item to the cluster where the total support of the item sets containing it is the highest. For instance, if the total support of the item sets in the green cluster containing the word "protein" ({protein, natural} and {protein, whey}) is higher than the total support of the item sets in the blue cluster containing the word "protein" ({protein, quality}), then we would assign the item "protein" to the green cluster as shown in Figure 6. The diagram in Figure 6 also shows the result of the clustering following a similar procedure for each of the aspects:

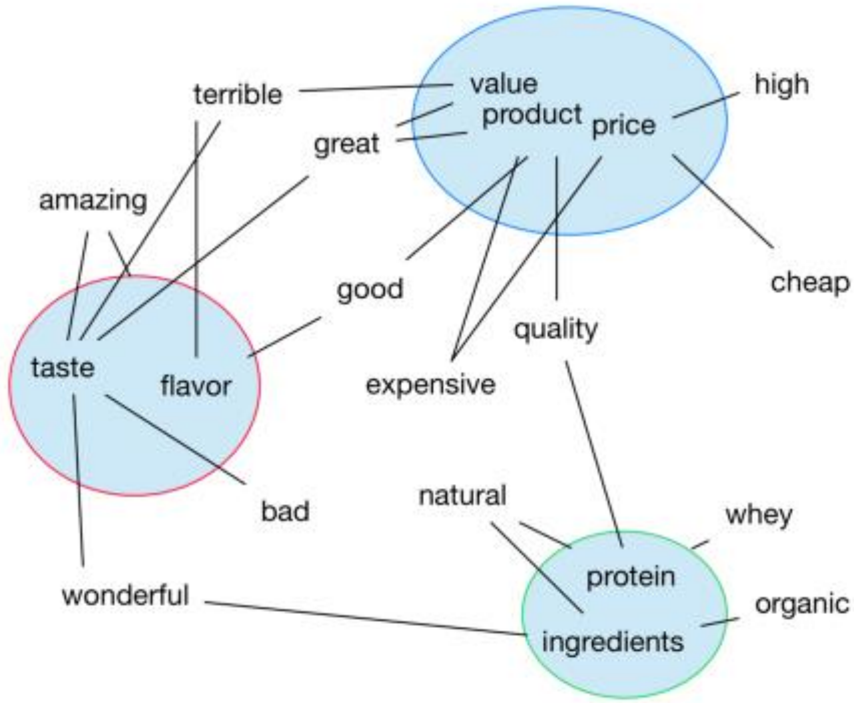

Figure 6 – Item Clusters Diagram

The same procedure is applied to the opinion words. In some cases, because the Apriori algorithm disregards infrequent item sets, some item sets are not included in the item set clustering and it is therefore difficult to assign some of the aspect and opinions within these item sets an aspect (or opinion) cluster. The current approach is to group such aspects (or opinions) into a separate cluster of their own.

Now that we have independent aspect and opinion clusters, we run the Apriori algorithm a second time, but this time creating association rules on the independent aspect and opinion clusters, not the words themselves.

Model Evaluation

We used the bootstrap and 10-fold cross-validation to find the optimal number of clusters for our model. The optimal number of clusters is that for which the average support, confidence and lift are highest and for which the prediction error is the lowest.

For evaluating our model, we applied the bootstrap procedure to 70% of the observations and used 10-fold cross validation in the remaining 30%. The bootstrap was used to verify the dependency of the support, confidence and lift on the number of clusters selected. A 1-15 range of clusters was selected to evaluate our model.

To validate our results, we applied 10-fold cross validation. Using the model's ability to predict opinions from aspects, we calculated the model's prediction error as a function of the number of clusters. Again, we used 1-15 clusters to evaluate our model.

Model Results

Although our needs analysis can be applied to both social media feeds and Amazon.com reviews, we found that social media feeds did not yield actionable or insightful results. Starting from the data collection – limited by the various API's data quotas, questionable API query functionality and lack of new and original content – followed by the data extraction – with poor results due to inconsistent sentence structures, poor grammar and unpredictable spellings – we decided it was best to focus our analysis specifically on the Amazon.com reviews.

Generation of Product Features

Starting with a list of 46 aspects, we were able to generate from the Amazon.com reviews an additional 158 for a total of 204 aspects. While some words were utterly irrelevant (such as the words "people" or "things"), there were others we had not thought about including but were nonetheless important (such as "aftertaste", "energy" or "carbs"). We decided to keep all the words, relevant and irrelevant, expecting those that are irrelevant to naturally have less effect on the results.

Aspect, Opinion Extraction

The results of the aspect opinion extraction seemed accurate and reflected, to a great extent, the needs expressed in the reviews. There were some issues tokenizing some reviews due to poor sentence punctuation, causing inaccuracies in POS tagging and the application of the opinion extraction rules. Other observed inaccuracies stemmed from the generalization of some of the opinion extraction rules. In particular, the application of Rule 9, which caused some opinions to be assigned to the wrong aspects.

Sentiment Analysis

SentiStrength, the Python application used to perform sentiment analysis is commonly used when dealing with short text such as social media posts or ecommerce reviews. To test the application's overall effectiveness in calculating sentiment, we used the overall star ratings of the reviews as a reference against which to compare the sentiment scores. Since the sentiment scores were calculated on a sentence-by-sentence basis and the overall ratings are provided on a per-review basis, it was necessary to add together the sentiment scores of the sentences within each review. That way, we could compare sentiment vs. rating by review. Next, because both metrics have a different scale, we were required to transform each variable to its standard normal form. To calculate the prediction error, we used the mean squared error as the statistic, which we implemented using the bootstrap to account for sampling variances. The results indicate a very small difference between the normalized ratings and sentiment scores, with a point estimate of 0.0004, standard error of 1.9308e-07 and bias of 7.2749e-07.

Aspect-Opinion Association Rules and Clustering

Once we extracted the aspect-opinion pairs, the application of the Apriori algorithm resulted in association rules with low support and confidence (average support of 0.0027 and average confidence of 0.184), which is to be expected due to noise and the wide variety of aspects and opinions about those aspects.

After clustering the results, we obtain much higher support and confidence levels, although this required us to select the optimal number of clusters in which to divide our aspects, opinions and needs. Following the evaluation procedures explained above we obtained the support, confidence and lift values for each number of clusters, obtaining the following results shown in Table 4 and plotted in figures 7 and 8.

Number of Clusters	Mean Support	Mean Confidence	Mean Lift
1	0.25	0.5	1.044895
2	0.12451577	0.354177	1.256182
3	0.07237307	0.2739761	1.369224
4	0.04741868	0.225319	1.366775
5	0.03345156	0.1918758	1.340756
6	0.02492308	0.1680886	1.346979
7	0.01971486	0.151926	1.369287
8	0.01615927	0.13973	1.421145
9	0.0137356	0.1313978	1.513165
10	0.01198511	0.1255193	1.622596
11	0.01068548	0.1214598	1.7388
12	0.009697174	0.118712	1.860129
13	0.008931312	0.1169777	1.9892
14	0.008361053	0.1163808	2.134144

15	0.007914605	0.1165015	2.304505

Table 4 – Bootstrap Results: Support, Confidence and Lift

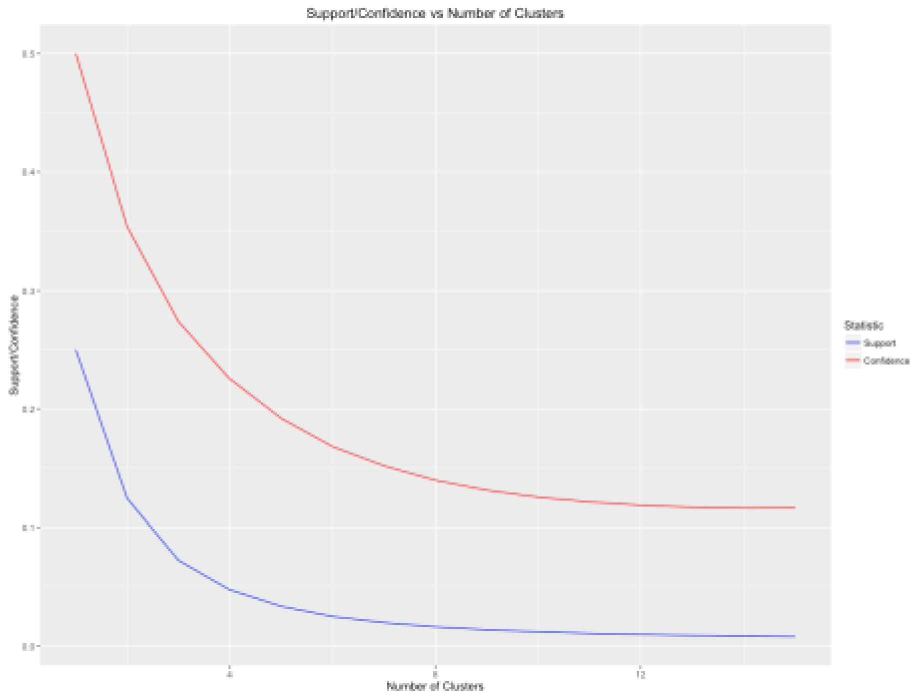

Figure 7 – Support/Confidence vs Number of Clusters

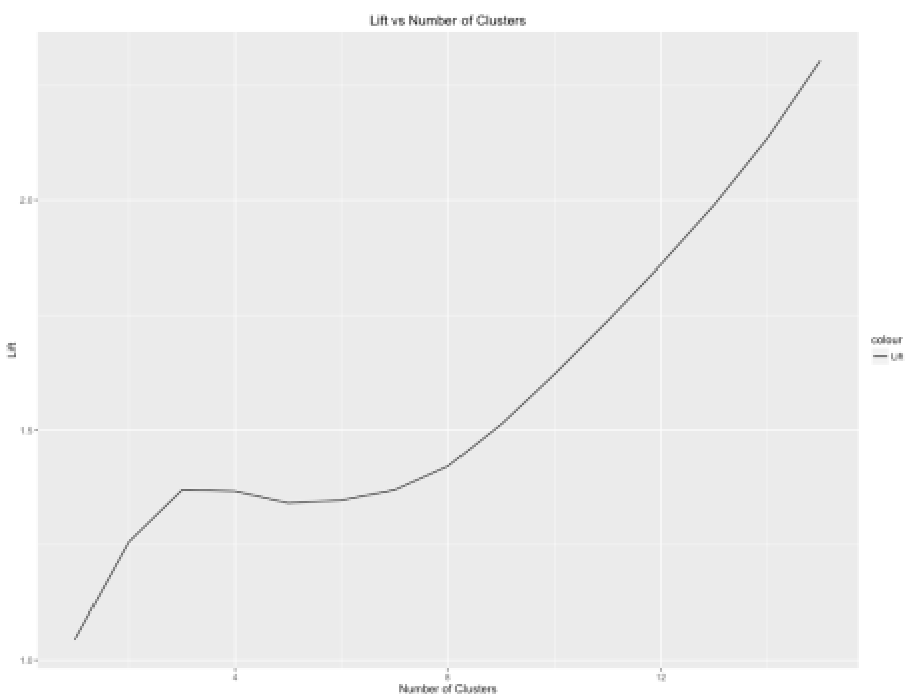

Figure 8 – Lift vs Number of Clusters

The lift seems to increase significantly as the number of clusters increases, reaching a local high peak at 3 clusters, a local low peak at 5 clusters and then increasing monotonically as the number of clusters increases. However, as the number of clusters increase, both the support and the confidence decrease asymptotically towards 0 and 0.1, respectively. Thus, we conclude the optimal number of clusters is 2. We also notice that the resulting average support and confidence has increased to 0.124 and 0.354, respectively.

To confirm these results a 10-fold cross-validation procedure was applied. Using the misclassification error of the resulting association rules as the performance metric, it was confirmed that opting for 2 or 3 clusters was optimal for this dataset. The following chart shows the relationship between the misclassification error and the number of clusters.

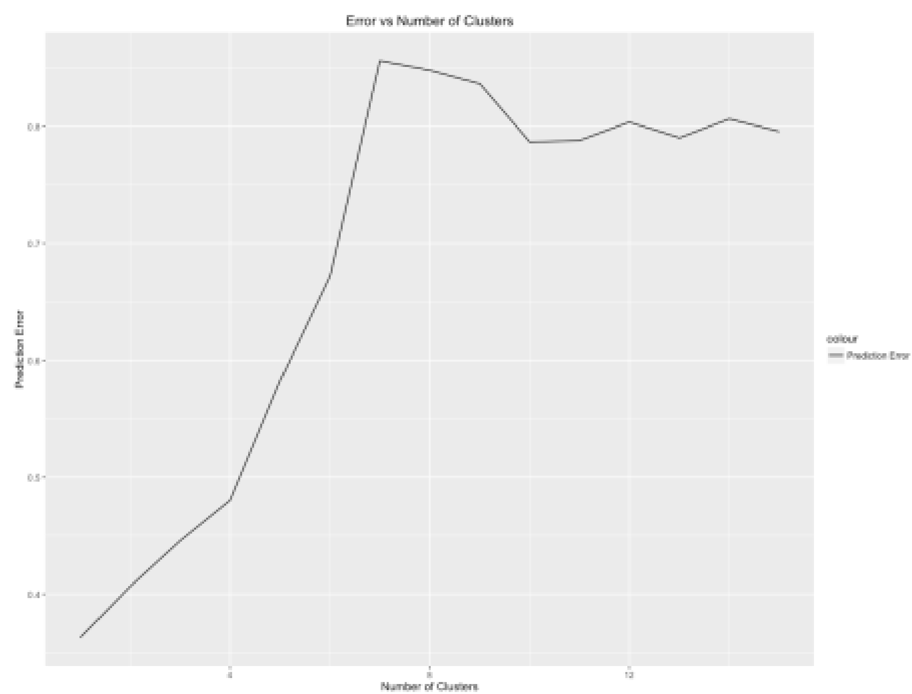

Figure 9 – Prediction Error vs Number of Clusters

Results

Findings

Once we have obtained our model results, we proceed to draw conclusions and find actionable insights that tend to any or all the issues derived from the market penetration problem. We use the 80/20 rule to look for the 20 percent of the data driving 80 percent of the results. We segment the data by cluster and by overall star rating and we examine the aspects, opinions, and aspect-opinion pairs within each category, looking for themes, calculating the average sentiment score and comparing frequencies. We use visualizations to help us make comparative judgements and make recommendations for Ample to consider.

Figure 10 shows the data segmented by aspect-opinion clusters and the mean sentiment score for each cluster. Surprisingly, the frequent item set-based clustering we performed, resulted in three clearly defined segments with significantly different mean sentiment scores. For ease of interpretation, a word cloud for each item set cluster was created, allowing us to find themes within the clusters. Cluster 1, having the highest positive sentiment score, seems to appreciate aspects such as taste, protein, product value and other nutritional elements. In cluster 0, where the sentiment score is moderate, there seems to be a greater focus on utility, showing elements such as “pre-workout drink”, “overall health”, “good diet”, “extra boost”, etc. Finally, in cluster 2, which has the lowest sentiment score, it is observed that there is a strong focus on taste, although a large portion of it is negative: we see elements such as “bad taste” and “horrible taste”.

Figure 10 – Sentiment by Item Set Clusters

Segmentation by star ratings revealed similar results, but it also revealed additional interesting insights. The bar graph in Figure 11 shows that ratings have a direct relationship with sentiment scores, with ratings of 5 and 4 stars representing high positive sentiment, ratings of 3 and 2 stars representing moderate positive sentiment, and ratings of 1 star representing negative sentiment score.

Figure 11 – Sentiment by Star Ratings

Once again, we used word clouds for each of these three sentiment groupings to find themes and interpret the results. We found that in the group with the highest sentiment, with ratings 4 and 5, there is high emphasis on taste, protein and product. The group with medium sentiment score, having ratings of 3 and 2, emphasizes similar aspects, but it has greater focus on opinions about the ingredients, with aspect-opinion pairs such as "artificial sweeteners", "whey protein" and "lactose intolerant". Finally, the group with the lowest sentiment score, having ratings of 1 shows complaints about the taste, but also about the ingredients and components, including "raw probiotics", "synthetic hormones" and "whey protein".

Based on these observations, it is easy to see that aspects such as taste, protein content, overall product appeal and value are key dimensions driving most of the sentiment discrepancies among the consumers. Taste seems to have a strong impact in whether the consumer embraces or rejects meal replacement products. On the other hand, complaints about product ingredients and components can be very detrimental to market penetration, and Ample will need to make sure its products have ingredients that are closest to their natural source and that there is no speculation about the artificial quality of their products.

Challenges

Syntactical and Grammatical Inconsistencies

Text analytics is an evolving area of research, an important topic in an age where the amount of digitally available text is increasing at vast speeds. Nonetheless, it is a topic that undergoes a significant amount of challenges due to inconsistencies in sentence structures, poor grammar, misspelling of words, idioms, shorthand expressions, etc.

Our project was not exempt from these difficulties. As we had intended at the outset, we collected text data from social media posts (Twitter, Youtube and Google Plus) but as we implemented our model we noticed it was significantly more difficult to extract text from these sources than from Amazon.com. One of the reasons is that the reviews from Amazon.com tend to contain more opinions about product aspects than social media posts, but we also noticed there were syntactical and grammatical differences that our model did not account for. Unable to cope with these differences within the time constrains of the project we decided to focus exclusively on the product reviews from Amazon.com.

Competitors and Review Factories

While we were able to capture much of the consumer's opinions through social media feeds and Amazon.com reviews, we have no guarantee that the data is completely unbiased and does not contain either reviews or social media posts from competitors nor activity from paid "review factories" (i.e., reviewers that were paid to submit positive reviews about the sponsoring company's products or negative reviews about a competitor of the sponsoring company).

Opportunities for Future Research

Future research might focus on fine tuning the model we used to extract aspect and opinions from text data. Benchmarking the model's accuracy and capability of extracting aspects and opinions from text against a dataset containing aspects and opinions that have been manually extracted would be the proper way to evaluate, modify and improve our text mining model. Doing the manual extraction of a sufficiently large body of aspects and opinions may require a considerable amount of time and effort, but we believe it will help improve the effectiveness of our project.

Another opportunity for future research might focus on finding different ways to properly perform the clustering of items (aspects and opinions). The current clustering approach is based on frequent item sets (aspect-opinion pairs describing needs), but clustering could be done based on item meanings, perhaps leading to better-defined themes in each item cluster and more parsimonious results overall.

Recommendations

Focus on Specific Product Types

Performing this analysis using data specifically related to a single type of product (such as weight loss products) within the category of meal replacements is likely to produce more insightful results. The reasoning is that, while some aspects of meal replacement products might be desirable for some type of products, it might be undesirable for other types of products. For example, high caloric content may appeal to consumers of weight gainer products but, it may repel consumers looking for weight loss products.

Limit the Amount of Aspects Extracted

Analyzing text for our solution is a computing-intensive process. We must extract every sentence where every aspect we wish to investigate is mentioned and then apply the text mining rules to extract aspects or opinions. Thus, it is recommended to limit the amount of extracted aspects in the first step of the analysis (Generation of Aspects). By the same token, it is wise to limit the amount of aspects to only the ones that are believed to be relevant for the analysis, to limit the amount of noise in the results.

Repeat the Analysis Periodically

Our dataset, while covering the right product group, dated back to 2014. If available, we would recommend to re-run the analysis with more up to date reviews data periodically, as tastes, preferences, and opinions about meal replacement products might shift over time.

Conclusion

The proposed project focuses on addressing the issue of market penetration by finding ways to systematically and cost-effectively listen to the needs of consumers, their preferences and opinions. Collecting and analyzing textual data from user-generated content in social media feeds, blogs and product reviews can reveal a variety of product issues and unmet consumer needs, which can become opportunities, and can lead into promising business strategies. Based on the results of our analysis our recommendation for Ample is to focus on improving product taste and on combatting the perception that the nutritional elements of the product are "synthetic". Other product aspects such as product value (price) and protein content are also important for successful market penetration.

Based on our decomposition of the issues preventing successful market penetration, our initial suspicion that product taste was an obstacle was correct. We also found out that price is not an issue as long as there is enough value in the product, that the ingredients are not artificial or synthetic and that there is an appropriate amount of protein content in the product. Contrary to what we expected, factors like a product's inability to satisfy appetite, the convenience of a to-go meal, weight loss support and health benefits were not found to play a significant role in customer needs. We do not rule out the possibility that these product needs will help Ample successfully penetrate the meal replacement market, but we certainly recommend Ample to pursue product marketing and product development strategies targeting taste appeal, non-artificial or non-synthetic ingredients, and appropriate protein content.

References

Ample. (n.d.). Retrieved October 02, 2016, from http://www.amplemeal.com/

Batagelj, V., Bock, H., Ferligoj, A., & Žiberna, A. (2006). *Data Science and Classification*.

Bing Liu. *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, May 2012.

Borgelt, Christian. "Apriori." *Apriori Documentation*. N.p., n.d. Web. 15 Oct. 2016.

"Extract Meaning from Your Text." *TextRazor - The Natural Language Processing API*. TextRazor, n.d. Web. 30 Nov. 2016.

Lee, Thomas Y. "Needs-based analysis of online customer reviews." *Proceedings of the ninth international conference on Electronic commerce*. ACM, 2007.

Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit *In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55-60.

Market penetration Definition in the Cambridge English Dictionary. (n.d.). Retrieved October 01, 2016, from http://dictionary.cambridge.org/us/dictionary/english/market-penetration

Miller, T. W. (2015) *Modeling Techniques in Predictive Analytics with Python and R: A Guide to Data Science*. NJ: Pearson FT Press.

J. McAuley, (2013). Amazon product data. Retrieved September 29, 2016, from http://jmcauley.ucsd.edu/data/amazon/

J. McAuley, C. Targett, J. Shi, A. van den Hengel. *SIGIR*, 2015

J. McAuley, R. Pandey, J. Leskovec. *Knowledge Discovery and Data Mining*, 2015

Nasukawa, Tetsuya, and Jeonghee Yi. "Sentiment analysis: Capturing favorability using natural language processing." *Proceedings of the 2nd international conference on Knowledge capture*. ACM, 2003.

Stanford CoreNLP. (n.d.). Retrieved October 21, 2016, from http://stanfordnlp.github.io/CoreNLP/

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social Web. *Journal of the American Society for Information Science and Technology*, 63(1), 163-173.

Using Social Media to Improve Product & Service Launches. (2014, April). *Gleanster: Deep Dive*. Retrieved October 2, 2016, from http://www.gleanster.com/report/using-social-media-to-improve-product-service-launches

Figure Captions

Table Captions

Appendix

See files attached.