Amazon Reviews, business analytics with sentiment analysis

Maria Soledad Elli
mselli@iu.edu
CS background.
Interests: data mining.

Yi-Fan Wang
wang624@iu.edu
HR background.
Interests: busyness analytics.

Abstract

Nowadays in a world where we see a mountain of data sets around digital world, Amazon is one of leading e-commerce companies which possess and analyze these customers' data to advance their service and revenue. In order to understand the power of text mining, we utilize these data sets to have a better understanding of the perspectives between stock price and customer comments. We also use machine learning techniques for fake review detection and trend patterns.

1 Introduction

The aim of this project is to extract sentiment from more than 2.7 million reviews and analyze the implications they have in the business area. The data set we used in our project is called *Amazon product data* and was provided by researchers from UCSD (McAuley et al , 2015). In order to acquire insightful business sharpness and the big picture of the whole information we acquired, we combine two original data sets: one is composed by customer reviews, and the other one contains product information. Furthermore, in terms of our goals for detecting user emotions from reviews, gender based on their names and review, and further fake reviews, we not only adapt *Textblob*, and *Genderizer* to advance our understanding towards these perspectives, but also build up our classifier to measure the system's accuracy. Afterwards, we started to analyze targeted perspectives or famous-brand-related accessories like Nokia, Apple, HTC to dig and interpret our interesting findings by different methods. We use Python and R tools to clean, extract, analyze and show results achieved in our work. The following figure 1 shows a representation of our work method.

This paper is organized as follow: in the following sections we will explain the methodology used in the machine learning stage where we extracted the sentiment from the reviews and then we will explain and analyze the results we obtained from our process.

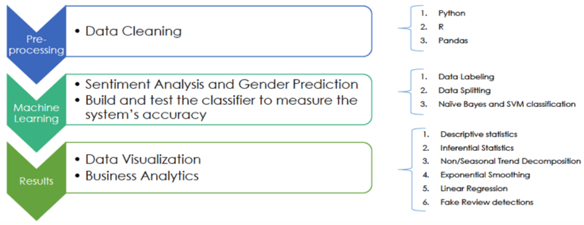

Figure 1: Work process

2 Sentiment Analysis on data

In order to achieve our main goals, it is imperative to do some sentiment analysis on the data set to extract people's opinion about the products they have bought. As far as we know, there is no published work about sentiment analysis in amazon reviews.

In terms of the data set, we have two big JSON files where the structure of the data set is as follows:

- Review structure
  - `reviewerID - ID of the reviewer, e.g. A2SUAM1J3GNN3B`
  - `asin - ID of the product, e.g. 0000013714`
  - `reviewerName - name of the reviewer`
  - `helpful - helpfulness rating of the review, e.g. 2/3`
  - `reviewText - text of the review`
  - `overall - rating of the product`

- `summary` - summary of the review
- `unixReviewTime` - time of the review (unix time)
- `reviewTime` - time of the review (raw)

- Product description structure
  - `asin` - ID of the product, e.g. 0000031852
  - `title` - name of the product
  - `price` - price in US dollars (at time of crawl)
  - `imUrl` - url of the product image
  - `related` - related products (also bought, also viewed, bought together, buy after viewing)
  - `salesRank` - sales rank information
  - `brand` - brand name
  - `categories` - list of categories the product belongs to

After combining these two files together, we labeled each review based on the polarity and subjectivity values obtained with the Textblob v0.11.0 package for python. This package seamed to be robust and it has very good reviews in terms of performance; a result that we could confirm with our experiments in section 4.7. The polarity and subjectivity levels returned by TextBlob are in an scale from [-1, 1] and [0, -1] respectively and because of this, we had to define some threshold for these values to set the label of each review. Thus, we considered that reviews with a polarity greater than 0.25 is positive, less than 0 will be negative and between o and 0.25 neutral. After we labeled the reviews, we extracted the gender of each review for further analysis. In this case we applied the python package Genderizer v0.1.2.3 which provides functions that not only by verifying the name, but also the text related to it. After the labeling process, we need to extract features from the reviews and build a classifier for future incoming reviews. The next subsections will explain how we tackle these problems.

2.1 Feature extraction

Since we have more than two million reviews, extracting features from all of them and building a classifier with that amount of samples it is computational expensive and, in some cases, even impossible. Because of this, we extracted a reduced amount of reviews of each category taking into account not only the polarity but also the rating value of that review. This is, we filtered the positive reviews by selecting the ones that have a polarity grater than 0.25 and a rating value greater or equal to 4. The same with the negative reviews but with a polarity less than 0 and rate value less or equal to 2 and for the neutral reviews, we filtered the data with the polarity values between 0 and 0.25. Since we are dealing with reviews and not with complex texts, the vocabulary used does not include many different words, so selecting the most 15000 representative samples of each category will be enough to represent the entire data set. After this filtering process, we used the bag-of-words approach for text. The most intuitive way to do this is by assigning a fixed integer id to each word occurring in any of the samples of the training set. Then, for each document i, we count the number of occurrences of each word w and store it in a dictionary Xi, j as the value of feature $\#j$ where j is the index of word w in the dictionary. Since the bag-of-words approach is a good start, there is an issue: larger reviews will have higher average count values than shorter reviews. To avoid this we can divide the occurrence of each word in a review by the total number of words in that review, these new features are called tf for Term Frequencies. Another improvement on top of the tf is to downscale weights for words that occur in many reviews in the data set and are therefore less informative than those that occur only in a smaller portion of the data set. This downscaling is called $tfidf$ for "Term Frequency times Inverse Document Frequency" (Baeza-Yates et al , 1999), (Manning et al , 2008). This is a well known method widely used by researchers in text mining. In some cases, only the 100 or even the 25 most frequent words are enough to describe the documents of a particular corpus.

2.2 Classification

As for the classification problem, we build a Multinomial Nave Bayes (MNB) and a Support Vector Machine (SVM) classifier (Joachims, 1998), (Wu et al , 2004) using the Scikitlearn python package. We trained both classifiers with 50% of the data and tested them with the other 50% of the data to calculate the accuracy. The final results are shown in table 1.

Method	Accuracy	Time
MNB	72.95%	0.1307 sec
SVM	80.11%	16 min 37.8846 sec

Table 1: Classifiers performance

As you can see, the accuracy for both cases is very high. Since there is no similar project already done, we cannot compare it with some previous work. It is worth mention that the processing time between the two algorithms is very different. This is because of the simplicity of Nave Bayes. This algorithm only uses simple arithmetic operations, while svm does not. As the number of samples increases, the more time it will take to svm to complete the classification process and in some cases it won't be able to finish it at all. The following figures 2 and 3 show the confusion matrix of each classifier and reflects the results obtained so far.

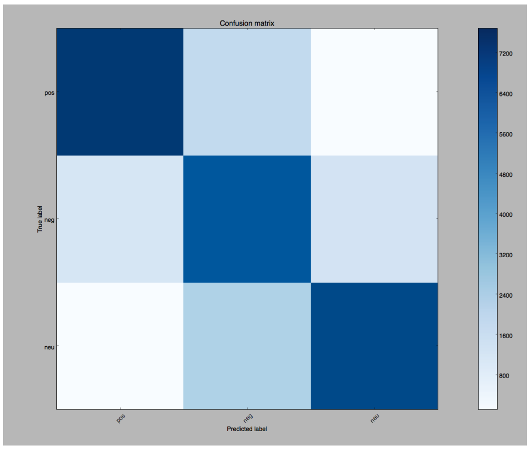

Figure 2: Multinomial Nave Bayes Confusion Matrix

3 Fake review detections

Based on the research of (Liu, 2012), the author concludes that negative outlier review, ratings with significant negative deviations from the average rating of a product, tend to be heavily spammed. Positive outlier reviews are not badly spammed. According to his conclusion, we decide to adapt and extend Bing's conclusion to detect possible fake reviews by the following method: Detect

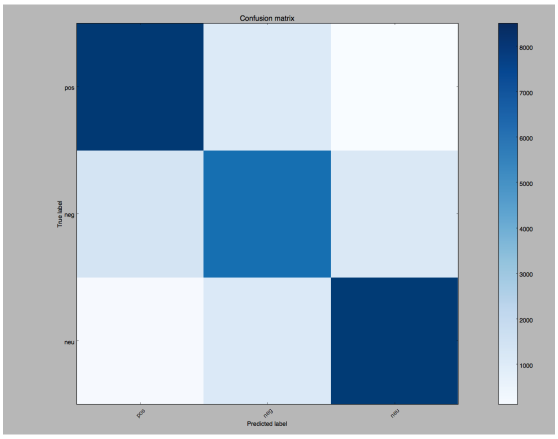

Figure 3: Support Vector Machine Confusion Matrix

discriminations between polarity and overall rating under the situation. In terms of this way, we choose polarity to be part of perspectives as this detection method in that the previous result of our linear regression shows the significant positive correlation between polarity and overall ranking, which also can narrow down the number of possible fake reviews. After filtering the possible fake reviews by these two methods, we will manually confirm these reviews by sampling to identify whether these are fake reviews. Here we are going to analyze this case with the following brief summary, ”Otterbox Defender Series Hybrid Case & Holster for iPhone 4 & 4S” which has 14961 reviews by our method - we design the filter with outliers of overall ranking and polarity. At this point we focus on rating values of significant negative deviations from the average rating of the product ranging from 2.5 to 1, and on the polarity value of significant negative deviations from the average polarity, ranging from 0.8907933 to 1. In other words, we try to find the reviews under this condition when customers gave extreme low grades on the product but their reviews are somewhat relatively positive. Figure 4 briefly summarize the total reviews of Otterbox Defender Series Hybrid Case. Table 2 shows the information of possible fake reviews by our method and the text of each review is as follows:

1. you know what. It has three layers, and for what? It does protect your phone against falls (that's why I gave it 2 stars instead of 1) but that's the best that can be said about it. The silicone gasket that wraps around the phone never stays in place, as well as the port cov-

ers. This product lets in a lot of dust and then traps it. Look for another product to protect your iPhone.

2. This cover fits perfect, but it has some type of film or oil or something that is on the screen protector that I can't get to go away. Otherwise I would have given this product a five.

3. There are gaps in the case, so I feel like my phone isn't as protected as it should be. It LOOKS great though!

4. The otterbox I purchased was not in the greatest shape when I got it. The screen has scratches all over it.

5. Would have contacted the seller but doesn't look like amazon gives you that option. Work in health care and bought this so I could clip it onto my scrubs after a week and a 1/2 the belt clip started to break. For a product that is supposed to hold up and protect doesn't add up to me. So I either got a factory defected one or its not the best quality product.

```
overallRanking    polarity            subjectivity    label
Min.   :1.000  Min.   :-1.00000   Min.   :0.0000   neg:1965
1st Qu.:4.000  1st Qu.: 0.08056   1st Qu.:0.4286   neu:5885
Median :5.000  Median : 0.23750   Median :0.5333   pos:7111
Mean   :4.186  Mean   : 0.25562   Mean   :0.5321
3rd Qu.:5.000  3rd Qu.: 0.42500   3rd Qu.:0.6417
Max.   :5.000  Max.   : 1.00000   Max.   :1.0000
```

Figure 4: Otterbox Defender Series Hybrid Case sentiment summary

Review#	Ranking	Polarity
[1]	2	1
[2]	2	1
[3]	2	1
[4]	1	1
[5]	1	1

Table 2: Fake reviews description

Based on the following definitions of types of spam and spamming: Type 1 (fake reviews): These are untruthful reviews that are written not based on the reviewers' genuine experiences of using the products or services, but are written with hidden motives.
Type 2 (reviews about brands only): These reviews do not comment on the specific products or services that they are supposed to review, but only comment on the brands or the manufacturers of the products.
Type 3 (non-reviews): These are not reviews. There are two main subtypes: (1) advertisements and (2) other irrelevant texts containing no opinions (e.g., questions, answers, and random texts). Strictly speaking, they are not opinion spam as they do not give user opinions.

Those 5 possible fake reviews don't match the preceding definitions of 3 type of reviews. However, we do see that some discriminations between the ratings and review texts, showing that some reviewers reflect lower ratings exaggeratedly but they were not that satisfied with the product based on their review texts. For example, "There are gaps in the case, so I feel like my phone isn't as protected as it should be. It LOOKS great though!", we can see that this comment ended up with positive conclusion, nevertheless this reviewer still gave 2 to this rating. Furthermore, we test these reviews by using Review Skeptic (RS) `http://reviewskeptic.com/` ,based on the research at Cornell University, to check whether these match their fake review detection methods. And we acquire the results from table 3.

Review #	Ranking	Polarity	RS
[1]	2	1	Truthful
[2]	2	1	Truthful
[3]	2	1	Deceptive
[4]	1	1	Deceptive
[5]	1	1	Truthful

Table 3: Fake reviews results

Although Review Skeptic's data sets are based on hotels' reviews, after we manually confirmed these possible fake reviews and test those with Review Skeptic, there is something worthy to dig further. As the researcher at Cornell University mentions this kind of fake review detections might be "first-round filter", we will adapt our classifier to compare these results in order to advance our detection method as our future work. Figure 5 shows a graphical view of the outliers detected in the reviews. The yellow points are the cases that match the fake review relation between polarity and rate value. An html file will be added to the project folder which contains the 3D graph of figure 5 for more details.

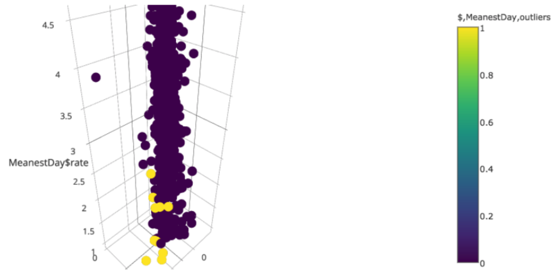

Figure 5: Outliers - Fake reviews

4 Business related results

4.1 Basic understanding of perspectives

After cleaning and processing data we acquired 2403356 customer reviews linked to the corresponding product. The data related to the reviews include the following fields: review ID, asin (product ID), reviewName, reviewText, overall ranking, summary, unix review time, review time, helpful, price, title, brand, polarity, subjectivity, label, and gender. With this new data set we can have a general understanding about the data related to our main objectives. Something that is worth mention is the flexibility that Amazon provides to their customers in terms of the reviews. The e-questionnaire they have to fill out after a purchase allows user to skip the text parts such as reviews, summary, etc and for us, that represent a missing value in our data set. The others fields like brand reveal NA values because of the incomplete original data related to the product. Figure 6 shows a summary of the raw data we obtained. Also, figure 7 shows the result of the processed data.

```
          reviewID                asin                    reviewName               reviewText        overallRanking
 A2NYK9KWFMJV4Y:    165   B005SUHPO6:  14961   Amazon Customer:  46755                  :    576   Min.   :1.000
 A18U49406IPPIJ:    125   B0090YGJ4I:   8495   Pen Name       :   9642   Love it   :    350   1st Qu.:3.000
 A3LDPF5FMB78ZZ:    121   B008OHNZI0:   7817   Chris          :   2739   Good      :    331   Median :5.000
 ABDR6IJ93HFIO :    108   B005SUHRVC:   6820   Mike           :   2427   Great     :    327   Mean   :3.831
 A2NOW4U7W3F7RI:    106   B008DJIIG8:   6114   John           :   2364   Love it! :    205   3rd Qu.:5.000
 A1EVV74UQYVKRY:    101   B009A5204K:   6025   David          :   2154   Excellent:    154   Max.   :5.000
 (Other)       :2402630   (Other)   :2353124   (Other)        :2337275   (Other)  :2401413   NA's   :173
          Summary            unixReviewTime        ReviewTime            helpful            Price
  Five Stars   :  25750   1380672000:   8316   Min.   :2000-03-24   0/0    :1780309   Min.   :  0.01
  Great        :  14207   1389052800:   7514   1st Qu.:2013-01-28   1/1    : 223097   1st Qu.:  3.95
  Great product:  11509   1404691200:   7360   Median :2013-09-04   0/1    :  99522   Median :  8.95
  Love it      :  11247   1404777600:   6820   Mean   :2013-06-14   2/2    :  58166   Mean   : 21.07
  Great case   :  10924   1388707200:   6615   3rd Qu.:2014-02-24   1/2    :  49343   3rd Qu.: 19.95
  Love it!     :  10195   1404950400:   6604   Max.   :2014-07-23   3/3    :  24003   Max.   :999.00
 (Other)       :2319524   (Other)   :2360127   NA's   :303         (Other): 168916

Title
  Otterbox Defender Series Hybrid Case &amp; Holster for iPhone 4 &amp; 4S  - Retail Packaging - Sun Yellow/Gunmetal Grey
:  14961
  OtterBox Defender Series Case for iPhone 5 (Discontinued by Manufacturer) ( Not for iPhone 5C) Retail Packaging Bolt - Gr
ey:   8495
  Tech Armor Ultimate 4-Way 360 Degree Privacy Screen Protector for Apple New iPhone 5, Latest Generation, 1-Pack
:   7817
  OtterBox Commuter Series Case for iPhone 4/4S  - Retail Packaging - Hot Pink/White
:   6820
  Caseology Slim Fit Flexible TPU Case Compatible with Samsung Galaxy S3 (Pink)
:   6114
  LG Electronics Tone+ HBS-730 Bluetooth Headset - Retail Packaging - Black
:   6025
 (Other)
:2353124
```

Figure 6: Raw data Summary

```
        Brand             polarity          subjectivity      label            label_int          gender
            :2397628   Min.   :-1.00000   Min.   :0.0000   neg: 306215   Min.   :-1.0000         : 126411
 Generic    :   1298   1st Qu.: 0.08392   1st Qu.:0.4458   neu: 963373   1st Qu.: 0.0000         : 170393
 Motorola   :    511   Median : 0.23686   Median :0.5500   pos:1133768   Median : 0.0000   female:1293216
 Cellet     :    357   Mean   : 0.24698   Mean   :0.5452                 Mean   : 0.3443   male  : 813336
 Jabra      :    333   3rd Qu.: 0.40667   3rd Qu.:0.6578                 3rd Qu.: 1.0000
 AccessoryOne:   283   Max.   : 1.00000   Max.   :1.0000                 Max.   : 1.0000
(Other)     :   2946
```

Figure 7: Processed data Summary

Based on this sentiment data summary, one can clearly find aggressive customers by review ID, popular products, etc. In addition, as for the output obtained with Textblob and Genderizer, Amazon commenters generally provide relatively positive reviews over 3 stars of 5. With regard to the gender prediction, aside form the noisy data, we have 60% of female reviewers, and we will compare the research result of the work exposed in (Hovy et al , 2015), especially in customer behavior filed.

4.2 Identify the frequency of words of comments/summaries on each brand

In order to have a big picture about the variety of comments on specific brands we targeted (such as Apple, Nokia, etc). Meanwhile, we selected most repeated words for seeking performance and opinions in products, we can also recognize which terms, especially in the summary comments perspective, customers mostly used and might be considered for advertisements. By conducting the word-cloud function, we can have a basic visualization on which words are mostly used by customers. For instance, Nokia's customers on Amazon commented on its products by some adjectives such as "poor", "excellent", and "nice". Comparing with the plot of the summary and the review text, we can clearly see that Nokia users prefer to comment on their products by some general terms. In figure 8 and 9 we can see the most frequent words used by customers for Apple products and accessories. The same details for Nokia are showed in figures 10 and 11. Interestingly, according to figure 11, there is a significant amount of Nokia users that mention about iPhone.

In regard to comments on Apple accessories, the wordcloud show mostly positive feedbacks on Apple-related accessories and a greater frequency of the word, "recommend" compared to Nokia. In terms of the summary comments, the word "great" comes out as the most frequent one. In this case, one can conclude that part of vendors in Amazon produce advanced quality of Apple's accessories that most fit to Apple customers' expecta-

Figure 8: Apple Summary Wordcloud

Figure 10: Nokia Summary Wordcloud

Figure 9: Apple Reviews Wordcloud

Figure 11: Nokia Reviews Wordcloud

tions. The function of visualizing these most frequent words in comments of each brand can help us to easily distinguish the overall user opinions for these brand's accessories.

4.3 Identify average ratings for each brand

According to our summary on each brand rating information showed in figure 12, with the benchmark of the average rate of all ratings, the rating of some brands like Nokia, Google, LG, and Motorola are above par. Also, HTC's average rating just matches the benchmark. The rest brands like Blackberry, Sony, and Apple are below par. With the boxplot graph displayed by figure 13, the data shows that only Nokia, Google, LG, Motorola have no ratings lower than 3. This plot clearly demonstrates that Nokia and Google have relatively better rankings than the rest brands.

4.4 Customer subjectivity

As mentioned before, we used TextBlob, one built-in function for processing textual data in Python, that gives an API for diving into common natural language processing (NLP) tasks like sentiment analysis and text classification. In this case, we decided to use TextBlob to analyze each text review and identify the polarity (positive/negative reviews) and the subjectivity (subjective/objective users). After this, we can see in figure 14 how these values are summarized for each brand. In

```
    All              Nok              BBRY             GOOG             HTC              LG               MSI
Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.0
1st Qu.:2.000   1st Qu.:3.000   1st Qu.:2.000   1st Qu.:3.000   1st Qu.:2.000   1st Qu.:3.000   1st Qu.:3.0
Median :4.000   Median :5.000   Median :4.000   Median :4.000   Median :4.000   Median :5.000   Median :4.0
Mean   :3.675   Mean   :3.862   Mean   :3.641   Mean   :3.888   Mean   :3.676   Mean   :3.767   Mean   :3.7
3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:5.0
Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.0
NA's   :11                                                                                      NA's   :6
    Sony             APPL
Min.   :1.000   Min.   :1.000
1st Qu.:2.000   1st Qu.:2.000
Median :4.000   Median :4.000
Mean   :3.642   Mean   :3.445
3rd Qu.:5.000   3rd Qu.:5.000
Max.   :5.000   Max.   :5.000
                NA's   :33
```

Figure 12: Brand rating summary

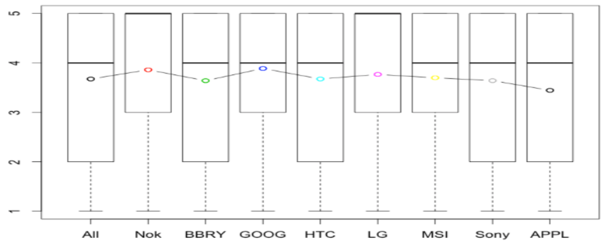

Figure 13: Brands boxplot for summary of ratings

the meanwhile, we will compare these outputs of Textblob with the overall rankings commented by customers, which could be assumed as real ratings at this stage to see whether emotion detection by Textblob can effectively reflect or match the rating behaviors.

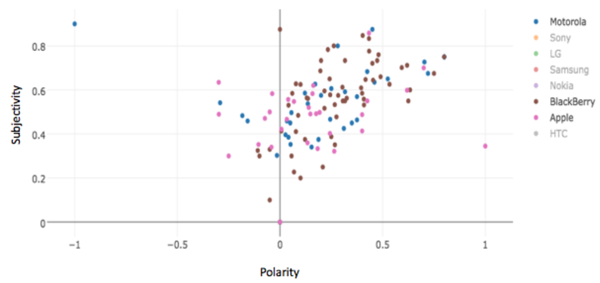

Figure 14: Motorola, BlackBerry, and Apple polarity and subjectivity distribution

4.5 Correlation between review's sentiment and customers

In this case, we try to see the correlation between price, reviews polarity, reviews subjectivity, and customer gender reflected through a linear regression model represented in figure 15. Although the 23.74% of the data is well predicted by our linear regression model, we can see that factors such as price, polarity, subjectivity, and gender reveal significant positive correlation with the actual rating behavior. Among these perspectives, the polarity of the reviews has more significant influence than other factors on the ranking value. Interestingly male customers seemly have slightly negative influence, the result matching the idea in (Hovy et al , 2015): men tend to vote slightly negative than women.

```
Call:
lm(formula = Data.Clean_labels$overallRanking ~ Data.Clean_labels$Price +
    Data.Clean_labels$polarity + Data.Clean_labels$subjectivity +
    Data.Clean_labels$gender)

Residuals:
    Min      1Q  Median      3Q     Max
-5.0075 -0.8219  0.2991  0.9598  4.6253

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    3.085e+00  4.400e-03 701.031  < 2e-16 ***
Data.Clean_labels$Price        2.888e-04  1.891e-05  15.275  < 2e-16 ***
Data.Clean_labels$polarity     2.719e+00  3.476e-03 782.159  < 2e-16 ***
Data.Clean_labels$subjectivity 7.863e-02  4.840e-03  16.248  < 2e-16 ***
Data.Clean_labels$genderfemale 2.696e-02  3.775e-03   7.140 9.31e-13 ***
Data.Clean_labels$gendermale  -1.843e-02  3.871e-03  -4.761 1.92e-06 ***
Data.Clean_labels$genderNone   2.464e-01  4.790e-03  51.447  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.278 on 2403176 degrees of freedom
  (173 observations deleted due to missingness)
Multiple R-squared:  0.2374,    Adjusted R-squared:  0.2374
F-statistic: 1.247e+05 on 6 and 2403176 DF,  p-value: < 2.2e-16
```

Figure 15: Linear Regression Model

4.6 Correlation between a brand and its pricing design

Here we are going to analyze how venders set up their accessories' price to induce customers to buy their products. At this point we are going to introduce one example of venders called Jabra which provides wireless and corded headsets for mobile phone users, contact centers and office-based users. And its customers include Apple, Sony and Nokia users. First of all, we assume these reviews represent consuming behaviors, one review thought as buying one product which was commented by one reviewer. In other words, we simplify the situation and won't consider any situation like comments without purchase. According to example for Jabra showed in figure 16, we can see how this brands' pricing strategy is defined on the right plot of figure 16. For instance, under Jabra's product line, Samsung with better selling record/ more reviews has more centralized pricing strategy range from \$3 to \$7. Not only Sony but HTC chose to provide higher price accessories. Furthermore, Apple and Blackberry competed harshly each other with providing similar price for customers.

Also, we can see in figure 17 how some brands' pricing strategy changed during the time, for example, Blackberry, a Canadian telecommunication and wireless equipment company, has changed its

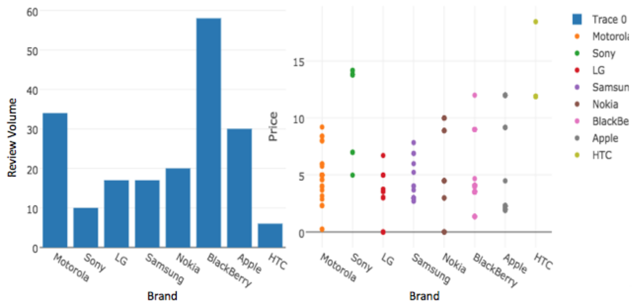

Figure 16: Pricing Sales Volume for Jabra

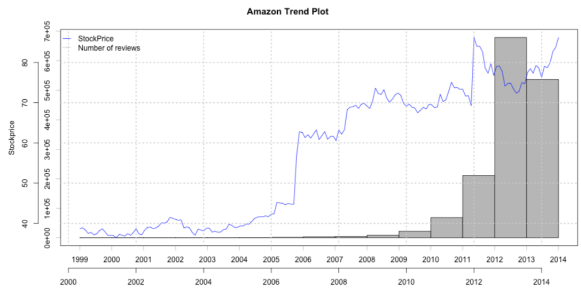

Figure 18: Amazon Trends

pricing strategy since 2010 to around $4 for its accessories. In contrast to Blackberry, Apple within wider pricing design has some customers who are more interested in lower price accessories, around $2. Thus, considering this data is still based on the customer reviews, we still need some additional information such as financial news to confirm our result.

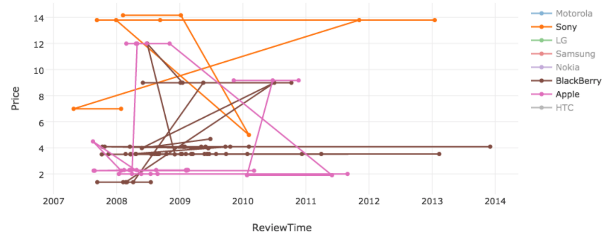

Figure 17: Pricing variations with time

4.7 Relation between stock price and average rating

Based on the research result of (Dickinson et al , 2015), researchers concluded that the correlation has been shown to be strongly positive in several companies, particularly Walmart and Microsoft which are primarily consumer facing corporations. In our case, we try to detect whether these related products' reviews are correlated to the brands' stock prices. Here we consolidate three companies' data to observe if there is any correlation.

The first plot in figure 18 includes all the reviews and the historical stock price from 1999 to 2014. We can see that the number of customer reviews increased slower than the increase of its stock price, however, Amazon's stock price and number of reviews still show a positive relationship.

In terms of the HTC plot displayed by figure 19, although its customer reviews has increased year by year, we can clearly see that its stock price, daily average rating, and daily average polarity have strong correlation, especially from 2011 to 2014. Generally, HTC customers' ratings share a similar trend with the polarity score on reviews assigned by TextBlob package, and even its rating seems to follow its stock price.

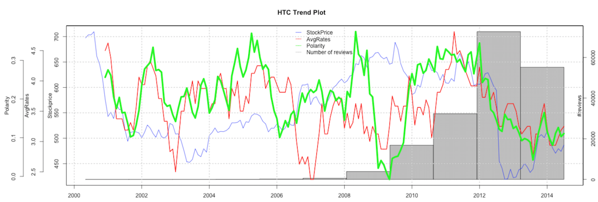

Figure 19: HTC Trends

In contrast to HTC's daily rating, we can see that the average rankings on Apple accessories in figure 20 looks above par most of time. Once again, we can see from 2011 to 2014, Apple's stock price, average rate, and polarity have comparable trend.

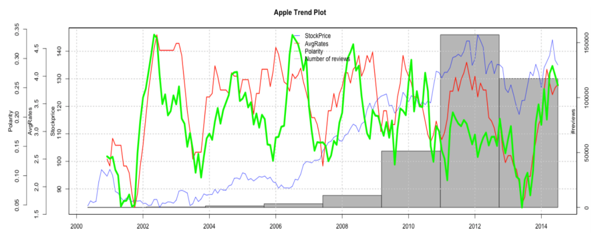

Figure 20: Apple Trends

Regarding Blackberry's plot in figure 21, its rating shows more bigger variance, even though its stock price had increased from 2002 to 2013. Its average rating and polarity shares similar pattern as well, which implies that actual rating behaviors mostly match their comments. Generally, Blackberry's number of reviews has similar trend with its stock price even though our data set only ranges from 2000-3-24 to 2014-7-2. Also from 2012 to 2014, its stock price, Average rate, and polarity have comparable trends.

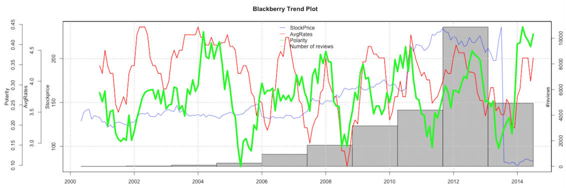

Figure 21: BlackBerry Trends

5 Conclusions

As for the machine learning process carried out in this project, we believe that all the tools we used demonstrated to be robust enough to achieve high values of accuracy. The Textblob package demonstrate to perform very well and it helped us to find fake reviews from customers, as explained in section 3. Regarding the feature extraction process, the top ten most used words are: *phone* (309929 times), *case* (155322 times), *battery* (104506 times), *great*(101257 times), *like* (83396 times), *good* (82753 times), *just* (79668 times), *product* (73719 times), *screen* (73618 times), *use* (72127 times). This result was extracted with the bag-of-words method and clearly reflects the scope of the texts in our data set.
Without more information aside from the data set, we conclude several following points based on our analytic results:

1. The contrasts of these brands' frequency of words reveal additional information what these aggressive customers think about like Nokia customers would compare their products with Apple-related products. However, as the examples of word-cloud we have also show the disadvantage of these most common reviews Such as good, great, and excellent, which cannot truly reflect what kind of details on accessories these brand can improve, if we stand in these companies' shoes, we will need to explore more negative feedbacks for improving these products.

2. Emotion detection can be useful in marketing segments which help corporates to distinguish what kind of current customers they have and what kind of potential customers they prefer in the future. Through the emotion distribution by subjectivity and polarity, we can have clear view on which brands' commenters tend to be more unsatisfied as well. But due to the limit of our computing efficiency, we cannot provide the comprehensive scatter plot at this stage.

3. In regard with detecting fake reviews, we have built up our own first round filter to investigate these possible fake reviews. However, according to our finding, these possible fake reviews without further emotion analyses could only be thought as unmatched rating behaviors with their ratings and their emotions on the comments. Thus, for the future work, we will consider to adapt classifiers to have more accurate findings.

4. Understanding each brand's pricing design requires lots of insightful data from the markets, here we give another perspective to dig to this pricing field to know how these vendors like Jabra design and decide their price for different brands' accessories. In the future, our findings should be compared with reputable market survey to confirm whether our customer review data set are representative enough to be considered as pricing strategy.

5. With our three examples from HTC, Apple, Blackberry, we all find that during the period from 2011 to 2014, their stock price, rating, and polarity share almost identical trends which interest us to acquire more information to understand why these customer reviews started to match the financial market.

As for the work division regarding to this project, we both discussed the tasks together and aport the same amount of work to the project. Although the business insights and theory were provided by Yi-Fan since his background is related to that area. The code used through out this work will be added to project's folder.

References

McAuley, Julian, et al. *Image-based recommendations on styles and substitutes* Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2015.

Hovy, Dirk, Anders Johannsen, and Anders Sgaard. *User review sites as a resource for large-scale sociolinguistic studies.*, Proceedings of the 24th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2015.

Dickinson, Brian, and Wei Hu. *Sentiment Analysis of Investor Opinions on Twitter.*, Social Networking 4.03 (2015): 62.

Liu, Bing. *Sentiment Analysis and Opinion Mining*, Synthesis Lectures on Human Language Technologies 5.1 (2012): 1-167.

Baeza-Yates, Ricardo, and Berthier Ribeiro-Neto. *Modern information retrieval.*, Vol. 463. New York: ACM press, 1999.

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schtze. *Introduction to information retrieval. Vol. 1.*, Cambridge: Cambridge university press, 2008.

Joachims, Thorsten *Text categorization with support vector machines: Learning with many relevant features.*, Springer Berlin Heidelberg, 1998.

Wu, Ting-Fan, Chih-Jen Lin, and Ruby C. Weng. *Probability estimates for multi-class classification by pairwise coupling.*, The Journal of Machine Learning Research 5 (2004): 975-1005.