Sentiment Analysis Summary Report

In this study, the sentiment analysis model of spaCy is implemented and performed on a set of customer review texts. Two similar customer review texts are also compared.

1. The language model employed: en_core_web_md

2. The dataset

2.1 A dataset of Consumer Reviews of Amazon Products is used in this study. The dataset consists of 34660 entries and 21 columns, as shown below.

```
RangeIndex: 34660 entries, 0 to 34659
Data columns (total 21 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   id                   34660 non-null  object 
 1   name                 27900 non-null  object 
 2   asins                34658 non-null  object 
 3   brand                34660 non-null  object 
 4   categories           34660 non-null  object 
 5   keys                 34660 non-null  object 
 6   manufacturer         34660 non-null  object 
 7   reviews.date         34621 non-null  object 
 8   reviews.dateAdded    24039 non-null  object 
 9   reviews.dateSeen     34660 non-null  object 
 10  reviews.didPurchase  1 non-null      object 
 11  reviews.doRecommend  34066 non-null  object 
 12  reviews.id           1 non-null      float64
 13  reviews.numHelpful   34131 non-null  float64
 14  reviews.rating       34627 non-null  float64
 15  reviews.sourceURLs   34660 non-null  object 
 16  reviews.text         34659 non-null  object 
 17  reviews.title        34654 non-null  object 
 18  reviews.userCity     0 non-null      float64
 19  reviews.userProvince 0 non-null      float64
 20  reviews.username     34653 non-null  object 
dtypes: float64(5), object(16)
memory usage: 5.6+ MB
None
```

2.2 The focus of interest in this study is the accuracy of the sentiment prediction of the language model on the review texts. In the study, reviews rating is used as a cross-reference of the accuracy of sentiment prediction. Thus, only the columns ‘reviews.rating’ and ‘reviews.texts’ are copied out and used in the analysis.

2.3 Cleaning performed on the data set:

1. In order not to miss any review texts, missing entries of reviews ratings, which are digits from 1 to 5, are grouped into 0s. So, a rating of 0 does not mean the lowest score but only a missing one.

2. As reviews ratings are expressed as a digit from 0 to 5, the data type of ‘reviews_rating’ is changed into integer type.

3. Missing entries are removed from the dataset.

4. Duplicated entries are removed from the dataset.
5. The cleaned data frame is re-indexed so that there will not be errors when a missing index is chosen randomly in the program.

2.4 After cleaning, the dataset comprises of 34659 entries, as shown below.

```
RangeIndex: 34659 entries, 0 to 34658
Data columns (total 2 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   reviews.rating  34659 non-null  int8
 1   reviews.text    34659 non-null  object
dtypes: int8(1), object(1)
memory usage: 304.7+ KB
None
```

3. Sampling and output

3.1 User is asked to input the total number of review texts to be chosen at random from the 34659 entries for the sentiment analysis. The minimum number set is 10 and the maximum number is one-seventh of the size of the dataset.

3.2 While all of the sampled review texts will be covered by the sentiment analysis, only a maximum of 3 in those which are inconsistent with the review rating and a maximum of 3 other review texts will be studied. Inconsistency means having a high review rating but negative polarity in the sentiment analysis on the review text, or vice versa.

3.3 One further review text is chosen at random to be a reference base text, which would be compared against all other review text samples to check for the similarity between the two. The top two most similar review texts with the reference base text will be printed out.

3.4 A violin plot graph will be prepared to show the spread of sentiment polarities in each of the review rating groups.

4. Text Preprocessing

4.1 Before performing the analysis, the sampled review texts are cleaned so that the sentiment analysis of the language model could be effectively implemented.

- Removing all stop words that would not affect the sentiment analysis of the language model from the text. It is found that some of the stop words used in spaCy such as 'not', 'very', 'more', and 'really' are used in the sentiment assessments, as indicated in a sample output below. So, those words are excluded from the stop word removal process.

```
Review text examined:
Small and lightweight. I wouldn't really recommend for teenagers or
adults more so for kids under 10
Sentiment measures:
   Polarity is 0.1500 and Polarity with stop words removed is -0.2500
   Words used in the assessment of the full text:
```

```
     positive words: [(['really'], 0.2), (['more'], 0.5)]
     negative words: [(['small'], -0.25)]
Missing words in sentiment assessment of text with stop words removed:
{'really', 'more'}

Review text examined:
This product so far has not disappointed. My children love to use it and
I like the ability to monitor control what content they see with ease.
Sentiment measures:
   Polarity is 0.3250 and Polarity with stop words removed is -0.0500
   Words used in the assessment of the full text:
     positive words: [(['far'], 0.1), (['not', 'disappointed'], 0.375),
(['love'], 0.5)]
Missing words in sentiment assessment of text with stop words removed:
{'not'}
```

- After investigation, the stop words that are excluded from the list are:
  'never', 'most', 'more', 'many', 'very', 'not', 'much', 'less', 'few', 'first', 'due', 'other', 'no', 'really', 'least', 'top', 'above', 'down', 'full', 'last', 'only', 'enough', 'such', 'back', 'only', 'several', 'next', 'mostly', 'own', 'whole', 'behind', 'becoming', 'former', 'third', 'same', 'various', 'latterly', 'latter', 'serious', 'further', 'formerly', 'ten', 'empty'
  The stop words excluding the ones above-mentioned are removed by replacing them with spaces.

- All whitespaces are then removed from the text by first splitting the text into a list of words using space as separator and then re-joining them into a text string.

- All punctuations and proper nouns are removed from the text by first applying the language analysis model to the text and then only retaining the text of tokens that are not punctuations and not proper nouns, using the punctuation checking method and checking the pos tags of the tokens in the language model. Proper nouns are removed to avoid proper nouns such as “smart” in “Jenny Smart” or “good” of “Good Prospect Co.” are included as assessed words in the sentiment analysis.

- Change the whole text into a sting of lower case letters by applying the lower() method to the string.

5. Results and Evaluation

5.1 Eight analysed review texts are shown here. For each text, its sentiment prediction as a whole and individual assessed words with polarity are evaluated. For ease of reading, the assessed words in the texts are highlighted in red if they are positive, in green if they are neutral, and in blue if they are negative.

TEXT A

- Result of Sentiment Analysis on Review Text A

```
Review rating: 2
Review text examined:
My 5 year old dropped this from a standing position (about a foot off the
ground) and it busted. I had to use the BB warranty to replace it the day
after purchase for a fall that the tablet should've been able to handle
```

```
that is designed for kids. I'm happy with the apps, but definitely not
thrilled with the lack of protection that is designed for kids.
Sentiment measures:
   Polarity is 0.2750
   Words used in the assessment of the full text:
     positive words: [(['old'], 0.1), (['able'], 0.5), (['happy'], 0.8)]
     negative words: [(['definitely', 'not', 'thrilled'], -0.3)]
```

- Evaluation of Sentiment Analysis on Review Text A

- The polarity given is higher than expected but should be around neutral since there are positive sentiment and negative sentiment expressed.
- While the words “definitely not thrilled” are correctly taken care of together as one assessment, the model did not recognise that “able” should be assessed as “should’ve been able” in other instance, which should be in negative polarity.
- On the other hand, “old”, goes with the word “age”, is a stop word here, and so it should not be included, but “lack of”, which expresses deficiency, should be included in the assessment to express a negative sentiment.

TEXT B

- Result of Sentiment Analysis on Review Text B

```
Review rating: 4
Review text examined:
My previous generation Kindle Paperwhite got scratched a little on the
screen, so I debated about picking up a new Kindle, but it was worth the
$120. The 7th generation Paperwhite has twice the memory, runs a bit
faster so it's noticeably quicker in response--making highlighting and
annotating easier, and the uneven screen lighting problem (the older
Kindle had "shadow" spots on the bottom edge of the screen) has been
solved.My only complaint is that Amazon continues to put ads on the
Kindle, which you have to pay $20 to get rid of. It's almost necessary
because I use the Kindle to view books in church. Very inappropriate when
the ad for a murder, crime, or romance novel pops up as a screensaver.
Sentiment measures:
   Polarity is -0.0051
   Words used in the assessment of the full text:
     positive words: [(['new'], 0.13636363636363635), (['worth'], 0.3),
(['older'], 0.16666666666666666), (['very'], 0.2)]
      neutral words: [(['only'], 0.0), (['necessary'], 0.0)]
     negative words: [(['previous'], -0.16666666666666666), (['little'],
-0.1875), (['uneven'], -0.2), (['complaint'], -0.3)]
```

- Evaluation of Sentiment Analysis on Review Text B

- Although some of the negative words are related to the Kindle previously got but not the one just acquired, it was actually a negative sentiment in the writer. So, the sentiment polarity of the text is accurate though not the whole of the sentiment was generated by the Amazon product just acquired.

- In the positive words assessed, “older” was not expressing positive sentiment here, and “very” actually went together with “inappropriate” to obtain a negative polarity rather.

TEXT C

- Result of Sentiment Analysis on Review Text C

```
Review rating: 1
Review text examined:
A lot of apps are incomplete. Doesn't run apps that great and has a lot
of apps that are generic.
Sentiment measures:
   Polarity is 0.4000
   Words used in the assessment of the full text:
     positive words: [(['great'], 0.8)]
      neutral words: [(['generic'], 0.0)]
```

- Evaluation of Sentiment Analysis on Review Text C

- The assessment of the text as positive is highly inappropriate.
- It was clear that “great” here did not convey a positive sentiment but it rather should be interpreted as “doesn’t run that great”.
- Moreover, “incomplete” should be included in the assessment.

TEXT D

- Result of Sentiment Analysis on Review Text D

```
Review rating: 2
Review text examined:
My kids use it for playing games only. Memory is too limited. If you want
to expand memory you have to buy special memory card from Amazon. Regular
cards won't work. I'll like the app free time to help protect the kids
from what they watch and play but it is difficult and time consuming to
get updates for the games they play. Amazon does not have the best user
friendly software for their tablets. They do what we get it for,
entertain the kids for awhile.
Sentiment measures:
   Polarity is 0.1951
   Words used in the assessment of the full text:
     positive words: [(['special'], 0.35714285714285715), (['free'],
0.4), (['best'], 1.0), (['friendly'], 0.375)]
      neutral words: [(['only'], 0.0), (['regular'], 0.0)]
     negative words: [(['limited'], -0.07142857142857142),
(['difficult'], -0.5)]
```

- Evaluation of Sentiment Analysis on Review Text D

- As the review text shows disappointment and frustration, the assessment of the text as positive is inappropriate.

- "special", in contrast with later expression "regular won't work", in this text conveys negative sentiment, i.e., hard to find or difficult to acquire, rather than a positive one in the context.
- "best" and "friendly" are positives, but a lack of them, conveyed by "does not have" previously, should change them into negatives.
- Moreover, "incomplete" should be included in the assessment.

TEXT E

- Result of Sentiment Analysis on Review Text E

```
Review rating: 4
Review text examined:
It is a tablet. Downloads kids apps fast. Only complaint is when I shut
it down the kid's user screen goes away. I do pull it back up but both my
child and I get annoyed. Some regular apps I have not been able to remove
from the regular screen. However, I would recommend.
Sentiment measures:
   Polarity is -0.0173
   Words used in the assessment of the full text:
     positive words: [(['fast'], 0.2), (['able'], 0.5)]
      neutral words: [(['only'], 0.0), (['back'], 0.0), (['regular'],
0.0), (['regular'], 0.0)]
     negative words: [(['complaint'], -0.3), (['down'], -
0.15555555555555559), (['annoyed'], -0.4)]
```

- Evaluation of Sentiment Analysis on Review Text E

- Although the review rating is 4, the sentiment prediction of the review text to be a bit negative is deemed correct as except fast download, the other parts of the text is expressed in negative way.
- The positive token "able" should be assessed with the negation, i.e., "not been able".

TEXT F

- Result of Sentiment Analysis on Review Text F

```
Review rating: 4
Review text examined:
affordable for the kids and quality is good not like other cheap tablet,
amazon always make a good product.
Sentiment measures:
   Polarity is 0.4187
   Words used in the assessment of the full text:
     positive words: [(['good'], 0.7), (['cheap'], 0.4), (['good'], 0.7)]
     negative words: [(['other'], -0.125)]
```

- Evaluation of Sentiment Analysis on Review Text F

- The overall sentiment prediction is deemed appropriate.

TEXT G

- Result of Sentiment Analysis on Review Text G

```
Review rating: 4
Review text examined:
Nice screen, got it at discounted price for my kid, replaced a smaller
one he broke. Tip add a child proff cover. The migration executive
experience amazing! Picked up from cloud and i was good to go!
Sentiment measures:
   Polarity is 0.4750
   Words used in the assessment of the full text:
     positive words: [(['nice'], 0.6), (['amazing'], 0.6000000000000001),
(['good'], 0.7)]
      neutral words: [(['smaller'], 0.0)]
```

- Evaluation of Sentiment Analysis on Review Text G

- The overall sentiment prediction is deemed appropriate.
- Attention should be paid on the mis-spelt word “proff”, which should be “proof”. It is unknown whether the corrected word would make a different prediction.

TEXT H

- Result of Sentiment Analysis on Review Text H

```
Review rating: 5
Review text examined:
Despite not being able to get one of these from Amazon until late June
due to apparent strong demand, Best Buy was able to get one to me on the
day of its release, April 27. Well done Best Buy! Having owned several
previous Kindles, including the most recent Paperwhite model, I was
unprepared for how small and light this new Oasis model is -- reading
about it is one thing, but actually holding one in your hand is something
else entirely. The Oasis and its battery-charging cover together weigh
only one ounce more than the latest Paperwhite with no case. The Oasis
_without_ its cover feels impossibly light at 4.6 ounces. Size-wise, the
Oasis is nearly a square shape at about 5" wide by 5.5" tall. In fact,
the shape is so different from traditional Kindle designs that I had to
compare it next to my Paperwhite to verify that the screens are actually
the same size (they are). It's the first Kindle I think I could easily
fit into a jacket pocket without any effort, For frequent travelers like
me, this is a key reason to buy one of these, as it's extremely compact
and easy to pack in a briefcase or purse. The screen is the first front-
lit Kindle I've ever seen where you simply cannot see any trace of the
LEDs from any angle. No dark spots, no bright spots...just a completely
uniform illumination from edge to edge in both directions. The "page"
also appears to be nearly flush with the top glass surface, helping
achieve the illusion of printed paper. It's sharp and clear, better than
any Kindle I've used. The reader can be easily separated from the battery
case by simply pulling them apart, since they are held together with
magnets. After doing so, this is where the Amazon engineering team's
achievements really become clear, as you just won't believe how
lightweight this thing is. It's easy to imagine holding onto it for hours
without any arm fatigue, and switching from hand to hand is a simple
matter of rotating it 180 degrees when switching. The display instantly
rotates, and the page turn buttons also "follow", with the top button
always used to advance the page, and the bottom button always used to go
backward. You can also use the screen to change pages as you always have
with earlier touch screen models. The battery on the reader has a lower
```

capacity than the cover's battery, so it runs down faster if you don't have it connected to the case. It's very easy to snap the reader back into its cover, which allows it to recharge while you continue to read. The power button and charge port are located in the upper top-right corner, more or less opposite from where they are on the Paperwhite. I find this location to be less awkward, as it's easier to reach on the top than it is on the bottom edge. The back surface of the reader (not the case) seems to easily attract fingerprints, and they are not so easy to clean off. You'd think that Amazon would have figured out a way to prevent this by now. It's a minor annoyance. Lots of reviewers in the tech press have complained loudly about this model's high price relative to other Kindle models (without ads, it's $310 vs. $139 for an ad-less Paperwhite). Is this a lot of money? Perhaps, but if you are a book addict and value an extremely lightweight reader that's easy on both your arms and your eyes, then I personally think it's worth it. A lot of thought and effort clearly went into designing and building this unit, and the price reflects that. Is it for everyone? Probably not, but there is obviously a market for it when you consider how many months you'll have to wait to get one on Amazon's website. If you are a serious reader, this device deserves a very close look -- I'm definitely keeping mine!
Sentiment measures:
Polarity is 0.1275
Words used in the assessment of the full text:
positive words: [(['able'], 0.5), (['apparent'], 0.05), (['strong'], 0.4333333333333333), (['able'], 0.5), (['most'], 0.5), (['light'], 0.4), (['new'], 0.13636363636363635), (['more'], 0.5), (['latest'], 0.5), (['impossibly', 'light'], 0.4), (['wise'], 0.7), (['nearly'], 0.1), (['first'], 0.25), (['easily', 'fit'], 0.4), (['frequent'], 0.1), (['easy'], 0.43333333333333335), (['first'], 0.25), (['no', 'dark'], 0.075), (['completely'], 0.1), (['nearly'], 0.1), (['top'], 0.5), (['clear'], 0.10000000000000002), (['better'], 0.5), (['easily'], 0.43333333333333335), (['really'], 0.2), (['clear'], 0.10000000000000002), (['easy'], 0.43333333333333335), (['top'], 0.5), (['very', 'easy'], 0.5633333333333334), (['top'], 0.5), (['right'], 0.2857142857142857), (['more'], 0.5), (['top'], 0.5), (['easily'], 0.43333333333333335), (['easy'], 0.43333333333333335), (['clean'], 0.3666666666666667), (['loudly'], 0.1), (['high'], 0.16), (['easy'], 0.43333333333333335), (['worth'], 0.3), (['clearly'], 0.10000000000000002), (['many'], 0.5), (['very'], 0.2)]
neutral words: [(['several'], 0.0), (['recent'], 0.0), (['actually'], 0.0), (['entirely'], 0.0), (['only'], 0.0), (['different'], 0.0), (['traditional'], 0.0), (['next'], 0.0), (['actually'], 0.0), (['same'], 0.0), (['key'], 0.0), (['simply'], 0.0), (['simply'], 0.0), (['simple'], 0.0), (['instantly'], 0.0), (['earlier'], 0.0), (['back'], 0.0), (['upper'], 0.0), (['opposite'], 0.0), (['back'], 0.0), (['relative'], 0.0), (['personally'], 0.0), (['obviously'], 0.0), (['definitely'], 0.0)]
negative words: [(['late'], -0.3), (['due'], -0.125), (['previous'], -0.16666666666666666), (['small'], -0.25), (['wide'], -0.1), (['extremely'], -0.125), (['no', 'bright'], -0.35000000000000003), (['sharp'], -0.125), (['down'], -0.15555555555555559), (['less'], -0.16666666666666666), (['less'], -0.16666666666666666), (['awkward'], -0.6), (['minor'], -0.05), (['complained'], -0.3), (['other'], -0.125), (['less'], -0.16666666666666666), (['extremely'], -0.125), (['serious'], -0.3333333333333333)]

- Evaluation of Sentiment Analysis on Review Text H

- It is correctly assessed as conveying a positive sentiment although a higher score is expected. Since the overall polarity is the average polarity of all the assessed words, here, the intensity of having a much higher score in positive words than in negative words is diluted by a large number of neutral words. If the neutral words does not involve in calculating the overall polarity, the overall polarity would be +0.177664. If there are 10 fewer positive words but 10 more neutral words,
- In this context of describing an LED screen, “No dark spots” and “no bright spots” should both be considered positive. But, the “No dark” is positive and “no bright” is negative here. Similarly, “small” in “small and light” should be positive rather than negative here.
- The occurrence of “top” and “right” here, accounted for about +2.3, actually do not express any sentiment but only indicate the positions of the button and screen.
- “wise” here is actually “size-wise”, not conveying and sentiment.
- “first” here indicates the order of acquiring, not the rank, and thus conveys no sentiment.
- The “sharp”, here describing the display quality, conveys a positive sentiment, but not a negative one, as in other cases.
- Similarly, “less” in “ad-less”, “extremely” in “extremely lightweight”, and “serious” in “serious reader” are all carrying positive sentiment, not negative.
- It is concluded that, due to a different discourse, many words carry sentiment in a totally different direction, which the language model is unable to differentiate.

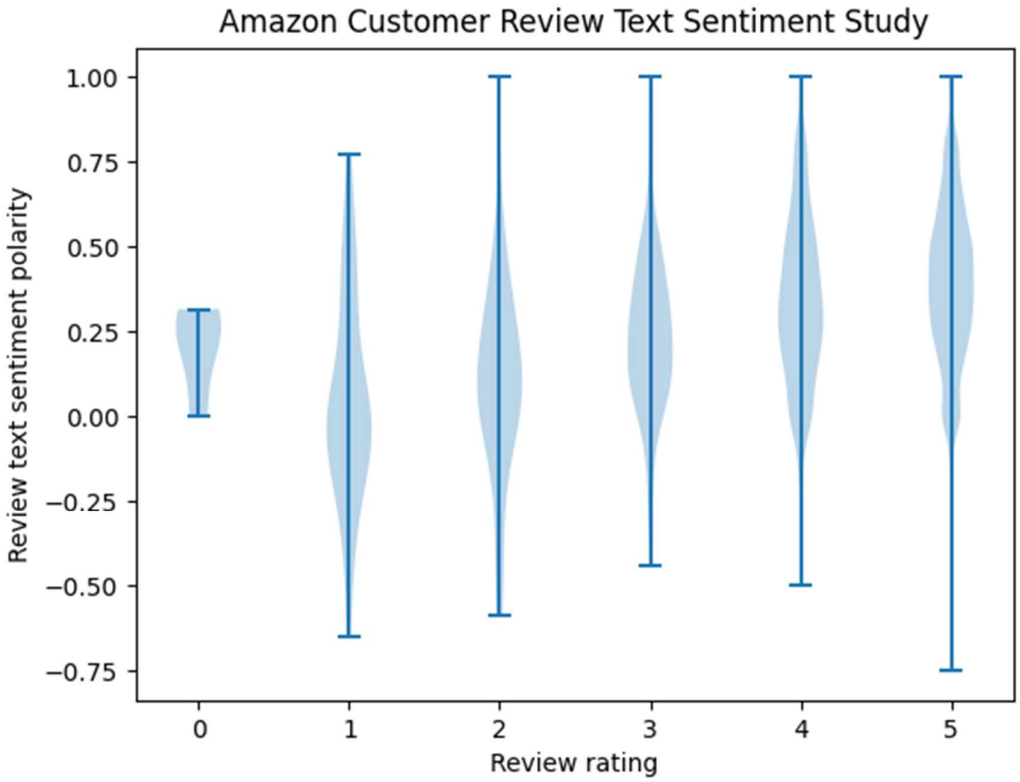
5.2 Sentiment polarities of two review texts having high similarity with a reference base review texts are shown below:

```
The following two review texts are having a similar score of
0.955316964115265

Review Text 1 (sentiment polarity = 0.4166666666666667) :
I love these things and this is the best iteration yet! If you like to
read, there is nothing better! Not only are the books cheaper, you can
get almost anything you want to read instantly. If you like to travel, it
is a must have - you can bring with as many books as you want (I take
mine backpacking all the time) and never have to worry about taking up
valuable space in your luggage.

Review Text 2 (sentiment polarity = 0.535) :
Amazon Echo is the best thing you do not know you need. It is nice to
have Prime along with it. The speaker fills the house with nice sets on
request. Go get one. And you will soon find you want one in other rooms.
Dots do nice here.
_____________________________________________________________________

The following two review texts are having a similar score of
0.9536585705919324
```

```
Review Text 1 (sentiment polarity = 0.4166666666666667) :
I love these things and this is the best iteration yet! If you like to
read, there is nothing better! Not only are the books cheaper, you can
get almost anything you want to read instantly. If you like to travel, it
is a must have - you can bring with as many books as you want (I take
mine backpacking all the time) and never have to worry about taking up
valuable space in your luggage.

Review Text 2 (sentiment polarity = 0.1774621212121212) :
The Amzon Echo is like a dream come true. Everyone in my house feels like
we live in some future movie. We use it everyday to get our news,
weather, cooking timer, control our lights, control our AC and listen to
music over Spotify, all with just our voice. It's useful to get quick
wiki info or even just for a joke. The kids love to play paper rock and
scissors with her. My only regret is I don't have another!
```

Having a high similarity between two texts does not mean that they have similar sentiment predictions. This means that the two models are trained to look for different aspects in a text.

5.3 The spread of polarity of review texts is plotted against the review rating, showing whether the sentiment predictions go along with the review rating. The result of applying the model to 3500 sample review texts is shown below.



- As shown in the graph, most of the sentiment prediction match the expectation inferred from the associated review ratings.

- However, the graph also shows that there are a small number of sentiment predictions deviating from the expectation. For example, the polarity of the review text by a person who rated 1 in the review is +0.75 and a person rated 5 in the review wrote a review with a polarity of -0.75 are unimaginable.
- Moreover, that, except 1, all other rating groups have review texts achieving the highest polarity of +1.0 shows also that the model has rooms for improvement.

6. Conclusion

6.1 The strengths of this language model in sentiment prediction are:

- simple and easy to use
- fast and efficient
- able to handle a group of words as one unit such as “definitely not thrilled” has polarity of -0.3 as a whole
- able to make sentiment prediction correctly in most cases

6.2 The weaknesses of this language model in sentiment prediction are:

- neutral words do not contribute to the sum of polarity but would lower the intensity level of the polarity by including them in the calculation of the overall polarity, which is an average of all the polarity measures
- unable to detect negation if it is not put immediately before the assessed word, e.g., in the case of “it is not that great as advertised”, or when the whole sentence is negated by the word not in the verb phrase
- unable to detect a change of sentiment polarity when a word is used in different context, e.g., ‘no bright’ in ‘no bright future’ conveys negative sentiment but in ‘no bright spots on an LED monitor” conveys positive sentiment
- unable to associate different sentiment polarities to a word with different meanings, e.g., ‘top’ in ‘top quality’ and in ‘top drawer’ has different meaning, the former conveys positive sentiment but the latter does not
- the model could not handle words with wrong spelling

6.3 The limitation of this language model is that it could only assess the sentiment of the writer of text as a whole but could not differentiate whether the sentiment is caused by or directed to a specific object, such as in this case it is not clear whether the sentiment shown in the review texts is related to the Amazon product just acquired or not.