NLP - Assignment 3

Sentiment Analysis of Amazon Product Reviews

The given data set of amazon contains the review of baby products which has nearly 160k of reviews from the year 2000 to 2014. The goal of the project is to perform sentiment analysis on all the review text and classify the positive and negative reviews.

Example of one review in the baby.txt:

reviewerID:A1HK2FQW6KXQB2
asin:097293751X
reviewerName:Amanda Johnsen "Amanda E. Johnsen"
helpful:[0, 0]
reviewText:Perfect for new parents. We were able to keep track of baby's feeding, sleep and diaper change schedule for the first two and a half months of her life. Made life easier when the doctor would ask questions about habits because we had it all right there!
overall:5.0
summary:Awesine
unixReviewTime:1373932800
reviewTime:07 16, 2013

```
In [4]:  rawText[:100000]

Out[4]:  'reviewerID:A1HK2FQW6KXQB2\nasin:097293751X\nreviewerName:Amanda Johnsen "Amanda E. Johnsen"\nhelpful:[0, 0]\nreviewText:Perf
         ect for new parents. We were able to keep track of baby\'s feeding, sleep and diaper change schedule for the first two and a
         half months of her life. Made life easier when the doctor would ask questions about habits because we had it all right there!
         \noverall:5.0\nsummary:Awesine\nunixReviewTime:1373932800\nreviewTime:07 16, 2013\n\nreviewerID:A19K65VY14D13R\nasin:09729375
         1X\nreviewerName:angela\nhelpful:[0, 0]\nreviewText:This book is such a life saver.  It has been so helpful to be able to go
         back to track trends, answer pediatrician questions, or communicate with each other when you are up at different times of the
         night with a newborn.  I think it is one of those things that everyone should be required to have before they leave the hospi
         tal.  We went through all the pages of the newborn version, then moved to the infant version, and will finish up the second i
         nfant book (third total) right as our baby turns 1.  See other things that are must haves for baby at [...]\noverall:5.0\nsum
         mary:Should be required for all new parents!\nunixReviewTime:1372464000\nreviewTime:06 29, 2013\n\nreviewerID:A2LL1TGG90977E
         \nasin:097293751X\nreviewerName:Carter\nhelpful:[0, 0]\nreviewText:Helps me know exactly how my babies day has gone with my m
         other in law watching him while I go to work.  It also has a section for her to write notes and let me know anything she may
         need.  I couldn\'t be happier with this book.\noverall:5.0\nsummary:Grandmother watching baby\nunixReviewTime:1395187200\nrev
         iewTime:03 19, 2014\n\nreviewerID:A5G19RYX8599E\nasin:097293751X\nreviewerName:cfpurplerose\nhelpful:[0, 0]\nreviewText:I bou
         ght this a few times for my older son and have bought it again for my newborn. This is super easy to use and helps me keep tr
         ack of his daily routine. When he started going to the sitter when I went back to work, it helped me know how his day went to
         better prepare me for how the evening would most likely go. When he was sick, it help me keep track of how many diapers a day
         he was producing to make sure he was getting dehydrated. The note sections to the side and bottom are useful too because his
         sitter writes in small notes about whether or not he liked his lunch or if the playtime included going for a walk, etc.Excell
         ent for moms who are wanting to keep track of their kids daily routine even though they are at work. Excellent for dads to ke
```

The process of classification consists of two major tasks

1) Data Pre-processing
2) Sentiment analysis using classifiers and word features.

Data Pre-processing:

The data preprocessing tasks has two main steps which includes data extraction and data cleaning.

a) **Data extraction:**

The data extraction step is used to extract only the necessary data which is used for sentiment analysis. Only one year (year = 2013) is used in this project for classification. So, all the reviews which are from 2013 are extracted separately and used for further extraction.

For sentiment analysis, only the review text of the reviews will be used. So, all the review text of the particular year will be extracted for further processing.

Steps for data extraction:

1) Load the text file and store it as a string
2) Split the whole string with '\n\n' as each review is separated by two new lines
3) On splitting the whole line, all the reviews can be stored as a list
4) Each element in the split can be split again to get all the attributes of review content
5) For the list of review content, we can retrieve check the review time for any given year and filter the review text of that particular year into reviews list which can be used for sentiment analysis and classification
6) On extracting, total of 62223 reviews were selected which belongs to year 2013.

```
In [10]: filteredReviewTexts = filterReviewTextsByYear('2013')

In [11]: filteredReviewTexts[:5]

Out[11]: ["Perfect for new parents. We were able to keep track of baby's feeding, sleep and diaper change schedule for the first two and
a half months of her life. Made life easier when the doctor would ask questions about habits because we had it all right ther
e!",
 'This book is such a life saver.  It has been so helpful to be able to go back to track trends, answer pediatrician questions,
or communicate with each other when you are up at different times of the night with a newborn.  I think it is one of those thin
gs that everyone should be required to have before they leave the hospital.  We went through all the pages of the newborn versi
on, then moved to the infant version, and will finish up the second infant book (third total) right as our baby turns 1.  See o
ther things that are must haves for baby at [...]',
 'I bought this a few times for my older son and have bought it again for my newborn. This is super easy to use and helps me ke
ep track of his daily routine. When he started going to the sitter when I went back to work, it helped me know how his day went
to better prepare me for how the evening would most likely go. When he was sick, it help me keep track of how many diapers a da
y he was producing to make sure he was getting dehydrated. The note sections to the side and bottom are useful too because his
sitter writes in small notes about whether or not he liked his lunch or if the playtime included going for a walk, etc.Excellen
t for moms who are wanting to keep track of their kids daily routine even though they are at work. Excellent for dads to keep t
rack as my husband can quickly forget what time he fed our son. LOL',
 'My 3 month old son spend half of his days with my mother and half with a neighbor while I worked.  I had them track his activ
ity (loosely) in this which allowed me to get an idea of how his schedule was developing and how much milk he was eating.  It w
as the best way to have some cohesion in his life while I was at work.',
 "This book is perfect!  I'm a first time new mom, and this book made it so easy to keep track of feedings, diaper changes, sle
ep.  Definitely would recommend this for new moms.  Plus it's small enough that I throw in the diaper back for doctor visits."]
```

b) Data cleaning:

Data cleaning is the main process of data preprocessing as it will improve the accuracy of the classification only if we have the proper cleaned data.

Steps involved in data cleaning:
1) Tokenize the reviews which are extracted in the previous step for cleaning the data.

```
In [13]: tokenizedReviewsList = tokenizeFilteredReviews()

In [14]: tokenizedReviewsList[0]

Out[14]: ['Perfect',
         'for',
         'new',
         'parents',
         'We',
         'were',
         'able',
         'to',
         'keep',
         'track',
         'of',
         "baby's",
         'feeding',
         'sleep',
         'and',
         'diaper',
         'change',
         'schedule',
         'for',
         'the',
```

Tokenized words

2) All these tokenized words must be converted into lower case for making the further process case-insensitive
3) The non-alphabets from the token list must be removed as they will not have any meaning or give any value for our next steps
4) The stop words must be filtered out from the list of tokens so that only meaning full words are used for classification and sentiment analysis

```
In [21]: preProcessedReviewTokens[:10]

Out[21]: [['perfect',
          'new',
          'parents',
          'able',
          'keep',
          'track',
          "baby's",
          'feeding',
          'sleep',
          'diaper',
          'change',
          'schedule',
          'first',
          'two',
          'half',
          'months',
          'life',
          'made',
          'life',
          'easier'
```

Preprocessed tokens

Part 2: Sentiment analysis and Classification

The sentence polarity corpus can be loaded for training. The entries which are loaded from the sentence polarity are taken as tuple which has the sentence and the positive/negative category

```
In [23]: sentences[:10]

Out[23]: [(['simplistic', ',', 'silly', 'and', 'tedious', '.'], 'neg'),
          (["it's",
           'so',
           'laddish',
           'and',
           'juvenile',
           ',',
           'only',
           'teenage',
           'boys',
           'could',
           'possibly',
           'find',
           'it',
           'funny',
           '.'],
          'neg'),
          (['exploitative',
           'and',
           'largely'
```

Word features:

The sentences that obtained from sentence polarity corpus can be shuffled randomly for getting better representation of both test and training sets. As these sentences are not clean, they are cleaned by removing stop words and non-alphabetic characters in the sentences. After cleaning, we take only the 2000 most used as the features for classification

```
In [26]: wordFeatures[:100]

Out[26]: ['film',
          'movie',
          'one',
          'like',
          'story',
          'much',
          'even',
          'good',
          'comedy',
          'time',
          'characters',
          'little',
          'way',
          'funny',
          'make',
          'enough',
          'never',
          'makes',
          'may',
```

Bag-of-word features:

The bag of features is created for both the tokens from sentences of sentence polarity and tokens from the reviews. The bag of feature is formed in the dictionary format by "contains" keyword with the word as key and a Boolean value as value which represent whether it is present or not.

```
In [91]:  featureSet1FromSentences[:200]

Out[91]:  [({'contains(film)': False,
            'contains(movie)': False,
            'contains(one)': False,
            'contains(like)': False,
            'contains(story)': False,
            'contains(much)': False,
            'contains(even)': False,
            'contains(good)': False,
            'contains(comedy)': False,
            'contains(time)': False,
            'contains(characters)': False,
            'contains(little)': False,
            'contains(way)': False,
            'contains(funny)': True,
            'contains(make)': False,
            'contains(enough)': False,
            'contains(never)': False,
            'contains(makes)': False,
            'contains(may)': False,
```

feature set from sentences from sentence polarity

```
In [92]:  featureSet1FromReviews[:200]

            'contains(much)': False,
            'contains(even)': False,
            'contains(good)': False,
            'contains(comedy)': False,
            'contains(time)': False,
            'contains(characters)': False,
            'contains(little)': False,
            'contains(way)': False,
            'contains(funny)': False,
            'contains(make)': False,
            'contains(enough)': False,
            'contains(never)': False,
            'contains(makes)': False,
            'contains(may)': False,
            'contains(us)': False,
            'contains(work)': False,
            'contains(best)': False,
            'contains(bad)': False,
            'contains(director)': False,
            'contains(love)': False,
```

Feature set from reviews

Classification by Naïve Bayes Classifier for feature set 1

The Naïve Bayes classification is used to classify the review texts whether it is positive or negative. Before classification, the sentences from sentence polarity can be used to train and test the classifier. The 90/10 split can be taken for training and testing classifier. All the data except first thousand can be used for training and the first thousand records can be used to test the accuracy of the classification. The most informative features for the classifiers can be printed for our information.

```
In [150]: pos_sent,neg_sent,accuracy= NaiveBayesClassification()
```

```
Most Informative Features
        contains(engrossing) = True            pos : neg    =     19.4 : 1.0
           contains(beauty) = True            pos : neg    =     17.4 : 1.0
         contains(provides) = True            pos : neg    =     16.8 : 1.0
         contains(mediocre) = True            neg : pos    =     16.5 : 1.0
          contains(generic) = True            neg : pos    =     16.5 : 1.0
          contains(routine) = True            neg : pos    =     14.5 : 1.0
             contains(flat) = True            neg : pos    =     13.6 : 1.0
        contains(refreshing) = True            pos : neg    =     13.5 : 1.0
         contains(wonderful) = True            pos : neg    =     12.8 : 1.0
           contains(boring) = True            neg : pos    =     12.6 : 1.0
             contains(warm) = True            pos : neg    =     12.4 : 1.0
             contains(dull) = True            neg : pos    =     12.4 : 1.0
         contains(absorbing) = True            pos : neg    =     12.2 : 1.0
      contains(refreshingly) = True            pos : neg    =     11.5 : 1.0
         contains(touching) = True            pos : neg    =     11.4 : 1.0
            contains(stale) = True            neg : pos    =     11.1 : 1.0
           contains(stupid) = True            neg : pos    =     11.1 : 1.0
           contains(devoid) = True            neg : pos    =     10.5 : 1.0
```

Most informative features of the classifier with feature set 1

The accuracy of the classifier can be calculated using NLTK classify accuracy functionality. By training the classifier with sentences from sentence polarity, the accuracy of 73.80 is obtained

```
In [171]: print('FEATURE SET 1 - Naive Bayes Classifier')
          print('ACCURACY: %.2f' %(accuracy*100))

          FEATURE SET 1 - Naive Bayes Classifier
          ACCURACY: 73.80
```

Accuracy of classification using sentences from sentence polarity

The positive and negative reviews can be classified using this trained classifier. This can be done by using NLTK classifier.classify functionality. And the results can be stored in the text file.

```
In [168]: pos_sent[:100]
```

```
Out[168]: ["Perfect for new parents. We were able to keep track of baby's feeding, sleep and diaper change schedule for the first two a
          nd a half months of her life. Made life easier when the doctor would ask questions about habits because we had it all right t
          here!",
           'I bought this a few times for my older son and have bought it again for my newborn. This is super easy to use and helps me
          keep track of his daily routine. When he started going to the sitter when I went back to work, it helped me know how his day
          went to better prepare me for how the evening would most likely go. When he was sick, it help me keep track of how many diape
          rs a day he was producing to make sure he was getting dehydrated. The note sections to the side and bottom are useful too bec
          ause his sitter writes in small notes about whether or not he liked his lunch or if the playtime included going for a walk, e
          tc.Excellent for moms who are wanting to keep track of their kids daily routine even though they are at work. Excellent for d
          ads to keep track as my husband can quickly forget what time he fed our son. LOL',
           'My 3 month old son spend half of his days with my mother and half with a neighbor while I worked.  I had them track his act
          ivity (loosely) in this which allowed me to get an idea of how his schedule was developing and how much milk he was eating.
          It was the best way to have some cohesion in his life while I was at work.',
           "This book is perfect!  I'm a first time new mom, and this book made it so easy to keep track of feedings, diaper changes, s
          leep.  Definitely would recommend this for new moms.  Plus it's small enough that I throw in the diaper back for doctor visit
          s.",
           "The Baby Tracker brand books are the absolute best of the trackers that are available. You can track naps, feedings, diaper
          changes, activities, and notes. They're easy to use and have a nice clean layout.I started using the Newborn Baby Tracker, wh
          ich tracks nights and days. It gave me a nice way to track feedings and note important milestones. Once I used up the entire
```

Classified positive sentences

```
In [169]: neg_sent[:100]

Out[169]: ['This book is such a life saver.  It has been so helpful to be able to go back to track trends, answer pediatrician question
s, or communicate with each other when you are up at different times of the night with a newborn.  I think it is one of those
things that everyone should be required to have before they leave the hospital.  We went through all the pages of the newborn
version, then moved to the infant version, and will finish up the second infant book (third total) right as our baby turns 1.
See other things that are must haves for baby at [...]',
 "I use this so that our babysitter (grandma) can keep track of what goes on during the day.  We've recorded so many mileston
es since we received this.  It is a 6 month diary and has room for notes.  I wish the hours started around 6 am and went unti
l 8 pm, but we just write over certain times to suit our purposes. The nice thing is that if we had a rough night and we are
super tired, we have something to refer to if we forgot when we last fed the baby or when he woke up. I also like that the fr
ont page has a place for emergency comments and a consent form in case your child needs medical care.  It also has a place at
the back of the book for immunizations.  I'll definitely get another one in 6 months when we finish this book.",
 "I like this log, but think it would work better with clearer AM & PM sections. Each page is 12 hours so you really need two
pages a day if your baby feeds or wets a lot in the early morning hours between midnight and 7am. We're cramming those in to
the 2 blank spaces above 7am right now.",
 "My wife and I have a six month old baby boy and around the 4-month mark my wife decided that she would return to work inste
ad of being a stay-at-home mom. We hired an in-home nanny to care for our little boy and the arrangement has worked quite wel
l ever since. Shortly after starting the arrangement we realized that we would need some sort of journal to track our little
boy's activities while he was with the nanny and we were working. We used a plain notebook for a period of weeks until we stu
mbled on the Baby Tracker - Daily Childcare Journal.--------------------LAYOUT AND USE:The Baby Tracker is really an excellen
```

Classified negative reviews

A	B
Positive reviews	**Negative reviews**
Perfect for new parents. We were able to keep track of baby's feeding, sleep and diaper change schedule for the first two and a half months of her life. Made life easier when the doctor would ask questions about habits because we had it all right there!	This book is such a life saver. It has been so helpful to be able to go back to track trends, answer pediatrician questions, or communicate with each other when you are up at different times of the night with a newborn. I think it is one of those things that everyone should be required to have before they leave the hospital. We went through all the pages of the newborn version, then moved to the infant version, and will finish up the second infant book (third total) right as our baby turns 1. See other things that are must haves for baby at [...]
I bought this a few times for my older son and have bought it again for my newborn. This is super easy to use and helps me keep track of his daily routine. When he started going to the sitter when I went back to work, it helped me know how his day went to better prepare me for how the evening would most likely go. When he was sick, it help me keep track of how many diapers a day he was producing to make sure he was getting dehydrated. The note sections to the side and bottom are useful too because his sitter writes in small notes about whether or not he liked his lunch or if the playtime included going for a walk, etc.Excellent for moms who are wanting to keep track of their kids daily routine even though they are at work. Excellent for dads to keep track as my husband can quickly forget what time he fed our son. LOL	I use this so that our babysitter (grandma) can keep track of what goes on during the day. We've recorded so many milestones since we received this. It is a 6 month diary and has room for notes. I wish the hours started around 6 am and went until 8 pm, but we just write over certain times to suit our purposes. The nice thing is that if we had a rough night and we are super tired, we have something to refer to if we forgot when we last fed the baby or when he woke up. I also like that the front page has a place for emergency comments and a consent form in case your child needs medical care. It also has a place at the back of the book for immunizations. I'll definitely get another one in 6 months when we finish this book.
My 3 month old son spend half of his days with my mother and half with a neighbor while I worked. I had them track his activity (loosely) in this which allowed me to get an idea of how his schedule was developing and how much milk he was eating. It was the best way to have some cohesion in his life while I was at work.	I like this log, but think it would work better with clearer AM & PM sections. Each page is 12 hours so you really need two pages a day if your baby feeds or wets a lot in the early morning hours between midnight and 7am. We're cramming those in to the 2 blank spaces above 7am right now.
This book is perfect! I'm a first time new mom, and this book made it so easy to keep track of feedings, diaper changes, sleep. Definitely would recommend this for new moms. Plus it's small enough that I throw in the diaper back for doctor visits.	My wife and I have a six month old baby boy and around the 4-month mark my wife decided that she would return to work instead of being a stay-at-home mom. We hired an in-home nanny to care for our little boy and the arrangement has worked quite well ever since. Shortly after starting the arrangement we realized that we would need some sort of journal to track our little boy's activities while he was with the nanny and we were working. We used a plain notebook for a period of weeks until we stumbled on the Baby Tracker - Daily Childcare Journal.--------------------LAYOUT AND USE:The Baby Tracker is really an excellent idea. The journal pages are clearly divided into columns

Positive vs negative sentence table

Kindly refer to the attached/submitted excel file named – 'classification1.xlsx' to see the whole list of positive and negative sentences.

classification1.xlsx

Feature set 2 using subjectivity lexicon:

The subjectivity lexicon can be loaded from the file 'subjclueslen1-HLTEMNPOS.tff'. This can be used to form the feature set 2 for classification. The diction can be formed by parsing through the lexicon with word as key and its strength, pos tag, stem information and polarity as the value.

Score count features in subjectivity:

The bag of features can be defined with the subjectivity. The dictionary of feature can be created as 'contains' keyword with words from both sentence polarity and reviews as keys and Boolean value with denotes its presence in the word dictionary respectively. If the word present in the dictionary that we formed with the subjectivity lexicon, then we can calculate and add the positive and negatives of the features.

```
In [33]: featureSetFromSentences2[:10]

Out[33]: [({'contains(film)': False,
          'contains(movie)': False,
          'contains(one)': False,
          'contains(like)': False,
          'contains(story)': False,
          'contains(much)': False,
          'contains(even)': False,
          'contains(good)': False,
          'contains(comedy)': False,
          'contains(time)': False,
          'contains(characters)': False,
          'contains(little)': False,
          'contains(way)': False,
          'contains(funny)': False,
          'contains(make)': False,
          'contains(enough)': False,
          'contains(never)': False,
          'contains(makes)': False,
          'contains(may)': False,
          'contains(us)': False,
```

Feature set 2 from sentences

```
In [35]: featureSetFromReviews2[:10]

Out[35]: [{'contains(film)': False,
          'positivecount': 6,
          'negativecount': 0,
          'contains(movie)': False,
          'contains(one)': False,
          'contains(like)': False,
          'contains(story)': False,
          'contains(much)': False,
          'contains(even)': False,
          'contains(good)': False,
          'contains(comedy)': False,
          'contains(time)': False,
          'contains(characters)': False,
          'contains(little)': False,
          'contains(way)': False,
          'contains(funny)': False,
          'contains(make)': False,
          'contains(enough)': False,
          'contains(never)': False,
          'contains(makes)': False,
```

Feature set 2 from reviews

Classification by Naïve Bayes Classifier for feature set 2

The Naïve Bayes classification is used to classify the review texts whether it is positive or negative. Before classification, the sentences from sentence polarity can be used to train and test the classifier. The 90/10 split can be taken for training and testing classifier. All the data except first thousand can be used for training and the first thousand records can be used to test the accuracy of the classification. The most informative features for the classifiers can be printed for our information.

```
In [37]:  pos_sentfor2,neg_sentfor2,accuracyfor2= NaiveBayesClassification2()

          Most Informative Features
                contains(engrossing) = True          pos : neg    =     19.7 : 1.0
                  contains(captures) = True          pos : neg    =     17.7 : 1.0
                    contains(boring) = True          neg : pos    =     17.3 : 1.0
                   contains(routine) = True          neg : pos    =     15.6 : 1.0
                  contains(mediocre) = True          neg : pos    =     15.0 : 1.0
                      contains(flat) = True          neg : pos    =     14.6 : 1.0
                contains(unexpected) = True          pos : neg    =     14.4 : 1.0
                   contains(generic) = True          neg : pos    =     14.3 : 1.0
                  contains(inventive) = True         pos : neg    =     13.7 : 1.0
                 contains(refreshing) = True         pos : neg    =     13.7 : 1.0
                      contains(dull) = True          neg : pos    =     12.8 : 1.0
                      contains(warm) = True          pos : neg    =     12.2 : 1.0
                  contains(wonderful) = True         pos : neg    =     11.8 : 1.0
               contains(refreshingly) = True         pos : neg    =     11.7 : 1.0
                   contains(delicate) = True         pos : neg    =     11.7 : 1.0
                     contains(stale) = True          neg : pos    =     11.6 : 1.0
                  contains(provides) = True          pos : neg    =     10.6 : 1.0
                      contains(ages) = True          pos : neg    =     10.4 : 1.0
                       contains(car) = True          neg : pos    =     10.3 : 1.0
                     contains(bears) = True          neg : pos    =     10.3 : 1.0
```

Most informative features of the classifier with feature set 2

The accuracy of the classifier can be calculated using NLTK classify accuracy functionality. By training the classifier with sentences from sentence polarity, the accuracy of 76.30 is obtained

```
In [38]:  print('FEATURE SET 2 - Naive Bayes Classifier')
          print('ACCURACY: %.2f' %(accuracyfor2*100))

          FEATURE SET 2 - Naive Bayes Classifier
          ACCURACY: 76.30
```

Accuracy of classification using subjectivity

The positive and negative reviews can be classified using this trained classifier. This can be done by using NLTK classifier.classify functionality. And the results can be stored in the text file.

```
In [42]: pos_sentfor2[:20]
```

```
Out[42]: ["Perfect for new parents. We were able to keep track of baby's feeding, sleep and diaper change schedule for the first two a
         nd a half months of her life. Made life easier when the doctor would ask questions about habits because we had it all right t
         here!",
          'I bought this a few times for my older son and have bought it again for my newborn. This is super easy to use and helps me
         keep track of his daily routine. When he started going to the sitter when I went back to work, it helped me know how his day
         went to better prepare me for how the evening would most likely go. When he was sick, it help me keep track of how many diape
         rs a day he was producing to make sure he was getting dehydrated. The note sections to the side and bottom are useful too bec
         ause his sitter writes in small notes about whether or not he liked his lunch or if the playtime included going for a walk, e
         tc.Excellent for moms who are wanting to keep track of their kids daily routine even though they are at work. Excellent for d
         ads to keep track as my husband can quickly forget what time he fed our son. LOL',
          'My 3 month old son spend half of his days with my mother and half with a neighbor while I worked.  I had them track his act
         ivity (loosely) in this which allowed me to get an idea of how his schedule was developing and how much milk he was eating.
         It was the best way to have some cohesion in his life while I was at work.',
          "This book is perfect!  I'm a first time new mom, and this book made it so easy to keep track of feedings, diaper changes, s
         leep.  Definitely would recommend this for new moms.  Plus it's small enough that I throw in the diaper back for doctor visit
         s.",
          "The Baby Tracker brand books are the absolute best of the trackers that are available. You can track naps, feedings, diaper
         changes, activities, and notes. They're easy to use and have a nice clean layout.I started using the Newborn Baby Tracker, wh
         ich tracks nights and days. It gave me a nice way to track feedings and note important milestones. Once I used up the entire
```

Classified positive sentences

```
In [43]: neg_sentfor2[:20]
```

```
Out[43]: ['This book is such a life saver.  It has been so helpful to be able to go back to track trends, answer pediatrician question
         s, or communicate with each other when you are up at different times of the night with a newborn.  I think it is one of those
         things that everyone should be required to have before they leave the hospital.  We went through all the pages of the newborn
         version, then moved to the infant version, and will finish up the second infant book (third total) right as our baby turns 1.
         See other things that are must haves for baby at [...]',
          "I use this so that our babysitter (grandma) can keep track of what goes on during the day.  We've recorded so many mileston
         es since we received this.  It is a 6 month diary and has room for notes.  I wish the hours started around 6 am and went unti
         l 8 pm, but we just write over certain times to suit our purposes. The nice thing is that if we had a rough night and we are
         super tired, we have something to refer to if we forgot when we last fed the baby or when he woke up. I also like that the fr
         ont page has a place for emergency comments and a consent form in case your child needs medical care.  It also has a place at
         the back of the book for immunizations.  I'll definitely get another one in 6 months when we finish this book.",
          "I like this log, but think it would work better with clearer AM & PM sections. Each page is 12 hours so you really need two
         pages a day if your baby feeds or wets a lot in the early morning hours between midnight and 7am. We're cramming those in to
         the 2 blank spaces above 7am right now.",
          "My wife and I have a six month old baby boy and around the 4-month mark my wife decided that she would return to work inste
         ad of being a stay-at-home mom. We hired an in-home nanny to care for our little boy and the arrangement has worked quite wel
         l ever since. Shortly after starting the arrangement we realized that we would need some sort of journal to track our little
         boy's activities while he was with the nanny and we were working. We used a plain notebook for a period of weeks until we stu
         mbled on the Baby Tracker - Daily Childcare Journal.--------------------LAYOUT AND USE:The Baby Tracker is really an excellen
```

Classified negative reviews

Positive reviews	Negative reviews
Perfect for new parents. We were able to keep track of baby's feeding, sleep and diaper change schedule for the first two and a half months of her life. Made life easier when the doctor would ask questions about habits because we had it all right there!	This book is such a life saver. It has been so helpful to be able to go back to track trends, answer pediatrician questions, or communicate with each other when you are up at different times of the night with a newborn. I think it is one of those things that everyone should be required to have before they leave the hospital. We went through all the pages of the newborn version, then moved to the infant version, and will finish up the second infant book (third total) right as our baby turns 1. See other things that are must haves for baby at [...]
I bought this a few times for my older son and have bought it again for my newborn. This is super easy to use and helps me keep track of his daily routine. When he started going to the sitter when I went back to work, it helped me know how his day went to better prepare me for how the evening would most likely go. When he was sick, it help me keep track of how many diapers a day he was producing to make sure he was getting dehydrated. The note sections to the side and bottom are useful too because his sitter writes in small notes about whether or not he liked his lunch or if the playtime included going for a walk, etc.Excellent for moms who are wanting to keep track of their kids daily routine even though they are at work. Excellent for dads to keep track as my husband can quickly forget what time he fed our son. LOL	I use this so that our babysitter (grandma) can keep track of what goes on during the day. We've recorded so many milestones since we received this. It is a 6 month diary and has room for notes. I wish the hours started around 6 am and went until 8 pm, but we just write over certain times to suit our purposes. The nice thing is that if we had a rough night and we are super tired, we have something to refer to if we forgot when we last fed the baby or when he woke up. I also like that the front page has a place for emergency comments and a consent form in case your child needs medical care. It also has a place at the back of the book for immunizations. I'll definitely get another one in 6 months when we finish this book.
My 3 month old son spend half of his days with my mother and half with a neighbor while I worked. I had them track his activity (loosely) in this which allowed me to get an idea of how his schedule was developing and how much milk he was eating. It was the best way to have some cohesion in his life while I was at work.	I like this log, but think it would work better with clearer AM & PM sections. Each page is 12 hours so you really need two pages a day if your baby feeds or wets a lot in the early morning hours between midnight and 7am. We're cramming those in to the 2 blank spaces above 7am right now.
This book is perfect! I'm a first time new mom, and this book made it so easy to keep track of feedings, diaper changes, sleep. Definitely would recommend this for new moms. Plus it's small enough that I throw in the diaper back for doctor visits.	My wife and I have a six month old baby boy and around the 4-month mark my wife decided that she would return to work instead of being a stay-at-home mom. We hired an in-home nanny to care for our little boy and the arrangement has worked quite well ever since. Shortly after starting the arrangement we realized that we would need some sort of journal to track our little boy's activities while he was with the nanny and we were working. We used a plain notebook for a period of weeks until we stumbled on the Baby Tracker - Daily Childcare Journal.--------------------LAYOUT AND USE:The Baby Tracker is really an excellent idea. The journal pages are clearly divided into columns for tracking feedings, nap-time, diaper changes, play-time as well as general areas for notes and tracking milestones. The legibility in the Baby Tracker is a huge improvement over our standard notebook

Positive vs negative sentence table

Kindly refer to the attached/submitted excel file named – 'classification2.xlsx' to see the whole list of positive and negative sentences

classification2.xlsx

Comparison between the results of feature set 1 and feature set 2:

The accuracy of the Naïve Bayes classifier is 73.80 percentage for the feature set 1 and it increased to 76.30 percentage while using the feature set 2. This difference depicts that the feature set 2 with subjectivity count features give a better accuracy than the feature set 1.

The negative and positive count feature which we did in the feature set 2 improved the accuracy of the classifier. Also, the feature set 1, just classify the data sets using the unigrams for classification which decreases the classification accuracy. But the classification using subjectivity not just taking words and classifying but it also takes the POS and stemming into consideration. This makes the classifier take a better decision on classification.

Count of classified positive and negative reviews:

In 2013, there are 62223 reviews in 2013.

By feature set 1:

      Positive reviews: 26382/62223   negative reviews: 35842/62223

By feature Set 2:

      Positive reviews:31523/62223   negative reviews: 30700/62223

Appendix:

Task 1: Data preprocessing

Loading data from baby.text

```
In [3]:  file = open('baby.txt')
         rawText = file.read()
         file.close()
```

```
In [4]:  rawText[:100000]
```

```
Out[4]:  'reviewerID:A1HK2FQW6KXQB2\nasin:097293751X\nreviewerName:Amanda Johnsen "Amanda E. Johnsen"\nhelpful:[0, 0]\nreviewText:Perf
         ect for new parents. We were able to keep track of baby\'s feeding, sleep and diaper change schedule for the first two and a
         half months of her life. Made life easier when the doctor would ask questions about habits because we had it all right there!
         \noverall:5.0\nsummary:Awesine\nunixReviewTime:1373932800\nreviewTime:07 16, 2013\n\nreviewerID:A19K65VY14D13R\nasin:09729375
         1X\nreviewerName:angela\nhelpful:[0, 0]\nreviewText:This book is such a life saver.  It has been so helpful to be able to go
         back to track trends, answer pediatrician questions, or communicate with each other when you are up at different times of the
         night with a newborn.  I think it is one of those things that everyone should be required to have before they leave the hospi
         tal.  We went through all the pages of the newborn version, then moved to the infant version, and will finish up the second i
         nfant book (third total) right as our baby turns 1.  See other things that are must haves for baby at [...]\noverall:5.0\nsum
         mary:Should be required for all new parents!\nunixReviewTime:1372464000\nreviewTime:06 29, 2013\n\nreviewerID:A2LL1TGG90977E
         \nasin:097293751X\nreviewerName:Carter\nhelpful:[0, 0]\nreviewText:Helps me know exactly how my babies day has gone with my m
         other in law watching him while I go to work.  It also has a section for her to write notes and let me know anything she may
         need.  I couldn\'t be happier with this book.\noverall:5.0\nsummary:Grandmother watching baby\nunixReviewTime:1395187200\nrev
         iewTime:03 19, 2014\n\nreviewerID:A5G19RYX8599E\nasin:097293751X\nreviewerName:cfpurplerose\nhelpful:[0, 0]\nreviewText:I bou
         ght this a few times for my older son and have bought it again for my newborn. This is super easy to use and helps me keep tr
         ack of his daily routine. When he started going to the sitter when I went back to work, it helped me know how his day went to
         better prepare me for how the evening would most likely go. When he was sick, it help me keep track of how many diapers a day
         he was producing to make sure he was getting dehydrated. The note sections to the side and bottom are useful too because his
         sitter writes in small notes about whether or not he liked his lunch or if the playtime included going for a walk, etc.Excell
         ent for moms who are wanting to keep track of their kids daily routine even though they are at work. Excellent for dads to ke
```

Filtering data by year:

```
In [9]:  def filterReviewTextsByYear(year):
             tlist=rawText.split('\n\n')
             filteredlist=[]
             for elem in tlist:
                 t1 = elem.split('\n')
                 if(t1[len(t1)-1].find(year)!=-1):
                     if t1[4].find('reviewText')!=-1:
                         t1[4]=t1[4].replace('reviewText:','')
                         filteredlist.append(t1[4])
                     elif t1[3].find('reviewText')!=-1:
                         t1[3]=t1[3].replace('reviewText:','')
                         filteredlist.append(t1[3])
             return filteredlist
```

```
In [10]:  filteredReviewTexts = filterReviewTextsByYear('2013')
```

```
In [11]:  filteredReviewTexts[:5]
```

```
Out[11]:  ["Perfect for new parents. We were able to keep track of baby's feeding, sleep and diaper change schedule for the first two and
          a half months of her life. Made life easier when the doctor would ask questions about habits because we had it all right ther
          e!",
          'This book is such a life saver.  It has been so helpful to be able to go back to track trends, answer pediatrician questions,
          or communicate with each other when you are up at different times of the night with a newborn.  I think it is one of those thin
          gs that everyone should be required to have before they leave the hospital.  We went through all the pages of the newborn versi
          on, then moved to the infant version, and will finish up the second infant book (third total) right as our baby turns 1.  See o
          ther things that are must haves for baby at [...]',
          'I bought this a few times for my older son and have bought it again for my newborn. This is super easy to use and helps me ke
          ep track of his daily routine. When he started going to the sitter when I went back to work, it helped me know how his day went
          to better prepare me for how the evening would most likely go. When he was sick, it help me keep track of how many diapers a da
          y he was producing to make sure he was getting dehydrated. The note sections to the side and bottom are useful too because his
          sitter writes in small notes about whether or not he liked his lunch or if the playtime included going for a walk, etc.Excellen
          t for moms who are wanting to keep track of their kids daily routine even though they are at work. Excellent for dads to keep t
          rack as my husband can quickly forget what time he fed our son. LOL',
          'My 3 month old son spend half of his days with my mother and half with a neighbor while I worked.  I had them track his activ
          ity (loosely) in this which allowed me to get an idea of how his schedule was developing and how much milk he was eating.  It w
          as the best way to have some cohesion in his life while I was at work.',
          "This book is perfect!  I'm a first time new mom, and this book made it so easy to keep track of feedings, diaper changes, sle
```

Tokenizing reviews for processing:

```
In [12]: def tokenizeFilteredReviews():
             tokenized_List = []
             PATTERN = r'''(?x) [a-zA-z]+'[a-zA-z]+ | [a-zA-z]+
                     '''
             tokenized_List=[nltk.regexp_tokenize(review_text,PATTERN) for review_text in filteredReviewTexts]
             return tokenized_List
```

```
In [13]: tokenizedReviewsList = tokenizeFilteredReviews()
```

```
In [14]: tokenizedReviewsList[0]
```

```
Out[14]: ['Perfect',
          'for',
          'new',
          'parents',
          'We',
          'were',
          'able',
          'to',
          'keep',
          'track',
          'of',
          "baby's",
          'feeding',
          'sleep',
          'and',
          'diaper',
          'change',
          'schedule',
          'for',
```

Preprocessing helper functions:

```
In [15]: def filterNonAlphabets(w):
             pattern = re.compile('^[^a-z]+$')
             if (pattern.match(w)):
                 return True
             else:
                 return False
```

```
In [16]: def toLower(list):
             lowerCaseList=[]
             for word in list:
                 lowerCaseList.append(word.lower())
             return lowerCaseList
```

```
In [17]: def filterNonAlphaFromList(List):
             nonAlphalist=[]
             for word in List:
                 if (filterNonAlphabets(word) == False):
                     nonAlphalist.append(word)
             return nonAlphalist
```

```
In [18]: def filterStopWordsFromList(List):
             listWOStopWords = []
             for word in List:
                 if word not in stopwords:
                     listWOStopWords.append(word)
             return listWOStopWords
```

Preprocessing:

```
In [19]: def preProcessing():
             preProcessedReviewTokens = []
             for innerReviewList in tokenizedReviewsList:
                 lowerCaseList= toLower(innerReviewList)
                 nonAlphaList= filterNonAlphaFromList(lowerCaseList);
                 listWOStopWords= filterStopWordsFromList(nonAlphaList)
                 preProcessedReviewTokens.append(listWOStopWords)
             return preProcessedReviewTokens
```

```
In [20]: preProcessedReviewTokens = preProcessing()
```

```
In [21]: preProcessedReviewTokens[:10]
```

```
Out[21]: [['perfect',
           'new',
           'parents',
           'able',
           'keep',
           'track',
           "baby's",
           'feeding',
           'sleep',
           'diaper',
           'change',
           'schedule',
           'first',
           'two',
           'half',
           'months',
           'life',
           'made',
           'life',
           'easier'
```

Task 2: Classification

Loading sentences:

```
In [22]: sentences = [(sentence, catg) for catg in sentence_polarity.categories() for sentence in sentence_polarity.sents(categories=catg)
```

```
In [23]: sentences[:10]
```

```
Out[23]: [(['simplistic', ',', 'silly', 'and', 'tedious', '.'], 'neg'),
          (["it's",
            'so',
            'laddish',
            'and',
            'juvenile',
            ',',
            'only',
            'teenage',
            'boys',
            'could',
            'possibly',
            'find',
            'it',
            'funny',
            '.'],
           'neg'),
          (['exploitative',
            'and',
            'largely'
```

Word feature1:

```
In [24]: def wordFeatures():
             random.shuffle(sentences)
             wordlist = [word for (sentence,catg) in sentences for word in sentence]
             stopWordsRemovedList=filterStopWordsFromList(wordlist)
             finalwordlist=filterNonAlphaFromList(stopWordsRemovedList)
             finalwords = nltk.FreqDist(finalwordlist)
             wordItems = finalwords.most_common(2000)
             wordFeatures = [ word for (word,frequency) in wordItems]
             return wordFeatures
```

```
In [25]: wordFeatures = wordFeatures()
```

```
In [26]: wordFeatures[:100]
```

```
Out[26]: ['film',
          'movie',
          'one',
          'like',
          'story',
          'much',
          'even',
          'good',
          'comedy',
          'time',
          'characters',
          'little',
          'way',
          'funny',
          'make',
          'enough',
          'never',
          'makes',
          'may',
```

Forming features for Naïve Bayes classifier:

```
In [131]: def getFeaturesForNB(sentence):
              # check if the sentence's word is present in the word features of above. If yes, then list as tuples
              sentence_words = set(sentence)
              features = {}
              for word in wordFeatures:
                  features['contains({})'.format(word)] = (word in sentence_words)
              return features
```

```
In [132]: def GetfeatureSetForNBSentences():
              featureSetFromSentences = [(getFeaturesForNB(sent), catg) for (sent,catg) in sentences]
              return featureSetFromSentences
```

```
In [133]: def GetfeatureSetForNBReviews():
              featureSetFromReviews = [getFeaturesForNB(reviewText) for reviewText in preProcessedReviewTokens]
              return featureSetFromReviews
```

```
In [134]: featureSet1FromSentences = GetfeatureSetForNBSentences()
```

```
In [138]: featureSet1FromReviews = GetfeatureSetForNBReviews()
```

Feature set 1 from sentences:

```
In [91]: featureSet1FromSentences[:200]

Out[91]: [({'contains(film)': False,
           'contains(movie)': False,
           'contains(one)': False,
           'contains(like)': False,
           'contains(story)': False,
           'contains(much)': False,
           'contains(even)': False,
           'contains(good)': False,
           'contains(comedy)': False,
           'contains(time)': False,
           'contains(characters)': False,
           'contains(little)': False,
           'contains(way)': False,
           'contains(funny)': True,
           'contains(make)': False,
           'contains(enough)': False,
           'contains(never)': False,
           'contains(makes)': False,
           'contains(may)': False,
```

Feature set 1 from reviews:

```
In [92]: featureSet1FromReviews[:200]

           'contains(much)': False,
           'contains(even)': False,
           'contains(good)': False,
           'contains(comedy)': False,
           'contains(time)': False,
           'contains(characters)': False,
           'contains(little)': False,
           'contains(way)': False,
           'contains(funny)': False,
           'contains(make)': False,
           'contains(enough)': False,
           'contains(never)': False,
           'contains(makes)': False,
           'contains(may)': False,
           'contains(us)': False,
           'contains(work)': False,
           'contains(best)': False,
           'contains(bad)': False,
           'contains(director)': False,
           'contains(love)': False,
```

Naïve Bayes classification for feature set 1:

```python
In [149]: sentiments=[]
          def NaiveBayesClassification():
              train_set = featureSet1FromSentences[1000:]
              test_set = featureSet1FromSentences[:1000]
              classifier = nltk.NaiveBayesClassifier.train(train_set)
              accuracy = nltk.classify.accuracy(classifier, test_set)
              classifier.show_most_informative_features(20)
              sentiments = [ classifier.classify(feature) for feature in featureSet1FromReviews]
              pos_sentences=[]
              neg_sentences=[]
              i=0
              for value in sentiments:
                  if(value == 'neg'):
                      neg_sentences.append(filteredReviewTexts[i])
                  else:
                      pos_sentences.append(filteredReviewTexts[i])
                  i+=1

              return pos_sentences,neg_sentences,accuracy
```

Most Informative features and accuracy of Naïve Bayes classification by feature set 1:

```
In [150]: pos_sent,neg_sent,accuracy= NaiveBayesClassification()

          Most Informative Features
                  contains(engrossing) = True            pos : neg    =     19.4 : 1.0
                     contains(beauty) = True            pos : neg    =     17.4 : 1.0
                   contains(provides) = True            pos : neg    =     16.8 : 1.0
                   contains(mediocre) = True            neg : pos    =     16.5 : 1.0
                    contains(generic) = True            neg : pos    =     16.5 : 1.0
                    contains(routine) = True            neg : pos    =     14.5 : 1.0
                       contains(flat) = True            neg : pos    =     13.6 : 1.0
                 contains(refreshing) = True            pos : neg    =     13.5 : 1.0
                  contains(wonderful) = True            pos : neg    =     12.8 : 1.0
                     contains(boring) = True            neg : pos    =     12.6 : 1.0
                       contains(warm) = True            pos : neg    =     12.4 : 1.0
                       contains(dull) = True            neg : pos    =     12.4 : 1.0
                   contains(absorbing) = True           pos : neg    =     12.2 : 1.0
               contains(refreshingly) = True            pos : neg    =     11.5 : 1.0
                   contains(touching) = True            pos : neg    =     11.4 : 1.0
                      contains(stale) = True            neg : pos    =     11.1 : 1.0
                     contains(stupid) = True            neg : pos    =     11.1 : 1.0
                      contains(devoid) = True           neg : pos    =     10.5 : 1.0
                    contains(tiresome) = True           neg : pos    =     10.5 : 1.0
                     contains(amazing) = True           pos : neg    =     10.2 : 1.0
```

```
In [171]: print('FEATURE SET 1 - Naive Bayes Classifier')
          print('ACCURACY: %.2f' %(accuracy*100))

          FEATURE SET 1 - Naive Bayes Classifier
          ACCURACY: 73.80
```

Classified positive sentences

```
In [168]: pos_sent[:100]
```

```
Out[168]: ["Perfect for new parents. We were able to keep track of baby's feeding, sleep and diaper change schedule for the first two a
          nd a half months of her life. Made life easier when the doctor would ask questions about habits because we had it all right t
          here!",
           'I bought this a few times for my older son and have bought it again for my newborn. This is super easy to use and helps me
          keep track of his daily routine. When he started going to the sitter when I went back to work, it helped me know how his day
          went to better prepare me for how the evening would most likely go. When he was sick, it help me keep track of how many diape
          rs a day he was producing to make sure he was getting dehydrated. The note sections to the side and bottom are useful too bec
          ause his sitter writes in small notes about whether or not he liked his lunch or if the playtime included going for a walk, e
          tc.Excellent for moms who are wanting to keep track of their kids daily routine even though they are at work. Excellent for d
          ads to keep track as my husband can quickly forget what time he fed our son. LOL',
           'My 3 month old son spend half of his days with my mother and half with a neighbor while I worked.  I had them track his act
          ivity (loosely) in this which allowed me to get an idea of how his schedule was developing and how much milk he was eating.
          It was the best way to have some cohesion in his life while I was at work.',
           "This book is perfect!  I'm a first time new mom, and this book made it so easy to keep track of feedings, diaper changes, s
          leep.  Definitely would recommend this for new moms.  Plus it's small enough that I throw in the diaper back for doctor visit
          s.",
           "The Baby Tracker brand books are the absolute best of the trackers that are available. You can track naps, feedings, diaper
          changes, activities, and notes. They're easy to use and have a nice clean layout.I started using the Newborn Baby Tracker, wh
          ich tracks nights and days. It gave me a nice way to track feedings and note important milestones. Once I used up the entire
```

Classified negative sentences:

```
In [169]: neg_sent[:100]
```

```
Out[169]: ['This book is such a life saver.  It has been so helpful to be able to go back to track trends, answer pediatrician question
          s, or communicate with each other when you are up at different times of the night with a newborn.  I think it is one of those
          things that everyone should be required to have before they leave the hospital.  We went through all the pages of the newborn
          version, then moved to the infant version, and will finish up the second infant book (third total) right as our baby turns 1.
          See other things that are must haves for baby at [...]',
          "I use this so that our babysitter (grandma) can keep track of what goes on during the day.  We've recorded so many mileston
          es since we received this.  It is a 6 month diary and has room for notes.  I wish the hours started around 6 am and went unti
          l 8 pm, but we just write over certain times to suit our purposes. The nice thing is that if we had a rough night and we are
          super tired, we have something to refer to if we forgot when we last fed the baby or when he woke up. I also like that the fr
          ont page has a place for emergency comments and a consent form in case your child needs medical care.  It also has a place at
          the back of the book for immunizations.  I'll definitely get another one in 6 months when we finish this book.",
          "I like this log, but think it would work better with clearer AM & PM sections. Each page is 12 hours so you really need two
          pages a day if your baby feeds or wets a lot in the early morning hours between midnight and 7am. We're cramming those in to
          the 2 blank spaces above 7am right now.",
          "My wife and I have a six month old baby boy and around the 4-month mark my wife decided that she would return to work inste
          ad of being a stay-at-home mom. We hired an in-home nanny to care for our little boy and the arrangement has worked quite wel
          l ever since. Shortly after starting the arrangement we realized that we would need some sort of journal to track our little
          boy's activities while he was with the nanny and we were working. We used a plain notebook for a period of weeks until we stu
          mbled on the Baby Tracker - Daily Childcare Journal.--------------------LAYOUT AND USE:The Baby Tracker is really an excellen
          t idea. The journal pages are cleanly divided into columns for tracking feedings, nap-time, diaper changes, play-time as well
```

```
In [167]: saveToFile('positive_Feature_Set1.txt',pos_sent)
          saveToFile('negative_Feature_Set1.txt',neg_sent)
```

Loading subjectivity lexicon from file:

```
In [27]: def readSubjectivity():
             file = open('subjclueslen1-HLTEMNLP05.tff', 'r')
             # initialize an empty dictionary
             sdict = { }
             for line in file:
                 data = line.split()
                 strength = data[0].split("=")[1]
                 word = data[2].split("=")[1]
                 posTag = data[3].split("=")[1]
                 stemmed = data[4].split("=")[1]
                 polarity = data[5].split("=")[1]
                 if (stemmed == 'y'):
                     sdict[word] = [strength, posTag, True, polarity]
                 else:
                     sdict[word] = [strength, posTag, False, polarity]
             return sdict
```

```
In [28]: wordDictionary = readSubjectivity()
```

Forming feature set 2 for classification using subjectivity lexicons:

```
In [29]: def getFeaturesForNB2(sentence):
             uwords = set(sentence)
             features = {}
             for word in wordFeatures:
                 features['contains(%s)' % word] = (word in uwords)
             strongNeg = 0
             weakNeg = 0
             weakPos = 0
             strongPos = 0
             for word in uwords:
                 if word in wordDictionary:
                     strength, posTag, isStemmed, polarity = wordDictionary[word]
                     if strength == 'weaksubj' and polarity == 'positive':
                         weakPos += 1
                     if strength == 'strongsubj' and polarity == 'positive':
                         strongPos += 1
                     if strength == 'weaksubj' and polarity == 'negative':
                         weakNeg += 1
                     if strength == 'strongsubj' and polarity == 'negative':
                         strongNeg += 1
                     features['positivecount'] = weakPos + (2 * strongPos)
                     features['negativecount'] = weakNeg + (2 * strongNeg)
             return features
```

```
In [30]: def GetfeatureSet2ForNBSentences():
             featureSetFromSentences = [(getFeaturesForNB2(sent), catg) for (sent,catg) in sentences]
             return featureSetFromSentences
```

```
In [31]: def GetfeatureSet2ForNBReviews():
             featureSetFromReviews = [getFeaturesForNB2(reviewText) for reviewText in preProcessedReviewTokens]
             return featureSetFromReviews
```

Feature set 2 from sentences:

```
In [32]: featureSetFromSentences2 = GetfeatureSet2ForNBSentences()
```

```
In [33]: featureSetFromSentences2[:10]
```

```
Out[33]: [({'contains(film)': False,
          'contains(movie)': False,
          'contains(one)': False,
          'contains(like)': False,
          'contains(story)': False,
          'contains(much)': False,
          'contains(even)': False,
          'contains(good)': False,
          'contains(comedy)': False,
          'contains(time)': False,
          'contains(characters)': False,
          'contains(little)': False,
          'contains(way)': False,
          'contains(funny)': False,
          'contains(make)': False,
          'contains(enough)': False,
          'contains(never)': False,
          'contains(makes)': False,
          'contains(may)': False,
```

Feature set 2 from reviews:

```
In [34]: featureSetFromReviews2 = GetfeatureSet2ForNBReviews()
```

```
In [35]: featureSetFromReviews2[:10]
```

```
Out[35]: [{'contains(film)': False,
          'positivecount': 6,
          'negativecount': 0,
          'contains(movie)': False,
          'contains(one)': False,
          'contains(like)': False,
          'contains(story)': False,
          'contains(much)': False,
          'contains(even)': False,
          'contains(good)': False,
          'contains(comedy)': False,
          'contains(time)': False,
          'contains(characters)': False,
          'contains(little)': False,
          'contains(way)': False,
          'contains(funny)': False,
          'contains(make)': False,
          'contains(enough)': False,
          'contains(never)': False,
          'contains(makes)': False,
```

Naïve Bayes classification for feature set 2:

```
In [36]: sentiments2=[]
         def NaiveBayesClassification2():
             train_set = featureSetFromSentences2[1000:]
             test_set = featureSetFromSentences2[:1000]
             classifier = nltk.NaiveBayesClassifier.train(train_set)
             accuracy = nltk.classify.accuracy(classifier, test_set)
             classifier.show_most_informative_features(20)
             sentiments2 = [ classifier.classify(feature) for feature in featureSetFromReviews2]
             pos_sentences=[]
             neg_sentences=[]
             i=0
             for value in sentiments2:
                 if(value == 'neg'):
                     neg_sentences.append(filteredReviewTexts[i])
                 else:
                     pos_sentences.append(filteredReviewTexts[i])
                 i+=1

             return pos_sentences,neg_sentences,accuracy
```

Most Informative features and accuracy of Naïve Bayes classification by feature set 2:

```
In [37]:  pos_sentfor2,neg_sentfor2,accuracyfor2= NaiveBayesClassification2()

          Most Informative Features
                contains(engrossing) = True           pos : neg   =     19.7 : 1.0
                 contains(captures) = True            pos : neg   =     17.7 : 1.0
                   contains(boring) = True            neg : pos   =     17.3 : 1.0
                  contains(routine) = True            neg : pos   =     15.6 : 1.0
                 contains(mediocre) = True            neg : pos   =     15.0 : 1.0
                     contains(flat) = True            neg : pos   =     14.6 : 1.0
               contains(unexpected) = True            pos : neg   =     14.4 : 1.0
                  contains(generic) = True            neg : pos   =     14.3 : 1.0
                 contains(inventive) = True           pos : neg   =     13.7 : 1.0
                contains(refreshing) = True           pos : neg   =     13.7 : 1.0
                     contains(dull) = True            neg : pos   =     12.8 : 1.0
                     contains(warm) = True            pos : neg   =     12.2 : 1.0
                 contains(wonderful) = True           pos : neg   =     11.8 : 1.0
              contains(refreshingly) = True           pos : neg   =     11.7 : 1.0
                 contains(delicate) = True            pos : neg   =     11.7 : 1.0
                    contains(stale) = True            neg : pos   =     11.6 : 1.0
                 contains(provides) = True            pos : neg   =     10.6 : 1.0
                     contains(ages) = True            pos : neg   =     10.4 : 1.0
                      contains(car) = True            neg : pos   =     10.3 : 1.0
                    contains(bears) = True            neg : pos   =     10.3 : 1.0
```

```
In [38]:  print('FEATURE SET 2 - Naive Bayes Classifier')
          print('ACCURACY: %.2f' %(accuracyfor2*100))

          FEATURE SET 2 - Naive Bayes Classifier
          ACCURACY: 76.30
```

Classified positive sentences:

```
In [42]:  pos_sentfor2[:20]
```

```
Out[42]: ["Perfect for new parents. We were able to keep track of baby's feeding, sleep and diaper change schedule for the first two a
         nd a half months of her life. Made life easier when the doctor would ask questions about habits because we had it all right t
         here!",
          'I bought this a few times for my older son and have bought it again for my newborn. This is super easy to use and helps me
         keep track of his daily routine. When he started going to the sitter when I went back to work, it helped me know how his day
         went to better prepare me for how the evening would most likely go. When he was sick, it help me keep track of how many diape
         rs a day he was producing to make sure he was getting dehydrated. The note sections to the side and bottom are useful too bec
         ause his sitter writes in small notes about whether or not he liked his lunch or if the playtime included going for a walk, e
         tc.Excellent for moms who are wanting to keep track of their kids daily routine even though they are at work. Excellent for d
         ads to keep track as my husband can quickly forget what time he fed our son. LOL',
          'My 3 month old son spend half of his days with my mother and half with a neighbor while I worked.  I had them track his act
         ivity (loosely) in this which allowed me to get an idea of how his schedule was developing and how much milk he was eating.
         It was the best way to have some cohesion in his life while I was at work.',
          "This book is perfect!  I'm a first time new mom, and this book made it so easy to keep track of feedings, diaper changes, s
         leep.  Definitely would recommend this for new moms.  Plus it's small enough that I throw in the diaper back for doctor visit
         s.",
          "The Baby Tracker brand books are the absolute best of the trackers that are available. You can track naps, feedings, diaper
         changes, activities, and notes. They're easy to use and have a nice clean layout.I started using the Newborn Baby Tracker, wh
         ich tracks nights and days. It gave me a nice way to track feedings and note important milestones. Once I used up the entire
         book, I bought this Baby Tracker, which tracks only days since my daughter is sleeping during the night now. Her caregiver no
```

Classified negative sentences:

```
In [43]:  neg_sentfor2[:20]
```

```
Out[43]: ['This book is such a life saver.  It has been so helpful to be able to go back to track trends, answer pediatrician question
         s, or communicate with each other when you are up at different times of the night with a newborn.  I think it is one of those
         things that everyone should be required to have before they leave the hospital.  We went through all the pages of the newborn
         version, then moved to the infant version, and will finish up the second infant book (third total) right as our baby turns 1.
         See other things that are must haves for baby at [...]',
          "I use this so that our babysitter (grandma) can keep track of what goes on during the day.  We've recorded so many mileston
         es since we received this.  It is a 6 month diary and has room for notes.  I wish the hours started around 6 am and went unti
         l 8 pm, but we just write over certain times to suit our purposes. The nice thing is that if we had a rough night and we are
         super tired, we have something to refer to if we forgot when we last fed the baby or when he woke up. I also like that the fr
         ont page has a place for emergency comments and a consent form in case your child needs medical care.  It also has a place at
         the back of the book for immunizations.  I'll definitely get another one in 6 months when we finish this book.",
          "I like this log, but think it would work better with clearer AM & PM sections. Each page is 12 hours so you really need two
         pages a day if your baby feeds or wets a lot in the early morning hours between midnight and 7am. We're cramming those in to
         the 2 blank spaces above 7am right now.",
          "My wife and I have a six month old baby boy and around the 4-month mark my wife decided that she would return to work inste
         ad of being a stay-at-home mom. We hired an in-home nanny to care for our little boy and the arrangement has worked quite wel
         l ever since. Shortly after starting the arrangement we realized that we would need some sort of journal to track our little
         boy's activities while he was with the nanny and we were working. We used a plain notebook for a period of weeks until we stu
         mbled on the Baby Tracker - Daily Childcare Journal.--------------------LAYOUT AND USE:The Baby Tracker is really an excellen
```