A Mini Project in

Product Review Sentimental Analysis

Submitted in partial fulfillment of
the requirements for the award of the degree of

Bachelor of Technology
In
Computer Science and Engineering

By

Reg. No.	Name Of Student
20134044	Shivam Mohan
20134086	Shreyash Hisariya
20134022	Prashant Agrawal
20134166	Ankit Kumar Sharma
20134164	Banda Prashanth Yadav

Project Group – CS26

To the

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

MOTILAL NEHRU NATIONAL INSTITUTE OF TECHNOLOGY

ALLAHABAD

April 2016

UNDERTAKING

I declare that the work presented in this project titled “**Product Review Analysis**", submitted to the Computer Science and Engineering Department, Motilal Nehru National Institute of Technology, Allahabad, for the award of the Bachelor of Technology degree in Computer Science & Engineering, is my original work. I have not plagiarized or submitted the same work for the award of any other degree. In case this undertaking is found incorrect, I accept that my degree may be unconditionally withdrawn.

April 2016.

CERTIFICATE

Certified that the work contained in the project titled “**Product Review Analysis**”, by *Shivam Mohan, Shreyash Hisariya, Prashant Agrawal, Ankit Kumar Sharma and Banda Prashanth Yadav* has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

(Er. Rupesh Dewang)

Computer Science and
Engineering Department

Motilal Nehru National Institute
of Technology Allahabad

Preface

On-line reviews on services and products are fast becoming an important factor while buying/using that product/service. India’s e-Commerce market focuses on the various sub-segments of the e-Commerce market and highlights the factors driving growth across these segments and the challenges.

India's e-commerce market was worth about $3.9 billion in 2009, it went up to $12.6 billion in 2013. In 2013, the e-retail segment was worth US$2.3 billion. About 70% of India's e-commerce market is travel related. According to Google India, there were 35 million online shoppers in India in 2014 and is expected to cross 100 million mark by end of year 2016. India's retail market is estimated at $470 billion in 2011 and is expected to grow to $675 Bn by 2016 and $850 Bn by 2020, – estimated CAGR of 10%.

Buying decisions on these websites are greatly influenced by the product reviews given by previous purchasers. In this regard, many reviews are written by people who have not actually purchased/used the product. These reviews may misguide a purchaser's decision regarding a particular product.

Acknowledgements

We would like to acknowledge our gratitude to the people who extended their support and cooperation and played an important role in the completion of this project. We express our sincere thanks to Er. Rupesh Dewang, for his guidance and encouragement which led to successful completion of this undertaking. Without his constant appraisals, motivation and efforts this task could not have been possible. We would also like to thank our friends and others who imparted great help at every step to make it a success. Lastly we would also like to thank our Head of Department, Prof. R.S. Yadav and our Director, Prof. P. Chakrabarti for giving us this opportunity of conduct this project and providing the resources that were essential for the completion of this project.

Contents

Chapter 1

Introduction

Nowadays, most products and services are listed online and welcome online feedback from their customers. The feedback provided is mostly in the form of user-ratings and reviews (which gives in-depth analysis of the rating given by a user). Ratings are quantified values hence does not explain details of the product/service. Reviews on the other hand are useful data which a potential customer uses in order to make a choice.

As the number of reviews of a particular product may be large, it is a tedious task to analyze each and every review manually. So here we came up with an idea to calculate the percentage of positive and negative reviews programmatically, which enables the buyers to compare and choose between different products in the market.

1.1 Motivation

The motivation for doing this project was primarily an interest in undertaking a challenging project. In the current scenario reviews play a very important role.

88% Of Consumers Trust Online Reviews As Much As Personal Recommendations Each new review written about a product on Company's site increases the amount of unique content site offers on that product , it will be seen as having higher relevance and be useful in deciding future strategies.

Reviews serve as a strong source of "word-of-mouth" communication for next customer in his purchase decision.

1.1.1 Some Wonderful Minds

Bing Liu, professor of opinion mining at the University of Illinois at Chicago, is working on detecting fake or deceptive opinions on social media platforms like re-views, Facebook, Twitter, Weibo, and forum discussion sites, and also is a author of sentiment analysis and opinion mining. Nitin Jindal is a professor at University of Illinois at Chicago. It is interesting to know that Jindal and Liu together had worked on many research papers together on detecting review spammers.

Chapter 2

Hypothesis

2.1 Initial Approach

2.1.1 Web Crawler

A Web crawler is a computer program that browses the World Wide Web in a methodical, automated manner to gather or collect particular set of data.

We designed a web crawler which could collect reviews of mobiles from an E-Commerce website (Snap-deal) which could be treated as a dataset for further analysis of reviews. We performed some preprocessing and remove the unwanted data present.

Next step would be classifying the product overall whether the mobile product is favorable or non-favorable to buy.

2.2 Data Collection

The content format of the **Snapdeal** website is 48 products per page which comprises 10 reviews per product per page. In order to obtain all the reviews for each product of a particular category (mobiles), a python script was written that automatically traverses each page and extracts the reviews.

We have obtained the HTML Source page by using the *Requests* library and then relevant tags are extracted using the *BeautifulSoup4* library. We traversed the links from the tags and extracted the reviews of the products. This process is repeated for all the products. We even made this to work with authenticated proxy.

2.3 Preprocessing

2.3.1 Punctuation Removal

Punctuations do not contribute anything to the meaning and context of words when one word is processed at a time. We wrote a Python script that removed punctuation from the documents and gave us a single document.

2.3.2 Stop words Removal

Sometimes, some extremely common words which would appear to be of little value in helping select, documents matching the user need are excluded from the vocabulary entirely. These words are called **stop words**. To remove such words, we used NLTK's built-in library.

2.3.3 Lemmatizing

The goal of lemmatization is to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form. For instance:

am, are, is => be
car, cars, car's, cars' => car

The result of this mapping of text will be something like:

The boy's cars are different colors =>
The boy car be differ color.

2.4 Sentiment Calculation

Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. It’s also known as opinion mining, deriving the opinion or attitude of a speaker. It is used to identify and extract subjective information in source materials. It aims to determine the overall attitude and contextual polarity of a document with respect to some topic.

We have calculated the overall polarity of a product considering all the collected reviews.

We have used two dictionaries, one for positive and other for negative words. These dictionaries contain words with their corresponding weightages, which helps in summing up the polarity of the review. By comparing each word from the reviews with these dictionaries, we determine the polarity of the products.

2.5 Probability Determination:

We have determined probability of a particular word, whether the word has been used in terms of positive or negative respect. We have performed this probability determination by using the Naïve Bayes theorem.

Naïve Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

Abstractly, naive Bayes is a conditional probability model: given a problem instance to be classified, represented by a vector ***x = (x₁, . . . , xₙ)*** representing some n features (independent variables), it assigns to this instance probabilities.

$$p(C_k|x_1, \dots, x_n)$$

for each of *K* possible outcomes or *classes*.

The problem with the above formulation is that if the number of features *n* is large or if a feature can take on a large number of values, then basing such a model on probability tables is infeasible. We therefore reformulate the model to make it more tractable. Using Bayes' theorem, the conditional probability can be decomposed as

$$p(C_k|\mathbf{x}) = \frac{p(C_k)\ p(\mathbf{x}|C_k)}{p(\mathbf{x})}$$

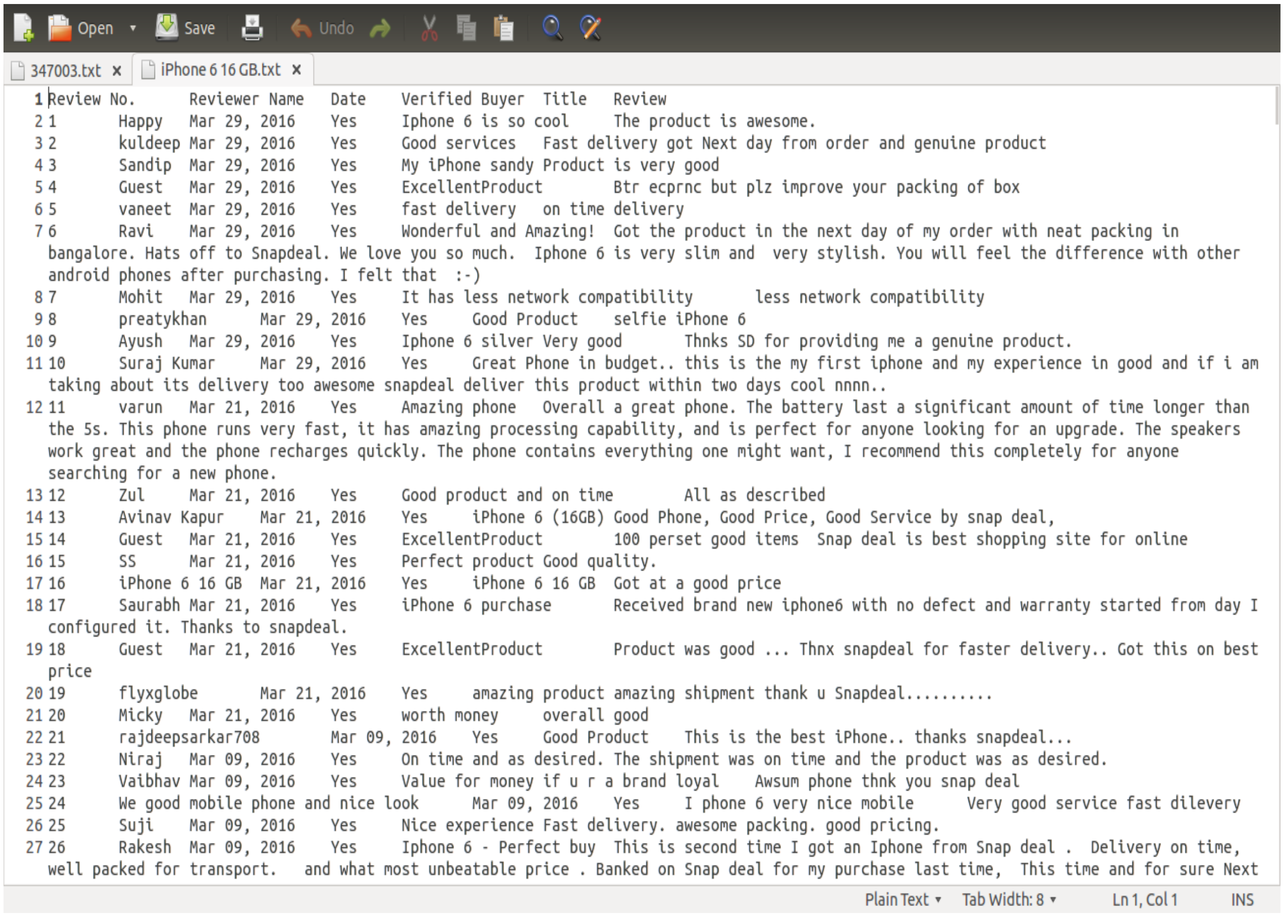
In plain English, using Bayesian probability terminology, the above equation can be written as

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

Chapter 3

Result

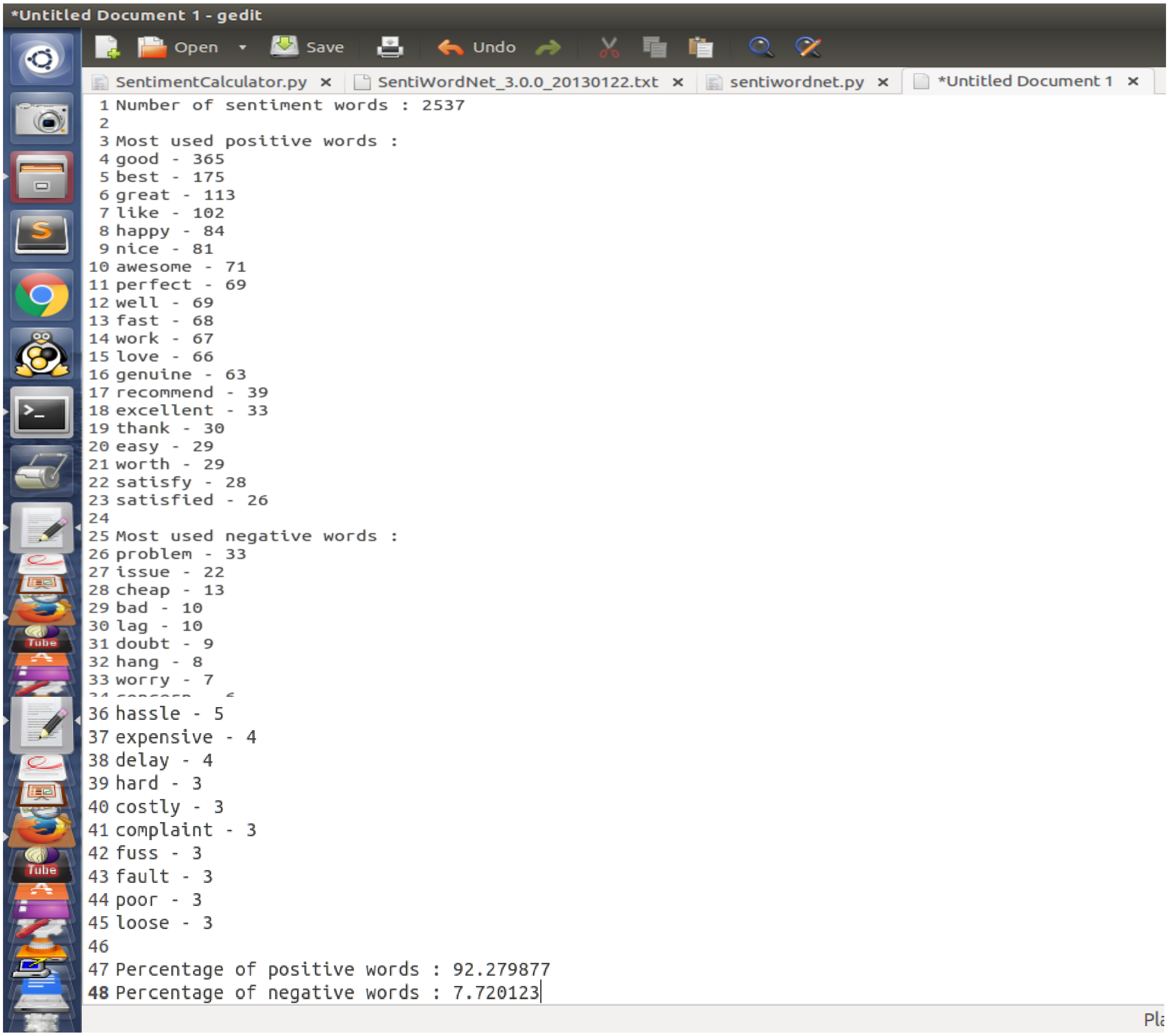
3.1 Sample Product Reviews:



```
Review No.      Reviewer Name   Date    Verified Buyer  Title   Review
1       Happy   Mar 29, 2016    Yes     Iphone 6 is so cool     The product is awesome.
2       kuldeep Mar 29, 2016    Yes     Good services   Fast delivery got Next day from order and genuine product
3       Sandip  Mar 29, 2016    Yes     My iPhone sandy Product is very good
4       Guest   Mar 29, 2016    Yes     ExcellentProduct        Btr ecprnc but plz improve your packing of box
5       vaneet  Mar 29, 2016    Yes     fast delivery   on time delivery
6       Ravi    Mar 29, 2016    Yes     Wonderful and Amazing!  Got the product in the next day of my order with neat packing in
bangalore. Hats off to Snapdeal. We love you so much.  Iphone 6 is very slim and  very stylish. You will feel the difference with other
android phones after purchasing. I felt that  :-)
7       Mohit   Mar 29, 2016    Yes     It has less network compatibility       less network compatibility
8       preatykhan      Mar 29, 2016    Yes     Good Product    selfie iPhone 6
9       Ayush   Mar 29, 2016    Yes     Iphone 6 silver Very good       Thnks SD for providing me a genuine product.
10      Suraj Kumar     Mar 29, 2016    Yes     Great Phone in budget.. this is the my first iphone and my experience in good and if i am
taking about its delivery too awesome snapdeal deliver this product within two days cool nnnn..
11      varun   Mar 21, 2016    Yes     Amazing phone   Overall a great phone. The battery last a significant amount of time longer than
the 5s. This phone runs very fast, it has amazing processing capability, and is perfect for anyone looking for an upgrade. The speakers
work great and the phone recharges quickly. The phone contains everything one might want, I recommend this completely for anyone
searching for a new phone.
12      Zul     Mar 21, 2016    Yes     Good product and on time        All as described
13      Avinav Kapur    Mar 21, 2016    Yes     iPhone 6 (16GB) Good Phone, Good Price, Good Service by snap deal,
14      Guest   Mar 21, 2016    Yes     ExcellentProduct        100 perset good items  Snap deal is best shopping site for online
15      SS      Mar 21, 2016    Yes     Perfect product Good quality.
16      iPhone 6 16 GB  Mar 21, 2016    Yes     iPhone 6 16 GB  Got at a good price
17      Saurabh Mar 21, 2016    Yes     iPhone 6 purchase       Received brand new iphone6 with no defect and warranty started from day I
configured it. Thanks to snapdeal.
18      Guest   Mar 21, 2016    Yes     ExcellentProduct        Product was good ... Thnx snapdeal for faster delivery.. Got this on best
price
19      flyxglobe       Mar 21, 2016    Yes     amazing product amazing shipment thank u Snapdeal..........
20      Micky   Mar 21, 2016    Yes     worth money     overall good
21      rajdeepsarkar708        Mar 09, 2016    Yes     Good Product    This is the best iPhone.. thanks snapdeal...
22      Niraj   Mar 09, 2016    Yes     On time and as desired. The shipment was on time and the product was as desired.
23      Vaibhav Mar 09, 2016    Yes     Value for money if u r a brand loyal    Awsum phone thnk you snap deal
24      We good mobile phone and nice look      Mar 09, 2016    Yes     I phone 6 very nice mobile      Very good service fast dilevery
25      Suji    Mar 09, 2016    Yes     Nice experience Fast delivery. awesome packing. good pricing.
26      Rakesh  Mar 09, 2016    Yes     Iphone 6 - Perfect buy  This is second time I got an Iphone from Snap deal .  Delivery on time,
well packed for transport.   and what most unbeatable price . Banked on Snap deal for my purchase last time,  This time and for sure Next
```

3.2 Sentiment Calculation

iPhone_5S_16GB_Space_Gray



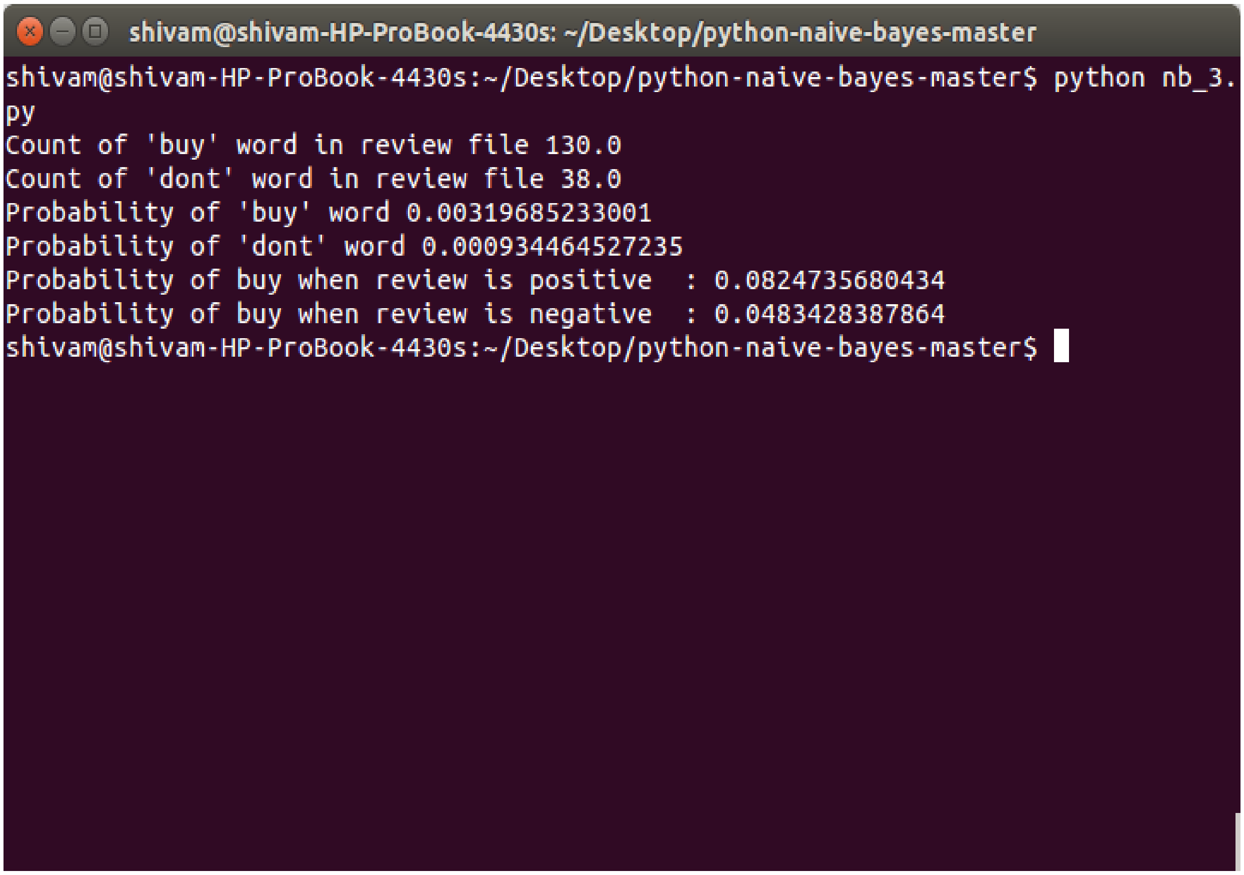
```
Number of sentiment words : 2537

Most used positive words :
good - 365
best - 175
great - 113
like - 102
happy - 84
nice - 81
awesome - 71
perfect - 69
well - 69
fast - 68
work - 67
love - 66
genuine - 63
recommend - 39
excellent - 33
thank - 30
easy - 29
worth - 29
satisfy - 28
satisfied - 26

Most used negative words :
problem - 33
issue - 22
cheap - 13
bad - 10
lag - 10
doubt - 9
hang - 8
worry - 7
concern - 6
hassle - 5
expensive - 4
delay - 4
hard - 3
costly - 3
complaint - 3
fuss - 3
fault - 3
poor - 3
loose - 3

Percentage of positive words : 92.279877
Percentage of negative words : 7.720123
```

3.3 Naïve Bayes Theorem Implementation



shivam@shivam-HP-ProBook-4430s: ~/Desktop/python-naive-bayes-master

```
shivam@shivam-HP-ProBook-4430s:~/Desktop/python-naive-bayes-master$ python nb_3.
py
Count of 'buy' word in review file 130.0
Count of 'dont' word in review file 38.0
Probability of 'buy' word 0.00319685233001
Probability of 'dont' word 0.000934464527235
Probability of buy when review is positive  : 0.0824735680434
Probability of buy when review is negative  : 0.0483428387864
shivam@shivam-HP-ProBook-4430s:~/Desktop/python-naive-bayes-master$
```

Chapter 4

Conclusion

So far we have successfully:

- Extracted mobile reviews from *Snapdeal* website.
- Converted them to excel format for ease of storage and implementing database model (if need arises).
- Calculated sentiment value of all mobile product reviews (score in percentage).
- Applied Naïve Bayes theorem to calculate conditional probability of occurrence of word in positive or negative sentiment.

Chapter 5

Limitations

- Inability to detect Sarcasm. Oxford dictionary defines sarcasm as the use of irony to mock or convey contempt.
  Example: *The price of iPhone is very cheap.*
- Inability to process slangs.
- Inability to expand general abbreviations.

Bibliography

- Myle Ott: Deceptive Opinion Spam Corpus v1.4. http://myleott.com/op spam/
- NLTK Documentation. http://www.nltk.org/
- BeautifulSoup4 Documentation: http://www.crummy.com/software/BeautifulSoup/bs4/doc
- Math Works Documentation.
- Nitin Jindal and Bing Liu. Opinion Spam and Analysis (2008)