

# Customer Feedback Analysis Using NLP

\* Subhabaha Pal

## Abstract

There are many ways to understand what customers think about certain products: usage surveys, surveys on focus groups, social media mining etc. However, these are time-consuming methods and need enough resources to get useful insights from them. The case is different for Amazon review. It provides a fast, accessible, and vast data resource for getting quick insights. It does both of these things and helps researchers to quickly explore what is going on and the data collection cost is effectively zero. Using NLP techniques on this gathered data provides quick insights into strengths and weakness of products. The paper discusses ways to use NLP in order generate useful insights about the strengths and weakness of a certain product. Apple iPhone 12 Pro Max has been studied here to present the effectiveness of NLP (text analytics) in order to generate useful business insights.

In this paper, it has been tried to present an efficient way which helps in obtaining insights from the consumer market like what people look for while buying products and what are the typical causes of frustration in the purchase process or in product. What are the typical pain points / causes of frustration in the purchase process or in a product itself?

**Keywords :** Customer feedback analysis, data scrapping, NLP, text analytics

## I. INTRODUCTION

The retail sector is observing continual and phenomenal transformations over the last decades with the advent of information technology. The application of IT in the retail sector, specifically e-commerce, has brought about astonishing customer experience and has opened a world of opportunities for both customers and sellers. The advent of IT in the online retail sector has changed the rules of customer engagement with different brands. The wide application of analytical techniques has greatly improved customer experience and played an important role in enhancing the sales of different online traders with optimum use of IT and analytical techniques.

Customer engagement plays a very crucial role for any business, be it online retail business or a regular brick-and-mortar business. Technology plays an important role during the times of launching new products, enhancing brand image, and connecting with customers. Companies take help of latest technologies which involve artificial intelligence, natural language processing etc. to gain a better understanding of customers and for driving personalized engagement. We know that customer opinion plays an important role

behind success and failure of any product and it is also needed to develop a new product. Retailers are slowly realizing the value that data mining can bring to their business by proving useful insights about their operations and customers. Opinion mining plays an important role to connect to understand customer's opinion about certain products and to drive personalized engagement [4]. The unstructured data like e-mails, feedbacks, social media etc. provides rich insights into any business. Natural Language Processing or NLP plays an important role in this journey [3].

Natural Language Processing (NLP) is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language, in particular how to program computers to process and analyze large amounts of natural language data [5]. NLP can be described as a Machine Learning technique that enables machines to process huge amounts of structured and unstructured customer data. NLP involves analysis of speech and text-based communication dataset. NLP is the main technology behind chatbots, virtual assistants, online translations, and many more services.

We can understand the usefulness of the NLP using the

---

Manuscript Received : November 12, 2020 ; Revised : December 14, 2020 ; Accepted: December 16, 2020. Published : February 5, 2021

\* S. Pal is Associate Professor with TAPMI School of Business, Manipal University Jaipur, Dehmi Kalan, Jaipur-Ajmer Express Highway, Jaipur, Rajasthan - 303 007. (Email : subhabaha.pal@jaipur.manipal.edu)

DOI: 10.17010/ijcs/2021/v6/i1/158186

business scenario where a customer wants to buy a particular red-coloured trek-shoe. The customer will search in an e-commerce website using the search query 'red trek shoe' or 'red shoe for trekking'. A search-engine which is not powered by NLP would show red-coloured shoes or trek-shoes and it will not be able to accurately interpret the language and will not be able to show the exact product wanted. The main reason behind this is that in that case, the Search Engine will not be able to accurately interpret the context to give necessary information [1]. This traditional search process is continuously trained and improved with NLP and it provides customers with images of the red trek shoe products as final outcome.

It is a fact that any retail e-commerce website targets to gain maximum customers by presenting preferred products and also try to retain customer loyalty. Loyalty is retained by not only making products available, but also by providing suitable discounts during certain period of the year or special discounts to regular customers for any purchase. In order to retain loyalty of customers, e-commerce portals needs to sell those products more which customers have liked more in the past and have the possibility of buying them again in the future. NLP plays a very important role in understanding customer sentiment. The customers provide positive or negative feedback about the products and some feedback scores (ratings). The ratings are enough to get a glimpse of what

customers think about the product. However, it is not enough to understand what features of the product the seller or manufacturer needs to work upon so that the products become acceptable to prospective customers. NLP shows the way to work upon major features.

This paper largely discusses the way how different NLP techniques like data scrapping, word cloud etc. can be used to get useful business insights which help in business decision making. The paper discusses ways to use NLP in order generate useful insights about the strengths and weakness of a certain product. The Apple iPhone 12 Pro Max has been studied here to present the effectiveness of NLP (text analytics) in order to generate useful business insights.

In this paper, we have tried to use NLP to get an understanding of customer feedback on certain products. There are many ways to understand what customers think about certain products : usage surveys, surveys on the focus groups, social media mining etc. However, these are time-consuming methods, and need enough resources to get useful insights from it. The case is different for Amazon review.

## II. METHODS OF DATA COLLECTION

There are many tools which can be used for scrapping

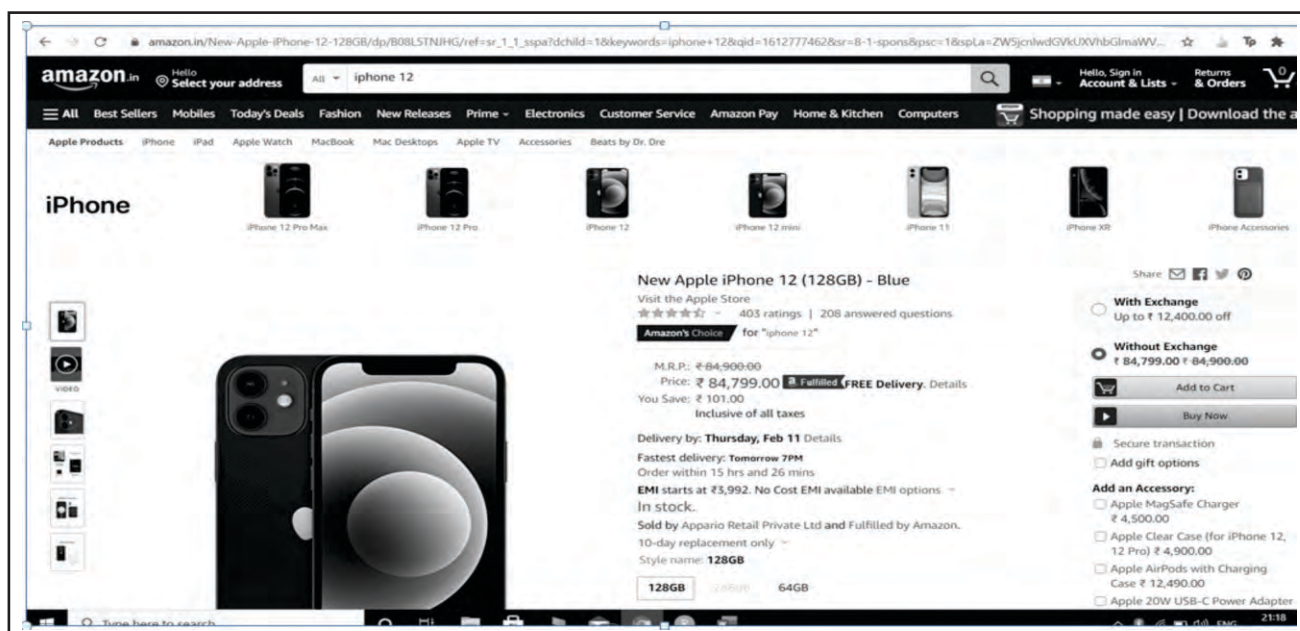


Fig. 1. Product Page in Amazon

Amazon Reviews – R and Python software can be used for this purpose. R is a commonly used tool that can be used for the purpose. Here, the data scrapping technique using R is discussed in detail.

The initial step is to stack the tidyverse and rvest bundles, as these are required for building the web-scrapping function (for example parsing html) and for general manipulation:

```
library(tidyverse)
library(rvest)
```

The following step is to get the ASIN (which represents Amazon Standard Identification Number) of the item on which the survey needs to be performed. This is adequately an item ID, which can normally be found

inside the URL of the item interface itself. ASINs are exceptional strings of 10 characters, which for books twould be equivalent to the ISBN number.

The next step is to write the main workhorse function for scrapping the reviews. In essence, what is tried to achieve is to download the HTML content from the Amazon review page, and then to use various html parsing and selector functions to organise the downloaded content into an easily manipulable format.

The read\_html() function from the xml2 bundle analyses the HTML content from a given URL, which can be assigned to an item in R and sort out some way to extricate content from the article.

The code in Fig. 2 is used for data scrapping:

After the main data-scrapping code has been setup, the lapply() can be used to loop through the page ranges.

```
3= scrape_amazon <- function(ASIN, page_num){
4
5   url_reviews <- paste0("https://www.amazon.in/New-Apple-iPhone-Pro-256GB/product-reviews/B08L5T31V6/ref=cm_cr_dp_d_show_all_btn?ie=UTF8&reviewerType=all_reviews", page_num)
6
7   doc <- read_html(url_reviews) # Assign results to 'doc'
8
9   # Review Title
10  doc %>%
11    html_nodes("[class='a-size-base a-link-normal review-title a-color-base review-title-content a-text-bold']") %>%
12    html_text() -> review_title
13
14  # Review Text
15  doc %>%
16    html_nodes("[class='a-size-base review-text review-text-content']") %>%
17    html_text() -> review_text
18
19  # Number of stars in review
20  doc %>%
21    html_nodes("[data-hook='review-star-rating']") %>%
22    html_text() -> review_star
23
24  # Return a tibble
25  tibble(review_title,
26         review_text,
27         review_star,
28         page = page_num) %>% return()
29 }
30 scrape_amazon(ASIN = "1612726383", page_num = 5) %>%
31   head()
```

Fig. 2. Data Scrapping Code

```
32
33
34 ASIN <- "1612726383" # Specify ASIN
35 page_range <- 1:10 # Let's say we want to scrape pages 1 to 10
36
37 # Create a table that scrambles page numbers using 'sample()'
38 # For randomising page reads!
39 match_key <- tibble(n = page_range,
40                     key = sample(page_range, length(page_range)))
41
42 lapply(page_range, function(i){
43   j <- match_key[match_key$n==i,]$key
44   message("Getting page ", i, " of ", length(page_range), "; Actual: page ", j) # Progress bar
45   Sys.sleep(3) # Take a three second break
46
47   if((i %% 3) == 0){ # After every three scrapes... take another two second break
48     message("Taking a break...") # Prints a 'taking a break' message on your console
49     Sys.sleep(2) # Take an additional two second break
50   }
51   scrape_amazon(ASIN = ASIN, page_num = j) # Scrape
52 } -> output_list
53
```

Fig. 3. Data Scrapping Code to Avoid Bot Detection

### III. WORD CLOUD CREATION

Data visualizations (like charts, graphs, info graphics and more) give businesses a valuable way to communicate important information at a glance. However, visualizations are difficult to make if text data is present. Word cloud is a technique for text data visualization and analysis. Word clouds work in a simple way: the more a specific word appears in a source of textual data (such as a speech, blog post, or database), the bigger and bolder it appears in the word cloud [2]. A word cloud is a collection, or cluster of words depicted in different sizes. The bigger and bolder the word appears, the more often it is mentioned within a given text and the more important it is. It is also known as tag clouds or text clouds. These are ideal ways to pull out the most pertinent parts of textual data, from blog posts to databases. They can also help business users compare and contrast two different pieces of text to find wording similarities between the two. The detailed method of developing word cloud is given next.

#### STEP 1: Retrieving the data and uploading the packages

To generate word clouds, word cloud package in R as

```
library(wordcloud)
library(RColorBrewer)
library(wordcloud2)

library(tm)

#Create a vector containing only the text
text <- output_list
# Create a corpus
docs <- Corpus(VectorSource(text))
```

Fig. 4. Word Cloud Creation

```
docs <- docs %>%
  tm_map(removeNumbers) %>%
  tm_map(removePunctuation) %>%
  tm_map(stripwhitespace)
docs <- tm_map(docs, content_transformer(tolower))
docs <- tm_map(docs, removewords, stopwords("english"))
docs <- tm_map(docs, removewords, c("nnnnnnnnnn", "iphone", "phone", "nnnnnnnn", "stars"))
```

Fig. 5. Word Cloud Creation

```
dtm <- TermDocumentMatrix(docs)
matrix <- as.matrix(dtm)
words <- sort(rowSums(matrix), decreasing=TRUE)
df <- data.frame(word = names(words), freq=words)
```

Fig. 6. Word Cloud Creation

well as the RcolorBrewer package for colours need to be downloaded. It is to be noted that there is also a wordcloud2 package with a slightly different design and fun applications. Both have been discussed here.

#### STEP 2: Text Data Cleansing

Data cleansing is extremely important for creation of word cloud. It is needed to remove numbers, special characters or punctuation marks from the text in order to get useful insights from the dataset. In addition to these, common stop words are also to be removed in order to produce meaningful results. It is also needed to avoid the most common frequent words such as “I” or “the” to appear in the word cloud. Fig. 4 to Fig. 6 show the code for Word Cloud creation. Fig. 5 shows how to do this using the tm package.

#### STEP 3: Creation a document-term-matrix

What someone wants to do as a next step is to have a data frame containing each word in the first column and its frequency in the second column.

This can be done by creating a document term matrix with the Term Document Matrix function from the tm package.

#### STEP 4: Generating the word cloud

The word cloud package is the most classic way to generate a word cloud. The following line of code shows how to properly set the arguments.

Fig. 8 shows an example word cloud which is created for





## V. CONCLUSION

We have seen from how we can get detailed insights about the perception of customers about a product from the Amazon Reviews textual data. This kind of insight greatly helps decision makers to design new products with improved features and more facilities. Thus, it can be concluded that NLP is playing a huge role in customer feedback analysis and helping companies retain customer loyalty by providing improved services.

## REFERENCES

- [1] S. Baruah, and S. Pal, "An Analytical approach to customer sentiments using NLP techniques and building a brand recommender based on popularity score," *Indian J. of Comput. Sci.*, vol. 4, no. 6, pp. 7–22, 2019. <http://dx.doi.org/10.17010/ijcs%2F2019%2Fv4%2Fi6%2F150421>
- [2] BoostLabs, "Word clouds & the value of simple visualizations," [Online]. Available: <https://boostlabs.com/blog/what-are-word-clouds-value-simple-visualizations/>
- [3] Y. Kang, Z. Cai, C. Tan, Q. Huang, and H. Liu, "Natural language processing (NLP) in management research: A literature review," *J. of Manage. Analytics*, vol. 7, no. 2, pp. 139–172, 2020. Doi: <https://doi.org/10.1080/23270012.2020.1756939>
- [4] D. Lee, O. Jeong, and S. Lee, "Opinion mining of customer feedback data on the web," in *Proc. of the 2nd Int. Conf. on Ubiquitous Inform. Manage. and Communication*, pp. 230–235, 2008. Doi: <https://doi.org/10.1145/1352793.1352842>
- [5] "Natural Language Processing," [Online]. Available: [https://en.wikipedia.org/wiki/Natural\\_language\\_processing](https://en.wikipedia.org/wiki/Natural_language_processing)

### About the Author

**Dr. Subhabaha Pal** is a seasoned Data Scientist and Academician with over 16 years of experience working in varied fields of Information Science and Analytics. He had been nominated among the top 20 Data Science and Machine Learning Academicians in India in 2018 by Analytics India Magazine. He completed Ph.D. from the University of Calcutta. He has taught Data Science at well-renowned institutions like Manipal University, T. A. Pai Management Institute, and International Institute of Digital Technologies among others. He had worked in senior software-related roles in organizations like Kuwait Petroleum Corporation and Manipal Global. He has around 40 research papers in the field of Data Science and Analytics and three books in the field of Data Science and Analytics to his credit and he has books in renowned publications to his credit. He is a hardcore Data Scientist who has delivered many data science projects to different SMEs and has worked in various domains.