SENTIMENT ANALYSIS ON AMAZON REVIEWS

**B. E. Information Technology**

By

**Group No. - 20 Akshay Gunjal 34**

**Shaun Dsilva 35**

**Davon Carvalho 36**

**Ms. Vandana Patil**

Assistant Professor



Department of Information Technology St. Francis Institute of Technology (Engineering College)

University of Mumbai 2020-2021

DECLARATION

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources.

We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in this submission.

We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

1.

(Signature)

(Akshay Gunjal, Roll Number 34)

2.

(Signature)

(Shaun Dsilva, Roll Number 35)

3.

(Signature)

(Davon Carvalho, Roll Number 36)

Date:

CERTIFICATE

This R programming Mini-project ***Sentiment Analysis on Amazon Reviews*** by ***Akshay Gunjal (34)***, ***Shaun Dsilva (35)***, ***Davon Carvalho (36)*** is complete in all respects and was successfully demonstrated on {Final external presentation date}.

Name :

Signature :

(Internal examiner)

Name :

Signature :

(External examiner)

Date:

Place: SFIT, Mumbai

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Content** | **Page No.** |
| 1. | Project Overview |  |
| 2. | Data Extraction |  |
| 3. | Exploratory Data Analysis |  |
| 4. | Application of Mining Algorithm |  |
| 5. | Data Visualization and Interpretation |  |
| 6. | Conclusion |  |
| 7. | Acknowledgement |  |

**Signature:**

**Ms. Vandana Patil**

**(Internal Guide)**

**Project Overview**

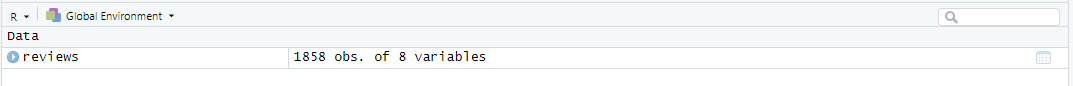
|  |  |  |
| --- | --- | --- |
| **PROJECT OVERVIEW** | | |
| **1.** | **Project Title -** | **Sentiment analysis on Amazon reviews** |
| **2.** | **Data set Name -** | Amazon product reviews data set |
| **3.** | **Introduction of Data set -** | The data set considered for analysis is the Amazon Product Review Data Set. The dataset consists of 8 columns for data such as:   1. Id- It contains the registered number for the following product 2. Profile name- It is the name of a person who has reviewed the product 3. Text- These are the comments about the product which are written by the reviewer 4. Date- Exact date when the review was given 5. Title- Written in bold statement which the reviewer want to highlight 6. Rating- The rating out of five which the customer gives about the product 7. Images- The image of the product uploaded by the reviewer |
| **4.** | **Length of Data set** | 12000 |
| **5.** | **Name of the source website -** | Amazon review exporter |
| **6.** | **URL -** | <https://mxnpro.ovh/>  <https://chrome.google.com/webstore/detail/amazon-reviews-exporter-c/njlppnciolcibljfdobcefcngiampidm/> |

**Data Extraction**

**Data Extraction**

* 1. **Import Data (.csv) ,**

The read.csv() function is used to read the file in table format and create a data frame from it. reviews <- read.csv(file.choose(),header=T)



* 1. **Viewing the data** 
     1. View(reviews)

The View() function is used to invoke a spreadsheet-style data viewer on a matrixlike R object.

Graphical user interface, text, application, email

Description automatically generated

* + 1. head(reviews)

Text

Description automatically generated

* 1. **Knowing the data** 
     1. dim(reviews)

The dim() function is used to retrieve or set the dimension of an object.

Text

Description automatically generated with medium confidence

* + 1. str(reviews)

The str() function is used to compactly display the internal structure of an R object

Text, letter

Description automatically generated

**Exploratory Data Analysis**

**Exploratory Data Analysis**

* **Text Mining using tm**

3.1 tm installation

install.packages(‘tm’)

3.2 library(tm)

3.3 corpus <- iconv(reviews$text)

Application, Word

Description automatically generated

3.4 inspect(corpus[1:5])

Text, letter

Description automatically generated

3.5 Cleaning the corpus to lower case

corpus <- tm\_map(corpus,tolower)

Text, letter

Description automatically generated

3.6 Cleaning the corpus to remove punctuations

corpus <- tm\_map(corpus, removePunctuation)

Text, letter

Description automatically generated

3.7 Cleaning the corpus to remove the numbers

corpus <- tm\_map(corpus, removeNumbers)

Text, letter

Description automatically generated

3.8 Cleaning the corpus to remove stop words.

corpus <- tm\_map(corpus, removeWords, stopwords("english"))

Text, letter

Description automatically generated

3.9 Cleaning the corpus to remove white spaces.

corpus <- tm\_map(corpus, stripWhitespace)

Text, letter

Description automatically generated

3.10 Storing the cleaned corpus in review\_final variable.

reviews\_final <- corpus

3.11 Creating a term document

tdm <- TermDocumentMatrix(reviews\_final)

tdm <- as.matrix(tdm)

tdm[1:10,1:5]

Table

Description automatically generated with medium confidence

3.12 Creating a bar plot for sum of tdm

w <- rowSums(tdm)

w <- subset(w, w>=25)

barplot(w, las = 2, col = "blue")

Graphical user interface, application

Description automatically generated

3.13 Cleaning the data again for specific non-sentimental words

corpus <- tm\_map(corpus, removeWords, c("given","bought","read","book","important","flow","ikigai","find","meaning"))

Graphical user interface

Description automatically generated

**Application of Mining Algorithm**

**Application of Mining/Analytics Algorithm**

**4.1 Sentiment Analysis**

There are a variety of methods and dictionaries that exist for evaluating the opinion or emotion in text. The tidy text package provides access to several sentiment lexicons. Three general-purpose lexicons are AFINN from Finn Arup Nielsen, Bing from Bing Liu and collaborators, and NRC from Saif Mohammad and Peter Turney. All three of these lexicons are based on unigrams, i.e., single words. These lexicons contain many English words, and the words are assigned scores for positive/negative sentiment, and also possibly emotions like joy, anger, sadness, and so forth. The NRC lexicon categorizes words in a binary fashion (“yes” / “no”) into categories of positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The Bing lexicon categorizes words in a binary fashion into positive and negative categories. The AFINN lexicon assigns words with a score that runs between -5 and 5, with negative scores indicating negative sentiment and positive scores indicating positive sentiment.

**NRC Sentiments**

The NRC Emotion Lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). The annotations were manually done by crowdsourcing.

4.2 Implementation of NRC Sentiments

sentiment\_data <- iconv(reviews$text)

s <- get\_nrc\_sentiment(sentiment\_data)

s[1:10,] #viewing sentiments for 10 records

Table

Description automatically generated

4.3 Creating a new variable in the sentiment table.

A new variable “score” is created in the sentiment table which is the difference of negative and positive sentiment score.

s$score <- s$positive - s$negative

s[1:10,]

Table

Description automatically generated

4.4 Saving the scores in a CSV file.

write.csv(x = s, file = "C:/Users/Shaun Dsilva/Desktop/SEM 8/R/Amazon Reviews/Final\_score\_ikigai.csv")

Graphical user interface, application, table, Excel

Description automatically generated

4.5 Checking the overall sentiments.

review\_score <- colSums(s[,])

print(review\_score)

A picture containing chart

Description automatically generated

**Data Visualization and Interpretation**

**Data Visualization and Interpretation**

5.1 Creating a Word Cloud of TDM

w <- sort(rowSums(tdm), decreasing = T)

set.seed(2000)

wordcloud(words = names(w),

freq = w,

max.words = 50,

random.order = T,

min.freq = 5,

colors = brewer.pal(25,"Dark2"),

scale = c(3,0.3))

Text

Description automatically generated

5.2 Creating a Bar plot of the sentiments

barplot(colSums(s),

las = 2,

col = rainbow(10),

ylab = 'Count',

main = 'Sentiment')

Chart, bar chart

Description automatically generated

5.3 Creating a 3D Pie chart of selected few sentiments

neg\_score <- colSums(s[9])

pos\_score <- colSums(s[10])

anger\_score <- colSums(s[1])

fear\_score <- colSums(s[4])

sadness\_score <- colSums(s[6])

joy\_score <- colSums(s[5])

x <- c(neg\_score,pos\_score,anger\_score,fear\_score,sadness\_score,joy\_score)

labels <- c("Negative", "Positive", "Anger", "Fear", "Sadness", "Joy")

pie3D(x,labels = labels,explode = 0.1, main = "Pie Chart of Sentiments ")

Chart

Description automatically generated

5.3 More to the list.

Adding more books / items from amazon reviews extractor and following the process all over again for these items. Then we merge all the CSV files of the sentiment scores together to get a single CSV file.

Table

Description automatically generated with low confidence

5.4 Visualizing data in Tableau

5.4.1 Filter for all books

Chart, histogram

Description automatically generated

5.4.2 Filter for book 1 (Death)

Chart, histogram

Description automatically generated

5.4.3 Filter for combination of Multiple books

Chart, histogram

Description automatically generated

**Conclusion**

**Conclusion:**

Sentiment analysis or opinion mining is one of the major tasks of NLP (Natural Language Processing). Sentiment analysis has gained much attention in recent years. In this paper, we aim to tackle the problem of sentiment polarity categorization, which is one of the fundamental problems of sentiment analysis. A general process for sentiment polarity categorization is proposed with detailed process descriptions. Data used in this study are online product reviews collected from Amazon.com. Experiments for both sentence-level categorization and review-level categorization are performed with promising outcomes. At last, we also give insight into our future work on sentiment analysis. Sentiment analysis or opinion mining is a field of study that analyses people’s sentiments, attitudes, or emotions towards certain entities. This paper tackles a fundamental problem of sentiment analysis, sentiment polarity categorization. Online product reviews from Amazon.com are selected as data used for this study. A sentiment polarity categorization process has been proposed along with detailed descriptions of each step. Experiments for both sentence-level categorization and review-level categorization have been performed.

**ACKNOWLEDGEMENT**

We would like to express our gratitude to our guide Mrs. Vandana Patil who gave us the opportunity to do a project on, “Sentiment analysis on Amazon Reviews using R”. Her constant guidance and prompt suggestions have helped us make progress in our project. We would also like to extend our gratitude to the Head of Department (INFT) Dr. Joanne Gomes and our Principal and Director for providing us with the necessary facilities required for the project.