Chapter 7
Sentiment Analysis

Abstract Sentiment or opinion analysis employs natural language processing to extract a significant pattern of knowledge from a large amount of textual data. It examines comments, opinions, emotions, beliefs, views, questions, preferences, attitudes, and requests communicated by the writer in a string of text. It extracts the writer's feelings in the form of subjectivity (objective and subjective), polarity (negative, positive, and neutral), and emotions (angry, happy, surprised, sad, jealous, and mixed). Thus, this chapter covers the theoretical framework and use cases of sentiment analysis in libraries. The chapter is followed by a case study showing the application of sentiment analysis in libraries using two different tools.

7.1 What Is Sentiment Analysis?

Sentiment analysis (also referred to as subjectivity analysis or opinion mining or emotion artificial intelligence) is a natural language processing (NLP) technique that identifies important patterns of information and features from a large text corpus. It analyzes thought, attitude, views, opinions, beliefs, comments, requests, questions, and preferences expressed by an author based on emotion rather than a reason in the form of text towards entities like services, issues, individuals, products, events, topics, organizations, and their attributes. It finds the author's overall emotion for a text where text can be blog posts, product reviews, online forums, speech, database sources, social media data, and documents. It usually consists of three elements depending on the context:

1. **Opinions or emotions**: An opinion is also referred to as polarity, whereas emotions can be qualitative such as sad, joy, angry, surprise, disgust, or happy or quantitative such as rating a movie on a scale of one to ten.
2. **Subject**: It refers to the subject of the discussion where one opinion can discuss more than one aspect of the same subject, for instance, the camera of the phone is great, but the battery life is disappointing.
3. **Opinion holder**: It refers to the author/person who expresses the opinion.

M. Lamba, M. Madhusudhan, *Text Mining for Information Professionals*,
https://doi.org/10.1007/978-3-030-85085-2_7

> March 17 2020 the journal Nature Medicine provided strong evidence against the "engineered in a lab" idea. The study found a key part of SARS-CoV-2, known as the spike protein, would almost certainly have emerged in nature and NOT as a lab creation!

a) Positive tweet

> Hi Mike what do you make of this paper? So much different info out there would like to get your take. Thanks.

b) Neutral tweet

> One year since one of the most flawed papers on the origins of SARS-CoV-2 was published……

c) Negative tweet

Fig. 7.1 Example showing different sentiments for a particular research paper. (**a**) Positive tweet. (**b**) Neutral tweet. (**c**) Negative tweet

Texts are thus categorized as *subjective* if they reflect opinion; *objective* if they express a fact; *positive* if they present a state of satisfaction, bliss, or happiness; *negative* if they present a state of dejection, disappointment, or sorrow; or *neutral* if they present a state that is neither negative nor positive. It classifies the author's feelings into polarity (positive, negative, or neutral) and subjectivity (objective or subjective). Polarity is measured on a scale of -1 to $+1$, where -1 means very negative, 0 means neutral, and $+1$ means very positive. On the other hand, subjectivity is measured on a scale of 0–1, where 0 means very objective, and 1 means very subjective. Sentiment analysis involves tasks such as sentiment classification, subjectivity classification, summarization of opinions, and sentiment extraction, among others. Figure 7.1 shows an example of three tweets segregated based on their sentiments for the same research paper.

7.1.1 Levels of Granularity

Sentiment analysis can be performed at the following levels of granularity:

1. **Document Level**: It looks at the whole document and tags each document with its sentiment. It finds polarity and subjectivity of each sentence or word and then combines them to find the polarity and subjectivity of the document where it

assumes that the individual document emphasizes a single object and consists of opinion from a single author.

2. **Sentence or Phrase Level**: It determines and tags opinions expressed in each sentence with their polarity and subjectivity. It determines the sentiment orientation for each word in the phrase or sentence and then merges them to get the sentiment of the entire phrase or sentence.
3. **Aspect or Feature Level**: It labels each word with its sentiment and identifies the entity towards which the sentiment is directed using techniques like dependency parsing. It first recognizes and extracts object feature that the opinion holder mentions; then finds if the opinion on feature was negative, neutral, or positive; and finally finds similar features.
4. **Word Level**: It uses the prior polarity of words at sentence and document levels and uses a dictionary- or corpus-based approach.

7.1.2 Approaches for Sentiment Analysis

7.1.2.1 Rule or Lexicon

It has a predefined list of words in a dictionary with valence scores manually created with rules by humans. The algorithms match the words from the lexicon to the words in the text and either sum or take the average scores for the total appearance of the sentiment for the sentence or document. It is quite fast but might fail at specific tasks because the polarity of words may change with the problem, which will not be reflected in a predefined dictionary. It can further be classified into:

1. Dictionary-based: Highly used words are collected and annotated manually, such as SentiWordNet, which used the terms from WordNet and assigned sentiment scores to them. The limitation of the dictionary-based method is that it cannot be used for domain- and context-specific orientations.
2. Corpus-based: It provides a dictionary for specific domains and tags sentiments to related words based on semantics such as WordNet or statistical technique such as latent semantic analysis (LSA).

7.1.2.2 Automatic or Supervised Machine Learning

It is an automated system based on machine learning (see Chap. 8) and heavily relies on historical labeled data with known sentiment to predict the sentiment of the new data (classification problem). It is quite powerful but might take a while to train the data. The most common features that are used in automatic sentiment classification are parts of speech, negations, opinion words and phrases, and term presence and their frequency. Figure 7.2 demonstrates the supervised machine learning process for sentiments.

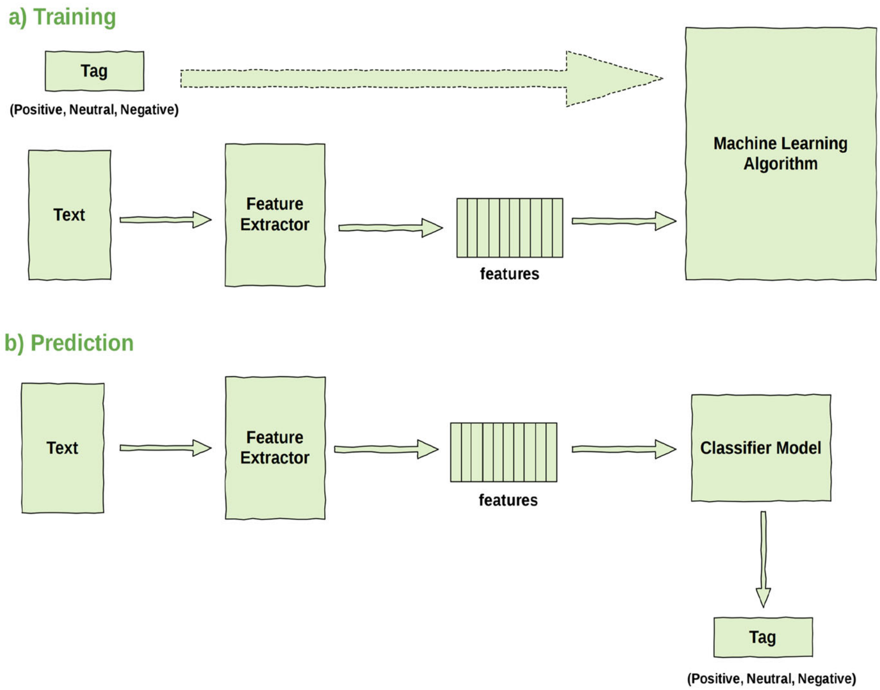


Fig. 7.2 Example showing the procedure of automatic tagging of sentiments. (**a**) Training. (**b**) Prediction

7.1.3 How to Perform Sentiment Analysis?

The typical steps to perform sentiment analysis are:

1. Pre-processing of data. It may include (i) removal of replies, mentions, URLs, hashtags, retweets, (ii) correction of spelling errors using a specific dictionary such as Hunspell dictionary, and (iii) replacing of abbreviations and shorthand notions using a specific dictionary such as SMS (Short Message Service) dictionary for social media data in addition to other pre-processing tasks covered in Sect. 3.3.
2. Feature extraction to extract the aspects from the processed data. This will be used to compute the polarity and subjectivity. This may include determining the n-gram with their frequency, part-of-speech tagging, identifying phrases and idioms, positioning of terms, negation, or syntactic patterns like collocations.
3. Performing sentiment analysis using an appropriate algorithm or open-source tool. The tools calculate the compound sentiment score for each article/post, giving an overall sentiment score.
4. Visualizing the sentiments.

There is a need to maintain order in a sentiment analysis process. Thus, a corpus is generally used instead of the bag-of-words (BOW) approach as it will fail to handle comparison and entity recognition. It is essential to determine sentiment analysis at a fine grain as the author may not like or dislike everything about a product or service and can have a point of view for several aspects that can be positive or negative.

7.1.4 Available Tools and Packages

There are various open-source tools such as `SentiStrength`,[1] `LightSide`.[2] `Sentiment Viz`,[3] `NodeXL`,[4] `AYLIEN`,[5] `MonkeyLearn`,[6] `RapidMiner`,[7] and `Orange`[8] available for non-programmers to perform sentiment analysis. Some of the popular packages in `R` to perform sentiment analysis include `syuzhet`,[9] `SentimentAnalysis`,[10] `mscstexta4r`,[11] `sentimentr`,[12] and `quanteda`,[13] whereas some popular packages to perform sentiment analysis in `Python` include `NLTK`,[14] `NLP Architect`,[15] `VADER`,[16] `Polyglot`,[17] `TextBlob`,[18] `Flair`,[19] and `CoreNLP`.[20] Most of the tools mentioned above are covered in detail in Chap. 10.

[1] http://sentistrength.wlv.ac.uk/.
[2] http://www.cs.cmu.edu/cprose/LightSIDE.html.
[3] https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/.
[4] https://nodexlgraphgallery.org/Pages/AboutNodeXL.aspx.
[5] https://aylien.com/blog/using-entity-level-sentiment-analysis-to-understand-news-content.
[6] https://monkeylearn.com/sentiment-analysis-online/.
[7] https://rapidminer.com/.
[8] https://orangedatamining.com/.
[9] https://cran.r-project.org/web/packages/syuzhet/index.html.
[10] https://cran.r-project.org/web/packages/SentimentAnalysis/index.html.
[11] https://cran.r-project.org/web/packages/mscstexta4r/index.html.
[12] https://cran.r-project.org/web/packages/sentimentr/index.html.
[13] https://cran.r-project.org/web/packages/quanteda/index.html.
[14] https://pypi.org/project/nltk/.
[15] https://intellabs.github.io/nlp-architect/sentiment.html.
[16] https://github.com/cjhutto/vaderSentiment.
[17] https://polyglot.readthedocs.io/en/latest/.
[18] https://pypi.org/project/textblob/.
[19] https://pypi.org/project/flair/.
[20] https://stanfordnlp.github.io/CoreNLP/index.html.

7.1.5 Applications

Sentiment analysis is commonly applied in the domain of reviews of customer services and products. It is usually performed to monitor customer feedback and understand their needs. Social media is the primary focal point of various sentiment analysis applications, for instance, tracking the reputation of a particular brand or real-time analysis of posts/tweets containing a particular query of interest. It also helps to present significant value to candidates running for different positions and helps the managers monitor how voters relate to their speeches, feel about various issues, and relate to the candidates' actions. Thus, sentiment analysis has wide applications in business, customer feedback, brand monitoring, reputation management, customer support, product analysis, market research, competitive research, voice of employee, voice of customer, financial marketing, and social media monitoring.

7.1.6 Advantages

1. Classifying data at a large scale based on their polarity
2. Real-time analysis
3. Comparatively few categories/attributes like polarity compared to text categorization
4. Having nondependent categories/attributes
5. Having a relation between topic, domain, and user as opposed to text categorization
6. Simple and efficient

7.1.7 Limitations

1. Complexity in determining the real meaning of the expressions expressed by the opinion holder.
2. Sarcasm, irony, and implication are common and hard to decipher.
3. Different words might have different polarity or subjectivity in different contexts.
4. Same sentence or phrases might have different meaning in different domains.
5. Sentiments can be negated in different ways and it is difficult to identify such negations.
6. Dependent on order.

7. Content-dependent opinion words cannot be processed.
8. Can result in under- or over-analyzed sentiments if the used dictionary is too sparse or exhaustive.

7.2 Sentiment Analysis and Libraries

Social media contains potentially high valuable insight into the thoughts and depositions of the general public. Many libraries turn to social media to involve their patrons in an online space as it offers real-time venues and channels of communication, information sharing for knowledge exchange, and interactive dialogue [1–6]. Various case studies on *library use* of different social media tools, viz., Pinterest, Flickr, YouTube, Facebook, and Twitter, have been studied. Several studies [5, 7–9] showed how Facebook could promote library services from a librarian's perspective. It has been shown that reference inquiries were mostly asked through Facebook when compared through phone, email, traditional face-to-face questions, instant messaging, and outside reference desk shifts [10]. In libraries, Facebook has been recognized for producing static hyperlinks to static library resources and community building, whereas Twitter has been used for communication and well-timed updates about current events and new resources [11, 12]. Sentiment analysis can help to identify influential players in the network of libraries on social media [13]. Libraries can leverage such rich data sources in making informed purchasing decisions and policies and gauge the public perception of the library's services. In libraries, sentiment analysis has been applied on library tweets [14–17], library Facebook posts[18], and tweets that might be of interest to subject librarians[19].

In libraries, sentiment analysis can be conducted on the text data retrieved from (i) social media platforms, (ii) digital or digitized survey feedback, (iii) live librarians' chat archive, (iv) reviews of books, or (v) comments on library management systems (LMS) to provide insight on the satisfaction level of the library users for the products and services offered and organized by the libraries which include workshops, seminars, and conferences. The success of any library directly depends on its patrons, so if patrons like the library's products and services, then the library is a success; otherwise the library needs to improve the products and services by making changes to them. To know if a library's products or services are successful or not, you need to analyze its patrons and one of the ways to analyze them is through sentiment analysis. Sentiment analysis computationally identifies and categorizes opinions from texts. By automatically analyzing users' survey feedback responses or their social media conversation, libraries can learn about the services and products that make their patrons happy or frustrated so that they can tailor new or modify their services and products to meet their customer's needs.

Vinit Kumar is an Assistant Professor at the Department of Library and Information Science, Babasaheb Bhimrao Ambedkar University, Lucknow, who performed the following study in association with **Maya Deori** (Research Scholar) and **Manoj Kumar Verma** (Associate Professor) from the Department of Library and Information Science, Mizoram University, Aizawl.

Story: Analysis of YouTube Video Contents on Koha and DSpace and Sentiment Analysis of Viewers' Comments

In the last decade, we have seen the adoption of the latest technological innovations in the services provided by the library and information centers. The library professionals have adopted tools ranging from integrated library management software for managing housekeeping operations to developing institutional repositories using digital content management software. Although these tools have been part of the curriculum of most library schools in India, library professionals struggle to learn the intricacies of these tools when they try to implement them practically. Professionals attempt to learn these tools by attending workshops and try to self-learn using the documentation of these tools. With the growing availability and abundance of tutorial videos about these tools on social media websites such as YouTube[a]*, professionals now have not only the opportunity to learn the features of these tools but also a chance to share their views and help each other in troubleshooting any difficulties they face while implementing these tools.*

We commenced a project to evaluate the characteristic features and opinion mining of the comments regarding the contents of videos uploaded on two tools top-rated among the library professionals of South Asian countries. We chose Koha as a representative for integrated library management software and DSpace for digital library software. Since there is no default guideline checking the quality and relevancy of the contents posted on YouTube, we analyzed the relevance and content quality satisfaction of the viewers by analyzing the sentiments expressed in the comments section of the videos.

We collected the metadata of the videos on the two selected tools using Webometric Analyst[b]*. Using the same software, we further extracted the comments on each relevant video and exported the metadata and comments in comma-separated value files. We further analyzed the dataset for the characteristic features of the videos using Google Sheets. For the text analysis of the sentiments, emotions, and intentions of the comments, we used Text Analysis API*[c] *from ParallelDots*[d] *that provides a commercial API for text analysis based on machine learning algorithms. We created a free account and deployed the Google Sheets add-on*[e] *supplied by the ParallelDots developers. The add-on returned the sentiments and emotions for each row of the Google Sheet data, which was further analyzed and visualized. One of the main reasons for using an API rather than an established implementation*

(continued)

of a text mining algorithm using some programming language was the seamless integration with Google Sheets and easy implementation. We further calculated the word frequency and other visualizations of the terms from the comments using Orange[f]*.*

We faced some difficulties while conducting this project, such as the dependability of the software Webometric Analyst to extract metadata about videos relating to Koha and DSpace. We could never confirm whether any videos were omitted from the retrieved set of videos as in manual searching; the retrieval results on YouTube are customized based on the preferences of the user searching for the videos. Similarly, the limitation for evaluating the number of comments in ParallelDots API for a free account is 1000, causing us to create a part-by-part analysis of the dataset deploying multiple free accounts. For analysis of a larger dataset, one has to buy any of the paid plans. Another limitation of the project was that since we deployed ParallelDots API, being a proprietary service, we could not evaluate the algorithm and had to depend on the output received from the API.

We assume that this project will be of use to the library professionals who are eager to learn and make informed decision about the relevance and quality of the contents of the videos on both Koha and DSpace and also to the professionals who are video content creators so they can get an overview of the content they are uploading and get feedback for their improvement. Further, there is a scope to conduct similar studies using sophisticated software on text mining.

We published the results of the project as a research article that may be cited as Deori, M., Kumar, V., and Verma, M.K. (2021), "Analysis of YouTube video contents on Koha and DSpace, and sentiment analysis of viewers' comments," Library Hi Tech. https://doi.org/10.1108/LHT-12-2020-0323.

References

[a] https://www.youtube.com
[b] http://lexiurl.wlv.ac.uk/
[c] https://komprehend.io/
[d] https://www.paralleldots.com/
[e] https://komprehend.io/add-on
[f] https://orangedatamining.com/

7.2.1 Use Cases

7.2.1.1 Scite

Scite[21] is a service based on citations that provides citation context and classifies them as supporting or contrasting. It is a sentiment analysis-based service that uses a deep learning model to automatically classify each citation.

7.2.1.2 Virtual Librarian Chat

Sentiment analysis of virtual chat will help to determine users' satisfaction level with the services provided by the library and its staff [20].

7.2.1.3 Altmetrics

Sentiment analysis can help to analyze the *qualitative* altmetric data to characterize the sentiments or feelings of the people sharing their thoughts about a particular topic. Thus, one can analyze authors' sentiments for posts on common themes and identify outlets that publish more positive or negative news/articles/posts. Analyzing sentiment analysis of social media data related to research papers at scale can give a sense of what people think about research and how they are engaging with it [21–23].

7.2.1.4 Marketing

Lamba and Madhusudhan [19] emphasized in their paper that "social media mining (SMM) provides a dynamically active platform to connect with libraries' users (both active and potential) to market the library's products and services." They suggested that "in today's digital environment, librarians should analyze user-generated content over social media for decision-making; understanding their users' needs; making predictions; conducting a voluntary survey and analyzing users' opinions towards a particular issue, product or service." They further introduced "a new way of conducting a *SWOT analysis* for libraries by performing sentiment analysis on library's social media data where positive posts can be considered as strengths; negative posts can be treated as weaknesses; the problems expressed by their users can be viewed as the opportunities; the sentiment analysis of other libraries which they view as competitors can be viewed as the threats posed to the functionality and utilization of the library" [19].

[21] https://scite.ai/.

7.3 Case Study: Sentiment Analysis of Documents Using Two Different Tools

7A: RapidMiner

About the Case Study This case study has been adapted from Lamba and Madhusudhan [19] to illustrate a sentiment analysis using RapidMiner.

Problem If you have text data such as tweets, virtual chat with a librarian, Facebook posts, and book reviews and want to provide temporal information service based on the topics trending on social media.

Goal To identify the sentiments and emotions of the public.

About the Tool Refer to Chap. 10, Sect. 10.2.3, to know more about `RapidMiner`.

Theory "Mining online opinion is a form of sentiment analysis that is treated as a difficult text classification task. It analyzes emotions, opinions, comments, views, beliefs, questions, requests, preferences, and attitudes expressed by the author in the form of text and determines the overall essence of the text expressed by the author" [19].

Methodology Figure 7.3 shows the workflow of sentiment analysis used in the RapidMiner platform. For retrieving data from Twitter, the API and Key were obtained by creating an app at the Twitter Developer[22] site and were then validated into the Twitter operator of RapidMiner (Fig. 7.4). Similarly, the API ID and Key were obtained from AYLIEN's Developer Portal[23] and were then validated into the AYLIEN operator of RapidMiner (Fig. 7.5).

Twenty bi-gram queries "related to the facets of productivity were searched on Twitter from 9th February 2018 to 21st February 2018 for a *recent or popular* type of tweets with a limit of 1500 tweets per query in the English language" [19]. The mining of tweets was followed by performing sentiment analysis using the AYLIEN operator. Therefore, 20 different Excel files[24] were downloaded from RapidMiner after each query search that consisted of the tweet, polarity (negative, positive, or neutral), and subjectivity (objective or subjective).

Results This study helped to comprehend the sentiment and emotions expressed by the public on Twitter for 20 different queries related to productivity. 6416 tweets and retweets were identified for 13 days' period. The following observations were made (Fig. 7.6):

[22] https://developer.twitter.com/en/apps.

[23] https://aylien.com/.

[24] https://github.com/textmining-infopros/chapter7/blob/master/7a_processed_dataset.rar.

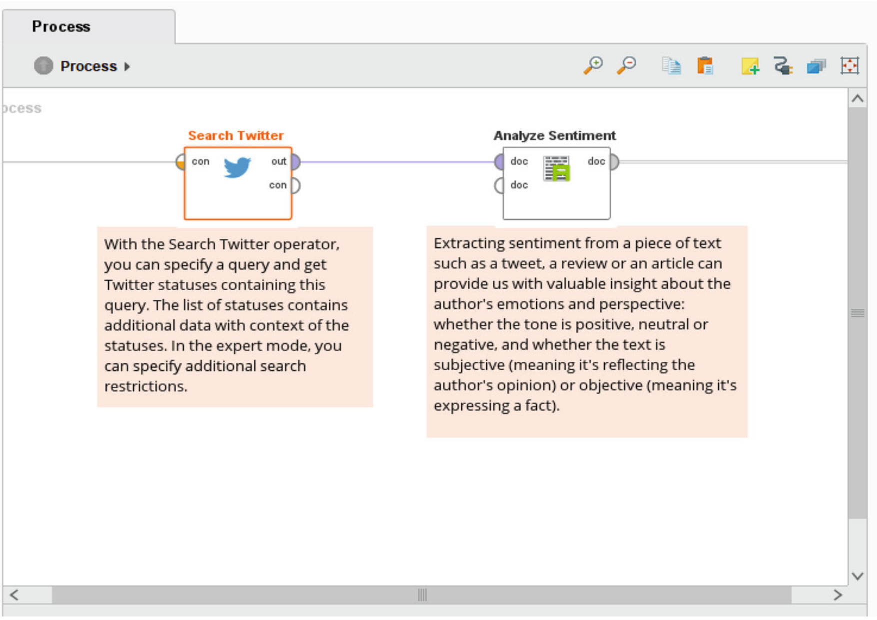


Fig. 7.3 Screenshot showing the workflow of sentiment analysis

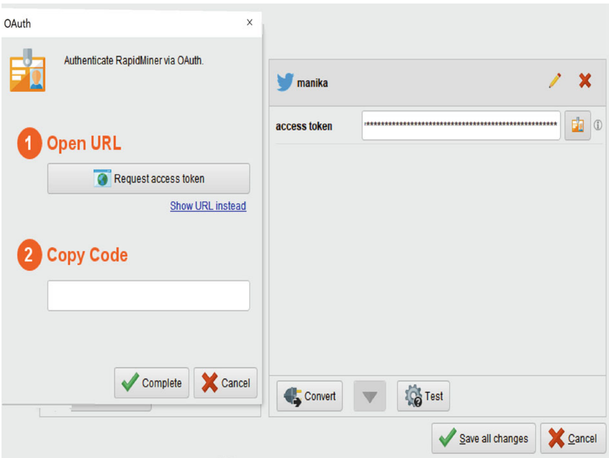


Fig. 7.4 Screenshot showing authentication in Twitter operator

- *Digital productivity* was the most popular facet, whereas *managerial productivity* was the least popular facet.
- Tweets for *low productivity* facet was the most negative, whereas *material productivity* facet had the most positive tweets.
- Tweets for Asian productivity facet were most neutral, whereas *material productivity* facet had the least neutral tweets.
- Tweets for *fish productivity* facet was most subjective, whereas *crop productivity* facet was the most objective.

7B: R

Problem If you have text data such as tweets, virtual chat with a librarian, Facebook posts, and book reviews and want to determine the satisfaction level of the users with the provided services and products.

Goal To identify the sentiments and emotions of the users.

Prerequisite Familiarity with the `R` language.

Virtual RStudio Server You can reproduce the analysis in the cloud without having to install any software or downloading the data. The computational environment runs using BinderHub. Use the link (https://mybinder.org/v2/gh/textmining-infopros/chapter7/master?urlpath=rstudio) to open an interactive virtual RStudio environment for hands-on practice. In the virtual environment, open the `sentiment_analysis.R` file to perform sentiment analysis.

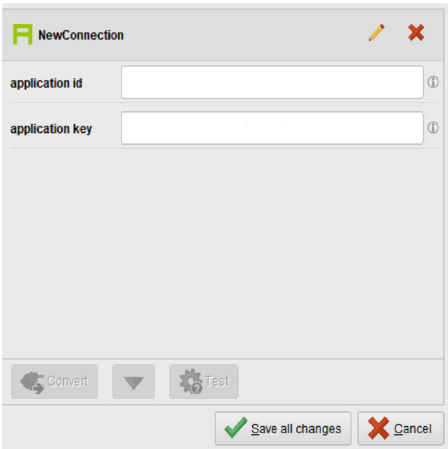
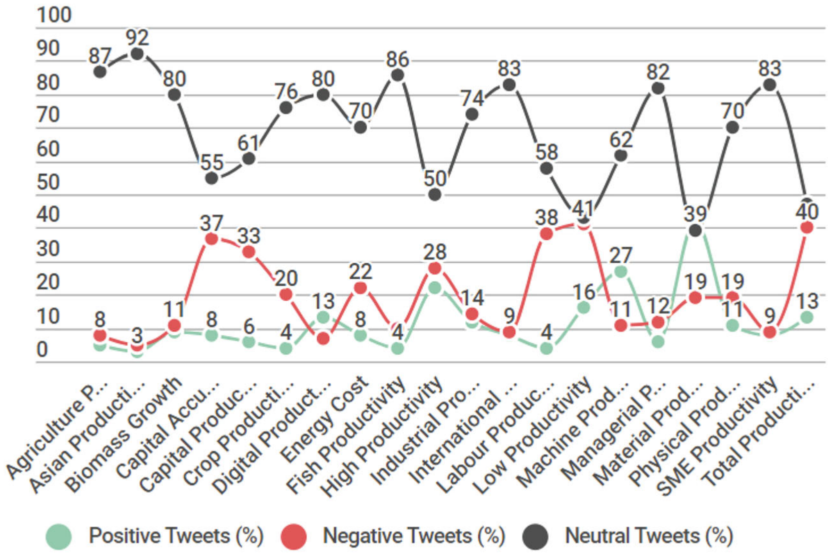
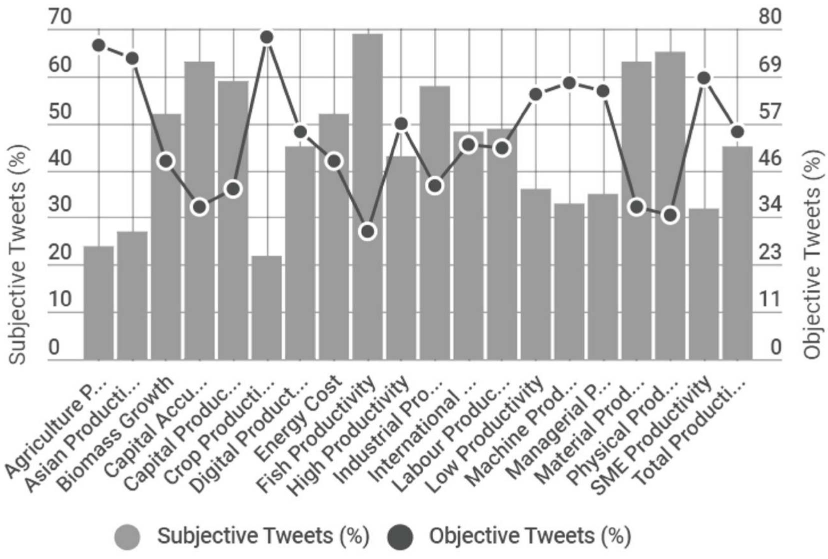


Fig. 7.5 Screenshot showing authentication in AYLIEN operator



(a) Polarity Percentage



(b) Subjectivity Percentage

Fig. 7.6 Percentage comparison between different polarities and subjectivities for the studied productivity facets (©2018 Springer Nature, all rights reserved—reprinted with permission from Springer Nature, published in Lamba and Madhusudhan [19]). (**a**) Polarity percentage. (**b**) Subjectivity percentage

Virtual Jupyter Notebook You can reproduce the analysis in the cloud without having to install any software or downloading the data. The computational environment runs using BinderHub. Use the link (https://mybinder.org/v2/gh/textmining-infopros/chapter7/master?filepath=Case_Study_7B.ipynb) to open an interactive virtual Jupyter Notebook for hands-on practice.

Dataset A CSV file[25] containing 5000 book reviews web-scrapped from Amazon in 2018. The data is available in public domain at https://www.kaggle.com/shrutimehta/amazon-book-reviews-webscraped.

About the Tool Refer to Chap. 10, Sect. 10.2.1, to know more about `R`.

Theory The `syuzhet` package of `R` was used to perform sentiment and emotion analysis. It has four sentiment dictionaries to extract sentiment in addition to the sentiment extraction tool developed by the NLP group of Stanford.

Methodology and Results The libraries and the dataset required to perform sentiment analysis in `R` were loaded.

```
#Load libraries
library(syuzhet)
library(tm)
library(twitteR)

  #Load dataset
data<- read.csv("https://raw.githubusercontent.com/textmining
-infopros/chapter7/master/7b_dataset.csv")
```

The `syuzhet` package works only on vectors. So, the data was converted to a vector.

```
vector <- as.vector(t(data))
```

For each book review, scores were determined for different emotions, where a score with value 0 implies the emotion is not associated with the review, and the score of value 1 means there is an association between the emotion and the review. Subsequently, a higher score indicates stronger emotion.

[25] https://github.com/textmining-infopros/chapter7/blob/master/7b_dataset.csv.

Table 7.1 Sentiment score for different sentiments

Sentiment score	Sentiment
< 0	Negative
$= 0$	Neutral
> 0	Positive

```
#Sentiment analysis
emotion.data <- get_nrc_sentiment(vector)
```

The following output gives a better representation of the book reviews with the associated emotions.

```
emotion.data2 <- cbind(data, emotion.data)
```

Sentiment scores were then computed for each book review using the built-in dictionary of the package that assigns sentiment score to different words. Table 7.1 shows the sentiment for the different range of sentiment scores.

```
sentiment.score <- get_sentiment(vector)
```

Reviews were then combined with both emotion and sentiment scores.

```
sentiment.data = cbind(sentiment.score, emotion.data2)
```

Positive, negative, and neutral reviews were then segregated and saved in three different CSV files.[26,27,28] Out of 5000 book reviews, 3587 were identified as positive, 1349 were identified as negative, and 64 were identified as neutral.

[26] https://github.com/textmining-infopros/chapter7/blob/master/positive_book_reviews.csv.
[27] https://github.com/textmining-infopros/chapter7/blob/master/negative_book_reviews.csv.
[28] https://github.com/textmining-infopros/chapter7/blob/master/neutral_book_reviews.csv.

```
#Getting positive, negative, and neutral reviews with associated scores
positive.reviews <- sentiment.data[which
(sentiment. data$sentiment.score > 0),]
write.csv(positive.reviews, "positive.reviews.csv")

negative.reviews <- sentiment.data[which
(sentiment.data$sentiment.score < 0),]
write.csv(negative.reviews, "negative.reviews.csv")

neutral.reviews <- sentiment.data[which
(sentiment.data$sentiment.score == 0),]
write.csv(neutral.reviews, "neutral.reviews.csv")
```

Lastly, a graph was plotted to visualize how the narrative is structured with the sentiments across the book reviews.

```
#Plot1: Percentage-Based Means
percent_vals <- get_percentage_values
(sentiment.score, bins=20)

plot(percent_vals,
type="l",
main="Amazon Book Reviews using Percentage-Based Means",
xlab="Narrative Time",
ylab="Emotional Valence",
col="red")
```

In Fig. 7.7, the x-axis presents the flow of time from start to end of the book reviews, and the y-axis presents the sentiments. In order to compare the trajectory of shapes, the text was divided into an equal number of chunks, and then the mean sentence

Fig. 7.7 Graph showing Amazon book reviews using percentage-based means

valence for each chunk was calculated. For this case study, the sentiments from the reviews were binned into 20 chunks where each chunk had 20 sentences.

The figure shows that the book reviews remain in the positive zone for all the 20 chunks. It dropped towards a comparatively less positive zone at many instances but never reached a neutral or negative zone. The limitation of the *percentage-based sentiment mean normalization* method is that in large texts, extreme emotional valence gets watered down, and the comparison between different books or texts becomes difficult.

To overcome the limitations of the *percentage-based sentiment mean normalization* method, the *discrete cosine transformation (DCT)* method was used as it gives an improved representation of edge values.

#Plot2: Discrete Cosine Transformation (DCT)
```
dct_values <- get_dct_transform(sentiment.score,
low_pass_size = 5,
x_reverse_len = 100,
scale_vals = F,
scale_range = T)

plot(dct_values,
type ="l",
main ="Amazon Book Reviews using Transformed Values",
xlab = "Narrative Time",
ylab = "Emotional Valence",
col = "red")
```

In Fig. 7.8, the x-axis presents the flow of time from start to end of the book reviews, and the y-axis presents the sentiments where 5 reviews were retained for low pass filtering, and 100 were returned. Figure 7.8 shows the transformed graph from the percentage-mean method. The reviews were of negative valence at the beginning that changed to positive valence and again dropped towards negative valence.

Moreover, eight different emotions, viz., anticipation, trust, joy, sadness, surprise, fear, anger, and disgust, were visualized using a bar plot.

#Plot3: Emotions Graph
```
barplot(sort(colSums(prop.table(emotion.data[, 1:8]))),
horiz=TRUE,
cex.names=0.7,
las=1,
main="Emotions in Amazon Book Reviews",
xlab = "Percentage")
```

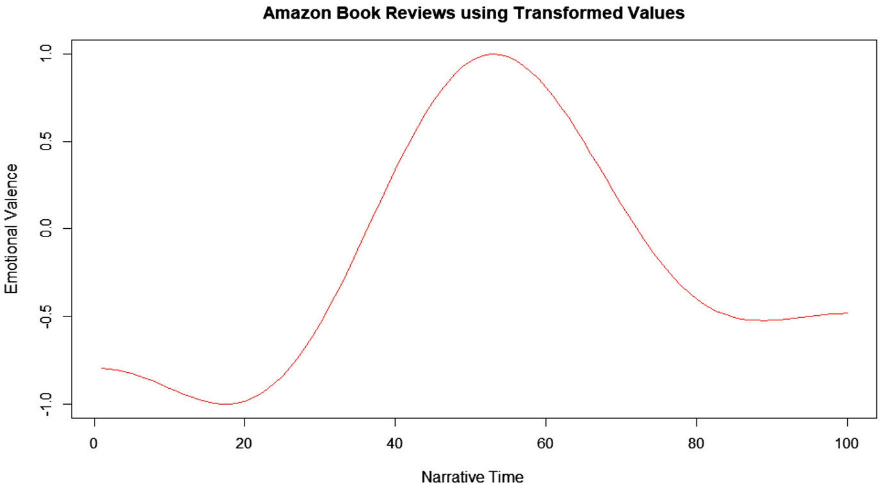


Fig. 7.8 Graph showing Amazon book reviews using discrete cosine transformation (DCT)

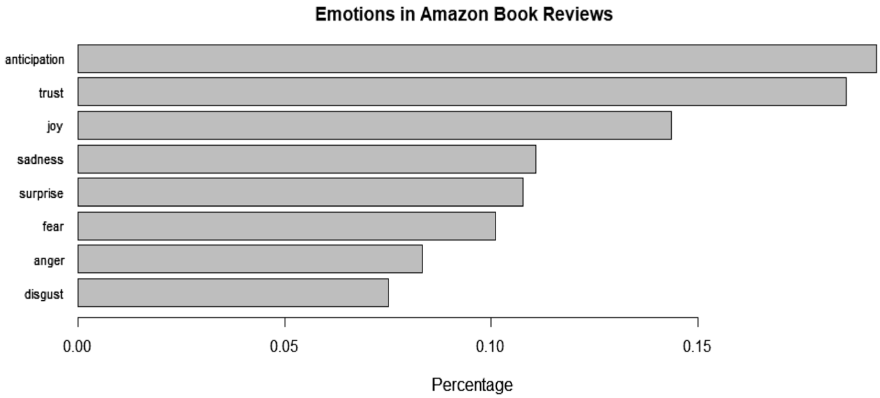


Fig. 7.9 Graph showing emotions for Amazon book reviews

Figure 7.9 shows that more than 15% of reviewers showed anticipation and trust emotion in their book reviews. Around 10% of the reviewers showed sadness, surprise, and fear emotions in their reviews, whereas the remaining reviewers showed anger and disgust.

References

1. Palmer S (2014) Characterizing university library use of social media a case study of Twitter and Facebook from Australia. J Acad Librarianship 40(6):611–619. https://doi.org/10.1016/j.acalib.2014.08.007
2. Burkhardt A (2010) Social media: a guide for college and university libraries. College Res Libraries News 71(1):10–24. https://crln.acrl.org/index.php/crlnews/article/view/8302
3. Collins G, Haines L (2012) Measuring libraries' use of YouTube as a promotional tool: an exploratory study and proposed best practices. J Web Librarianship 6(1):5–31. http://dx.doi.org/10.1080/19322909.2012.641789
4. Aziz NA, Chia YB, Loh H (2010) Sowing the seeds: towards reaping a harvest using social web applications in Nanyang Technological University Library. In: World library and information congress: 76th IFLA general conference and assembly, Gothenburg. https://www.ifla.org/past-wlic/2010/245-aziz-en.pdf. Accessed 24 May 2021
5. Aharony N (2012) Facebook use in libraries: an exploratory analysis. Aslib Proc 64:358–372. https://doi.org/10.1108/00012531211244725
6. Kim Y, Abbas J (2010) Adoption of Library 2.0 functionalities by academic libraries and users: a knowledge management perspective. J Acad Librarianship 36(3):211–218. http://dx.doi.org/10.1016/j.acalib.2010.03.003
7. Landis C (2007) Social networking sites: getting friendly with our users. College Res Libraries News 68(11):709–712. https://doi.org/10.5860/crln.68.11.7907
8. Matthews B (2006) Do you Facebook: networking with students online. College Res Libraries News 67(5):306–307. https://doi.org/10.5860/crln.67.5.7622
9. Miller SE, Jensen LA (2007) Connecting and communicating with students on Facebook. Comput Libraries 27(8):18–22
10. Mack D, Behler A, Roberts B, Rimland E (2007) Reaching students with Facebook: data and best practices. Electron J Acad Special Librarianship 8(2). https://digitalcommons.unl.edu/ejasljournal/85/
11. Chen DY-T, Chu SK-W, Xu S-Q (2012) How do libraries use social networking sites to interact with users. Proc Amer Soc Inf Sci Technol 49:1–10. https://doi.org/10.1002/meet.14504901085
12. Salisbury L, Laincz J, Smith J (2012) Science and Technology Undergraduate Students' Use of the Internet, Cell Phones and Social Networking Sites to Access Library Information. University Libraries Faculty Publications and Presentations. https://scholarworks.uark.edu/libpub/10. Accessed 24 May 2021
13. Yep J, Brown M, Fagliarone G, Shulman J (2017) Influential players in Twitter networks of libraries at primarily undergraduate institutions. J Acad Librarianship 43:193–200. https://doi.org/10.1016/j.acalib.2017.03.005
14. Lund BD (2020) Assessing library topics using sentiment analysis in R: a discussion and code sample. Public Serv Quarterly 16(2):112–123. https://doi.org/10.1080/15228959.2020.1731402
15. Patra SK (2019) How Indian libraries tweet? Word frequency and sentiment analysis of library tweets. Ann Library Inf Stud 66(4):131–139. http://op.niscair.res.in/index.php/ALIS/article/view/26636/465477307
16. Al-Daihani SM, Abrahams A (2016) A text mining analysis of academic libraries' Tweets. J Acad Librarianship 42(2):135–143. https://doi.org/10.1016/j.acalib.2015.12.014
17. Stewart B, Walker J (2018) Build it and they will come? Patron engagement via Twitter at historically black college and university libraries. J Acad Librarianship 44:118–124. https://doi.org/10.1016/j.acalib.2017.09.016
18. Al-Daihani SM, Abrahams A (2018) Analysis of academic libraries' Facebook posts: text and data analytics. J Acad Librarianship 44:216–225. https://doi.org/10.1016/j.acalib.2018.02.004

19. Lamba M, Madhusudhan M (2018) Application of sentiment analysis in libraries to provide temporal information service: a case study on various facets of productivity. Soc Netw Anal Min 8:63. https://doi.org/10.1007/s13278-018-0541-y
20. Logan J, Barrett K, Pagotto S (2019) Dissatisfaction in chat reference users: a transcript analysis study. College Res Libraries 80:925. https://doi.org/10.5860/crl.80.7.925
21. Friedrich N, Bowman TD, Stock WG, Haustein S (2015) Adapting sentiment analysis for tweets linking to scientific papers. In: 15th international society of scientometrics and informetrics conference (ISSI 2015), Istanbul. https://arxiv.org/abs/1507.01967. Accessed 21 May 2021
22. Hassan S-U, Saleem A, Soroya SH, et al (2020) Sentiment analysis of tweets through altmetrics: a machine learning approach. J Inf Sci 0165551520930917. https://doi.org/10.1177/0165551520930917
23. Hassan S-U, Aljohani NR, Tarar UI, et al (2020) Exploiting Tweet Sentiments in Altmetrics Large-Scale Data. https://arxiv.org/abs/2008.13023. Accessed 21 May 2021