Publisher

UNIVERSITY OF SOUTH FLORIDA M3 CENTER

Advances in Hospitality and Tourism Information Technology

Editors

DR. CIHAN COBANOGLU, DR. SEDEN DOGAN, DR. KATERINA BEREZINA, & DR. GALEN COLLINS

ISBN 978-1-7321275-8-6

Co-Editors

- **Dr. Cihan Cobanoglu**, University of South Florida, USA
- **Dr. Seden Dogan**, Ondokuz Mayis University, Turkey
- **Dr. Katerina Berezina**, University of Mississippi, USA
- **Dr. Galen Collins,** Northern Arizona University, USA

Editorial Assistants

- **Luana Nanu,** Auburn University, USA
- **Khuraman Shahtakhtinskaya,** University of South Florida, USA

ADVANCES IN HOSPITALITY AND TOURISM INFORMATION TECHNOLOGY

For all chapters, please visit: https://www.m3center.org/ahtit

ISBN 978-1-7321275-8-6

© University of South Florida M3 Publishing 2021

This imprint is published by University of South Florida M3 Publishing

The registered company address is: 8350 N Tamiami Trail, Sarasota, FL 34243 USA

Chapter 16

Tourism Products and Sentiment Analysis

Ibrahim Akin Ozen

Nevsehir Haci Bektas Veli University, Turkey

SUMMARY

Reviews about tourism products in online environments are an important data source for tourism businesses, destination managements and tourists. Tourist reviews online are completely unbiased reviews created voluntarily by tourists. Therefore, important feedback is provided for tourism businesses and destinations in the evaluation of tourism products. Collecting and analyzing tourist comments and transforming them into strategic information will create an important competitive power. Sentiment analysis, which is a sub-field of text mining, is a field of study that analyzes people's ideas and thoughts about tourism products and services from text-based comments. Sentiment analysis can be applied at the document level, sentence level and aspect-based sentiment levels. In this chapter, sentiment analysis methods are examined by using data collected from online platforms where tourism products are evaluated. Analyzing online tourist reviews using text mining methods will provide important opportunities for stakeholders in the tourism sector. These opportunities can be explained in terms of destination managements, tourism businesses and tourists. Understanding and interpreting the destination for destination management will thus provide opportunities for the creation of the brand value and image of the destination. It will also make an important contribution to the determination of tourist needs in the destination and meeting these needs. In terms of tourism businesses, evaluating the products and services they offer to tourists will create an opportunity to manage customer relations by discovering the negative and positive aspects. In addition, tourism businesses will have the opportunity to develop or improve their products and services in order to gain product and price advantage by evaluating their competitors. Tourists will be able to use it to make better travel plans.

Introduction

Technological developments cause changes in tourist behaviors and also in the way tourist travels are made. While the tourists make their travel plans, they constantly interact with online blogs and social media during and after the travel, especially thanks to mobile connections.

Potential tourists thinking about traveling to a destination start by first dreaming of their travel plans. At this stage, it is inspired by both online and offline sources of information. In the next stage, tourists plan for their travels. Planning provides the opportunity for tourists to define and compare transportation, accommodation and destination activities using various sources of information (social media, forums, blogs). In the next stage, tourists choose the most suitable supplier for their travels and make a reservation decision. The realization of travel, namely touristic experience, may be directly related to the selection made in the previous stages. The last stage of the journey is seen as the stage of evaluating the experiences of the tourists. At this stage, tourists share their positive or negative thoughts on their online experiences. Digital texts created in these new tourist processes are seen as an important opportunity for tourists and suppliers to get to know and understand each other. These opportunities should be evaluated in terms of suppliers, namely those offering touristic products (tourism enterprises, destination management), and tourists.

Suppliers, such as destinations, tourism businesses and promotional agencies, want to learn what tourists think about themselves (image, perceptions, brand awareness, etc.). The suppliers aim to understand and model the preferences, opinions, intentions of potential tourists both at an individual and mass level. This understanding will help businesses improve their products and services, facilitate access to potential tourists and ensure that they are relevant at different stages of the travel process.

Tourists need useful and relevant information to avoid problems and to organize their travel in the best way possible. In this way, they make risk-free purchases in accordance with their preferences and enjoy their travels and finally share their experiences in digital media. These shares are gaining importance day by day in parallel with the developments in information and communication technologies in social networks and e-commerce sites. In the literature, these textual content shares are referred to as user-generated content (UGC) or Electronic word-of-mouth (e-WOM). The evaluation of the textual content created by the tourist using scientific methods can be done with text mining methods, which is a subfield of data mining.

Text Mining

Today's technological developments and the decrease in the cost of access to technology have led people of all ages to create more digital texts. People in virtual environments prefer to use digital texts, abbreviations and figures (emoji) to express themselves. These complex digital texts and expressions make it difficult to understand the people and feelings that exist in virtual environments. Therefore, efforts to understand people from digital texts are seen as one of the most important problems of today. Text mining, seen as a sub-discipline of data mining, makes an important contribution to efforts to extract understandable, interpretable information from unstructured digital texts. According to some definitions in the literature on text mining;

Text mining is the process of editing irregular texts into digital format and extracting qualified information. It is also known as text data mining or knowledge discovery from textual databases (Delen & Crossland, 2008).

Text mining; It is a new and exciting research area used to solve today's information bombardment problem using data mining, machine learning, natural language processing (NLP), information retrieval (IR) and information management techniques (Feldman & Sanger, 2007).

According to the definitions in the paragraphs above, the main task of text mining is seen as the process of revealing useful information by making use of the texts in web environments. However, the diversity of web environments and the collection of high-quality textual data is a difficult process to collect and process.

Today user-generated content in areas such as Web 2.0 supported forums, blogs, social networking sites creates huge data sets. Extracting the necessary information from these high data sets is a very difficult process to perform with traditional information systems. As a result, text mining is becoming an increasingly important field of study to obtain understandable information by processing unstructured large text data.

There are many software and online tools for text mining. These software and tools can be classified as open access and commercial software. Table 2 shows these tools, software and functions.

3

Table 1. Text Analysis Software and Tools.

Software/Tool Name	Type	Functions	Web site
Aylien	Online and Api support	Sentiment Analysis Topic Extraction Text Clustering Text Classification Text Summarization	aylien.com
Meaning cloud	Online and Api support	Sentiment Analysis Topic Extraction Text Clustering Text Classification Text Summarization	meaningcloud.com
Basis Technology Rosette	Online and Api support	Sentiment Analysis Linguistic analysis, Statistical modeling, and machine learning	www.rosette.com
Optiwisdom	Online and Api support	Market Analysis Sentiment Analysis Social media analysis	optiwisdom.com
Voyant	Online	Document Analysis Trend Analysis Word Link Text visualization	voyant-tools.org
Pulsar	Online	Social media analysis Trend Analysis	www.pulsarplatform.com
Semrush	Online	Market Analysis Social media analysis Keyword analysis	semrush.com
Leximancer	Online and Commercial Software	Content Analysis	info.leximancer.com
SentiStrength	Online	Sentiment Analysis	http://sentistrength.wlv.ac.uk/
Rapidminer	Open Access Software	Data mining and text analytics, text processing	Rapidminer.com
Knime	Open Access Software	Data mining and text analytics, text processing	Knime.com
Weka	Open Access Software	Data mining and text analytics, text processing	https://sourceforge.net/projects/weka/
SAS Text Miner	Commercial Software	Text processing and analysis, Document theme discovery	www.sas.com/
R Programming	Open Access Software	Data transformation and text analysis	www.rproject.org/
IBM SPSS Predictive Analytics	Commercial Software	Data mining and text mining,	developer.ibm.com/predictiveanalytics

Text Mining Methods

Although text mining is basically seen as a subfield of data mining using mathematical and statistical methods, it contains important differences in terms of the data set they use. While data mining focuses mostly on structural data in relational databases, text mining focuses more on non-structural texts written in natural language. Text mining has a common working area with many disciplines to analyze data consisting of natural language texts (Miner et al., 2012). The disciplines associated with text mining are shown in Figure 1.

4
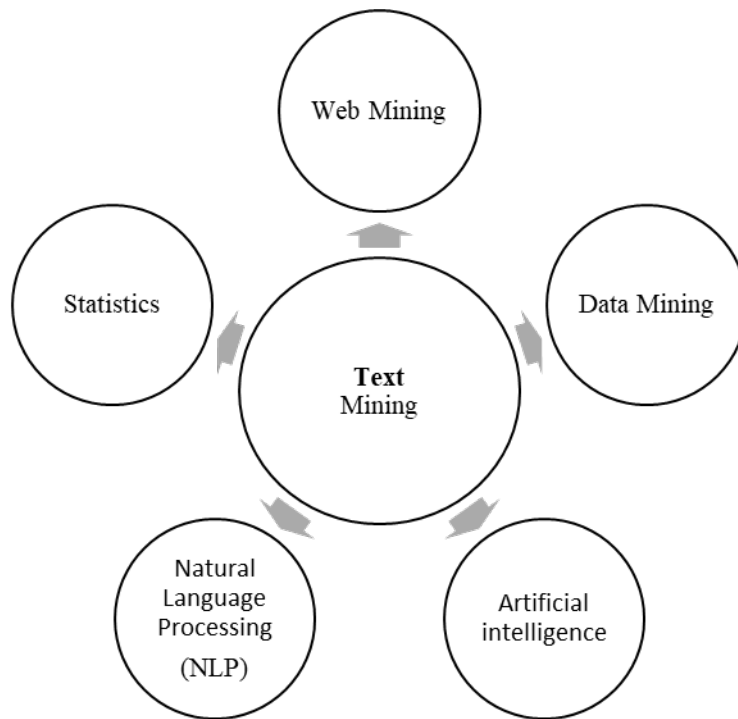
Figure 1. The disciplines associated with text mining.

Text mining utilizes disciplines and uses text classification, text clustering, topic extraction, sentiment analysis or opinion mining methods (Li et al., 2019).

Text Classification: It aims to classify text according to predetermined titles in texts written in spoken language. In text classification, it is used in many practical applications such as indexing documents, filtering, automatically hierarchically arranging web pages depending on a controlled vocabulary (Kim, 2014).

Text Clustering: A grouping of similar documents without a predetermined set of categories. In other words, clustering is the process of grouping the texts by considering their similarities among themselves (Aggarwal & Zhai, 2012).

Topic Extraction: It is used to extract topics or opinions of the person who wrote the text from large-scale text data. This method gives some information to the decision maker and helps to identify associated feelings (Hou-Feng, 2011).

Another text mining method is sentiment analysis. Since sentiment analysis is the main subject of this study, it will be examined in detail in the following sections.

5

Sentiment Analysis (Opinion Mining)

Sentiment analysis (SA) or opinion mining (OM) is a text mining method that is closely related to the natural language processing (NLP) discipline, which has a wide range of research areas.

The concepts Sentiment Analysis (SA), Opinion Mining (OM),) and Subjectivity Analysis are broadly used as synonyms. According to some researchers, these concepts are explained as follows:

SA aims to extract emotional expressions in texts. It is also used in studies of extracting mood, opinion, and more complex emotions from texts (Seker, 2016).

SA or OM is a study that analyzes people's ideas and thoughts about assets such as products and services in the text (Agarwal & Mittal, 2016).

As can be seen from the above definitions, SA or OM is often used interchangeably. However, OM makes polarity detection, whereas SA focuses on emotion recognition. Because the identification of sentiment is often exploited for detecting polarity, the two fields are usually combined under the same umbrella or even used as synonyms(Cambria et al., 2013).

SA researches is categorized as document level (Wilson et al., 2005), sentence level (Meena & Prabhakar, 2007) and aspect / feature-level sentiment analyzes (Jo & Oh, 2011).

Document-level Sentiment Analysis: Document-level sentiment analysis classifies an assessment document as containing positive or negative polarity. It views a record as an individual unit (Agarwal & Mittal, 2016). This level is linked to the job called document-level sentiment classification. Nevertheless, if a document presents several sentences coping with different aspects or entities, then your sentence level is more desirable (Serrano-Guerrero et al., 2015).

Sentence-level Sentiment Analysis: Sentence-level sentiment analysis is closely related to document-level sentiment analysis. Because the document has a structure consisting of sentences (Wilson et al., 2005). Sentiment analysis at the sentence level expresses the opinion or sentiment expressed in the sentence.

Both the document-level and the sentence-level sentiment analysis do not detect what exactly people liked and did not like (Agarwal & Mittal, 2016).

Aspect-based Sentiment Analysis (ABSA): In sentiment analysis, it is often not enough to determine the opinion or polarity of the opinion that is mentioned in a text alone. It is also necessary to find out on which entities the opinion is expressed. Aspect-based sentiment analysis is the process of finding out for which entities the opinions in a document or sentence are specified (Liu, 2012). Aspect based sentiment analysis is also called feature-level opinion mining (Hu &

6

Liu, 2004; Thet et al., 2010). A text example applied aspect-based sentiment analysis can be given as follows;

For instance, "**although the service isn't that good, I still like the food**"; with this example, "service" and "food" are two entities regarding which thoughts and opinions are indicated. Aspect-based sentiment analysis model, first of all, recognizes these entities, and furthermore, opinions regarding these entities will be identified (Hu & Liu, 2004; Liu, 2012).

Sentiment Analysis Techniques

There are two main approaches in current scientific studies in the field of sentiment analysis. These are machine learning and lexicon-based sentiment analysis approaches.

The main purpose of both approaches is based on the classification of sentiments (positive, negative and neutral) obtained from texts containing sentiment. Many alternative classification techniques can be used while analyzing sentiment. These techniques are shown in Figure 2.
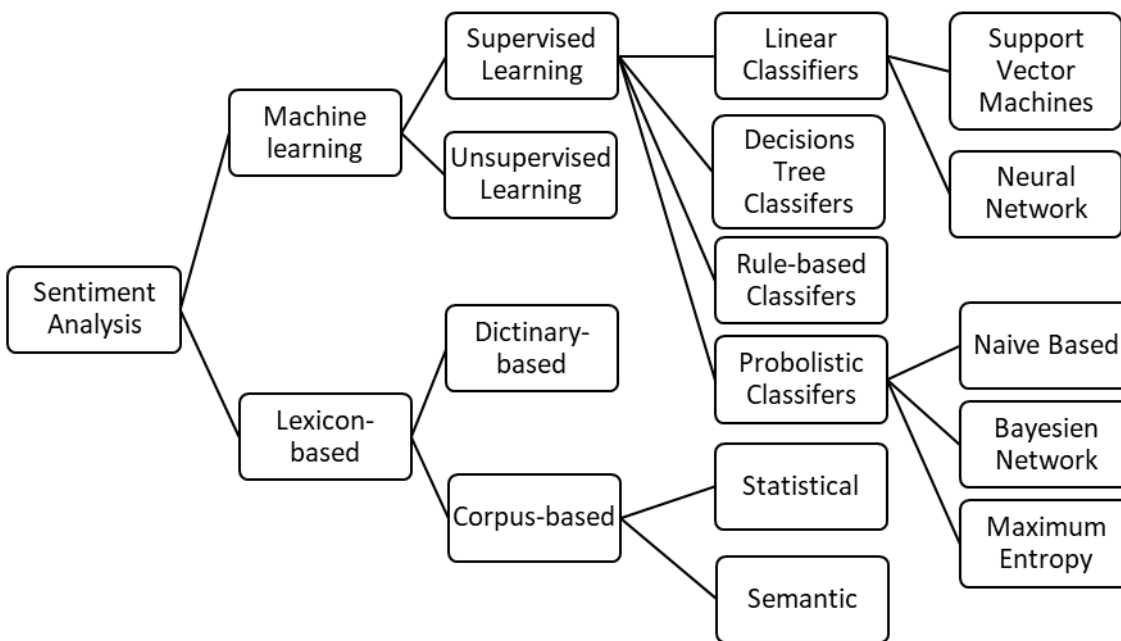
Figure 2. Sentiment Analysis Classification Techniques.

Source: (Medhat et al., 2014)

Machine Learning Approaches

In machine learning methods, algorithms and linguistic features are used to make text classification. The machine learning approach is examined in two categories as supervised and unsupervised learning methods.

The supervised learning method aims to model the function that includes these data and results, using data that has already been observed and whose results are known (labeled). Unsupervised learning includes methods based on the discovery of patterns hidden from unlabeled data, unlike supervised learning (Agarwal & Mittal, 2016). Regarding supervised methods, Maximum Entropy (ME), Support Vector Machines (SVM), Naive Bayes (NB) are a few of the most typical approaches used (Chenlo & Losada, 2014; Rushdi Saleh et al., 2011). While semi-supervised and unsupervised methods are suggested when it's impossible with a preliminary group of tagged documents/opinions to categorize the others of things (Xianghua et al., 2013). Besides, hybrid methods, combining supervised and unsupervised techniques, or semi-supervised techniques even, can be utilized to categorize sentiments.

Lexicon-Based Approaches

Two sub classifications can be found here: Dictionary-based and Corpus-based approaches.

Dictionary-based Approach: The dictionary-based approach depends on uncovering opinion seed words, and then searches the dictionary of their synonyms and antonyms. The dictionary-based approach focuses on compiling word lists based on synonyms and antonyms for every word. This process begins by obtaining a small set of words with known positive or negative emotion and an algorithm is utilized to increase this list using on-line dictionaries (Buzova et al., 2019). For this purpose, many dictionaries that give positive, negative or neutral opinions to words or phrases on the internet have been developed. In this study, some of the known online dictionary databases are presented in the Table 1. In this study, we will look at most known dictionary databases.

Table 2. Online dictionary databases.

Name	Number of words and phrases	Analysis Group	Opinion words	Sentiment Score	Web
Liu and Hu lexicon	Around 6800	Words and phrases	Positive, negative and subjectivity	-1 and +1	https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon
SentiWord Net v3	117000 synonymous	Nouns, verbs, adjectives and adverbs	Positive, negative and objectivity	-1 and +1	https://sentiwordnet.isti.cnr.it/
Natural Language Processing (SentiWords)	Around 155,000 English	Words and phrases	Positive, negative	-5 and +5	http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010
WordStat	9164 negative - 4847 positive English	Word patterns	Positive, negative	Rule Based	https://provalisresearch.com/products/content-analysis-software/wordstat-dictionary/sentiment-dictionaries/
SenticNet	50000	Natural language concepts	Positive-negative	-1 and +1	https://sentic.net/
The Whissell Dictionary of Affect in Language	348000	Spoken English	Word	Standard deviation	https://www.god-helmet.com/wp/whissel-dictionary-of-affect/index.htm

Source: (Davydova, 2017)

Corpus-based Approaches: Corpus-based approaches mainly depend on the method to see the polarity of the words (Agarwal & Mittal, 2016). These dictionaries are produced from a mates of seed thoughts and opinions cost that grows throughout the operation of incidental status by implementation of the employment of either listing or semantic techniques (Serrano-Guerrero et al., 2015). Making use of the corpus-based strategy alone isn't as effectual as the dictionary- based approach since it is hard to get ready an enormous ensemble to cover almost all English terms, but this method includes a main advantage that will help to discover domain and context particular opinion phrases and their orientations utilizing a domain name corpus (Medhat et al., 2014).

Sentiment Analysis in Tourism

Today, tourism businesses market their products and services using various online distribution channels. On the one hand, the diversity of these distribution channels provides access to more customers in marketing activities, but on the other hand it creates difficulties in collecting and interpreting customer reviews.

In the tourism industry, the analysis of online tourist reviews is regarded as one of the ways to evaluate the products and services offered by businesses and to understand the needs of tourists. Online tourist reviews are increasing day by day in parallel with the developments in information and communication technologies in social networks and e-commerce sites. However, it is a difficult process to reveal emotional expressions in such a huge amount of data. In many cases, sentiments are hidden in the huge and natural language content produced by customers.

The main approach to sentiment analysis of touristic products is based on determining the sentiment polarity in customer reviews. Sentiment polarity (positive, negative, neutral) is determined by sentiment scores. Sentiment polarity and score in touristic products are explained in the following section.

Sentiment Polarity in Touristic Products

While evaluating customer reviews on touristic products, predetermined sentiment scores of words expressing sentiment are used. Sentiment scores also determine the sentiment polarity. This approach is expressed as a dictionary-based sentiment analysis technique.

The sentiment score ranges from -1 to +1, and whether the sentence or document contains positive or negative sentiments is determined by the total score value. The score value of each word in the text is added together. If the resulting value is greater than 0, the sentence has a positive sentiment; if it is less than 0, the sentence has a negative sentiment. A sentence with a total score of 0 is considered neutral. These sentences are said not to contain sentiments.

The sentiment polarity in touristic products can be explained using the sentiment triangle. Figure 3 shows examples of customer reviews and their positions in the sentiment triangle. Accordingly,

9

customer reviews 1, 2 and 3 are positioned in the sentiment triangle according to the sentence level sentiment analysis result.

Review 1: Excellent Cuban Food!!! First came the ceviche, unbelievably good!! → **+1 positive**

Review 2: I went to Cuba. Everything we ate in Cuba was much better. What more can I say. → **0 neutral**

Review 3: Last night my friend and I went to the Cuba restaurant. There were unbelievably bad foods. → **-1 negative**
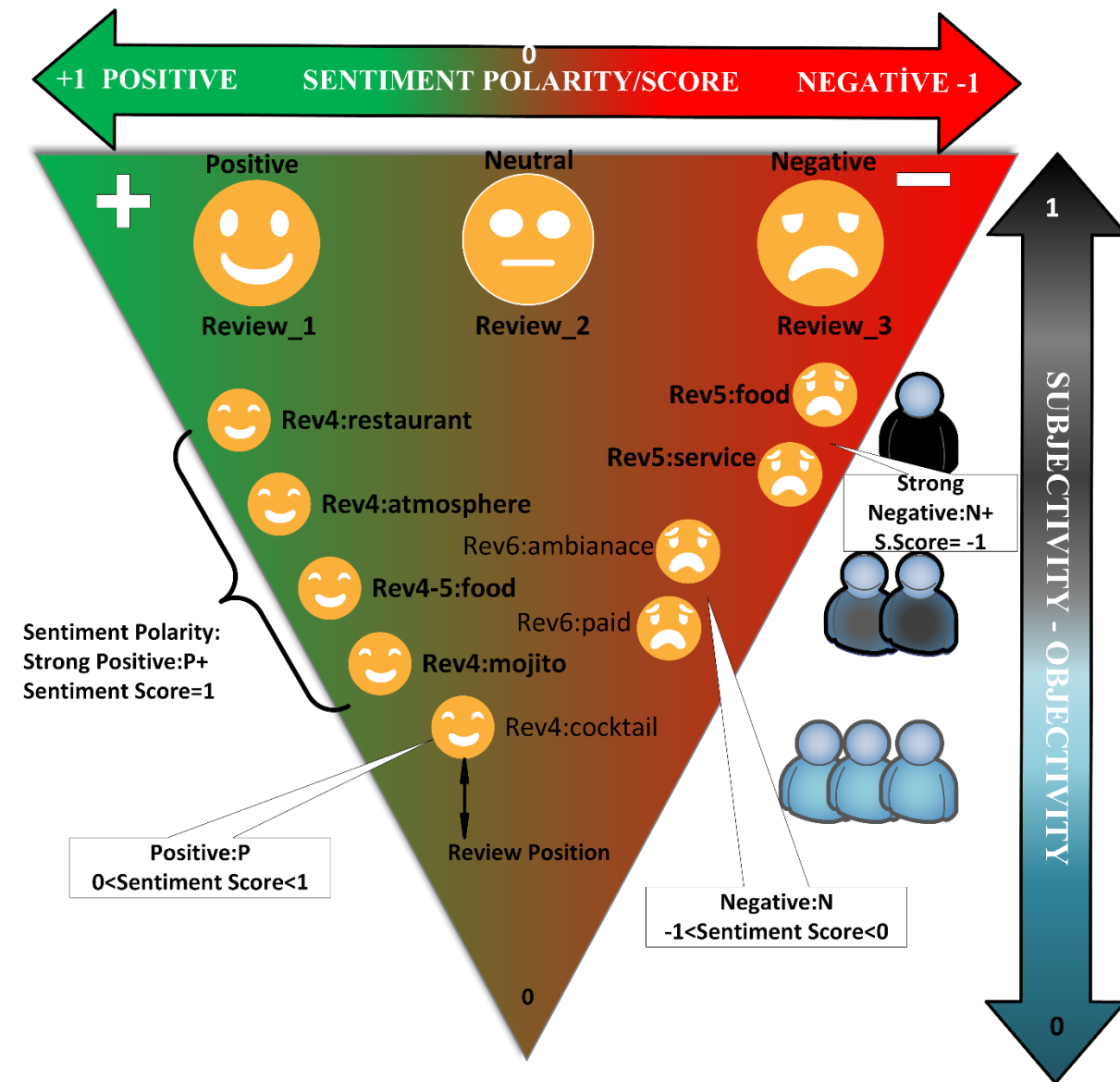
Figure 3. Sentiment Triangle.

Source: Adapted from Mostafa, (2013)

Mostly, customer reviews may not consist of simple sentences. In this case, sentence- level sentiment analysis is insufficient. Examples of customer reviews that reflect this situation are presented below (Review 4, 5, 6).

Review 4: Great little restaurant, wonderful atmosphere and the food is excellent!! don't forget to try a cocktail their mojitos they are the best!! → **restaurant:P+ | atmosphere:P+ | food:P+ | cocktail:P | mojito:P+**

Review 5: Food was very good. Service was painfully slow. Took forever to place your order. Drinks took forever. → **food:P+ | service:N+**

Review 6: I paid 100 euros for a really flavourless food and not so delightful ambiance. → **paid: N | ambiance: N | food: N+**

In the customer reviews given above, it is necessary to determine for which properties of touristic products are evaluated.

The ABSA method is used to determine the sentiment polarity according to product features. ABSA analysis results for customer reviews (Review 4, 5, 6) and their positions in the sentiment triangle are shown in the Figure 3.

SA also determines the subjective or objective value (0 or 1) of the sentence or document. If positive or negative reviews are generalized for the same product or service in customer reviews, it is objectively determined and given a value of 0.

Evaluating touristic products using sentiment analysis techniques requires a system modeling. The following section describes the recommended systems and processes for the sentiment analysis system.

Sentiment Analysis System in Touristic Products

The evaluation of tourism products by extracting hidden sentiments from customer reviews requires a system modeling.

The system to be developed should include natural language processing, text mining and sentiment analysis processes. Thanks to the system, customer perceptions of touristic products can be detected instantly.

In Figure 4, a system proposal for sentiment analysis using online tourist reviews related to touristic products and services is presented.
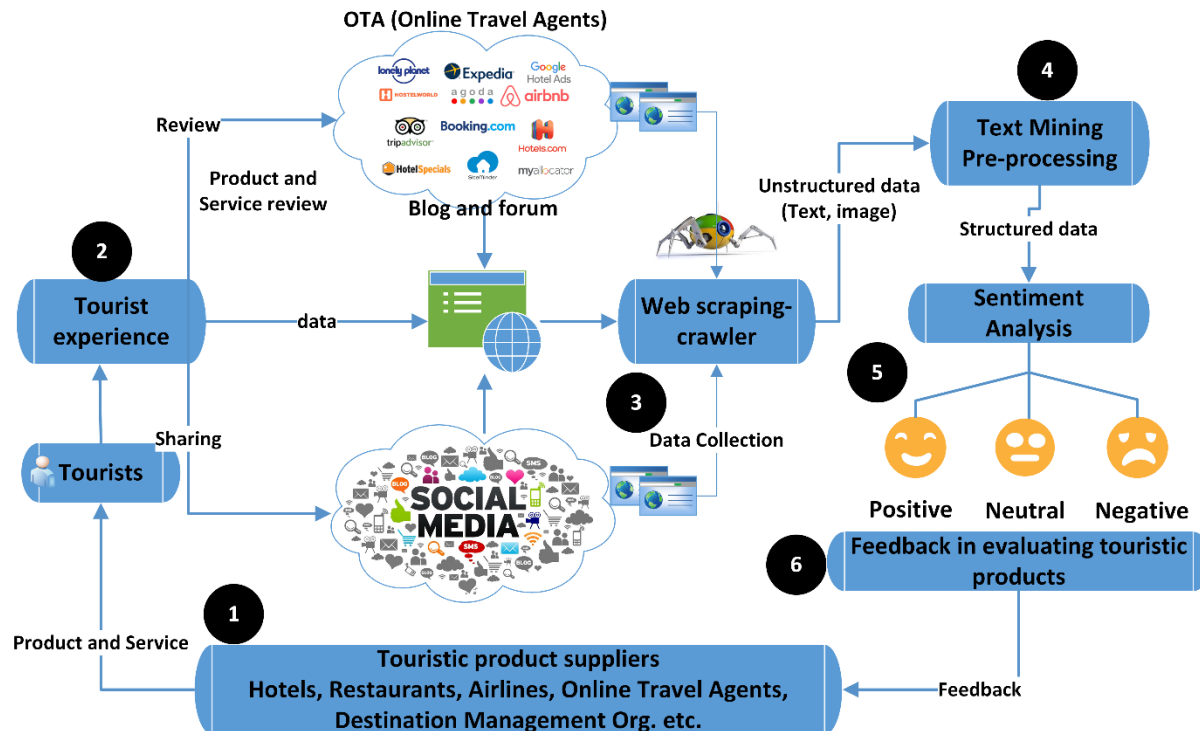
11

Figure 4. Sentiment Analysis System recommendation for touristic products.

The system model presented in Figure 4 consists of 6 stage. These stages will be explained below.

Stage 1: Touristic product suppliers seek to generate high income by offering their products and services to tourists in the tourism market. However, how the products and services offered to the market are perceived by tourists is of vital importance. It provides the opportunity to evaluate the products and services of tourism businesses and destinations by getting feedback from tourists. Using new analysis methods rather than traditional methods (survey, call center, etc.) in the evaluation of touristic products yields better results. In this respect, sentiment analysis provides the opportunity to evaluate the touristic products during and after their use.

Stage 2: Tourists consume touristic products by experience. Tourists share their comments about their experiences on social media (eg Facebook, Twitter, Instagram etc.), blogs and forums (eg TripAdvisor, Yelp, Expedia etc.) or e-commerce sites (Booking, Airbnb, etc.). These reviews and shares are created by the tourists completely at their own request and without any force. Therefore, tourism businesses and destinations can collect reliable feedback on their products and services.

Stage 3: Special software is used in data collection process because it changes according to the web environment. Web scraping or web crawling technologies are used for data collection from the web environment. These technologies collect the texts about tourist reviews on web pages and convert them into necessary files for sentiment analysis programs. The textual content collected at this stage has an unstructured feature.

12

Stage 4: Text preprocessing is one of the most important tasks of text mining. At this stage, unnecessary words and signs in the texts obtained from web pages are removed. The texts to be analyzed are converted into structured data and prepared for sentiment analysis. This stage is very important as the complete and correct data set will affect the data quality and thus the analysis result.

Stage 5: Sentiment analysis is applied to the textual content collected. Positive, negative or neutral sentiments about the business or destination are revealed. These sentiments can consist of subjective and objective evaluations. ABSA technique is used for detailed sentiment detection about touristic products or services.

Stage 6: Sentiment analysis results obtained at this stage are evaluated by visualization. Sentiment analysis results create an important feedback for tourism businesses or destinations.

These feedbacks help identify the strengths and weaknesses of tourism businesses and destinations. It also contributes to the identification, measurement and evaluation of market opportunities. Thanks to the feedback of tourists, businesses provide the opportunity to improve their products, develop new products, manage customer relations and understand their competitors. For destination management organizations, tourist feedback can be used in activities such as creating a destination image, determining brand value and determining tourist needs.

Text Mining in Tourism Literature

In the tourism industry, online customer reviews analysis is seen as one of the methods for evaluating the products and services offered by businesses and understanding the needs of tourists. These reviews are available on social networks and e-commerce sites. Evaluation of customer reviews using scientific methods, text mining methods are used. In the studies of text mining in tourism literature, customer reviews for accommodation or food & beverage businesses are used. Usually, customer reviews are collected and analyzed from Tripadvisor.com, Booking.com, Expedia.com, Airbnb.com and Yelp.com sites. Text mining studies for accommodation businesses are generally studies aimed at determining satisfaction and dissatisfaction with the products and services offered.

2510 customer reviews on the TripAdvisor site for hotels in Sarasota, Florida were analyzed using text mining methods (Berezina et al., 2016). The research findings revealed some common categories used in both positive and negative reviews, such as workplace (eg hotel, restaurant and club), room, furniture, members, and sports. The study results also showed that satisfied customers who are willing to recommend a hotel to others refer more often than customers who are not satisfied with the intangible aspects of their stay, such as staff members. On the other hand, dissatisfied customers spoke more often about concrete aspects of hotel accommodation such as furniture and finance.

13

For Antalya halal concept hotels, customer complaints on the booking.com site were analyzed using the aspect-based sentiment analysis technique (Özen, 2019). According to the research findings; It has been determined that the issues causing customer dissatisfaction stem from loud noise and staff. It has been observed that the reason for the customers' negative view of the staff at the hotels is related to their insufficient English speaking.

Text mining studies on food & beverage businesses mostly consist of studies on the evaluation of the products and services offered in restaurants in a touristic region.

TripAdvisor customer reviews for restaurants serving dinner in Macau were analyzed according to content analysis and taxonomy (Lei & Law, 2015). Research results showed that overall customer satisfaction in the dining experience in Macau was positive. Interestingly, expensive restaurants in Macau showed a higher percentage of negative reviews than cheaper restaurants.

For the restaurant serving home-cooked meals in the Cappadocia region, tripadvisor customer reviews perspective based emotion analysis was applied (Özen & İlhan, 2020). According to the results of the analysis, it was seen that the most popular products in the menus consisted of dessert, tea, dolma and soup. In addition, it was determined that the friendly attitude of the restaurant staff and the restaurant atmosphere were appreciated by the tourists.

Text Mining With RapidMiner

This article has been prepared based on the 9.5 version of RapidMiner and it is a new version as of the date of writing. Rapid Miner is community-founder software and it's free. There are also academic and professional versions. While the academic version is free, the professional version is paid. RapidMiner includes a lot of tools related to data mining, text mining. Some of these tools are automatically included when loading RapidMiner. Some tools can be added after the software is installed. The following sections will explain how to install the RapidMiner software and add text mining tools to the software.

- Downloading RapidMiner Studio

Rapidminer software is downloaded from https://rapidminer.com/products/studio/

When RapidMiner Studio installation is started, membership must be created for activation and license (Fig. 5).

Figure 5. RapidMiner Membership.

Source: RapidMiner, (2014)

License key is generated after membership activation. The license key can be viewed in the license options section by logging in at https://my.rapidminer.com/. Academic license is not paid and has a limit of 10,000 lines.
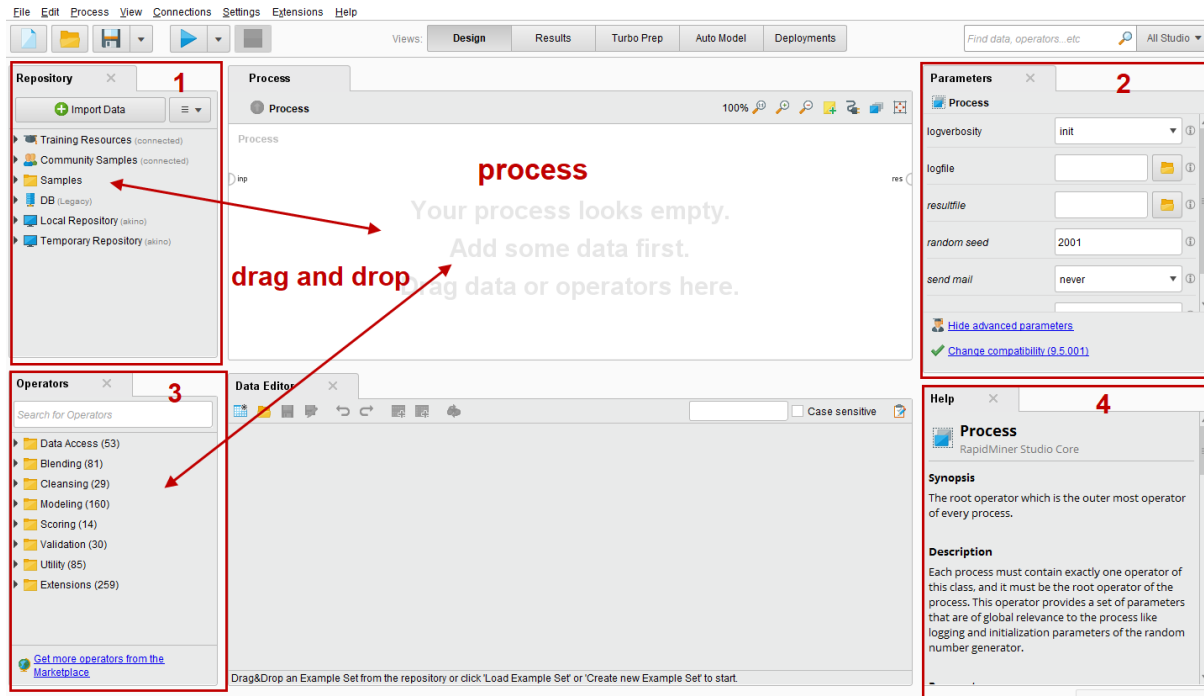
When RapidMiner Studio is opened for the first time, it starts working with the start screen. Some ready-made templates can be selected from this screen (Fig. 6).
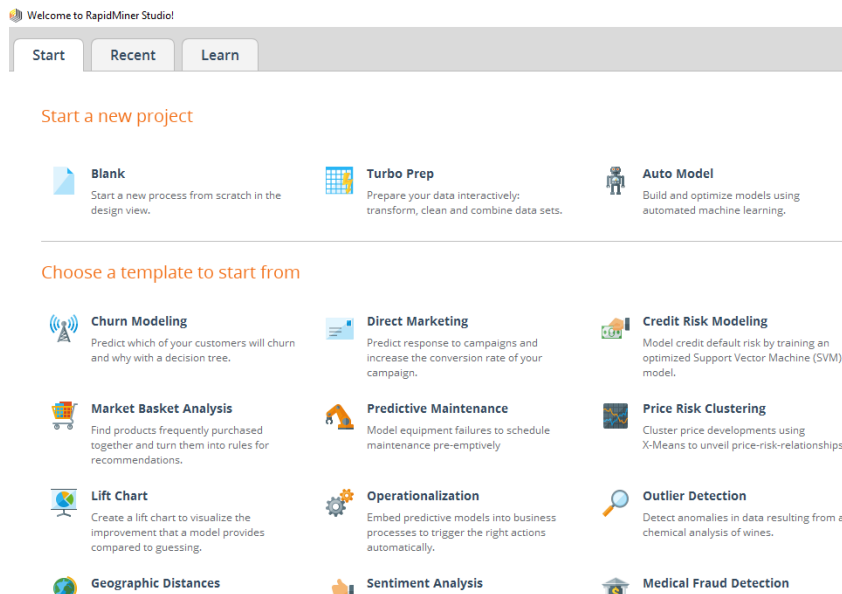
Figure 6. RapidMiner Start Screen.

Source: RapidMiner, (2014)

15

After selecting the blank operation box, the screen appears (Fig.7)

Figure 7. RapidMiner Blank Screen.

Source: RapidMiner, (2014)

The "1_Repository" window contains sample data. When RapidMiner adds data from outside, the repository part is used. The parameters of the operators are presented in the "2_parameters" section. The "3_Operators" window contains operators related to data mining. "4_Help" contains help with operators. On the other hand, operations are performed in the empty area where "Process" is written in the middle of the screen (Fig.7).

Text Preprocessing and Sentiment Analysis With RapidMiner

In the above sections, the installation and basic concepts of RapidMiner software are explained. In this section, using customer reviews about touristic products, models created in RapidMiner and text mining and sentiment analysis examples are presented.

RapidMiner Application 1

Goals of the Application:

To have basic knowledge about text mining techniques "text preprocessing" and "sentiment score" using the model created in RapidMiner.

16

Purpose of Application:

In this application, it is aimed to determine the sentiment score of the person who wrote the document for each "tweet" by collecting the twitter messages. In addition, using the "text preprocessing" technique in twitter messages, weighted keywords were determined. For this purpose, Turkey has recently written about twitter comments are used.

Application Data Set (text):

In application, Twitter's text "Turkey holiday" written in English tweet recently about 1,000 was collected. Tweets in the last gathering "Turkey holiday" word was questioned.

Application Model:

Operators included in RapidMiner 9.5 version are used. Application model and RapidMiner operation screen are presented in figure 8.
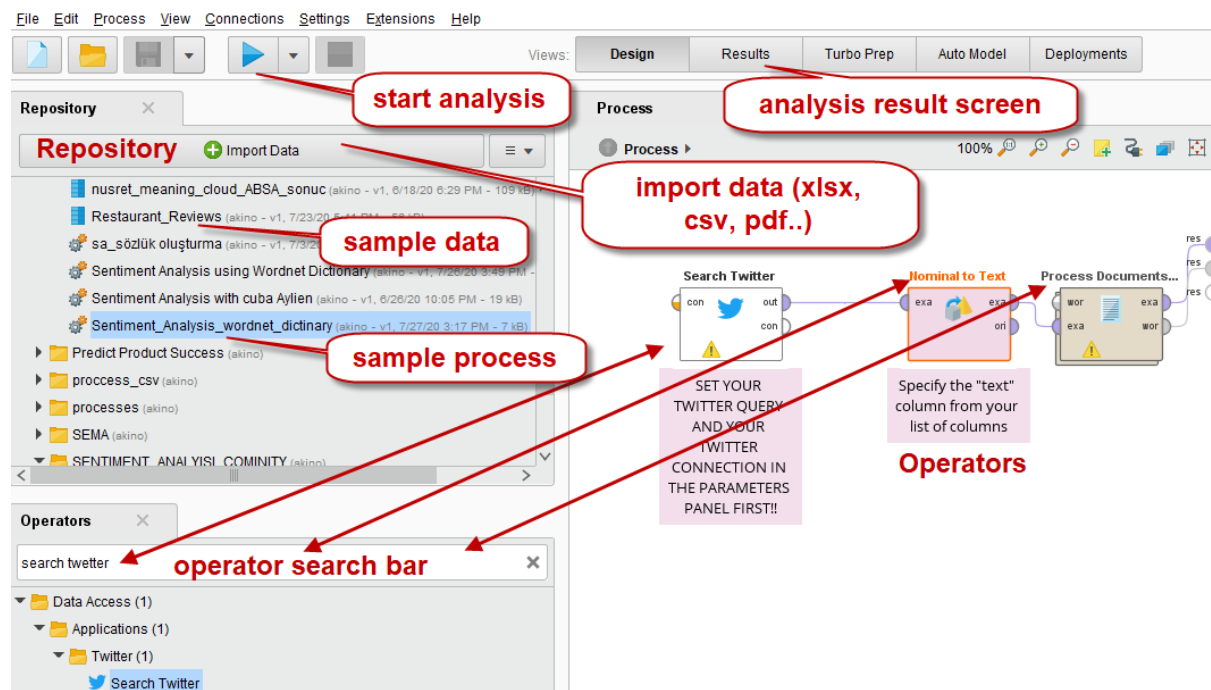

Figure 8. RapidMiner Application Model.
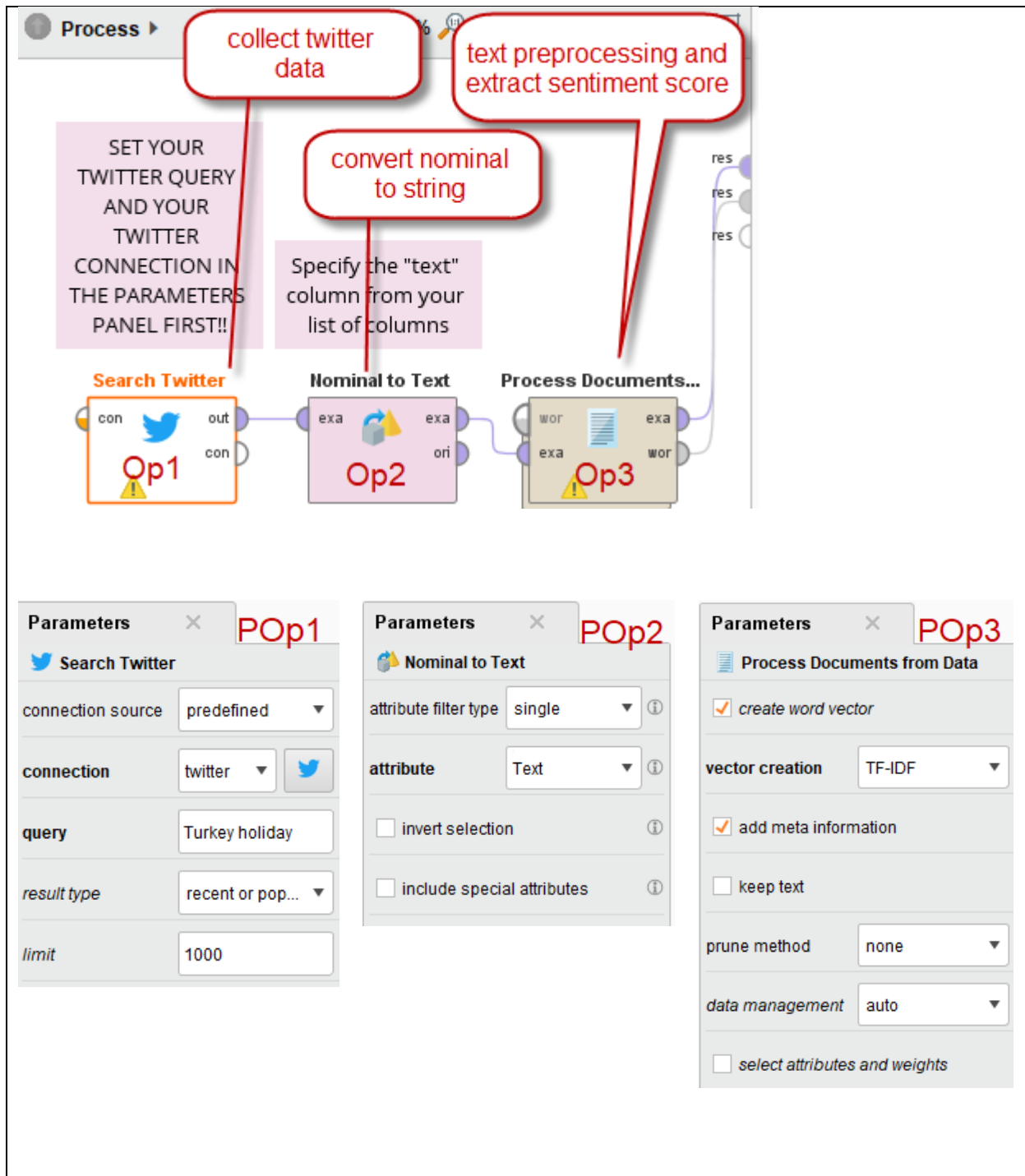
Source: RapidMiner, (2014)

Figure 9. RapidMiner Application Model Details.

Source: Adapted from RapidMiner software by the author.

The operators used in the model and their parameters are presented in the Figure 9. The functions of each operator will be described below, respectively.

Search Twitter (Op1): With the Search Twitter operator, you can specify a query and get Twitter statuses containing this query. The list of statuses contains additional data with context of the statuses. In the expert mode, you can specify additional search restrictions.

Select a Twitter connection to specify the Twitter account for the Twitter API access. Specify at least a query to search Twitter for it. There are advanced parameters you can use to specify additional search restrictions. For example, you can limit the search results to a language.

Parameters of Twitter Search Operator (POp1)

- **Connection source:** This parameter indicates how the connection should be specified. It gives you two options: repository and predefined. The parameter is not visible if the connection input port is connected.
- **Connection:** This parameter is only available when the connection source parameter is set to predefined. The connection details for the Twitter connection can be specified here. If you have already configured a Twitter connection, you can select it from the drop-down list. If you have not configured a Twitter connection yet, select the icon to the right of the drop-down list. Create a new Twitter connection in the Manage Connections box.
- **Query:** The term that should be searched. In this application, **"Turkey holiday"** phrase was used.
- **Limit:** The limit on the number of tweets to return. In this application, **1000** tweets were collected.

Nominal to Text (Op2): The Nominal to Text operator converts all nominal attributes to string attributes. Each nominal value is simply used as a string value of the new attribute. If the value is missing in the nominal attribute, the new value will also be missing (RapidMiner, 2014).

Parameters of Nominal to Text Operator (POp2)

- **Attribute filter type:** This parameter allows you to select the attribute selection filter; the method you want to use for selecting attributes on which you want to apply nominal to text conversion. Single, this option allows selection of a single attribute.
- **Attribute:** The desired attribute can be selected from this option. The attribute name can be selected from the drop-down box of attribute parameter if the meta data is known. In this application, "Text" is selected because Twitter messages have **"Text"** attribute (RapidMiner, 2014).

19

Process Documents From Data (POp3): It is the most important operator of text mining. It performs the cleaning of the noisy data and the calculation of the word weights in the data set (tweet text). Clearing the dataset will be explained in the subprocess of this operator (Fig. -Op4, Op5, Op6, Op7). Word weighting is calculated by creating a word-vector.

The word-vector model is based on the classification of words according to their frequency in the document. By considering the frequency of the words in the document, the word weights are calculated in the entire document. Word weighting is a coefficient that measures the importance of the word in the document. Determines the contribution of the words to the classification of the analyzed document.

When calculating the word weight with TF_IDF (Term Frequency- Inverse Document Frequency), it is calculated by considering the documents in the entire data set, not just the document the word goes through. While the weights of the words that are repeated too often fall in the documents in the data set, the weights of the rare words in the documents increase. If a word does not appear in other documents in a dataset and is in a document, it has the determining feature for that document (Soucy & Mineau, 2005).

Parameters of Process Documents From Data (POp3)

- **Create word vector:** If checked, the tokens of a document will be used to generate a vector numerically representing the document.
- **Vector creation**: (**TF_IDF**)

Sub-operators of Process Documents From Data (Op3): At this stage, the sub-operators of the "process from data" operator will be explained. Sub-operators can be created by double-clicking this operator (Fig. 10). The functions of the sub-operators are described below.
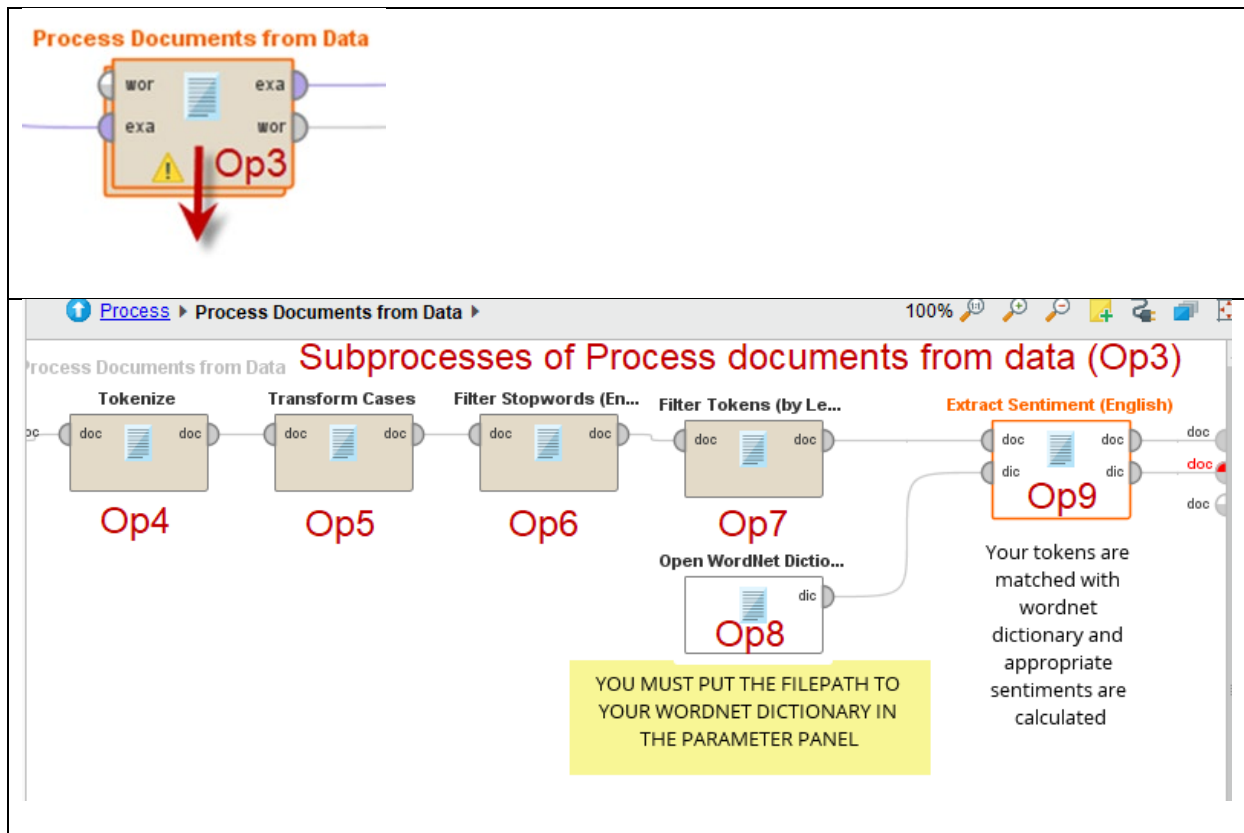
Figure 10. Sub-operators of Process documents from data (Op3).

Source: Adapted from RapidMiner software by the author.

The operators used in the model and their parameters are presented in the figure 10. The functions of each operator will be described below, respectively.

Tokenize (Op4): This operator splits the text of a document into a sequence of tokens. There are several options how to specify the splitting points. Either you may use all **non-letter** character, what is the default settings. This will result in tokens consisting of one single word, what's the most appropriate option before finally building the word vector.

Transform Cases (Op5): This operator transforms all characters in a document to either lower case or upper case, respectively.

Filter Stopwords (Op6): This operator filters English stopwords from a document by removing every token which equals a stopword from the built-in stopword list. Please note that, for this operator to work properly, every token should represent a single **English** word only. To obtain a document with each token representing a single word, you may tokenize a document by applying the Tokenize operator beforehand.

21

Filter Tokens (by Length) (Op7): This operator filters tokens based on their length (i.e. the number of characters they contain). In this application, words less than three and larger than twenty-five characters were removed.

Extract Sentiment English (Op8): This operator uses a WordNet 3.0 and a SentiWordNet 3.0.0 database to extract sentiment of an input document. If the WordNet dictionary is not the same version as SentiWordNet, that is version 3.0, the operator will not work as expected. The sentiment value is in range [-1.0,1.0] where -1.0 means very negative and 1.0 means very positive. The sentiment value is added to the meta data of the document. WordNet and SentiWordNet are connected by Synset IDs. To calculate sentiment of a document we calculate sentiment of each word, where the first meaning of a word has the most influence on a sentiment and each next meaning has less influence on a sentiment. Document sentiment is then calculated as the average value of all word sentiments.

Open WordNet Dictionary (Op9): This operator provides dictionary input to (Op 8). Thanks to this operator, sentiment scores of words are obtained. Parameter "directory" option is used in this operator. The "dict" folder in the wordnet dictionary installed on our computer is defined. The Wordnet dictionary can be downloaded from the "https://sentiwordnet.isti.cnr.it/" link.

Evaluation of Sentiment Analysis Output

The result of the sentiment analysis is presented in the Figure 11. The table resulting from the analysis and its properties are explained below.
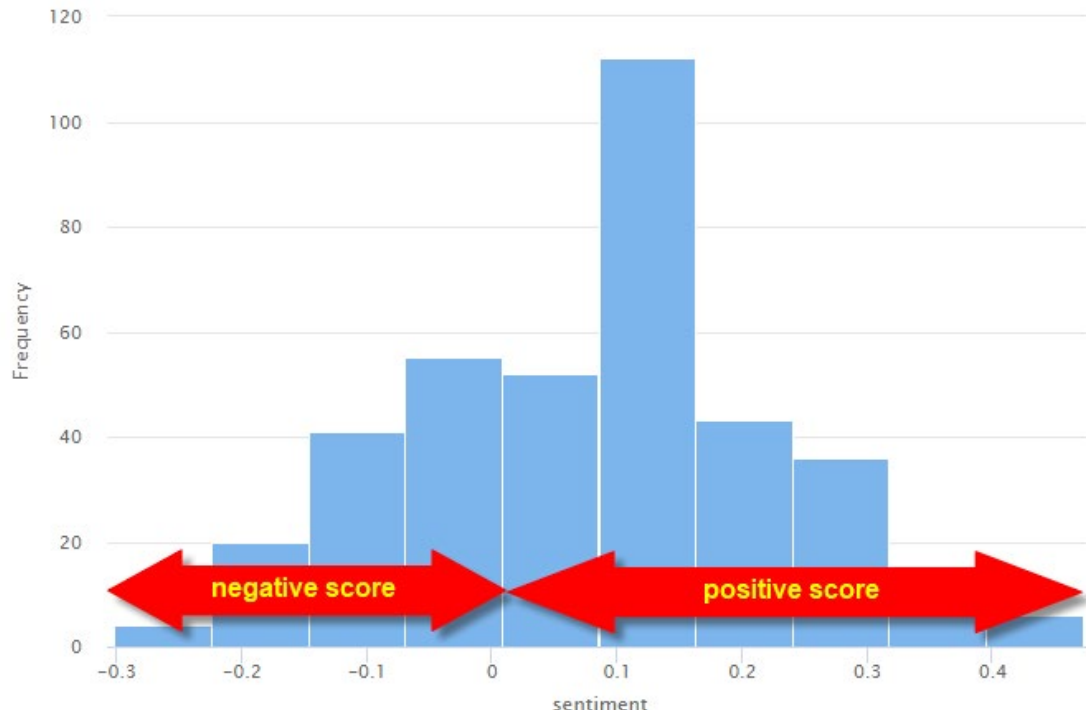

Figure 11. Sentiment Analysis Result.

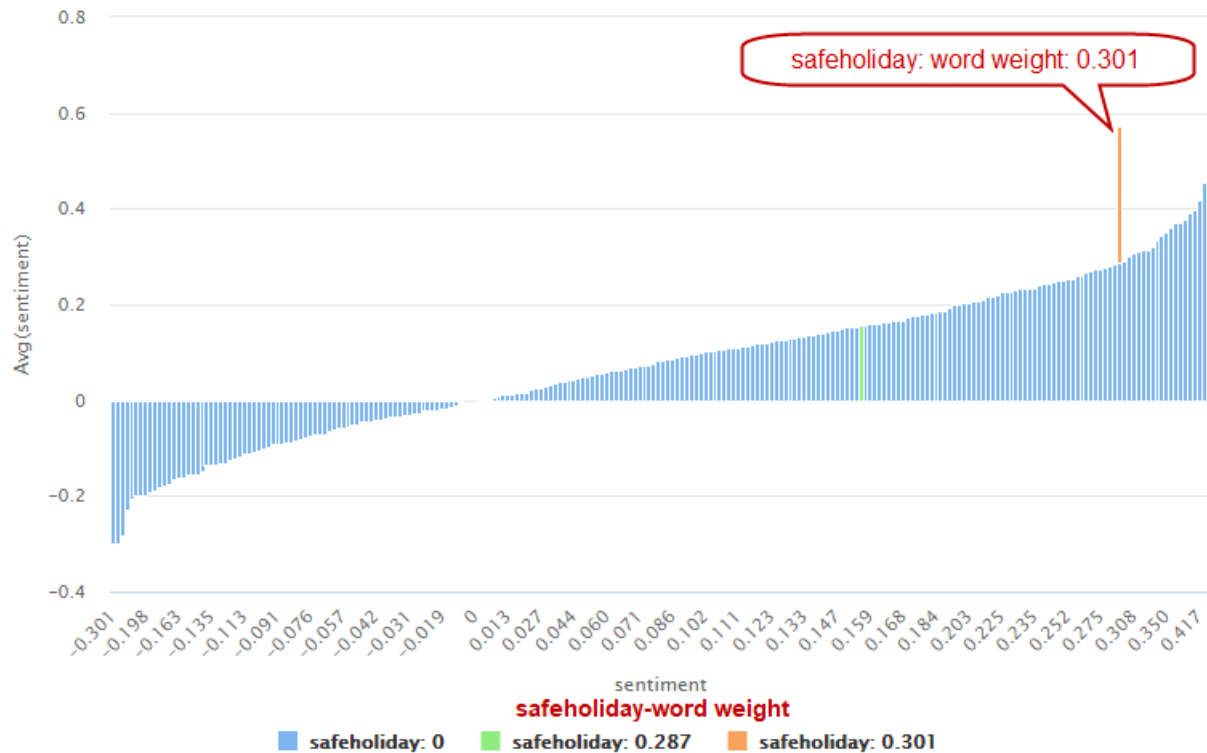*Source:* Adapted from RapidMiner software by the author.

According to the figure 11, "sentiment score" is calculated for each tweet. According to the score values, (-1-0) interval indicates negative, (+ 1-0) interval indicates positive sentiment. Tweets with a score of "0" indicate neutral sentiment. The wordnet dictionary was used to calculate the score values of the tweets. Accordingly, the distribution graph of sentiment scores for 1000 tweets is presented in the figure 12.

Figure 12. Sentiment Score Distribution.

Source: Adapted from RapidMiner software by the author.

According to the figure 12, it is seen that there is a positive sentiment about "Turkey holiday". In the application, word weights were calculated by preprocessing the text. Term weight and sentiment score of "safeholiday" term is presented in the figure 13.

23

Figure 13. Sentiment Score and Word Weight of the Term "safeholiday".

Source: Adapted from RapidMiner software by the author.

According to the figure 13, the score of the term "safeholiday" is seen to be positive. Twitter users have positive sentiments about "Turkey holiday". In addition, it is seen that they perceive it as a "safe holiday".

Sentiment Analysis With RapidMiner

RapidMiner Application 2

Goals of the Application:

To have basic information about sentiment analysis which is one of the text mining techniques. To be able to evaluate the results of sentiment analysis. In addition, to have knowledge about the technique of automatic collection of textual data from the web (web scraping).

Purpose of Application:

It is sentiment detection about products and services using customer reviews. For this purpose, a restaurant serving Cuba food was selected on TripAdvisor. Sentiment analysis was applied to customer reviews of the restaurant. As a result of the analysis, customers' feelings (positive,

24

negative, neutral) about the restaurant in general were determined. In addition, objective or subjective features of the comments were also determined.

Application Data Set (text):

In this application, customer reviews of the Cuba restaurant in Miami-Florida were used. Texts of customer reviews were collected automatically using the web scraping technique. The data set collected consists of customer reviews of 217 customers.

Application Model:

Two operators included in RapidMiner 9.5 version are used. Application model and RapidMiner process screen are presented in Figure 4.

Collection of Application Data (Web Scraping)

Customer reviews of the Cuba restaurant were collected from the TripAdvisor site using a web scraping technique. In order to collect customer reviews, the page containing the comments of the relevant business was opened from the TripAdvisor site (Fig.14).
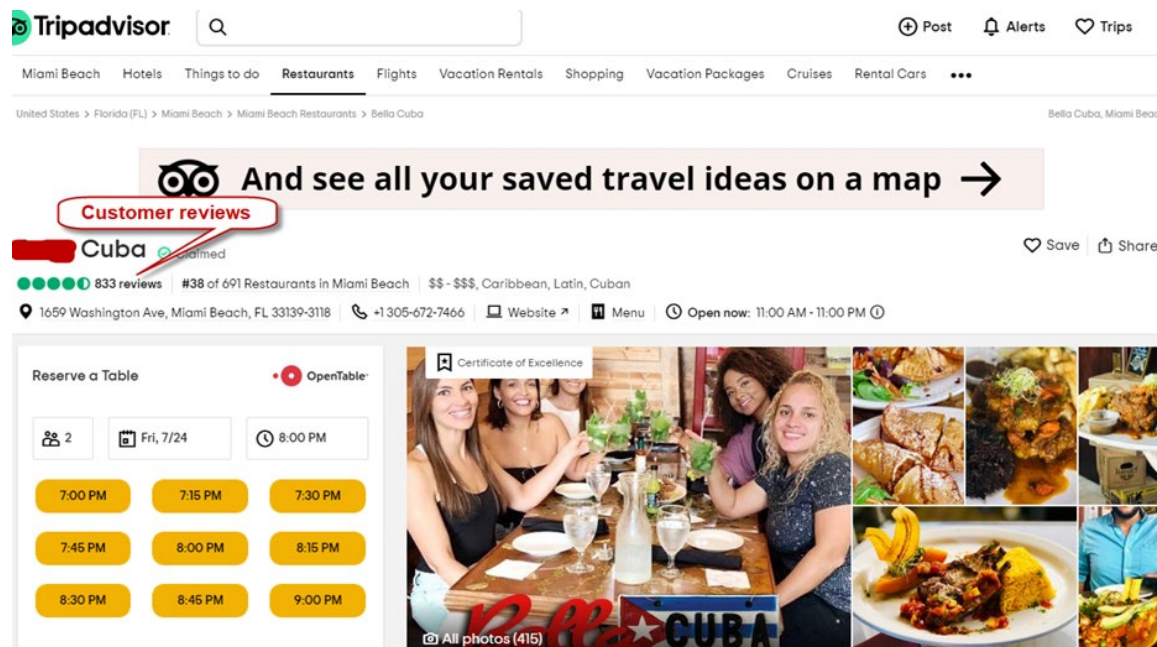

Figure 14. TripAdvisor site of the Cuba restaurant.

Source: Adapted from TripAdvisor site by the author.

The web scraping technique allows the automatic collection of texts on websites (Fig.15). In this application, Google chrome plugin "instant data scraper" extension is used. The Chrome plug-in

25

can be downloaded from "https://chrome.google.com/webstore/detail/instant-data-scraper/ofaokhiedipichpaobibbnahnkdoiiah".
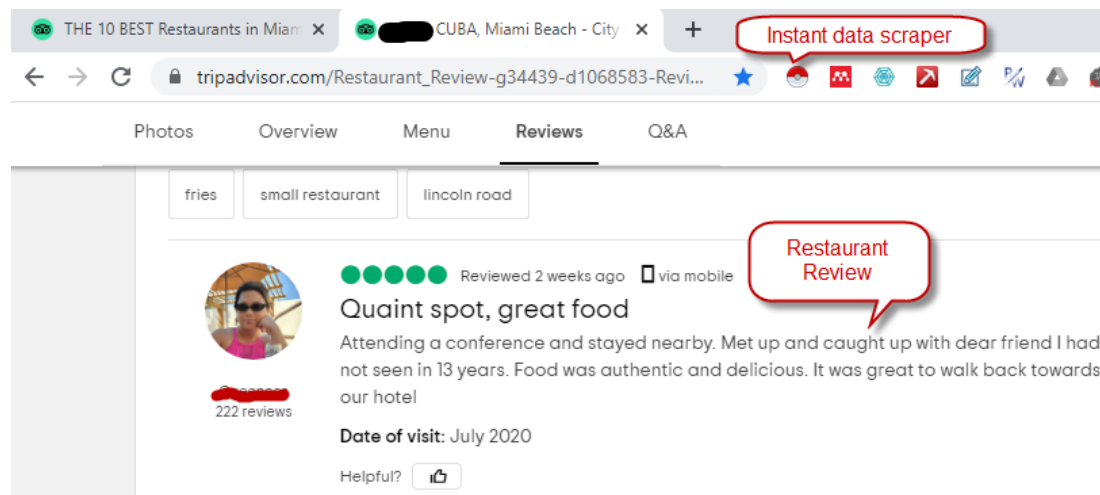

Figure 15. Sample customer review for Cuba restaurant. And web scraping plugin.

Source: Adapted from TripAdvisor site by the author.

The "instant data scraper" plug-in was run to collect customer reviews. The screenshot of the plug-in is presented in Figure 16.
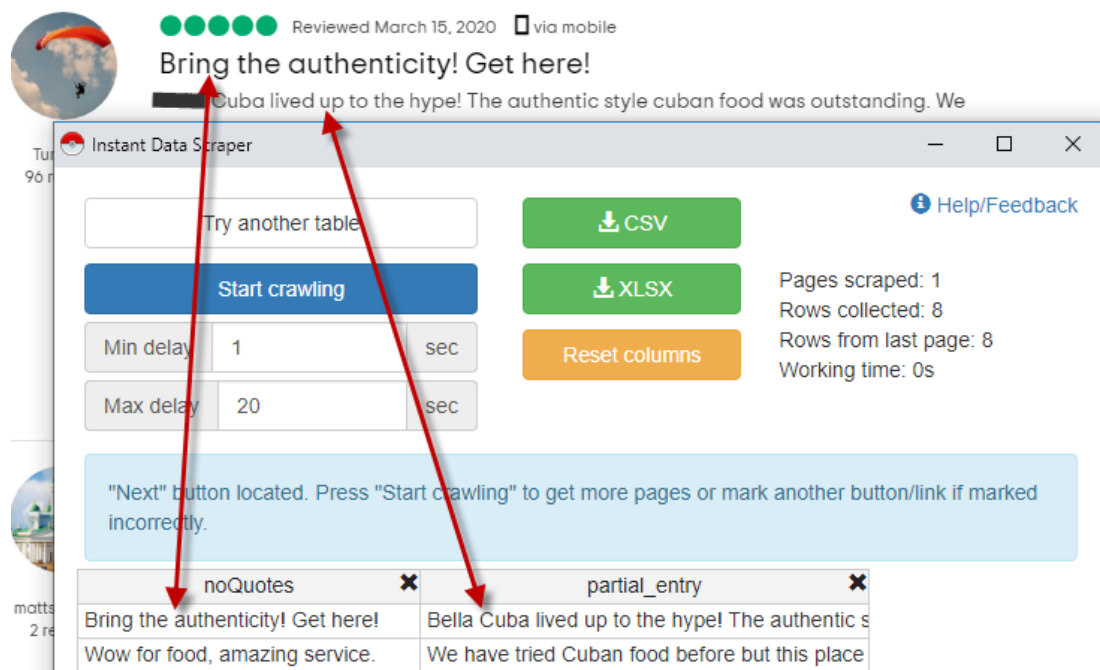

Figure 16. Data Scraper Plug-In.

Source: Adapted from TripAdvisor site by the author.

Customer reviews are automatically collected when the "start crawling" button is pressed. The collected texts were saved as an excel file by clicking the "XLSX" button in figure 16. The screenshot of the saved excel file is presented in Figure 17.

A4		✕ ✓	fx	Excellent service, food, and Cuban ambience. We went for the Drunk chicke

	A	B	C	D	E	F	G	H	I	J
1	Review ▾									
2	We have tried Cuban food before but this place is out of the world, the food is delicious, we asked our waiter fo
3	My best friend came from LA for the weekend and she wanted Cuban food, I took her for lunch to this small rest
4	Excellent service, food, and Cuban ambience. We went for the Drunk chicken rice & slow roast suckling pig. Bot
5	Just what we needed after a long walk on the beach. Very good food. Prompt and professional service.Coconut
6	We visited this restaurant a total of 3 times while in South Beach. They are incredibly welcoming, and the food i
7	I was in this restaurant last week and loved it. Excellent drinks and service. Excellent selection of dishes on the

Figure 17. Excel File for Customer Reviews

The obtained excel file was loaded into the RapidMiner software for sentiment analysis (Figure 18).
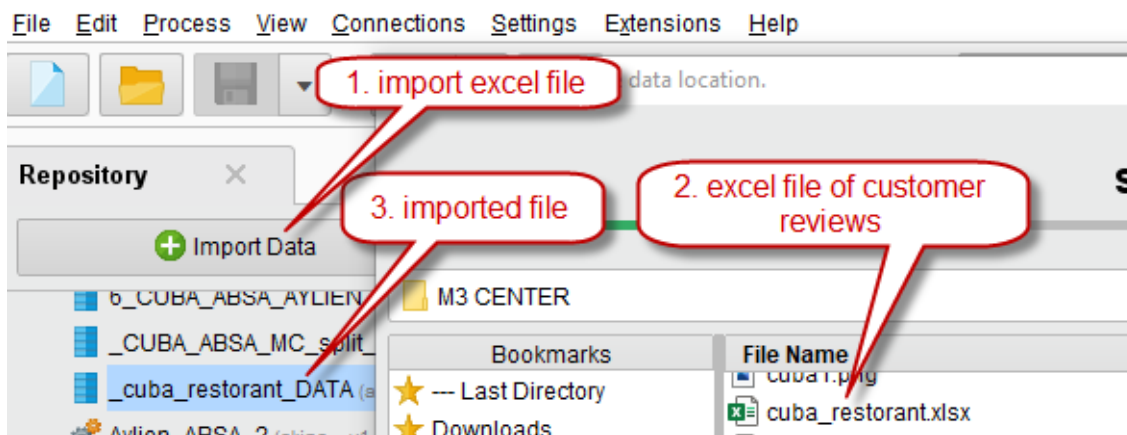

Figure 18. RapidMiner Data Import.

Source: Adapted from RapidMiner software by the author.

Sentiment analysis was applied to customer reviews obtained at this stage. The "MeaningCloud" plugin has been loaded into the RapidMiner software to perform sentiment analysis (Figure 19).
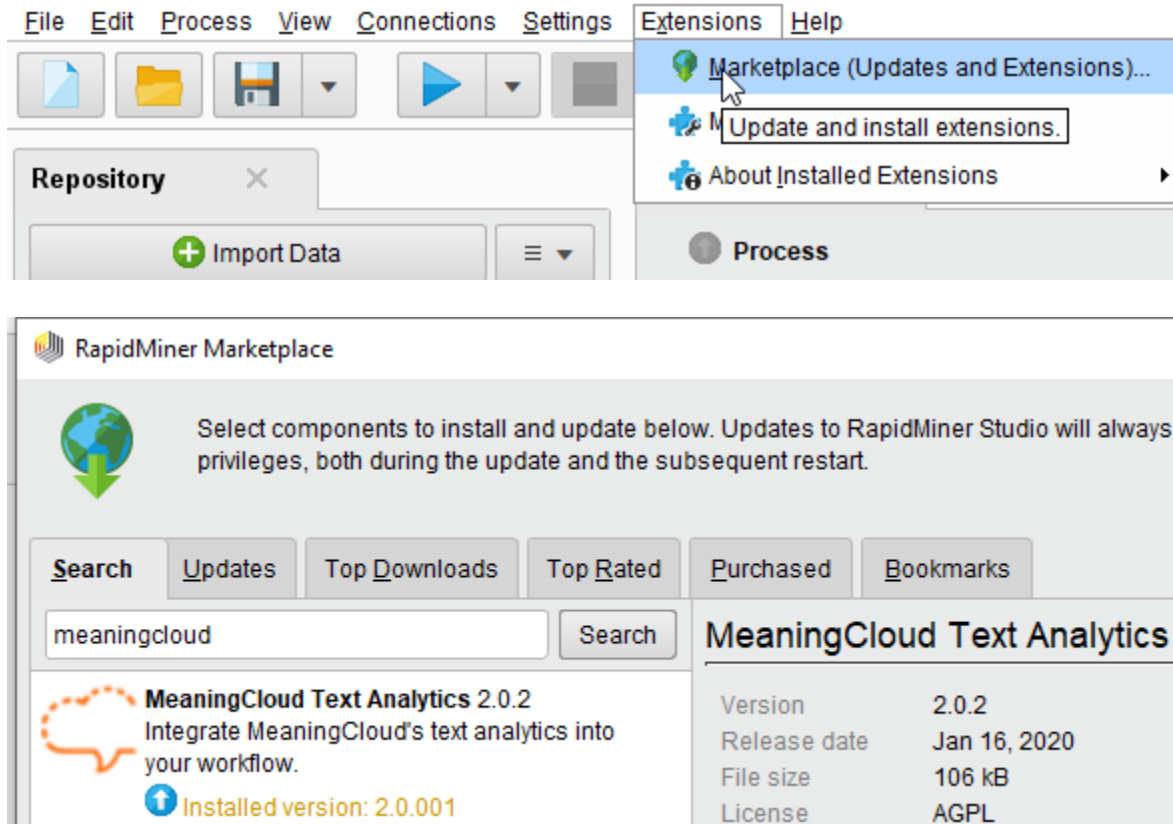
27

Figure 19. Installing RapidMiner MeaningCloud Plugin.

Source: Adapted from RapidMiner software by the author.

For sentiment analysis, customer reviews and sentiment analysis plugins have been moved to the RapidMiner process screen. The data set and the sentiment extension are linked (Fig.20).
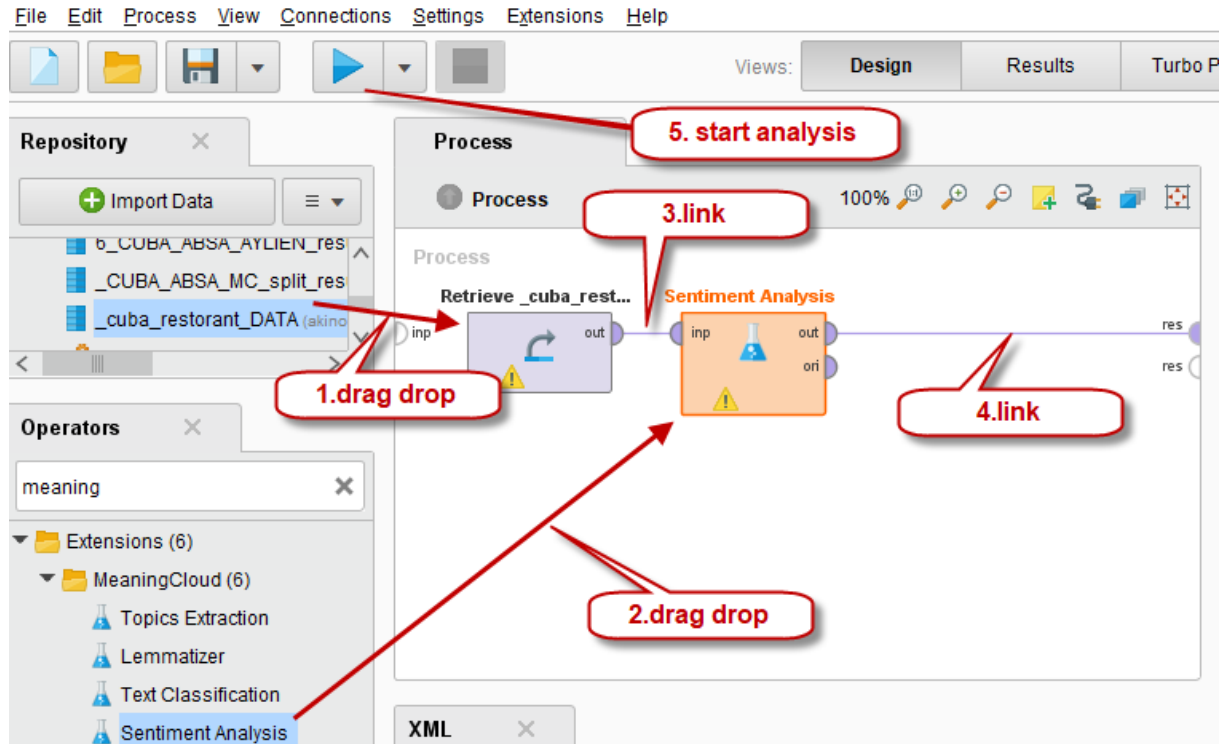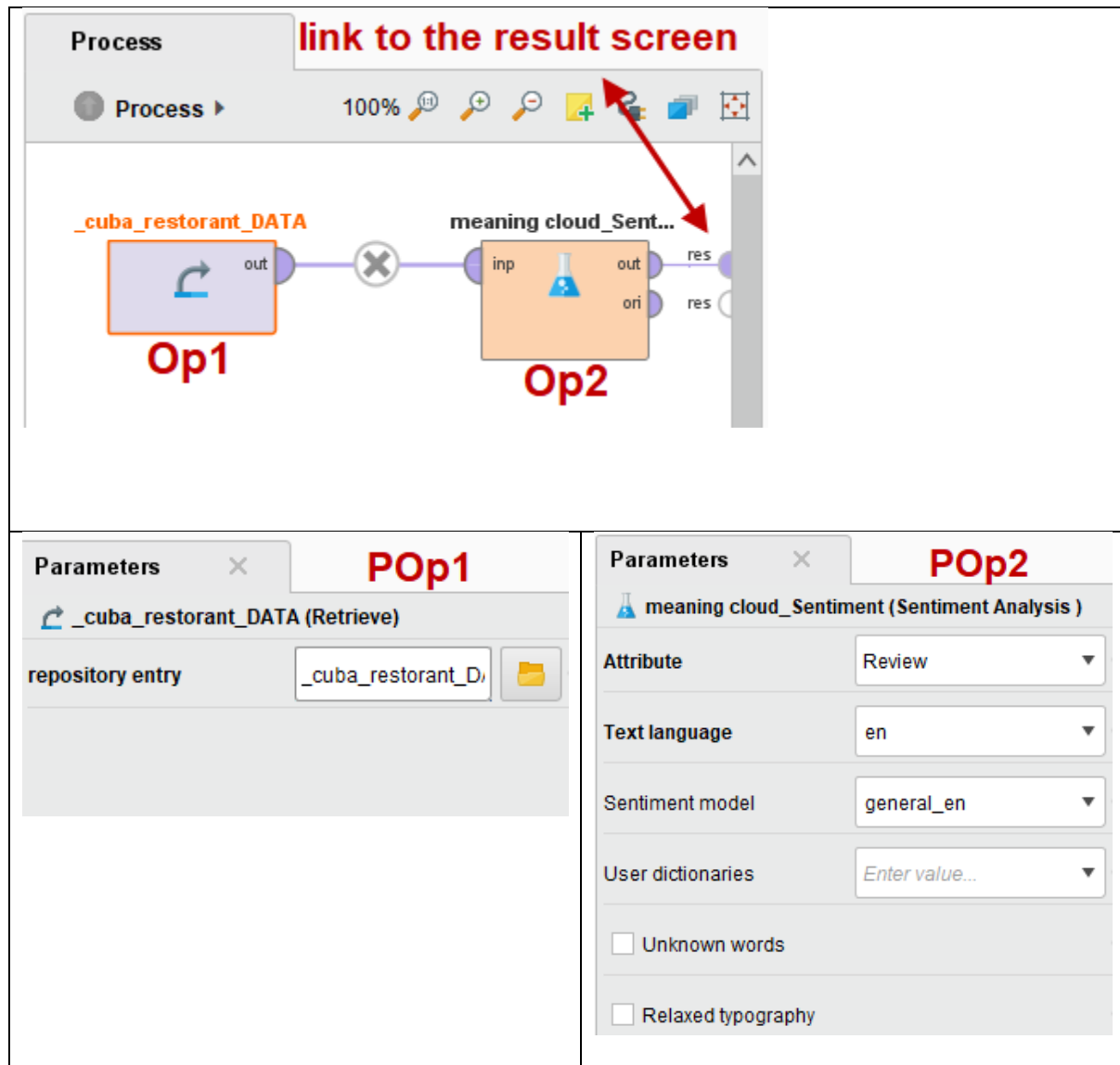
Figure 20. RapidMiner Process Screen and Sentiment Analysis Model.

Source: Adapted from RapidMiner software by the author.

RapidMiner generated sentiment analysis model operators and their parameters (Fig.21)

Figure 21. Sentiment Analysis Model Operators and Parameters.

Source: Adapted from RapidMiner software by the author.

The operators used in the model and their parameters are presented in the figure 21. The functions of each operator will be described below, respectively.

Retrieve (Op1): The Retrieve Operator loads a RapidMiner Object into the Process. This Object is often an Example Set but it can also be a Collection or a Model. Retrieving data this way also provides the meta data of the RapidMiner Object.

30

Parameters of Retrieve (POp1)

- **repository entry**

The path to the RapidMiner Object which should be loaded. This parameter references an entry in the repository, which will be returned as output of this Operator. Repository locations are resolved relative to the Repository folder containing the current Process. Folders in the Repository are separated by a forward slash ('/'). A '..' references the parent folder. A leading forward slash references the root folder of the Repository containing the current Process. A leading double forward slash ('//') is interpreted as an absolute path starting with the name of a Repository. The list below shows the different methods:

MeaningCloud Sentiment Analysis (Op2): Analyze the sentiment of social networks, reviews or customer satisfaction surveys. Have you ever wondered what people say about you, your company or your products in social networks? Have you ever tried to analyze the tens of thousands of free-text answers of a customer satisfaction survey? Sentiment Analysis (also known as Opinion Mining) consists of the application of natural language processing, text analytics and computational linguistics to identify and extract subjective information from various types of content. More information available at https://www.meaningcloud.com/developer/sentiment-analysis/doc.

Parameters of MeaningCloud Sentiment Analysis (POp2)

- **Attribute:** The attribute that contains the text to process.
- **Text language:** Language of the text. Possible values: en, es, fr, it, pt, ca
- **Sentiment model:** Sentiment model to use. The values available will appear once the license key is correctly set in the Settings>Preferences section. You may use the generic sentiment model for the given language ("general" value) or another user-defined model. More information at www.meaningcloud.com/developer/sentiment-analysis/doc/request#sentiment-models-language.
- **User dictionaries:** List of user dictionaries to use in the request, separated by pipes (|). This list corresponds to the "ud" API parameter, described at www.meaningcloud.com/developer/topics-extraction/doc/request.
- **Unknown words:** Smart detection of unknown words, using heuristics for text recognition and disambiguation.
- **Relaxed typography:** Deal with relaxed (informal) typography. This is mostly useful for dealing with social media messages written in informal style.

Evaluation of Sentiment Analysis Output

The result of the sentiment analysis is presented in the Figure 22. The table resulting from the analysis and its properties are explained below.

31

Sentiment analysis result screen

Review	polarity...	confidence(...	agreement(Re...	subjectivity(...	irony(Review)
Great little re...	P+	92	AGREEMENT	SUBJECTIVE	NONIRONIC
There was a ...	P	86	DISAGREEMENT	SUBJECTIVE	NONIRONIC
In our search ...	P	94	DISAGREEMENT	SUBJECTIVE	NONIRONIC
Authentic cub...	NEU	86	DISAGREEMENT	SUBJECTIVE	NONIRONIC
I came in last...	P	100	AGREEMENT	OBJECTIVE	NONIRONIC
After many ye...	N	86	DISAGREEMENT	SUBJECTIVE	NONIRONIC
We had a Su...	P	76	DISAGREEMENT	SUBJECTIVE	IRONIC
Dry rice and b...	N+	92	AGREEMENT	SUBJECTIVE	NONIRONIC

Figure 22. Sentiment Analysis Result Screen.

Source: Adapted from RapidMiner software by the author.

Review: Customer reviews of Cuban restaurant

Polarity: This field indicates polarity in reviews. The possible values are the following: P+: strong positive, P: positive, NEU: neutral, N: negative, N+: strong negative, NONE: without sentiment

Confidence: This field represents the confidence associated with the sentiment analysis performed on the text. Its value is an integer number in the 0-100 range.

Agreement: This field marks the agreement between the sentiments detected in the text, the sentence or the segment it refers to. It has two possible values: **Agreement:** the different elements have the same polarity. **Disagreement:** there is disagreement between the different elements' polarity.

Subjectivity: This field marks the subjectivity of the text. It has two possible values: **Objective:** The text does not have any subjectivity marks. **Subjective:** The text has subjective marks.

Irony: This field indicates the irony of the text. It has two possible values: **Nonironic**: the text does not have ironic marks **Ironic**: the text has ironic marks

Statistics of Sentiment Analysis Results

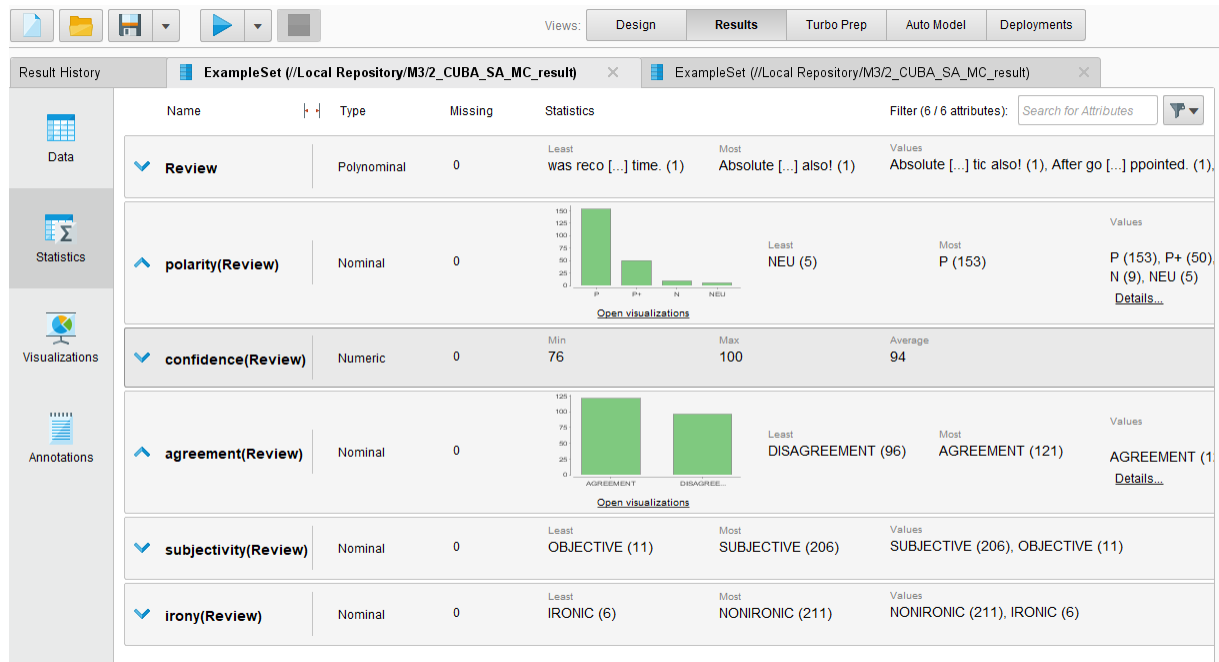The statistical data obtained as a result of the analysis are presented in the figure 23.

Figure 23. Statistics of Sentiment Analysis Results.

Source: Adapted from RapidMiner software by the author.

Visualization of Sentiment Analysis Results

- **Visualization of Sentiment Polarity:** 217 customer reviews of the Cuban restaurant were evaluated with sentiment analysis. The customers appear to have 153 positive, 50 strong positive, 9 negative and 5 neutral feelings about the Cuban restaurant (Fig.24).

Figure 24. Visualization of Sentiment Polarity.

Source: Adapted from RapidMiner software by the author.

- **Visualization of Subjectivity:** 217 customer reviews of the Cuban restaurant were evaluated with sentiment analysis. Customers have 211 subjective and 6 objective opinions about the Cuban restaurant. The objective opinions of customers have 6 positive polarities. The subjective opinions of customers have 193 positive, 13 negative and 5 neutral polarities (Fig. 25).
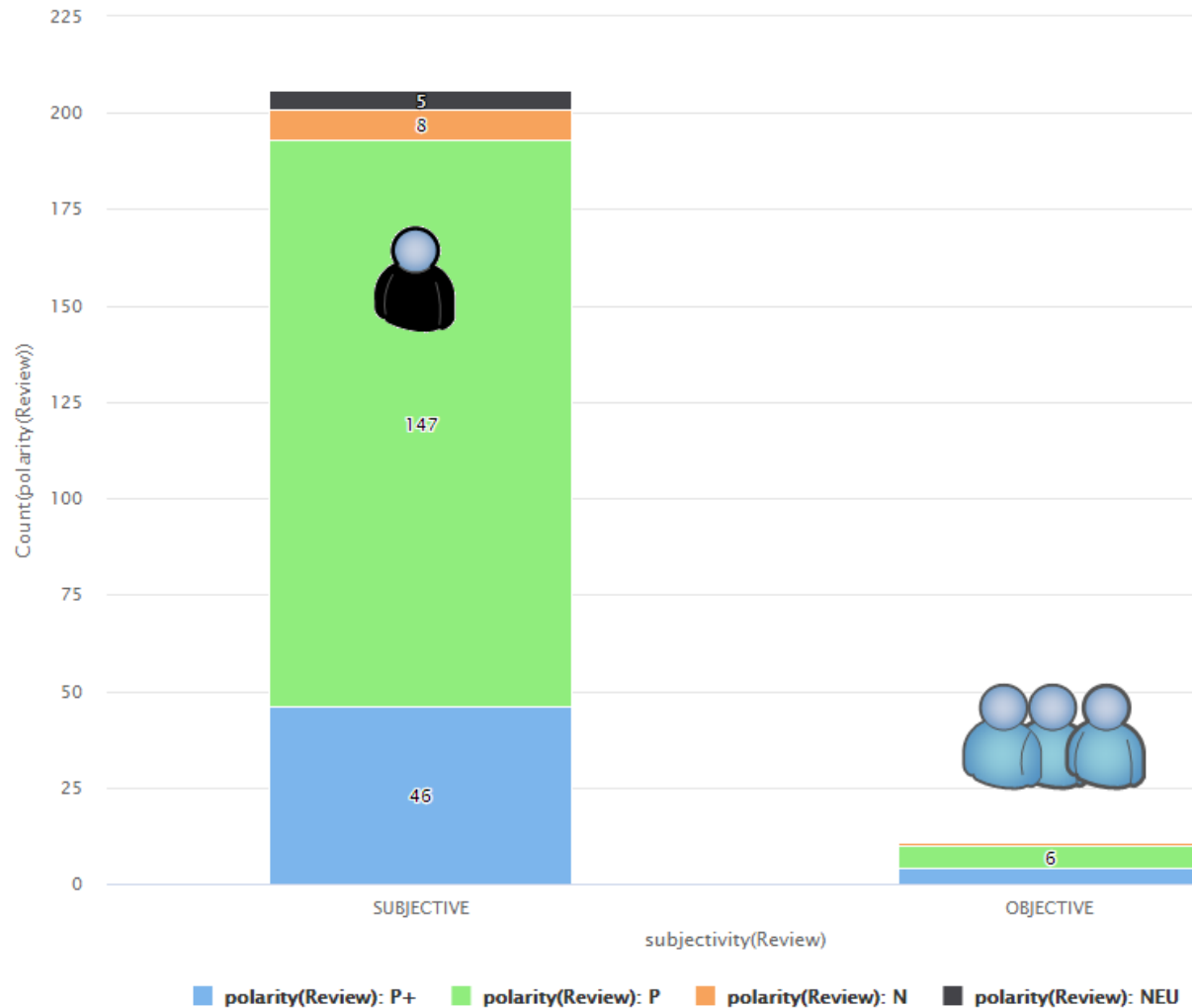
Figure 25. Visualization of Sentiment Subjectivity.

Source: Adapted from RapidMiner software by the author.

Important Note: Most customers have positive feelings about the Cuban restaurant. But we do not know what customers like or not. The answer to this question will be given in the next application.

Aspect Based Sentiment Analysis With RapidMiner

RapidMiner Application 3

Goals of the Application:

To have basic information about aspect-based sentiment analysis which is one of the text mining techniques. To be able to evaluate the results of aspect-based sentiment analysis.

35

Purpose of Application:

To make detailed sentiment determinations about products and services consumed by customers. For this purpose, a restaurant serving Cuba dishes was selected on the "TripAdvisor" site. Aspect based sentiment analysis was applied to customer reviews of the restaurant. As a result of the analysis, sentiment (positive, negative, neutral) about each of the restaurant products and services was determined.

Application Data Set (text):

In this application, customer reviews of the Cuba restaurant in Miami-Florida were used. Texts of customer reviews were collected automatically using the web scraping technique. The data set collected consists of customer reviews of 217 customers.

Application Model:

Operators included in RapidMiner 9.5 version are used. Application model and RapidMiner process screen are presented in Figure 26. The operators used in the model and their parameters are presented in the Figure 26. The functions of each operator will be described below, respectively.
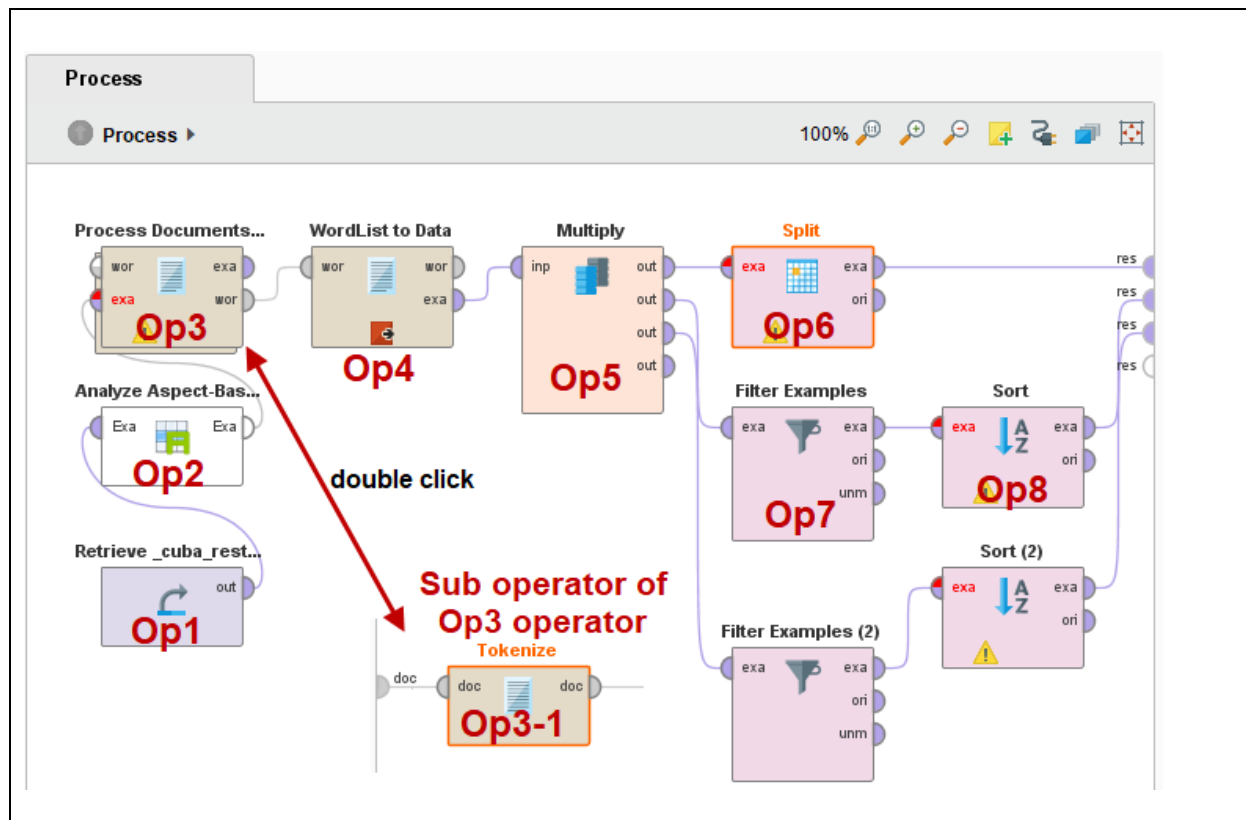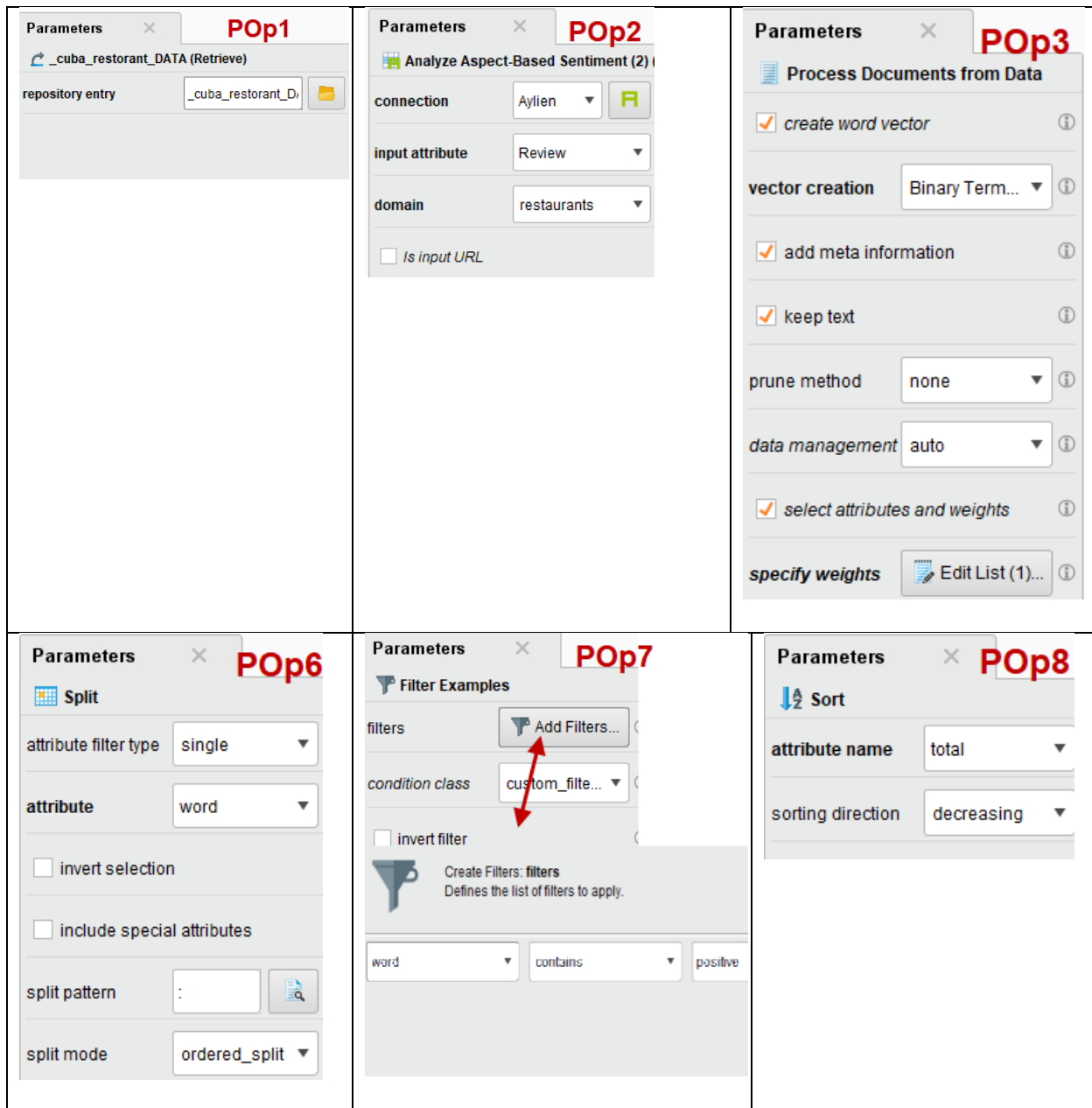
Figure 26. Aspect-Based Sentiment Analysis Model Operators and Parameters.

Source: Adapted from RapidMiner software by the author.

Retrieve (Op1): The Retrieve Operator loads a RapidMiner Object into the Process. This Object is often an ExampleSet but it can also be a Collection or a Model. Retrieving data this way also provides the meta data of the RapidMiner Object.

Analyze Aspect-Based Sentiment (Op2): Given a review for a product or service, Aspect-Based Sentiment analyzes the sentiment of the review towards each of the aspects of the product or review that are mentioned in it.

37

Parameters of Aylien Aspect-Based Sentiment (POp2)

- **Connection:** The connection details for the AYLIEN Text Analysis API have to be specified. If you have already configured a connection, you can select it from the drop-down list. If you have not configured one yet, select the icon to the right of the drop-down list and create a new connection in the Manage Connections box. Please note, the application id and application key are required.
- **Input attribute:** The name of the attribute whose content should be analyzed. In this application, customer reviews are selected.
- **Domain:** Specifies the domain. In this application, the domain restaurant was chosen.

In the model, the data table obtained after Aspect-Based Sentiment Analysis is presented in the Figure 27.

Row ...	aspects	Aspect Based Sentiment analysis results	Review	Customer reviews of Cuban restaurant
1	food:positive, staff:positive			We have tried Cuban food before but this place is ...
2	food:positive, menu:positive, staff:positive			My best friend came from LA for the weekend and ...
3	ambience:positive, food:positive, staff:positive			Excellent service, food, and Cuban ambience. We ...
4	food:positive, menu:positive, staff:positive			Just what we needed after a long walk on the beac...
5	food:positive, drinks:positive			We visited this restaurant a total of 3 times while in...
6	menu:positive, drinks:positive, staff:positive			I was in this restaurant last week and loved it. Exce...
7	food:neutral, menu:positive			Look I'm sure that the rest of the menu is great BU...
8	food:positive, ambience:positive, value:positive, staff:positive			Delicious meal. Best mojitos ever! So many appeti...

Dimensions obtained from customer reviews

Figure 27. Aspect-Based Sentiment Analysis Result.

Source: Adapted from RapidMiner software by the author.

Process Documents From Data (Op3): Generates word vectors from string attributes.

Tokenize (Op3-1): This operator splits the text of a document into a sequence of tokens. There are several options how to specify the splitting points. Either you may use all non-letter character, what is the default settings. This will result in tokens consisting of one single word, what's the most appropriate option before finally building the word vector.

WordList to Data (Op4): This operator builds a data set from a word list. The data set contains a row for each word and attributes for the word itself, the number of documents in which it occurred, the number of labeled documents in which it occurred and for each class number it occurred in a document of this class. The operator might be useful in order to filter a word list before reporting it.

In the model, the data table obtained after this operator (Op4) is presented in the Figure 28.

Figure 28. Aspect-Based Sentiment Analysis Result (WordList).

Source: Adapted from RapidMiner software by the author.

Multiply (Op5): The Operator takes the RapidMiner Object from the input port and delivers copies of it to the output ports. Each connected port creates an independent copy. So changing one copy has no effect on other copies.

Split (Op6): The Split operator creates new attributes from the selected nominal attributes by splitting the nominal values into parts according to the split criterion which is specified through the split pattern parameter in form of a regular expression. This operator provides two different modes for splitting; the desired mode can be selected by the split mode parameter. The two splitting modes are explained with an imaginary ExampleSet with a nominal attribute named 'att' assuming that the split pattern parameter is set to ',' (comma). Suppose the ExampleSet has three examples: value1 value2, value3 value3.

Filter Examples (Op7): The Operator returns those Examples that match the given condition. The conditions are defined by the user. Several pre-defined conditions also exist as advanced options.

Sort (Op8): This operator sorts the input ExampleSet in ascending or descending order according to a single attribute.

Visualization of Aspect Based Sentiment Analysis Results

Visualization of Positive Sentiment Polarity: 217 customer reviews of the Cuban restaurant were evaluated with aspect-based sentiment analysis. 13 aspects were identified that reflect the positive emotions of customers.

39

Positive Aspect Distribution of Customers: food:positive 189, staff:positive 137, value:positive 51, drinks:positive 48, ambience:positive 46, menu:positive 33, location:positive 27, busyness:positive 17, desserts:positive 12, reservations:positive 5, quietness:positive 4, cleanliness:positive 3, facilities:positive 2 (Figure 29).
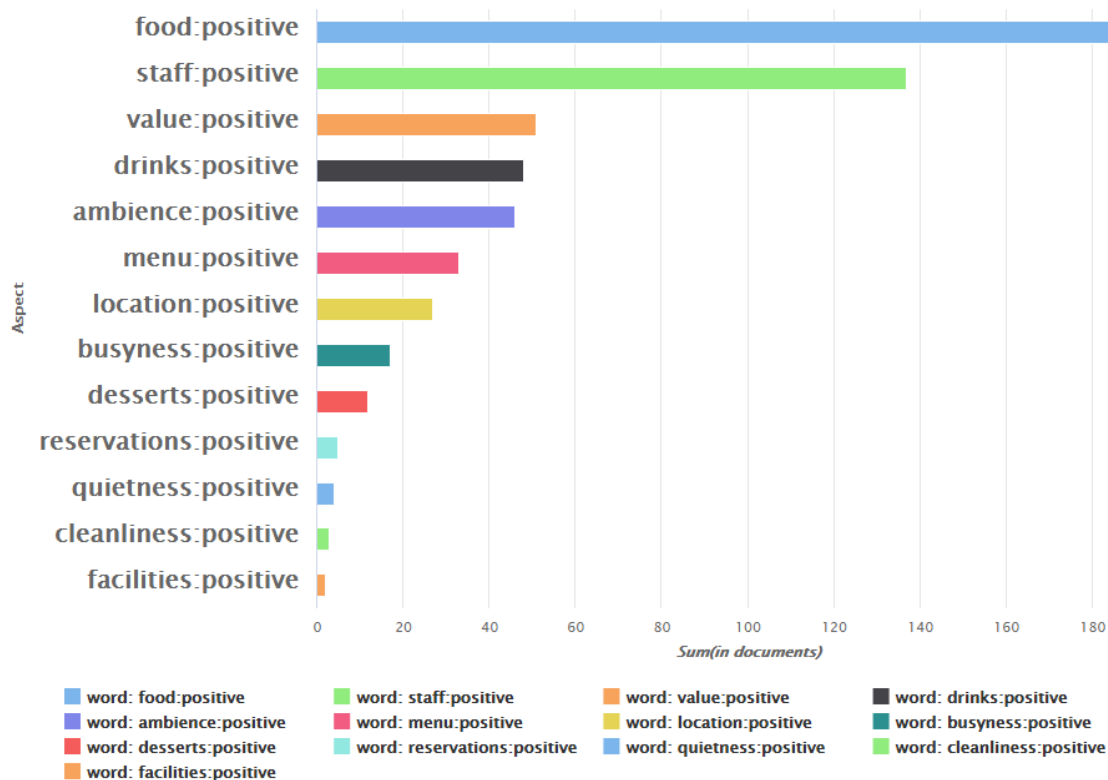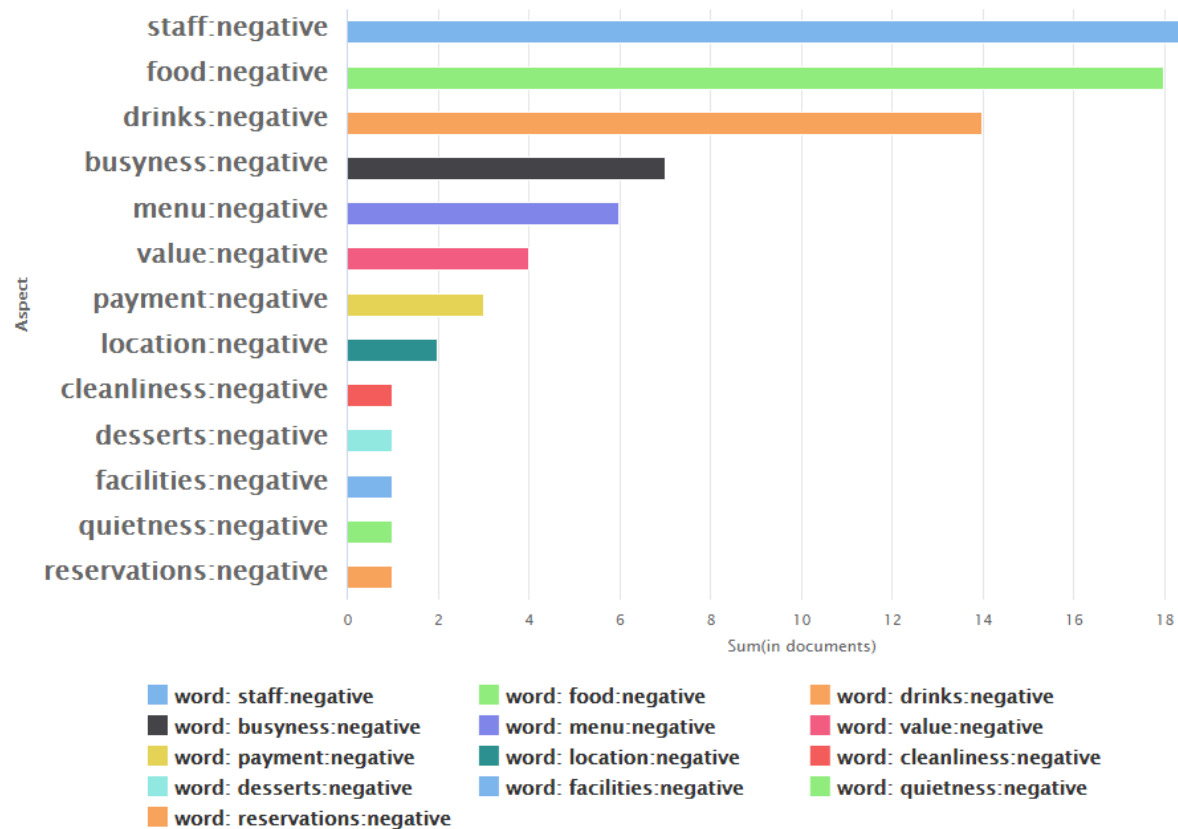

Figure 29. Positive Aspect Distribution of Customers.

Source: Adapted from RapidMiner software by the author.

Visualization of Negative Sentiment Polarity: 217 customer reviews of the Cuban restaurant were evaluated with aspect-based sentiment analysis. 13 aspects were identified that reflect the negative emotions of customers.

Negative Aspect Distribution of Customers: staff:negative 19, food:negative 18, drinks:negative 14, busyness:negative 7, menu:negative 6, value:negative 4, payment:negative 3, location:negative 2, cleanliness:negative 1, desserts:negative 1, facilities:negative 1, quietness:negative 1, reservations:negative 1 (Figure 30).

Figure 30. Negative Aspect Distribution of Customers.

Source: Adapted from RapidMiner software by the author

According to the aspect-based sentiment analysis results, it is seen that the customers have a significant positive sentiment about the Cuba restaurant. Food and staff dimensions appear to contain more positive sentiment than other dimensions. The negative sentiment of customers about the Cuba restaurant is significantly lower. It has been determined that negative sentiments are related to the staff and food.

Visualization of Positive and Negative Sentiment Polarity: Positive sentiment and negative sentiment were compared in customer reviews for Cuba Restaurant. Negative sentiments are seen to be very low compared to positive sentiments.

41

The comparison of positive and negative sentiments is presented in the Figure 31.
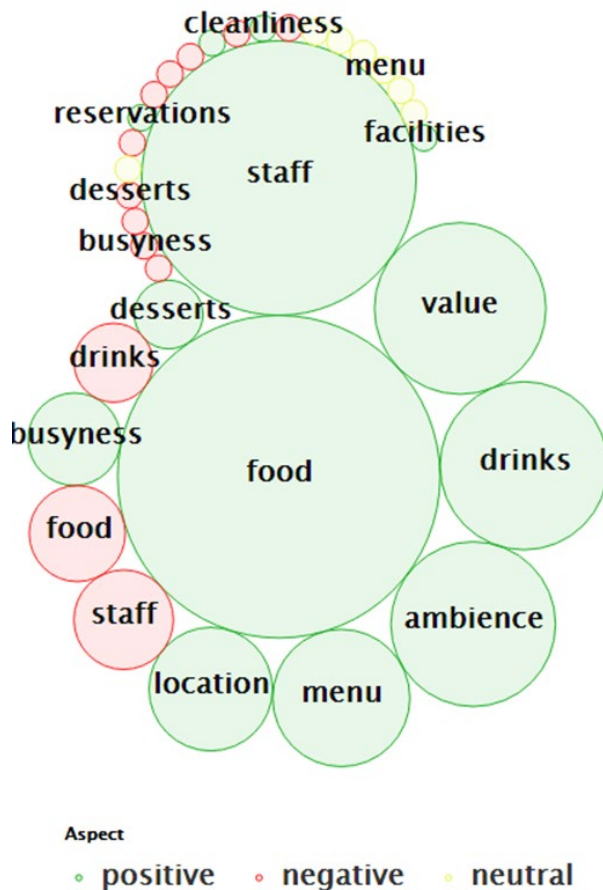
Figure 31. Visualization of Positive and Negative Sentiment Polarity.

Source: Adapted from RapidMiner software by the author.

Conclusion

In the first part of this study, text mining is discussed. In the second part, text mining methods and their applications in the tourism sector are explained with examples.

In example applications, the methods of determining sentiments from online tourist reviews were examined. Determination of sentiments from textual content is done using sentiment analysis (alternatively known as opinion mining) techniques. Sentiment analysis is applied at document-level, sentence-level and aspect-based sentiment levels.

Aspect-based sentiment level is thought to be a more effective technique than other levels (Medhat et al., 2014; Nasim & Haider, 2017). The most important reason for this is that the text analysis using ABSA technique has the ability to determine sentiments for each entity/event in the document.

Tourism is an information-intensive sector (Doolin et al., 2002). The tourism sector should consider online tourist reviews to improve their products and services. Extracting meaningful relationships from these reviews provides an important source of information for all stakeholders. Tourist reviews from other online platforms can be analyzed by considering other stakeholders in the tourism sector.

Online tourist reviews and opportunities for stakeholders can be listed as follows;

Opportunities for destinations;

- Opportunity to understand and interpret the destination

- Opportunity to understand and interpret tourist needs

Opportunities for tourism businesses;

- Opportunity to evaluate products and services offered to tourists: Discovering negative sentiments, discovering positive sentiments, opportunity to manage customer relationships.

- Opportunity to understand and evaluate competitors: Providing product and service advantage, product and service development or improvement, providing a price advantage (price determination).

Opportunities for tourists;

- Opportunity to make travel plans.

References

Agarwal, B., & Mittal, N. (2016). *Prominent Feature Extraction for Sentiment Analysis* (N. Mittal (ed.)). Springer International Publishing. https://doi.org/10.1007/978-3-319-25343-5

Aggarwal, C. C., & Zhai, C. (2012). A survey of text clustering algorithms. In *Mining text data* (pp. 77–128). Springer.

Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding Satisfied and Dissatisfied Hotel Customers: Text Mining of Online Hotel Reviews. *Journal of Hospitality Marketing and Management*, *25*(1), 1–24. https://doi.org/10.1080/19368623.2015.983631

Buzova, D., Sanz-Blas, S., & Cervera-Taulet, A. (2019). Does culture affect sentiments expressed in cruise tours' eWOM? *The Service Industries Journal*, *39*(2), 154–173. https://doi.org/10.1080/02642069.2018.1476497

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, *28*(2), 15–21. https://doi.org/10.1109/MIS.2013.30

Chenlo, J. M., & Losada, D. E. (2014). An empirical study of sentence features for subjectivity and polarity classification. *Information Sciences*. https://doi.org/10.1016/j.ins.2014.05.009

Davydova, O. (2017). *Sentiment Analysis Tools Overview, Part 1. Positive and Negative Words Databases*. Medium.Com. https://medium.com/@datamonsters/sentiment-analysis-tools-overview-part-1-positive-and-negative-words-databases-ae35431a470c

Delen, D., & Crossland, M. D. (2008). Seeding the survey and analysis of research literature with text mining. *Expert Systems with Applications*, *34*(3), 1707–1720. https://doi.org/10.1016/j.eswa.2007.01.035

Doolin, B., Burgess, L., & Cooper, J. (2002). Evaluating the use of the Web for tourism marketing: A case study

43

from New Zealand. *Tourism Management*. https://doi.org/10.1016/S0261-5177(02)00014-6

Feldman, R., & Sanger, J. (2007). Introduction to text mining. *The Text Mining Handbook: Advanced Approaches to Analyzing Unstructured Data*, 1–10.

Hou-Feng, X. G. W. (2011). The development of topic models in natural language processing. *Chinese Journal of Computers*, *8*, 9.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '04*, 168. https://doi.org/10.1145/1014052.1014073

Jo, Y., & Oh, A. H. (2011). Aspect and sentiment unification model for online review analysis. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining - WSDM '11*, 815. https://doi.org/10.1145/1935826.1935932

Kim, Y. (2014). Convolutional neural networks for sentence classification. *ArXiv Preprint ArXiv:1408.5882*.

Lei, S., & Law, R. (2015). Content analysis of TripAdvisor reviews on restaurants: A case study of Macau. *Journal of Tourism*, *16*(1), 17–28. http://search.ebscohost.com/login.aspx?direct=true&db=hjh&AN=111435361&site=ehost-live

Li, Q., Li, S., Zhang, S., Hu, J., & Hu, J. (2019). A review of text corpus-based tourism big data mining. *Applied Sciences (Switzerland)*, *9*(16). https://doi.org/10.3390/app9163300

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011

Meena, A., & Prabhakar, T. V. (2007). Sentence Level Sentiment Analysis in the Presence of Conjuncts Using Linguistic Analysis. In *Advances in Information Retrieval* (pp. 573–580). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-71496-5_53

Miner, G., Elder IV, J., Fast, A., Hill, T., Nisbet, R., & Delen, D. (2012). *Practical text mining and statistical analysis for non-structured text data applications*. Academic Press.

Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, *40*(10), 4241–4251. https://doi.org/10.1016/j.eswa.2013.01.019

Nasim, Z., & Haider, S. (2017). ABSA Toolkit: An Open Source Tool for Aspect Based Sentiment Analysis. *International Journal on Artificial Intelligence Tools*, *26*(06), 1750023. https://doi.org/10.1142/s0218213017500233

Özen, İ. A. (2019). Müşteri Şikâyetlerinde Fikir Madenciliği: Antalya Helal Oteller Üzerine Bir Araştırma. *2.Uluslararası Helal Turizm Kongresi*.

Özen, İ. A., & İlhan, İ. (2020). Opinion Mining in Tourism: A Study on "Cappadocia Home Cooking" Restaurant. In E. Çeltek (Ed.), *Handbook of Research on Smart Technology Applications in the Tourism Industry* (pp. 43–64). IGI Global.

RapidMiner. (2014). *Getting started with RapidMiner Studio - RapidMiner Documentation*. RapidMiner. https://docs.rapidminer.com/latest/studio/getting-started/

Rushdi Saleh, M., Martín-Valdivia, M. T., Montejo-Ráez, A., & Ureña-López, L. A. (2011). Experiments with SVM to classify opinions in different domains. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2011.05.070

Seker, S. E. (2016). Duygu Analizi (Sentimental Analysis). *YBS Ansiklopedisi*, *3*(3), 21–36.

Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*, *311*, 18–38. https://doi.org/10.1016/j.ins.2015.03.040

Soucy, P., & Mineau, G. W. (2005). Beyond TFIDF weighting for text categorization in the vector space model. *IJCAI International Joint Conference on Artificial Intelligence*, 1130–1135.

Thet, T. T., Na, J. C., & Khoo, C. S. G. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. *Journal of Information Science*. https://doi.org/10.1177/0165551510388123

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT '05*, 347–354. https://doi.org/10.3115/1220575.1220619

Xianghua, F., Guo, L., Yanyan, G., & Zhiqiang, W. (2013). Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. *Knowledge-Based Systems*. https://doi.org/10.1016/j.knosys.2012.08.003

44