TEXT MINING FOR ASPECT BASED SENTIMENT ANALYSIS ON CUSTOMER REVIEW: A CASE STUDY IN THE HOTEL INDUSTRY

Fitra A. Bachtiar, Wirdhayanti Paulina, Alfi Nur Rusydi

Faculty of Computer Science, Brawijaya University, fitra.bachtiar@ub.ac.id

ABSTRACT

The development of the role of the OTA (Online Travel Agent) site has become one of the E-WOM (Electronic Word of Mouth) media in addition to its main function as a platform for ticket reservations to encourage stakeholders in the hotel industry to utilize E-WOM for business continuity. One of the guest houses in Malang realized the importance of E-WOM because 90 percent of the booking process originated from the OTA website. However, the process of processing customer reviews only focuses on physical reviews, namely Guest Reviews. Meanwhile, information from online sources can have a more significant impact on E-WOM. One of the techniques of text mining is sentiment analysis which can be used to process and group text reviews. Sentiment analysis can be done to determine the sentiment of opinions on customer reviews to determine customer satisfaction with guest house services that aim to produce a positive E-WOM. Sentiment analysis is carried out at the aspect level using aspects of location, room, food, price, and service. The text of the review used in Indonesian originates from the sites Agoda.com, Expedia, Pegi-Pegi, Booking.Com, TripAdvisor and has a timeline from 2012 to 2019. This research yields findings in the form of customer satisfaction analysis of the five aspects where food aspects have urgency to be addressed and corrected immediately. Evaluation of the classification results also proves the effectiveness of the SVM method from Naïve Bayes

Key words: Guest House, E-WOM, Sentiment Analysis.

1. INTRODUCTION

The rapid growth of technology has encouraged the development of the role of the OTA (Online Travel Agent) site as one of the E-WOM (Electronic Word Of Mouth) media in addition to its main function as a platform for ticket reservations. Westbrook (1987) states that all informal communication aimed at consumers through internet-based technology related to the use or characteristics of a product, service, or provider is called E-WOM. Through these sites, customers are expected to provide reviews / reviews about what they feel and experience after a visit at the place they are going to either a hotel, restaurant, amusement vehicle and so on. Positive e-WOM can be generated through good customer reviews, while good reviews can be generated from satisfying customer experience for the services and accommodations produced.

One of the guest houses in Malang realized the importance of E-WOM in its business continuity because 90 percent of the booking process originated from the OTA website. Currently, the guest house has been listed on several OTA sites, namely TripAdvisor, Booking.com, Expedia, Agoda and Pegi-Pegi. However, the process of processing customer reviews only focuses on physical reviews, namely Guest Reviews. Meanwhile, information from online sources can have a more significant impact on E-WOM. The process of processing customer reviews becomes ineffective because it only focuses on one source, while the evaluation of guest house management services needs to be on target. In addition to these problems, the wide range of hotel attributes makes it difficult for stakeholders to determine aspects that have urgency to be addressed immediately.

A method that can be used to process and group text reviews is a sentiment analysis. Sentiment analysis or opinion mining is a computational study of people's opinions, sentiments, and emotions through entities and attributes that are expressed in text form (Liu, 2012). This sentiment analysis can classify the polarity of the text in sentences or documents to find out whether the opinions on the sentence or document are positive or negative. Sentiment analysis can be done to determine the sentiment of opinions on customer reviews to determine customer satisfaction with guest house services that aim to produce a positive E-WOM.

Ekawati and Khodra (2017) uses sentiment analysis on restaurant reviews to help restaurant owners improve the quality of their products and services. Sentiment analysis is carried out at the aspect level using aspects of food, service, price, and place. El-Jawad *et al.* (2018) describes several stages in sentiment analysis namely the data collection phase, the preprocessing phase, the term weighting phase, the classification phase, and the evaluation phase. Sentiment analysis will be carried out in 5 stages using the SVM method based on research Bhavitha *et al.* (2017) who analyzes comparisons of the techniques used in sentiment analysis. Researchers compared the lexicon based approach with the machine learning approach. Lexicon based has an average accuracy of 70% where

Machine Learning has an average accuracy that is much better above 80% and among the machine learning classifier results Support Vector Machine has the best accuracy compared to other classifiers. Miao *et al.* (2018) compared 3 machine learning methods, namely SVM, KNN, and Naïve Bayes for making a Chinese text news classification system. SVM has advantages in the value of Recall, Precision, and Recall although it takes longer than both methods because of the iteration process. Meanwhile, KNN and Naïve Bayes produce values that are not much different. Finally, Shi and Li (2011) uses the SVM method to compare TF-IDF with Frequency in the Term Weighting process in Sentiment Analysis. TF-IDF proved to be more effective with a Recall value of 89.2%, Precision 85.2%, and F1-Score 87.2%.

This study will discuss the use of sentiment analysis will be carried out at the selected aspect level to group reviews into 5 aspects and determine the sentiment of customer reviews by applying stages in text mining using machine learning classification methods. These five aspects were chosen based on research by Dolnicar and Otter (2003). The aspects used are the location, room, price, food and services chosen according to the needs of the organization. The results of this study are in the form of findings that can help stakeholders understand what are the customer complaints so that the decision making process to determine the services that need to be repaired and addressed becomes more effective and targeted. This paper is divided into several sections, namely Introduction, Literature Review, Methodology, Experiment, Analysis, and Conclusion.

2. LITERATUR REVIEW

2.1 Sentiment Analysis

Sentiment analysis or opinion mining is an umbrella of branches of study such as opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, mining review, etc. Sentiment analysis is a field of study that analyzes opinions, praise sentiments, one's emotions towards entities such as products, services, organizations, events, problems, and attributes of entities (Liu, 2012). Sentiment analysis is divided into 3 levels, namely: Document Level, Sentence Level, and Entity Level (aspect).

2.2 Support Vector Machine

The SVM algorithm aims to find Maximum Marginal Hyperplane (MMH) using support vectors and margins. MMH is the best hyperplane with the largest margin distance used to separate data maximally and accurately for each class. Margin can be defined as the shortest distance of a hyperplane to one side of the margin is the same as the hyperplane distance to the other side of the margin, provided that both margins are in a parallel position with the hyperplane (Han *et al.*, 2012).
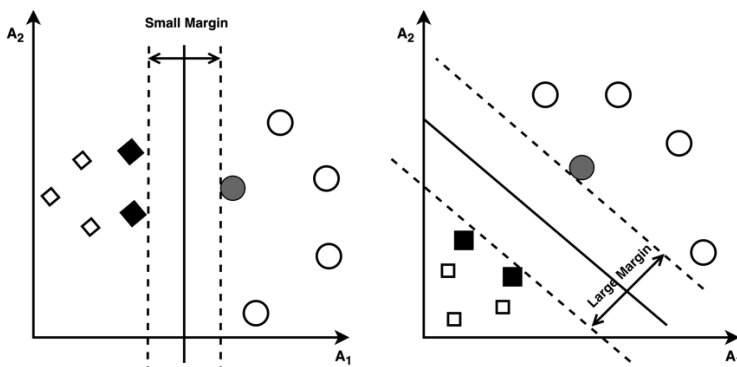

Fig. 1. Small Hyperplane Vs Optimal Hyperplane
Source : Han *et al.* (2012)

If there is a dataset in the form (X1, y1), (X2, y2), (X3, y3), ... , (Xi, yi) where Xi is tuple training and yi is the class label with $i = 1 .... N\ Xi \in Rd$ and $yi \in \{-1,1\}$. Every yi can choose one of two values either +1 or -1 SVM will form a classifier as shown in the Equation (1) as follows:

$$f(x_i) = \begin{cases} \geq 0, y_{i=} +1 \\ < 0, y_{i=} -1 \end{cases} \tag{1}$$

In SVM a hyperplane will be described in the following equation:

$$\boldsymbol{W}.\boldsymbol{X} + b = 0 \tag{2}$$

Based on equations (2), *W* is scalar Weight, *n* is attribute, *b* is scalar value or called bias, and *X* is training data set or training tuples.

106

2.3 Naïve Bayes

Naive Bayes is an algorithm put forward by a British scientist Thomas Bayes which is a classification algorithm using probability and statistical methods. This algorithm predicts opportunities based on past experience to be used in the future so it is known as the Bayes Theorem (Berrar, 2018). This study uses a multinomial model that takes into account the frequency of each word that appears in the document (Manning *et al.*, 2008). For example there are documents *d* and class *c*. To calculate the class of document *d*, it can be calculated using the formula:

$$P(c|document\ term\ d) = P(c)\ x\ P(t_1|c)\ x\ P(t_2|c)\ x\ P(t_3|c)\ x \dots x\ P(t_n|c) \qquad (3)$$

Based on Equation (3), P (*c*) is prior probability of class *c*, *tn* is word document d n[th], P (*c* | *term document d*) is the probability of *a* document including class *c* and P (*tn* | *c*) = N-word probability with class *c*. The probability of prior class c is determined by the formula:

$$P(c) = \frac{N_c}{N} \qquad (4)$$

Based on Equation (4), *Nc* is number of class *c* in all documents, *N* is number of all documents. The *n*-word probability is determined using the laplacian smoothing technique:

$$P(t_n\ |\ c) = \frac{count(tn,c)+1}{count(c)+|V|} \qquad (5)$$

In Equation (5), count *(tn, c)* is number of terms *tn* found in all training data with category *c*, count (*c*) is number of terms in all training data with category *c*, *V* is number of all templates in the training data

2.4 TF-IDF

TF-IDF Term Weighting is a weighting that is often used and is a combination of Term Frequency and Inverse Document Frequency. TF-IDF consists of frequency terms and inverse documents obtained from dividing the total number of documents to the number of documents that have these terms (Feldman and Sanger, 2007).

$$n_i(d) = df_i * \log D/df_i \qquad (6)$$

Based on Equation (6), $n_i(d)$ is the weight of the term *t* in the document *d*, df_i is frequency of occurrence of the term *t* in document *d*, and $\log \frac{D}{df_i}$ is inverse document frequency value of the term *t*.

2.5 Confusion Matrix

Confusion Matrix contains information about the performance of a classification system that is evaluated using the data or metrics contained in the Confusion matrix. Confusion Matrix analyzes how well the classification has been done on the actual class and the predicted class.

Table 1. Confusion Matrix.

		Prediction	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Source: Han *et al.* (2012).

Confusion matrix represents the level of accuracy of the classification process that has been done. Accuracy shows the proportion of the number of true predictions.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \qquad (7)$$

Precision is the proportion of correctly identified labeling (Completeness), the formula for finding Precision:

$$Precision = \frac{TP}{TP+TP} \qquad (8)$$

Recall is the proportion of information that can be found from a label, the search formula for Recall:

$$Recall = \frac{TP}{TP+FN} \qquad (9)$$

Precision and Recall can be used to get the proportion of other measurements namely F1-Score. F1-Score is the harmonic mean of the calculation of Precision and Recall, the formula to find F1-Score:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \qquad (10)$$
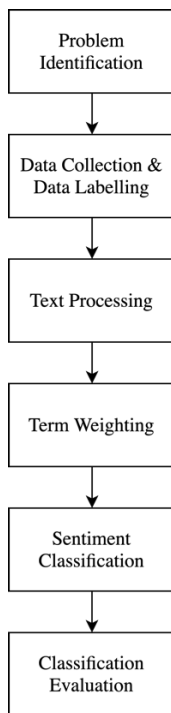
3. METHODOLOGY

Fig. 2.   Research Methodology

The research phase begins by identifying the problem through an interview process with one of the stakeholders to determine the aspects needed in the Sentiment Analysis. Next, the data collection stage is carried out on the OTA (Online Travel Agent) website using webscraping tools to extract customer review data that will be used in the Sentiment Analysis process. Then the data design is done which includes the process of categorizing data and labeling data manually. The next stage is preprocessing text which aims to prepare data before entering the term weighting stage by using the NLTK and Literature modules in Python. Some of the stages in the preprocessing text are formalization and translation, cleansing, tokenizing, case folding, stemming, and stopword removal. Furthermore, the term weighting phase or term weighting will produce a review data that has word weight using the Sckit-learn module in Python. The sentiment classification stage requires two data: review data which already has labels and word weights. This stage is divided into two parts, where the first stage aims to choose the best model that will be used in the second stage, namely the classification of sentiments for each aspect. Model selection is done by applying Stratified k-Fold Cross Validation and classification is done by classifier machine learning. Finally, the analysis and evaluation of the classification results are carried out using F1-Score, Precision, and Recall metrics and the Accuracy calculation. Evaluation is done by utilizing the Scikit-learn module in the Python Programming Language. See Figure 2 for details.

4. EXPERIMENT

4.1   Data Collecting and Labelling

This study uses a webscraping method with webcraper.io tools to extract variables containing information for sentiment analysis. The data used is the text of customer reviews on the TripAdvisor, Booking.com, Expedia, Agoda and Pegi-Pegi sites. Customer reviews that are used in Indonesian and have a period of 2012-2019. Total data collected was 1,561 consisting of 435 texts from the Agoda.com site, 37 texts from the Expedia site, 27 texts from the Pegi-Pegi site, 235 texts from the Booking.Com site, and 827 texts from the TripAdvisor site. The example of customer review can be seen in Figure 3. After that, the data that has been collected will be labeled manually based on the five aspects that have been selected namely location, room, food, price, and service as well as the polarity of sentiment aspects that are positive and negative. Data labeling is based on guidelines in Dolnicar and Otter (2003) and adjusted back to the Kertanegara to eliminate subjectivity. Data that already has a label is ready to enter the next stage, namely preprocessing. Table 2 shows the polarity of sentiment aspect guidance labeling.
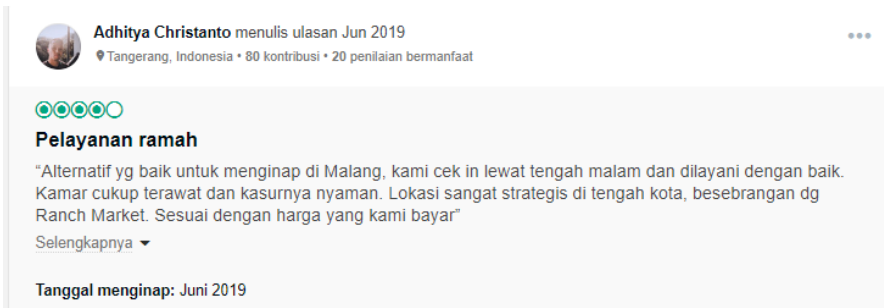
108

Fig. 3.   Customer Review (Source: Tripadvisor)

Table 2.   Polarity of Sentiment Aspects.

Aspect	Positive	Negative
Service	hotel meets customer expectations regarding the services they have	hotel does not meet customer expectations regarding the services they have
Price	hotel can set prices according to the value obtained	hotel cannot set prices according to the value obtained
Room	hotel can fulfill matters related to hotel room conditions	hotel cannot fulfill matters related to hotel room conditions
Food	hotel meets the customer's expectations regarding the food and drinks served	hotel cannot meet the customer's expectations regarding the food and drinks served
Location	the hotel can meet customer expectations related to the location of the hotel and its surroundings.	hotel cannot meet the customer's expectations regarding the condition of the location of the hotel and its surroundings

4.2   Text Preprocessing

Text Preprocessing is a step taken to prepare data before being analyzed in the sentiment classification process. The implementation of Text Preprocessing uses text variables of customer reviews on Kertanegara Premium Guest House through the NLTK and Sastrawi modules in Python. Formalization is the stage of changing words into standard forms in accordance with KBBI and Translation is the stage for translating words from foreign languages into Indonesian. Both stages are carried out manually by adjusting the KBBI Standard Form. Cleansing aims to eliminate elements that are not needed in the sentiment analysis process. These elements consist of punctuation, numbers, html tags. Tokenizing aims to separate the text or sentence review into pieces of words. Case Folding aims to change the words produced in the tokenizing process into lowercase characters. Stemming aims to eliminate word affixes in the customer review text so that the basic words are obtained. Stopword Removal aims to eliminate words that have little effect on sentiment analysis or words contained in a stopword list. Furthermore, the data enters into the Term Weighting phase which aims to provide weights that measure how important the value of a term is to the document. This study uses TF-IDF (Term Frequency-Inverse Document Frequency) which will display the word weight which increases according to the appearance of the word and IDF calculation produces a weight corresponding to the level of uniqueness of the word in the word index. This stage utilizes the Scikit-learn module in Python. Table 3 shows a representation customer reviews before and after the text preprocessing process.

Table 3.   Text Preprocessing.

Before	After
Suasana tenang dan kamar yang sangat nyaman membuat saya merasa sedang beristirahat di rumah sendiri. Rasanya ingin terus berada di sana menghabiskan waktu bersama orang yang saya cintai. Tidak mengecewakan	['suasana', 'tenang', 'kamar', 'sangat', 'nyaman', 'buat', 'rasa', 'sedang', 'istirahat', 'rumah', 'sendiri', 'rasa', 'terus', 'sana', 'habis', 'waktu', 'sama', 'orang', 'cinta', 'tidak', 'kecewa']

4.3   Sentiment Classification

Sentiment classification uses review text that has labels and weights. There are 2 stages in the Classification phase, namely the stage of model selection and the stage of classification model implementation. The model selection phase aims to choose the model that will produce the best accuracy. This stage uses K-Fold Cross Validation to

limit the problem of overfitting data in compiling data into training and testing, there is a single variance or overlap in the distribution of training data and testing data so that the model does not lose significant data for the modeling and testing process. The k value used in this stage is 6 so the dataset is divided into 6 parts. Next, the model implementation stage uses SVM and Naïve Bayes classification algorithms. All stages in classification utilize the Scikit-learn module at Python.

5. ANALYSIS

The highest number of reviews is in the aspect of rooms with a value of 1070 as shown in Table 4. This shows that the customer is very concerned about the condition and quality of rooms provided by the Guest House. Room cleanliness, amenities, bathroom conditions are things that are often discussed by customers during the stay. Comparison of positive and negative sentiment can be seen in Figure 4 and 5, respectively. The food aspect has the biggest ratio between positive and negative reviews. This shows that this aspect has urgency to be addressed immediately. The results of a trend analysis of the five aspects also show that food aspects have not been addressed optimally with increasing graphs in negative sentiment in 2019. Meanwhile, the other four aspects experienced an increase in the number of positive reviews in 2019. In 2013 and 2017 the highest average for negative sentiment. This shows that stakeholders have reformed and improved the management of the guest house so that the graph of positive sentiment increases rapidly in 2019 except in the aspect of food. The variety and taste of food that does not meet the standards of a guest house is often complained by customers.

Table 4.   Sentiment Classification.

Aspect	Positive	Negative	Total
Location	819	56	875
Room	860	210	1070
Food	200	105	305
Price	414	47	461
Service	816	69	885

Table 5.   Evaluation of classifier.

Aspect	Classifier	Precision	Recall	F1-Score	Accuracy
Location	SVM	0,88	0,94	0,91	0,93
	Naïve Bayes	0,91	0,92	0,92	0,92
Room	SVM	0,65	0,80	0,72	0,80
	Naïve Bayes	0,70	0,78	0,72	0,78
Food	SVM	0,78	0,68	0,57	0,68
	Naïve Bayes	0,55	0,64	0,55	0,64
Price	SVM	0,82	0,91	0,86	0,90
	Naïve Bayes	0,82	0,86	0,84	0,85
Service	SVM	0,86	0,93	0,89	0,92
	Naïve Bayes	0,87	0,88	0,88	0,88

The study also produced a comparison between the two classification algorithms, SVM and Naïve Bayes, which uses TF-IDF term weighting. The classification algorithm testing utilizes the classification report function in the Python Sckit-learn. Table 5 shows that overall SVM has a better Accuracy value than Naïve Bayes on all five aspects. This proves that SVM has better effectiveness than Naïve Bayes. The value of Precision, Recall and F1-Score almost reached 70 percent in all aspects except the food aspect. This can be caused by the process of labeling data review manually or an unbalanced dataset comparison.
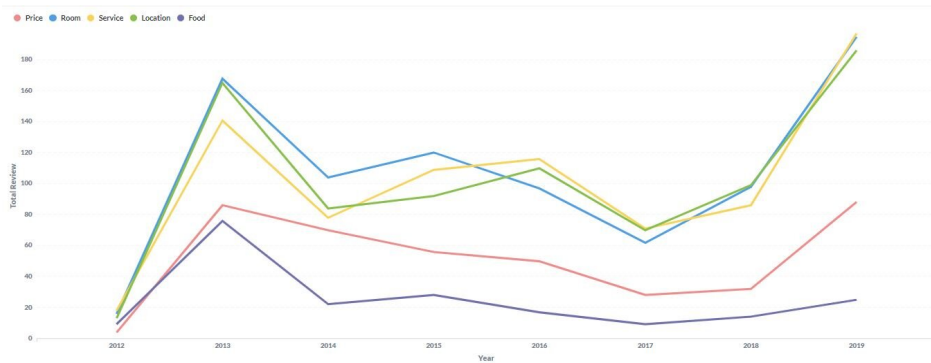
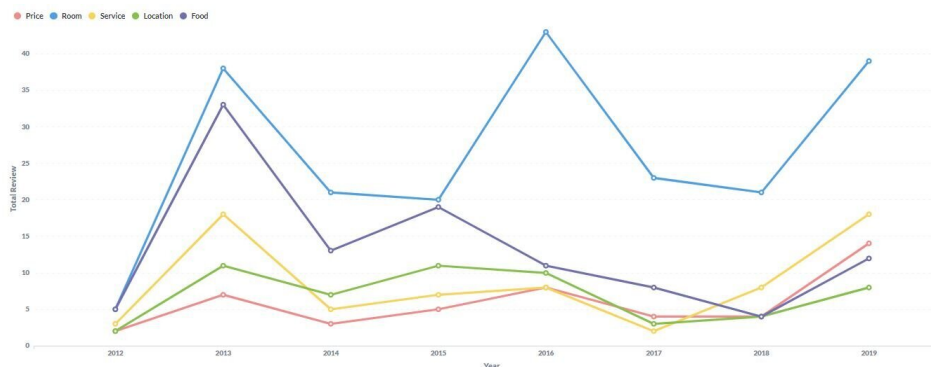**Fig. 4.   Trend Analysis of Positive Sentiment**

Fig. 5.   Trend Analysis of Negative Sentiment

6.   CONCLUSIONS

This study discusses the use of sentiment analysis at the aspect level on customer reviews as a guest house that aims to assist stakeholders in improving and improving management evaluations. Sentiment analysis conducted on the five aspects resulted in the finding that the food aspect had an urgency to be immediately addressed by the stakeholders. Evaluation of sentiment classification results in the SVM class of Accuracy, Precision, Recall, and F1-Score better than Naive Bayes which proves that SVM has better effectiveness. From this study it can be concluded that the guest house customers are satisfied with the management and accommodation of the guest house, but it is good for the guest house to continue to monitor and improve its services to all aspects, especially food aspects to produce a positive E-WOM.

REFERENCES

Berrar, D. (2018). Bayes' Theorem and Naive Bayes Classifier. *Encyclopedia of Bioinformatics and Computational Biology*, 1, 403-412.

Bhavitha, B.K., Rodrigues, A.P., and Chiplunkar, N.N. (2017). Comparative Study of Machine Learning Techniques in Sentimental Analysis, In: *International Conference on Inventive Communication and Computational Technologies (ICICCT 2017)*, 216-221. IEEE: New York NY.

Dolnicar, S., and Otter, T. (2003). Which Hotel attributes Matter? A review of previous and a framework for future research, In: *Proceedings of the 9th Annual Conference of the Asia Pacific Tourism Association (APTA)*, Griffin, T., and Harris, R. (Eds.), 1, 176-188. APTA: Busan, South Korea.

Ekawati, D., and Khodra, M.L. (2017). Aspect-based Sentiment Analysis for Indonesian Restaurant Reviews, In: *2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA)*, 1-6. Curan Associates: Red Hook NY.

El-Jawad, M.H.A., Hodhod, R., and Omar, Y.M. (2018). Sentiment Analysis of Social Media Networks Using Machine Learning, In: *2018 14th International Computer Engineering Conference (ICENCO)*, 174-176. IEEE: New York NY.

Feldman, R., and Sanger J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data.* New York: Cambridge University Press.

Han, J., Kamber, M., and Pei, J. (2012). *Data Mining: concepts and techniques.* Elsevier/Morgan: Amsterdam.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers: Williston VT.

Manning, C.D., Raghavan, P., and Schütze, H. (Eds.) (2008). *An Introduction to Information Retrieval.* Cambrigde University Press: Cambridge.

Miao, F., Zhang, P., Jin, L. and Wu, H. (2018). Chinese News Text Classification Based on Machine learning algorithm, In: *2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 48-51. IEEE: New York NY.

Shi, H.X. and Li, X.J. (2011). A Sentiment Analysis Model for Hotel Reviews Based On Supervised Learning, In: *Proceedings of the 2011 International Conference on Machine Learning and Cybernetics, Guilin, 10-13 July, 2011*, 10-13. IEEE: New York NY.

Westbrook, R.A. (1987). Product/Consumption-Based Affective Responses and Postpurchase Processes. *Journal of Marketing Research*, 24(3), 258-270.