Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 161 (2019) 805–812

Procedia Computer Science

www.elsevier.com/locate/procedia

The Fifth Information Systems International Conference 2019

New Filtering Scheme Based on Term Weighting to Improve Object Based Opinion Mining on Tourism Product Reviews

Ahimsa Denhas Afrizal*, Nur Aini Rakhmawati, Aris Tjahyanto

Department of Information Systems, Institut Teknologi Sepuluh Nopember Kampus ITS Sukolilo-Surabaya 60111, Indonesia

Abstract

Reviews are an essential thing in tourism industry. Opinion mining used for processing a massive amount of review data, so it can be more useful for the industry. The utilization of filtering can improve the feature extraction result from object based on opinion mining and can improve opinion classification result generally. However, there is no proven method yet to develop filter data automatically. This work applies several term weighting methods such as TF-IDF mTFIDF and BM25 to develop filter data automatically. The result from this research is the best term weighting method for developing filter data, that can improve the feature extraction and opinion mining relatively. TFIDF become the best term weighting method applied for filter data combined with the most frequent objects, The accuracy is 37.98%, the precision is 50.69%, the recall is 44,28%, and F-measure 47.27% for hotel data. Meanwhile, for restaurant data, the accuracy is 37.98%, precision is 50.69%, recall is 44.28%, and F-measure 47.27%.

Keywords: Opinion mining; feature extraction; filtering; term-weighting

1. Introduction

Nowadays, tourism is a potential industry. As a renewable resource industry, tourism has a significant contribution to foreign exchange income. The main product of the tourism industry, besides tourism spot itself, is a hotel and restaurant service. These services become the primary support for tourism spot, and sometimes become the tourism

* Corresponding author. Tel.: +62-813-3035-0401.
E-mail address: ahimsa.afrizal@gmail.com

10.1016/j.procs.2019.11.186

spot itself. Hotel and restaurant usually use reviewer website to promote their services, due to the customer behavior to read the review about the service before they decided to book it [1, 2]. These website using user-generated content, and the user can give their opinion either it good or bad about the service they booked. There are a lot of website using this model, such as zomato.com, foody.id, or tripadvisor.com.

On the other hand, there is a lot of users uploading their review to this platform, and the amount is increasing significantly [3]. It becomes hard for the user to decide after reading a lot of reviews. Not only the user, but the business owner also had difficulties in proceeding all of the reviews data from the platform.

Based on that problem, a system that can extract the review to be useful information is needed [4]. It can be beneficial for the user to use it; thus they can decide quickly based on the review or opinion on the review website. It also can help the business owner to determine whether they have to improve the service, based on the feedback from the user more quickly.

Website review extraction becomes the main focus for several years, as a part of opinion mining research field. Researchers are trying to solve human weakness for processing a lot of opinions and extract them so that it can become a common understanding of some topics [5, 6]. To solve that problem, Marrese-Taylor develops an approach to improve Liu models [7], called aspect-based opinion mining, and implement it in the tourism domain. This method applied because sometimes a reviewer in the tourism domain not explicitly commenting about physical appearances like bed material, food ingredient, or the building materials. The reviewer sometimes mentions about intangible aspects like staff hospitality, restaurant ambience, or interior layout. This method never implemented in the tourism domain.

Maresse-Taylor method can decide the semantic orientation of the aspect effectively, showed by precision and recall measurement 90% on the average. On the other hand, the result for feature extraction performance is not good, its only 30% for the hotel, and only 40% for the restaurant. It can happen like that because the user sometimes mentioned non-hotel –restaurant aspect, like the weather, time, day, or city [1]. Those words may be relevant, but it is not the desirable input. Sishephaputra uses filters to improve extraction result [2].

Using filters, Sishephaputra can improve the extraction result from Maresse-Taylor experiment. His filter can increase frequent object extraction result, from 45.7 % to 64,69% for hotel, and from 44,82% to 64,61% for restaurant. This improvement happened to in common object and infrequent object extraction result, which increase from 22,33% to 63,02% for a hotel review, and from 21,6 % to 65,4% for the restaurant. All the sentiment classification results also showed more dominant improvement than the result that comes from object extraction without filters.

The main problem in the Sisephaputra research is, there was no method to pick up terms that considered as a filter or not. In the previous study, the filter terms were pick up manually by an expert from a lot of sources. Therefore, it will depend on the expert to list the words. Main contribution of this research is to propose a new method to pick up the filter terms from a review corpus.

2. Research method

This research uses the following methods: extract review data for hotel and restaurant from TripAdvisor review, pre-processing method, filtering method, object-based opinion mining method, a classification method, and classification performance measurement method.

Fig. 1(a) explained the method step by step. The first method is extracting review data for hotel and restaurant from TripAdvisor. We use the same dataset used by Maresse-Taylor and Sisephaputra, involving 100 restaurant review data and 100 hotel review data from TripAdvisor. This review dataset, manually tagged by expert to define its opinion orientation whether a positive, negative, or neutral review. For better assessment, we invite six managers from each industry, with more than 5 years' experience. The second method is performing the pre-processing method, including tokenizing, POS Tagging, Chunking, Stopping, and Stemming using NLTK Library in Python. Third, a filtering method applied in the system. This method merely removes the terms that not included in the Filter Data List. Sishephaputra manually set the list from an expert, and collect 256 Data Filter for the restaurant, and 300 data Filter for the hotel. There was no specific method to obtain the Filter Data List. Consequently, the filtering method is our primary focus in this experiment. The fourth method is object-based opinion mining, consist of object extraction, and sentence orientation. Fifth, is the classification method, using Naïve Bayes, Random Forest, and Support Vector Machine. Sixth, analyse the performance of the object extraction, sentence orientation, and text classification.

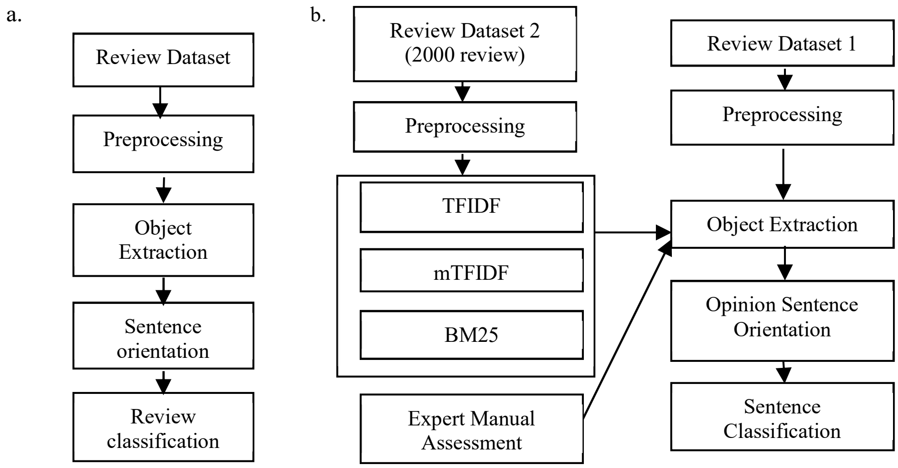

Fig. 1. (a) Research Method; (b) Proposed Filtering Method.

In the previous research, Sisephaputra makes a list of data filter manually chosen by an expert. There was no specific method explained in that research, hence we offer and try to find the best approach to select the data filter.

We proposed two approaches to collect the Filter Data List. First, manual assessment by the expert. Three hotel manager and three restaurant owner/manager in Surabaya, asses the Filter Data List from Sishephaputra using Hotel Attribute [7], and Restaurant Attribute [8]. The Hotel Attribute consists of Value, Location, Sleep Quality, Room, Cleanliness, and Service. While the Restaurant Attribute is Food, Value, Atmosphere, and Service. The given terms in the Filter Data List is being rated by the experts, with maximum value 6 for the hotel, and maximum value 4 for the restaurant based on the attribute. This filter data list will be a new benchmark for this research because it has better justification than Sishephaputra filter data list.

Second approach, we use term weighting to extract the best terms in the dataset therefore we can compile the Filter Data List automatically. We used TripAdvisor hotel review dataset [9] and TripAdvisor restaurant review dataset [10], and only 2000 random review processed in this approach. To find the best term weighting method, a preprocessing method applied, and then we use three different term weighting method. Implementation of Filter Data List into the Object Extraction step showed in Fig 1(b).

First, we use TFIDF [11], because this method is widely used to term weighting in different fields. Second, we use mTFIDF method [12], a modified TFIDF that calculated the absence of the terms in another document. Because of its characteristic, mTFIDF is useful for calculating term weighting in a short document like a review. Third, we use BM25, a term weighting approach used in search engine query [13] this method uses the newest term weighting formula that improves TFIDF weighting. We use the Filter Data List from each method and compare the result with the Filter Data List from previous research. Table 1 showing extracted filter data sample from TFIDF method.

Table 1. Extracted Filter Data.

	Hotel			Restaurant	
No	Terms	TFIDF Value	No	Terms	TFIDF Value
1	hotel	218.64	1	good	133.36
2	room	161.54	2	food	120.42
3	staff	105.13	3	great	101.88
4	locat	103.23	4	servic	97.66
5	great	96.47	5	place	71.08

After Filter Data List compiled automatically from each term weighting method, object-based opinion mining applied. This opinion mining method consists of these following steps. First is object extraction, we take the combination of common object and frequent object with filter data, because of common object and filter data combination gain higher result in previous research. Second, object sentiment orientation method. In this step, the orientation of the object extracted from each sentence define whether they are positive, neutral, or negative comment [2]. Third, as mentioned before, we classify the review based on their orientation, using three different classification method.

2.1. System overview

Sishephaputra rebuilds Object-Based Opinion Mining System proposed by Maresse-Taylor, and we use it for our research. As mentioned before, we add some new filter data which is the result of four different methods into the system. For this research purpose, we develop some additional system, such as TFIDF weighting using Python Sklearn Library, mTFIDF weighting, and BM25 weighting. We are also implementing the pre-processing system separate from another method to make it flexible.

2.2. Data preparation

We use the same dataset from Maresse-Taylor that used by Sishephaputra too. 100 sentence review about Restaurant, and 100 sentence review about Hotel. This data taken from TripAdvisor and individually reviewing hotel and restaurant in Lake City, Chile. An expert tags these sentence manually and decides whether the sentence is positive, neutral, or negative opinion. The second dataset is a review dataset from TripAdvisor, consist of 2000 review, each for the restaurant and hotel. We take the data from world.data, and it's chosen randomly.

2.3. Pre-process

We use the same preprocess step used by Maresse-Taylor and Sisephaputra. The first step is tokenizing; all of the data was tokenized into terms. The second step is the part-of-speech tagging. In this step, we tag all of the term using the NLTK Library. For example, the system will tag the term with nouns (NN), adjective (JJ), or verb (V). It consists of 36 different tags. In the Chunking process, we take only noun phrase with tag (NN) (NN), or (JJ) (NN). Stopwords removed in Stopping process, and the noun phrase stemmed using Porter Stemmer in the last step.

2.4. Term weighting

Term weighting is a standard method to do aspect extraction. This method tries to find an explicit aspect like noun phrase from a review in a specific domain. This approach is adequate because when a user gives their review, they use a similar phrase [14]. Popescu et al. [15] deleting all non-aspect phrase to increase the accuracy of aspect extraction. Blair & Goldenshon [16] calculating noun phrase frequency from opinion sentence. Ku et al. [17] use TF-IDF to find the frequent term in a review. Scaffidi et al. [18] try to identify aspect from phrase frequency in the review. These previous methods show that term weighting can be a useful approach to extracting element. In this research, we use the term weighting approach to choose a data filter automatically from the corpus. Each method explained below.

2.4.1. TFIDF

TF-IDF weighting is a common term weighting method in information retrieval and widely used for a lot of application [19]. We can determine the weight of each phrase in the corpus using a combination between term frequency and inverse document frequency in each document. As showed in the reference paper, Tf refers to term frequency, the raw frequency of the term in each document. N refers to the total number of documents in the corpus, and df is the number of the document which term t appear.

2.4.2. mTFIDF

mTF is a modified term frequency weighting offered by Sabbah et al. [12]. The main idea of this method includes the missing term weight into the term frequency formula. The missing term defines as a term that appears in the corpus, but not appear in the specific document. In this new formula, the missing term from the document and the amount of the document that does not contain the term, affect the total weight of the term. We can use this approach in a document with short length like a review.

In this scheme, the proportion of the total count of a term in all documents (denoted by Tt) to the total token count of the corpus (denoted by Tc) is considered. However, since Tc is usually much higher than Tt, the log of the inversed proportion is applied with the aid of square root of Tc for scaling the effect of massive difference between Tt and Tc. Also, the proportion of the length of the document (length) to Tc is considered in the normalization, when calculating the weight of term t in document d. The proposed modified Term Frequency is denoted by $mTFt$, however, the formula of $mTFt$, $d$is expressed as showed in the paper [12].

As in the standard TF formula, $tf\ t$, represents the raw frequency of term t in document d (which is commonly the count of occurrences of the token t). However, Tt stands for the total number of raw frequency of term t in all documents in the collection, while Tc and length are the total number of distinctive terms in the collection (known as Feature Space size) and the length of document d, respectively. However, the length of a document in this formula is the number of distinctive tokens in the document.

2.4.3. BM25

BM25 formula or Okapi BM25 developed from 2-Poisson probability model by Robertson and Walker. This formula used by search engine to give ranking to a document, based on term relevance from search query [20]. BM25 used for document weighting, different from term weighting like TF-IDF. But, inside BM25 formula, we can use the new term weighting scheme that different than TF-IDF.

2.5. Data classification

The method used in this research is SVM, Naive Bayes, and Random Forest. Support Vector Machine (SVM) is a classification algorithm that aims to find the function of the separator (hyperplane) with the largest margin, to separate the two data sets optimally. Text data matches the SVM algorithm as it tends to have a high dimension. The SVM algorithm is resilient and has accurate results [20].

The Naive Bayes classifier is a simple probabilistic classifier based on the Bayes theorem. The naive Bayes classifier has the assumption that the effect of an attribute value is independent of the value of another attribute. It is performed to simplify the calculation, and because of that assumption, it is considered naive. The advantages of this algorithm are easy to implement and can work with little training data [21].

Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. This method uses a random selection of features to split each node yields error rates that compare favorably to AdaBoost but are more robust concerning noise. Internal estimates monitor error, strength, and correlation and these are used to show the response to increasing the number of features used in the splitting. Internal estimates are also used to measure variable importance. These ideas are also applicable to regression [22].

3. Result and analysis

We use Filter Data List from Sisephaputra research, consist of 410 filter data for the hotel, and 288 filter data for the restaurant. We let some expert give a score based on Hotel attribute [7] and Restaurant attribute [8] to the filter data list and ask them to add more term to the list. As a result, Expert recommends 366 filter data for the hotel and 227 filter data for the restaurant. This list used for benchmarking another filter data list resulted from the term weighting method. Filter data list also extracted automatically from another review dataset using TFIDF,mTFIDF, and BM25 method, collecting bigram words. Base on the weight of the term, two different frequency has taken to limit the data. The result is taken and implemented as a filter in object-based opinion mining, seen in Table 1 for hotel and Table 2 for the restaurant.

Table 2. Hotel Object Extraction.

Object Extraction Type	Accuracy	Precision	Recall	F-Measure
Frequent Expert	38.74%	64.49%	43.52%	51.97%
Frequent TFIDF >5	37.57%	48.87%	44.35%	46.5%
Frequent TFIDF>10	37.98%	50.69%	44.28%	47.27%
Frequent mTFIDF >30	37.31%	50.31%	43.06%	46.4%
Frequent mTFIDF>10	37.20%	50.20%	42.28%	45.27%
Frequent BM25 >35	37.45%	50.37%	43.12%	46.7%
Frequent BM25>40	37.30%	50.35%	43.09%	46.5%

Table 3. Restaurant Object Extraction.

Object Extraction Type	Accuracy	Precision	Recall	F-Measure
Frequent Expert	38.74%	64.49%	43.52%	51.97%
Frequent TFIDF >5	37.57%	48.87%	44.35%	46.5%
Frequent TFIDF>10	37.98%	50.69%	44.28%	47.27%
Frequent mTFIDF >30	37.31%	50.31%	43.06%	46.4%
Frequent mTFIDF>10	37.20%	50.20%	42.28%	45.27%
Frequent BM25 >35	37.45%	50.37%	43.12%	46.7%
Frequent BM25>40	37.30%	50.35%	43.09%	46.5%

In the object extraction process (see Table 4 to Table 7), we took one better result from each method, which means TFIDF that have weight bigger than 10, mTFDIF that have weight bigger than 30, and BM25 bigger than 35. It showed that TFIDF gets a better result than mTFIDF and BM25 approach, but not really significant. TFIDF, BM25, and mTFIDF was a term weighting based on term frequency and document frequency combination, with some modification for a special case. All of the formula depends on the term frequency and neglect the semantic factors inside the text. That formula characteristic similarity causing the insignificant result differences among them. The object extraction result affects the classification process. We found that all term weighting approach gives similar classification result. Hotel category process showed that mTFIDF and BM25 give a better result than TFIDF, but there was no significant difference. The similar result came from restaurant data, with mTFIDF and BM25 give a better result than TFIDF, but there was no significant difference.

Table 4. Classification Results Expert.

Classifier	Hotel				
	Acc	Prec	Rec	F-Meas	ROC
Naïve Bayes	92.5 %	93%	92.6%	92.3%	97.9%
SVM	81.87%	70%	81.9%	75.3%	81.6%
RF	93.6%	93.7%	93.7%	93.6%	98.6%
	Restaurant				
Naïve Bayes	91.8%	92.2%	91.8%	91.1%	97.3%
SVM	86.5%	78.7%	86.6%	82%	85.5%
RF	94.6%	94.6%	94.7%	94.5%	98.6%

Table 5. Classification Results TFIDF.

Classifier	Hotel				
	Acc	Prec	Rec	F-Meas	ROC
Naïve Bayes	91.1%	91.17%	91.2%	90.9%	96.8%
SVM	80.2%	68.4%	80.2%	73.2%	79.6%
RF	91.06%	91.3%	91.1%	91.1%	98.1%
	Restaurant				
Naïve Bayes	91.06%	91.4%	91.1%	90.3%	96.7%
SVM	85.74%	77.5%	85.7%	80.9%	83.9%
RF	93.5%	93.5%	93.5%	93.3%	97.9%

Table 6. Classification Results mTFIDF.

Classifier	Hotel				
	Acc	Prec	Rec	F-Meas	ROC
Naïve Bayes	91.3%	92%	91.4%	91.1%	97.1%
SVM	80.3%	68.7%	80.4%	73.4%	79.7%
RF	92.2%	92.4%	92.2%	92.2%	98.4%
	Restaurant				
Naïve Bayes	91.06%	91.3%	91.1%	90.3%	96.6%
SVM	85.53%	77%	85.5%	80.6%	83.8%
RF	93.6%	93.6%	93.6%	93.3%	98%

Table 7. Classification Results BM25.

Classifier	Hotel				
	Acc	Prec	Rec	F-Meas	ROC
Naïve Bayes	91.3%	91.9%	91.3%	91%	97%
SVM	80.2%	68.7%	80.3%	73.3%	79.7%
RF	92%	92.2%	92.1%	92.1%	98.3%
	Restaurant				
Naïve Bayes	91.27%	91.4%	91.3%	90.6%	96.6%
SVM	85.6%	77.3%	85.6%	80.8%	83.9%
RF	93.5%	93.6%	93.5%	93.2%	97.7%

4. Conclusion and future research

In In this research, we try to find a method to gather a filter data list to improve object based opinion mining. We implement three different term weighting method to improve previous research which didn't propose any method to gather filter data list. The result showed that TFIDF is the best term weighting method for object extraction, but not significantly exceed mTFIDF and BM2. We expect this result before because TFIDF used widely in many research come out with the best result. Likewise, BM25 and mTFIDF formula also based on term frequency and document frequency. As a result of that, mTFIDF and BM25 give better result in the classification process with minimal differences. In conclusion, this work has successfully proposed the method to gather filter data list to improve object based opinion mining.

For future research, the main objective is to find another method to gather the filter data, such as semantic similarity, and compare the result with filter data from term weighting. We also suggest to let the expert reassess each term weighting result using the same criteria they used before, thus we can have expert justification over the automatic result comprehensively.

References

[1] Marrese-Taylor, E., J. D. Velásquez, and F. Bravo-Marquez. (2014) "A Novel Deterministic Approach for Aspect-Based Opinion Mining in Tourism Products Reviews." *Expert Systems with Applications*.

[2] Marrese-Taylor, E., J. D. Velásquez, F. Bravo-Marquez, and Y. Matsuo. (2013) "Identifying Customer Preferences About Tourism Products Using an Aspect-Based Opinion Mining Approach", in *Proceedings of The Knowledge-Based and Intelligent Information and Engineering Systems*, KES. pp. 182–191.

[3] Bucur, C. (2015) "Using Opinion Mining Techniques in Tourism", in *Procedia Economics and Finance* **23**: 1666 – 1673.

[4] Sisephaputra, B., and A. Tjahyanto. (2016) "The Utilization of Filter On Object-Based Opinion Mining In Tourism Product Reviews."

[5] Velásquez, J. D., and P. González. (2010) "Expanding the Possibilities of Deliberation: 1109." *The Use of Data Mining for Strengthening Democracy With An Application To*. p. 1110.

[6] Liu B., M. Hu, and J. Cheng. (2005)" Opinion Observer: Analyzing and Comparing Opinions On The Web", in *Proceedings of the 14th International Conference on World Wide Web,* ACM. pp. 342–351.

[7] Rhee, H., Timothy, Yang, and S. Byung. (2015) "Does Hotel Attribute Importance Differ by Hotel? Focusing on Hotel Star-Classification and Customers Overal Rating." *Computer in Human Behaviour* **50**: 576-587.

[8] Rhee, H., Timothy, Yang, S. Byung, and K. Kim. (2016) "Exploring the Comparative Salience of Restaurant Attributes : A Conjoint Analysis Approach." *International Journal of Information Management*.

[9] Zenodo.org. *Tripadvisor Hotel Review Dataset*. Available from: https://zenodo.org/record/1219899#.W6M5YnszbIW. [Accessed May 2018].

[10] Data.world. *Restaurant Review Dataset from Tripadvisor*. Available from: https://data.world/promptcloud/london-based-restaurants-reviews-on-tripadvisor-co-uk. [Accessed May 2018].

[11] Chianga, D.A., H.C. Keha, H.H. Huanga, and D. Chyrb. (2008)"The Chinese Text Categorization System with Association Rule and Category Priority." *Expert Systems with Applications* **35 (1-2)**: 102-110.

[12] Sabbah Thabit, Selamat Ali, Selamat Md Hafiz, Al-Anzi, S. Fawaz, Viedma Enrique Herrera, Krejcar Ondrej, Fujita, and Hamido. (2017) "Modified Frequency-Based Term Weighting Schemes for Text Classification." *Applied Soft Computing Journal.*

[13] Robertson Stephen, and Zaragoza Hugo. (2009) "The Probabilistic Relevance Framework: BM25 and Beyond." *Foundations and TrendsR in Information Retrieval* **3 (4)**: 333–389.

[14] Hu, M., and B. Liu. (2004a) "Mining and Summarizing Customer Reviews", in *Proceedings of The Tenth ACM SIGKDD International Conference On Knowledge Discovery and Data Mining, ACM*. pp. 168–177.

[15] Popescu, A., and O. Etzioni. (2005) "Extracting Product Features and Opinions from Reviews", in Proceedings of The Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Association for Computational Linguistics. pp. 339–346.

[16] Blair-Goldensohn Sasha, Kerry Hannan, Ryan McDonald, Tyler Neylon, George A. Reis, and Jeff Reynar. (2008) "Building a Sentiment Summarizer for Local Service Reviews", in *Proceedings of WWW-2008 Workshop on NLP in The Information Explosion Era*.

[17] Ku, L.-W, Y.-T. Liang, and H.-H. Chen. (2006) "Opinion Extraction, Summarization and Tracking In News and Blog Corpora", in *AAAI Spring Symposium: Computational Approaches to Analyzing* Weblogs. pp. 100–107.

[18] Scaffidi Christopher, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman Ng, and Chun Jin. (2007)"Red Opal: Product-Feature Scoring from Reviews", in *Proceedings of Twelfth ACM Conference on Electronic Commerce* (ec-2007).

[19] Chianga, D.-A, H.-C. Keha, H.-H. Huanga, and D. Chyrb. (2008). "The Chinese Text Categorization System with Association Rule and Category Priority." *Expert Systems with Applications* **35 (1-2)**: 102-110.

[20] Shah, F. P., and V. Patel. (2016) "A Review On Feature Selection and Feature Extraction for Text Classification", in *Wireless Communications, Signal Processing an Networking (WiSPNET), International Conference*. pp. 2264–2268.

[21] Domingos, P., and M.J. Pazzani. (1997) "On the Optimality of The Simple Bayesian Classifier Under Zeroone Loss." *Machine Learning* **29 (2/3)**: 103–130.

[22] Breiman, Leo. (2001) *Machine Learning*, Kluwer Academic Publishers, Netherland.