

Implementation of Integrated Bayes Formula and Support Vector Machine for Analysing Airline's Passengers Review

Aditya Tegar Satria^{1*}, Mustafid², Dinar Mutiara Kusumo Nugraheni¹

¹Master Program of Information System, School of Postgraduate Studies, Diponegoro University, Semarang, Indonesia

²Department of Statistics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia

³Department of Informatics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia

Abstract. Nowadays, the utilization of Internet of Things (IoT) is commonly used in the tourism industry, including aviation, where passengers of flight services can rate their satisfaction levels towards the product and service they use by writing their reviews in the form of text-based data on many popular websites. These passenger reviews are collections of potential big data and can be analyzed in order to extract meaningful informations. Some text mining algorithms are already in common use, including the Bayes formula and Support Vector Machine methods. This research proposes an implementation of the Bayes and SVM methods where these algorithms will operate independently yet integrated with other modules such as input data, text pre-processing and shows output result concisely in one single information system. The proposed system was successfully delivered 1000 documents of passenger reviews as input data, then after implemented the pre-processing method, the Bayes formula was used to classify the document reviews into 5 categories, including plane condition, flight comfort, staff service, food and entertainment, and price. While simultaneously, the positive and negative sentiment contained in the review document was analyzed with SVM method and shows the accuracy score of 83.6% for a training to testing set ratio of 50:50, while 82.75% accuracy for the 60:40 ratio, and 83.3% accuracy for the 70:30 ratio. This research shows that two different text mining algorithms can be implemented simultaneously in a effective and efficient way, while still providing an accurate and satisfying performance results in one integrated information system.

Keywords. text mining; Bayes formula; Support Vector Machine; integrated system; sentiment analysis;

* Corresponding author: tegarsatria@students.undip.ac.id

1 Introduction

Text mining is one of the processes in data mining that plays an important role in obtaining data pattern, trend, relationship and potential knowledge that could be analyzed from text-based data [1]. Several common algorithms that are used in text mining includes Naïve Bayes Classifier and Support Vector Machine (SVM), which can conduct classification sentiment analysis with a high level of accuracy [2]. In general, there are already a large number of studies that apply methods considered to be the best for text classification, wherein these methods are used to analyze data separately, using a different text pre-processing method and in which the output shown will be according to the method used. For example, the Bayes method will be used for the classification and clustering of result, while SVM will be used to conduct the positive or negative sentiment analysis.

This research suggests the implementation of two text mining algorithm methods, Bayes formula and SVM, to process the passenger reviews from the TripAdvisor website as a case study. These two algorithmic methods operate independently, yet simultaneously, according to their own functions but are integrated in one information system where other modules such as input data, text pre-processing and algorithm modules of Bayes and SVM are available in this one information system. The first step to acquiring the dataset for input can be done either real-time by the scrapping method, or by gathering existing data that has already been saved. Next, the dataset input is processed by the text pre-processing module to produce a clean data to then be processed by the Bayes method for the purpose of classifying the review's content into the determined categories. The text pre-processing module is also used for vectorization through the Bag-of-Words method for the processing of classification sentiment with SVM to analyze the positive and negative sentiment contained in the review documents. The objective of this research is to discover useful informations from input data by analysing the classification result from these two algorithm methods that will provide a complete information extraction in the form of review category and the sentiment inside of it, and will be displayed concisely by the system.

2 Methodology

2.1 Text pre-processing and bag-of-words model

In text mining, text pre-processing method is required as an initial stage to prepare dataset before the algorithm would able to process it [3]. The pre-processing method particularly used in this research is comprised of these stages: tokenization, stop word, stemming and lemmatization. Before using machine learning algorithm for training data, it is necessary to present the text document as a feature vector. A commonly employed model in Natural Language Processing (NLP) to conduct vectorization is the Bag-of-Words (BoW) model [4]. The main purpose of this model is to create a vocabulary, which is a collection of different words that appear in the document and each word is associated with its frequency of appearance. The word order in this vocabulary may be ignored. To illustrated, here are two document data, D_1 and D_2 , in a training set data:

[D_1]: *“Each plane has its own seats.”*

[D_2]: *“Every airlines has its own passengers.”*

Based on the data of the two documents above, the vocabulary can be written as such: $V = \{\text{each: 1, plane: 1, has: 2, its: 2, own: 2, seats: 1, every: 1, airlines: 1, passengers: 1}\}$. The vocabulary that is now formed is implemented as vectorization, shown in Table 1.

Table 1. Bag-of-Words Model

Input	each	plane	...	passengers
X_{D1}	1	1	...	0
X_{D2}	0	0	...	1
$\sum X_{D1}$ X_{D2}	1	1	...	1

The implementation of the text document vectorization as shown above will become an important key in the implementation of the next algorithms used to analyze text data.

2.2 Bayes formula for categories classification

Bayes formula is one of the more popular algorithm method used in text mining because of its easy use, relatively quick processing time, simple structure, easy implementation and high level of effectivity [5][6]. The Bayes algorithm can be used to find the highest probability value and conduct classification of tested text data in the most suitable category [7]. The basic Bayes formular for text classification [8] can be written as such:

$$Pr(Category|Word) = \frac{Pr(Word|Category) \cdot Pr(Category)}{Pr(Word)} \tag{1}$$

The equation above can be rewritten as such:

$$Pr(Ci|Wj) = \frac{Pr(Wj|Ci) \cdot Pr(Ci)}{Pr(Wj)} \tag{2}$$

Ci is the specified category ($C_1, C_2 \dots C_k$), while Wj is the frequency that a word appears in document X ($W_1, W_2 \dots W_k$). The equation $Pr(Ci|Wj)$ is the probability of word Wj in document X that is classified in category Ci , and $Pr(Wj|Ci)$ is the conditional probability for the amount of Wj data in category Ci . Probability $Pr(Ci)$ is the value for the probability of category Ci and $Pr(Wj)$ is the probability value of word Wj in the document.

Equation (2) is the main calculation to calculate the probability of document X , where the value of $Pr(Ci)$ as *prior probability*, can be calculated with the equation:

$$Pr(Ci) = \frac{\text{total number of words in category } Ci}{\text{total number of words in document } X} \tag{3}$$

Meanwhile the probability value of word Wj in document X or $Pr(Wj)$ is calculated with the equation:

$$Pr(Wj) = \frac{\text{frequency of word } Wj \text{ in document } X}{\text{total number of words in document } X} \tag{4}$$

The same method can be implemented to find the total amount of the frequency for all words in all categories in the training data base. Next, to calculate the likelihood of a certain category (Ci) that is inside a certain word (wj), then the appearance list of wj in Ci in the training data base is found and then al words in Ci are summed to find the value of $Pr(Wj|Ci)$, as demonstrated by the following equation:

$$Pr(Wj|Ci) = \frac{\text{frequency of word } Wj \text{ in category } Ci}{\text{total number of words in category } Ci} \tag{5}$$

From the previous explanation, then, the entire parameter of Bayes formula for text classification has been completed, including the value of prior probability, $Pr(C_i)$, likelihood, $Pr(W_j/C_i)$ and evidence, $Pr(W_j)$, thus producing the posterior probability, $Pr(C_i/W_j)$, in every word of the input document. The value of posterior probability for every word that is notated in a certain category is then entered to the classification table according to the specified category, as shown in Table 2.

Table 2. Category classification with Bayes formula

Probability of word (W_j)	Probability of Category C_1	Probability of Category C_2	Probability of Category C_k
W_1	$Pr(C_1 W_1)$	$Pr(C_2 W_1)$	$Pr(C_k W_1)$
W_2	$Pr(C_1 W_2)$	$Pr(C_2 W_2)$	$Pr(C_k W_2)$
.....
W_k	$Pr(C_1 W_k)$	$Pr(C_2 W_k)$	$Pr(C_k W_k)$
Total	$\sum Pr(C_1 W_k)$	$\sum Pr(C_2 W_k)$	$\sum Pr(C_k W_k)$
\sum Probability of Document	$\frac{\sum Pr(C_1 W_k)}{n}$	$\frac{\sum Pr(C_2 W_k)}{n}$	$\frac{\sum Pr(C_k W_k)}{n}$

After all columns of probability are completed, then C_i can be calculated by dividing the total columns of probability with the length of query n , using the equation:

$$Pr(\text{Document } X) = \frac{\sum Pr(C_i|w_1, w_2, w_3 \dots w_k)}{n} \tag{6}$$

$$n = \sum_{j=1}^k W_j$$

As an example, the value of probability of document X included in a certain category (C_i) can be calculated by finding the conditional probability of frequency of words in document X (W_j), in accordance with the word label contained in Category C_i , calculated with equation $Pr(C_i/W_j)$. The list of category C_i is arranged: $C_1, C_2, C_3, \dots C_k$, whereas the appearance of word W_j in document X is arranged: $W_1, W_2, W_3 \dots W_n$. The word appearance and probability table on Table 1 above has produced the probability value of a document in each category.

2.3 Support vector machine for sentiment analysis

Support vector machine (SVM) is a classification algorithm that uses the best method for finding a decision boundary to perform data mapping from the input space or parameter space into a feature space with higher dimensionality [2]. Several researches have proven that compared to other classification techniques, SVM is capable of presenting classification results with a decent level of accuracy [8]. SVM requires the transformation of raw text data into feature vector before it can be processed by the algorithm, and this process can be done using the Bag-of-Words model.

In classifying into two positive and negative classes in order to then do sentiment analysis in a document, the SVM algorithm is responsible to determine the optimal separating hyperplane and maximize the margin between the two different classes

[9]. SVM functions to solve that linear problem in order to maximized decision boundary. Figure 1 shows the calculation of optimal separating hyperplane in SVM.

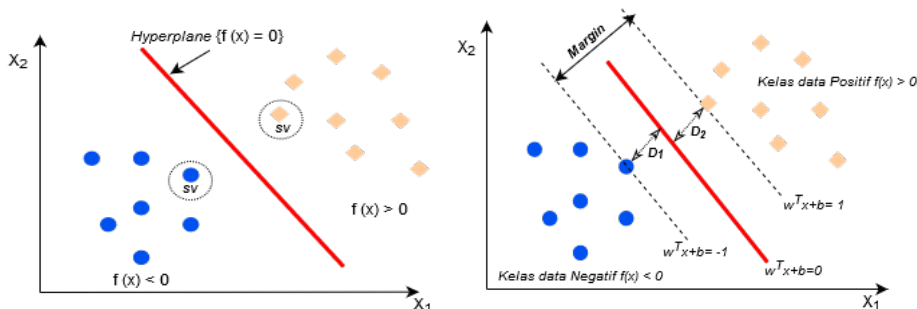


Figure 1. Optimal Separating Hyperplane

By using SVM as a linear classification to divide two different data classes, assuming that the dataset used has label x_i and feature y_i as class labels with $y_i \in \{-1, 1\}$, for $i = 1, 2, 3 \dots n$, where n is the number of data, with parameter w (weight) and b (bias), then hyperplane can be obtained with the equation:

$$f(x) = w^T x + b \tag{7}$$

To solve the linear problem of two classes, equation (7) can be written as such:

$$y = \text{sign}(w_1 x_1 + w_2 x_2 + b) \tag{8}$$

Note:

- w = weight
- x = vector input
- b = bias
- y = class label

if $y \geq 0$ then the class is 1 (positive), and if $y < 0$ the class is -1 (negative). To calculate distance D of a data point x from hyperplane, use equation:

$$D = \frac{|w^T x + b|}{|w|} \tag{9}$$

Thus, the margin of support vector is calculated using the equation:

$$\begin{aligned} \frac{w}{\|w\|} \cdot (x_+ - x_-) &= \frac{w^T (x_+ - x_-)}{\|w\|} \\ &= \frac{w^T \left(\left(\frac{+1-b}{w^T} \right) - \left(\frac{-1-b}{w^T} \right) \right)}{\|w\|} = \frac{2}{\|w\|} \end{aligned} \tag{10}$$

Fundamentally, the equation to find hyperplane can be used to solve the classification problem of two dimensionalities, such as sentimental analysis in a text document into two categories, positive or negative. On the other hand, a non-linear classification problem requires a more complex solution, which is to enter Kernel function in the mapping of feature vector into a higher dimensional space. The functions of Kernel trick commonly used in SVM includes *Polynomial*, *Gaussian radial basis function (Gaussian RBF)* and *Sigmoid*. SVM is considered computationally more efficient and quicker to process as it only uses a partial of the document presented as feature vector in order to do algorithmic learning [5].

2.4 Evaluation of classification performances

After data results from the classified method has reached, it is necessary to calculate the accuracy of those algorithm. One method to measure accuracy level is through cross tabulation or confusion matrix [5]. Confusion matrix is required to measure the performance and assess accuracy level of classification results. Confusion matrix is comprised of matrix row information and classification columns which respectively represent the actual data class and the class prediction obtained by algorithmic calculation. Table 3 shows an example of confusion matrix that is the classification result of 2 binary classes, class of 0 and class of 1.

Table 3. Confusion matrix of 2 classes

f_{ij}		Class Prediction (j)	
		Class = 0	Class = 1
Actual Class (i)	Class = 0	True Positive (TP) $\rightarrow f_{00}$	False Negative (FN) $\rightarrow f_{01}$
	Class = 1	False Positive (FP) $\rightarrow f_{10}$	True Negative (TN) $\rightarrow f_{11}$

In confusion matrix, every cell f_{ij} states the information of data number from class i , which is predicted into class j . As an example, a true positive (TP) is produced if the actual data class of 0 is mapped correctly in prediction class of 0, whereas a false negative (FN) is produced if the actual data class of 1 is mapped incorrectly in prediction class of 0. False negative (FN) is produced if the actual data class of 0 is mapped incorrectly in prediction class of 1, and a true positive (TP) is produced if the actual data class of 1 is mapped correctly in prediction class of 1. From the information resulted by confusion matrix, it is now possible to find the value of *Precision*, *Recall*, *Accuracy* and *F-Measure* to then determine the best classification accuracy, as shown below:

Precision is the level of accuracy and exactness for the estimated proportion of correct positive cases, formulated by the equation:

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

Recall is the level of accuracy and exactness for the estimated proportion of positive cases that are identified correctly, formulated by the equation:

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

Accuracy is the level of accuracy and exactness for the proportion of total correct predictions, formulated by the equation:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{13}$$

In some conditions it is possible to find a situation when there is a disproportion in the values of *precision*, *recall* dan *accuracy*. Therefore, the *F-Measure* (*F1 Score*) is required to find a balance using the equation:

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{14}$$

By using confusion matrix, it is possible to represent the predicted value and the actual condition of the data produced by machine learning algorithm.

3 Implementation

The proposed integrated information system consists of 3 modules: input, process and output. The input module consist of input data set used in this system, which is taken from the comments of Singapore Airlines passengers on the TripAdvisor website. The process module consist of text pre-processing and the algorithms model, then, the classification results shows in the output module. Figure 2 shows the proposed system framework.

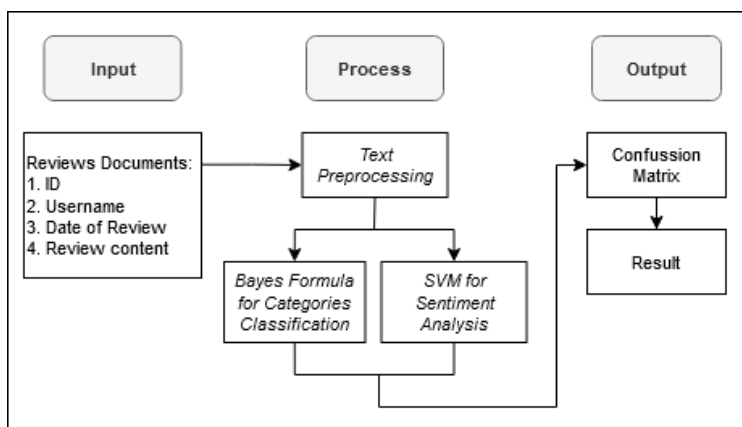


Figure 2. Integrated information system’s framework

The acquisition of data set can be done in real-time through the scrapping and filtering method, or by taking a collection of existing data set that has already been previously prepared. As an example, in this particular research, a data set of 1000 reviews written in English between December 2019 and March 2020 has been taken. Before being processed by the algorithm, the input data will pass through the text pre-processing module for the purpose of data cleaning, tokenization and text vectorization. Next, Bayes and SVM will process the clean data text simultaneously, yet still independently. For the Bayes model, the available input data will be classified into 5 category labels, which are: (1) *Airplane*; reviews of the condition of the airplane, (2) *Comfort*; reviews of comfort during the flight, (3) *Staff Services*; reviews of the flight attendants service, (4) *Food & Entertainment*; reviews of in-flight food and entertainment; dan (5) *Price*; reviews of ticket price and price in comparison with the service. For the Bayes algorithm to run, it has to receive training in the form of rules, which are the unique collection of words forming the desired category labels. Table 4 shows the example of rules contained in each category (C_i) for the Bayes algorithm classification process.

Table 4. List of rule and category label for Bayes algorithm

No	Category (C_i)	Words Label (W_j)
C1	Airplane	Plane, aeroplane, aircraft, airbus, boeing, new, old, Seat, space, legroom, room, chair
C2	Comfort	Reliable, stable, safe, safety, induction, Punctual, schedule, delay, cancel, late, resechedule, miss, quick, time, baggage, ground, lounge, terminal,
C3	Staff Services	Crew, staff, steward, stewardess
C4	Food & Entertainment	Food, beverage, drink, coffee, snack, meal, dinner, lunch, Entertainment, movie, tv, music, wifi, internet
C5	Price	Price, value, money, ticket, discount, promotion, pay, cost, cheap, expensive

Simultaneously, the SVM algorithm uses the vectorization result from the BoW model in the text pre-processing stage and will conduct a sentiment analysis to determine the meaning of each testing data document, if they are valued positive or negative.

4 Result and Discussion

4.1 Text pre-processing results

The first step undertaken by the processing system, once the input data of 1000 review documents ($D_1, D_2, D_3 \dots D_{1000}$) are collected, is data cleaning process by text pre-processing module. This is done by implementing a simple processes which include stop word removal, tokenization and stemming. Previously, input data set is given manual labels that encompass categories according to the predetermined rule category for Bayes processing, and sentiment label as a parameter prediction labels of positive and negative which refer to the Lexicon dictionary. This is done to shape the model of the training data for Bayes and SVM algorithms. Sample results of this text pre-processing method are showed in the following Table 5.

Table 5. Document text pre-processing result

Documents	Review Contents	Pre-processing	Predicted Labelling	Predicted Sentiment
D ₁	Flew on SQ A380 to LHR from SIN in October 2019. Had the privilege to sit on the newly refurbished 1st Class/Suite and it was absolutely fantastic! The suite itself with its seperate chair and bed was really nice. Service from the crew was superb and food was absolutely delicious.	[fly sq sin october had privilege sit new class suite fantastic separate chair bed real nice service crew superb food absolute delicious]	Airplane ; Comfort ; staff services ; F&E	Positive
D ₂	My husband and I could not be happier with the service, cleanliness, staff cheerfulness, food, and just about everything else on our flight in late February.	[husband happy service clean cheer food flight late february]	Comfort ; staff services ; F&E	Positive
....
D ₁₀₀₀	We paid extra to get good seating location. Food & beverage service slow and not much available.	[pay extra good seat locate food beverage service slow available]	Comfort ; F&E	Negative

The result data from this pre-processing step will then be processed by Bayes and SVM algorithm model. Because the final accuracy of classification result can be affected by the number of training document and the manual labelling process, it is therefore necessary to update the data label and rule continuously during the system evaluation stage.

4.2 Categories classification results

Furthermore, based on the available result of text pre-processing using the Bayes algorithm calculation in Table 2, the document probability is obtained as shown in Table 6:

Table 6. Bayes algorithm classification results

Input Doc	Prob. C1	Prob. C2	Prob. C3	Prob. C4	Prob. C5
D1	0.625	0.125	0.125	0.125	0
D2	0.25	0.5	0.25	0	0
D3	0.571428571	0.142857143	0	0.285714286	0
...
D1000	0.5	0	0.166666667	0.166666667	0.166666667
$\sum D_n$	0,063	0,023	0,036	0,075	0,005

Based on the category classification result of Bayes algorithm for 1000 tested documents in Table 6, the result is obtained from all of those tested document ($\sum D_n$), in which the highest category classification is C4 (Food & Entertainment) with a probability of 0.075 or a 37% percentage out of all testing documents. It is followed by category C1 (Airplane) with a probability of 0.063 or a 31% percentage in second place, and category C3 (staff services) with a probability of 0.036 or 18% in third. While the next category is C2 (comfort) with a probability of 0.023 or a 11% percentage, and lastly category C5 with a probability of 0.005 or a 3% percentage out of all documents. From this obtained data, then, the system displays the results information in Table 6 above as a bar chart, for the user’s ease of reading, as shown in Figure 4 below.

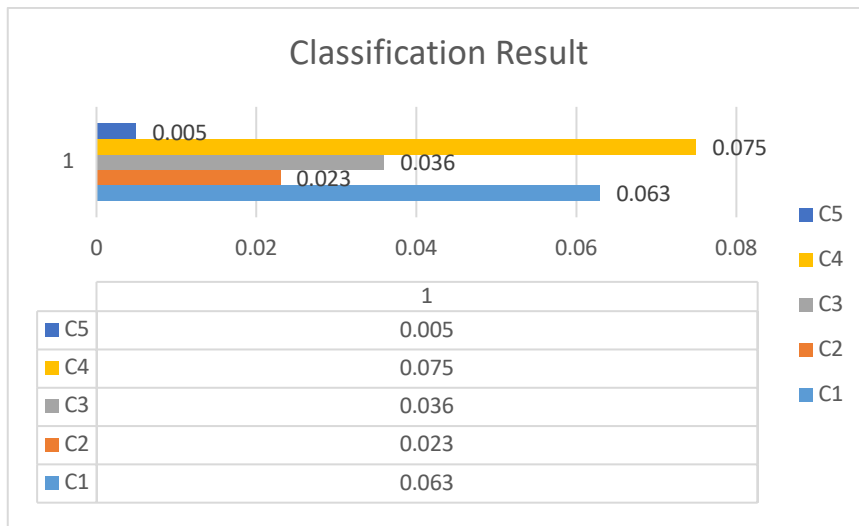


Figure 3. System output of Bayes algorithm classification results

This result shows that the Bayes algorithm module in the proposed system has successfully presented categories contained from the testing documents in accordance to the rules that have been set.

4.3 Sentiment analysis result

Simultaneously, using the available text pre-processing results data, the system parallelly processing the input data with the SVM algorithm to conduct sentiment analysis. Furthermore, confusion matrix will find out the accuracy level of the classification results as described below. The input document (*Dn*) of 1000 data yields 764 positive manually labelled data and 236 negative data labels. Then the training data to testing data is determined using ratios of 50:50, 60:40 and 70:30. Next, SVM algorithm uses the confusion matrix table to measure the classification accuracy level, producing the results shown in the following tables.

Table 7. Confusion Matrix for Data Ratio 50:50

Training - Testing Data Ratio	Matrix Confusion		Prediction	
			Positive	Negative
50:50	Actual	Positive	345	37
		Negative	45	73
	Precision		0,884615	0,663636
	Recall		0,903141	0,618644
	F-Measure		0,893782	0,640351
	Accuracy		0,836	

Table 8. Confusion Matrix for Data Ratio 60:40

Training - Testing Data Ratio	Matrix Confusion		Prediction	
			Positive	Negative
60:40	Actual	Positive	267	38
		Negative	31	64
	Precision		0,895973	0,627451
	Recall		0,87541	0,673684
	F-Measure		0,885572	0,649746
	Accuracy		0,8275	

Table 9. Confusion Matrix for Data Ratio 70:30

Training - Testing Data Ratio	Matrix Confusion		Prediction	
			Positive	Negative
70:30	Actual	Positive	207	26
		Negative	24	43
	Precision		0,896104	0,623188
	Recall		0,888412	0,641791
	F-Measure		0,892241	0,632353
	Accuracy		0,83333333	

These tables are shown in the output system with this result as follow: the 50:50 training data to testing data ratio yields a classification accuracy level of $0.836 \times 100\% = 83.6\%$, with the 60:40 ratio an accuracy level of $0.8275 \times 100\% = 82.75\%$, and the 70:30 ratio an accuracy level of $0.833 \times 100\% = 83.3\%$. This is an empirical proof that classification using the SVM method can give a high and stable level of accuracy with an average of 83%.

5 Conclusion

This research has proposed the implementation of 2 algorithm methods, Bayes formula and SVM, that are integrated with other modules such as input data module, text pre-processing module and classification results output in a integrated information system to analyse the review documents of Singapore Airlines passengers on the TripAdvisor website. The input data module can be used to take input data set, to then be processed with text pre-processing module using the implementation of BoW model for vectorization. The Bayes algorithm module is used to produce the classification result of passenger reviews according to the category of review content, as divided into 5 categories. The percentage out of all 1000 documents of data input, based on the calculated probability for each category, is made up as such: 31% of the total document is reviews on airplane condition, 11% on comfort, 18% on staff services, 37% on food & entertainment, and 3% on price. The conclusion of this result shows that most passenger reviews are concerning the in-flight food and entertainment.

Simultaneously, the SVM algorithm module implementation for sentiment analysis has produced a positive and negative label classification with a stable accuracy level, in which the 50:50 training data to testing data ratio has an accuracy of 83.6%, the 60:40 data ratio has an accuracy of 82.75% and the 70:30 data ratio has an accuracy of 83.3%. The proposed integrated information system for analysing review documents in this particular research can be used to process any other text-based reviews data, and can be developed in the next research in order to obtain an even higher level of accuracy, which can be done by conducting a more intensive training on the system, for instance by updating the data rule used by Bayes classification and refining the manual labelling process as training data for SVM algorithm.

Acknowledgment

This research was successfully conducted with the exceptional support and assistances from various parties. The first author would like to thank the Ministry of Research, Technology and Higher Education of Republic of Indonesia for provided the funding assistance of this research under PasTi Scholarship Program. And also to thank Professor Mustafid for provide insight and guidance that greatly assisted the research, and Dinar Mutiara K.N. Ph.D, for the comments that greatly improved the manuscript. We would also like to show our gratitude to colleagues who gave their supports and contributions.

References

1. A. Kao, S. Poteet, *Natural Language Processing and Text Mining* (Springer, Washington, 2006)
2. N. Ur-Rahman, J.A. Harding, *Textual data mining for industrial knowledge management and text classification: A business oriented approach*, *Expert Systems with Applications* **39**, 5, 4729–4739 (2012)
3. G. Miner, A. Fast, D. Delen, T. Hill, J. Elder, B. Nisbet, *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Application* (Elsevier, Oxford, 2012)

4. Y. Zhang, R. Jin, Z.H. Zhou, Understanding bag-of-words model: A statistical framework, *International Journal of Machine Learning and Cybernetics*, **1**, 1–4, 43–52 (2010)
5. J. Han, M. Kamber, *Data Mining: Concepts and Techniques Second Edition* (Morgan Kaufmann, New York, 2006)
6. S. Taheri, M. Mammadov, Learning The Naïve Bayes Classifier With Optimization Models, *Internasional Journal of Applied Mathematics and Computer Science*, 787–795 (2013)
7. R. Feldman, J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data* (Cambridge University Press, New York, 2007)
8. D. Isa, L.H. Lee, V.P. Kallimani, R. RajKumar, Text document Preprocessing with the Bayes Formula for Classification Using the Support vector machine, *IEEE Transactions on Knowledge and Data Engineering* **20**, 9, 1264-1272 (2008)
9. V. Vapnik, *The Nature of Statistical Learning Theory* (Springer, New York, 1995)