Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11564
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Supervised Model for Sentiment Analysis Based on Hotel Review Clusters using RapidMiner

Revin Novian Juliadi[1)*], Yan Puspitarani[2)]
[1)2)] Widyatama University, Bandung, Indonesia
[1)]revin.novian@widyatama.ac.id, [2)]yan.puspitarani@widyatama.ac.id

Submitted : July 15, 2022 | **Accepted** : Aug 1, 2022 | **Published** : Aug 1, 2022

Abstract: Customer feedback in the modern era like today is mostly presented in the form of digital reviews, including customer feedback at an inn or hotel, customer feedback is very valuable data where from this data the management can find out, identify and analyze the customer experience and what they need. With customer feedback in the form of digital reviews, it will allow a lot of data that can be obtained by hotel management and will provide many benefits if the data is processed correctly. To take advantage of large text review data, a combination of data mining and natural language processing techniques was chosen to process text in depth and efficiently. Text mining in the form of creating an opinion mining model using the Naïve Bayes classification algorithm is applied to find information and measure the main sentiments expressed in the reviewed text dataset, then the application of K-Means text grouping aims to group texts and get information about the main topics discussed from the content of the review dataset text in each group. By applying the constructed sentiment analysis model, approximately 90.90% accuracy results were obtained in reading texts and measuring sentiments related to hotel customer feedback data.

Keywords: K-means, Naïve Bayes, opinion mining, RapidMiner, sentiment analysis, text clustering, text mining

INTRODUCTION

Knowing and analyzing customer needs is a challenge in quality management. An organization can use different forms of means to identify and analyze those needs but the most commonly used forms are complaint analysis, attributive satisfaction measurement, and statistical analysis (Križo, Madzík, Vilgová, & Sirotiaková, 2018).

The modern era of digitalization or digital transformation has been intensively applied in masse, digitalization has begun to be applied to many things, including one of which is applied to customer feedback in various sectors and fields in human life, including the lodging and hotel sector. Digitalization also provides advantages by making everything more agile, flexible, and develop more easily, besides that digitalization also allows organizations to be able to make effective and efficient decisions to create a better customer experience because of the ease of interacting with customers ("What is Digital Transformation - Salesforce," n.d.). In other words, digitalization is also implemented to increase transparency and customer satisfaction (Mergel, Edelmann, & Haug, 2019).

The implementation of digitalization in customer feedback, allows organizations to collect review data on a large scale in the form of unstructured text. With a large data scale, manually checking to find the main idea of the review data is not possible. Therefore, natural language processing techniques in the form of sentiment analysis or opinion mining were chosen as a process to extract information from data in the form of text commonly called text mining, and gain an understanding of attitudes, opinions, and emotions expressed in data text reviews from customers regarding products and services provided by the organization (Redhu, Srivastava, Bansal, & Gupta, 2018). In the hotel sector, hotel management can take advantage of large-scale review data from hotel customers to maintain and improve the service facilities provided by the hotel specifically, in addition to other benefits, namely that it will make it easier for management to make decisions.

* Author Corresponding

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11564
e-ISSN : 2541-2019
p-ISSN : 2541-044X

LITERATURE REVIEW

Several previous literature studies have discussed sentiment analysis carried out on the hotel reviews dataset, with the results obtained from the experiment being in the form of predictive accuracy from the model used, for example in the literature (Shi & Li, 2011) the results of experiments carried out supervised using the SVM (support vector machine) algorithm as a classification algorithm and the TF-IDF (term frequency-inverse document frequency) cross-validation method applied to provide precision accuracy results of around 86%. In another literature example (Zvarevashe & Olugbara, 2018), an experiment was conducted on a sentiment analysis using the Naïve Bayes multinominal (NBM), Sequential minimal optimization (SMO), compliment Naïve Bayes (CNB), and composite hypercubes on iterated random projection (CHIRP) classification algorithms to conduct training and testing on the hotel review dataset whose results were compared with results of around 75% - 81% accuracy. Also, the literature (Pharisee, Sibaroni, & Faraby, 2019) discussing the same experiment uses a multinominal Naïve Bayes classifier algorithm where the experiment is done into three scenarios by comparing the performance gains of the results of preprocessing data with an accuracy of about 91%. Another literature study (Ghorpade & Ragha, 2012) conducted a sentiment classification experiment on the hotel reviews dataset using a Naïve Bayes algorithm that focused on extracting information with text analysis that had been carried out to reduce the loss of text information and the results obtained 96% accuracy achieved. It can be known from some literature that opinion mining can be accurately applied to hotel review datasets using various techniques and types of algorithms, Naïve Bayes is also proven to have a fairly high level of prediction accuracy.

From some of the literature mentioned above, the majority discusses the percentage of precision and accuracy obtained from the model applied to the dataset, but there is no deeper discussion of text mining so that the main idea or main topic discussed in the review dataset or vice versa, as in literature (Zhang & Yu, 2017) that experiments on sentiment analysis but is more focused on the discussion of word vector clustering. And in the literature (Hu, Chen, & Chou, 2017) discusses studies and experiments on text summarization techniques in sentences in the hotel review dataset to find the most informative sentences by applying the k-medoid text clustering technique.

Therefore, the studies and experiments carried out in this journal and what distinguishes it from other literature, namely in this journal will discuss further text mining, namely the construction of supervised models for opinion mining and text clustering so that the results of sentiment analysis predictions and the main topics that have been carried out by clusters by models applied to datasets and all experiments are carried out visual programming using applications.

METHOD

The entire method in this study was carried out starting with planning by making a workflow of work steps to be carried out, then experiments were carried out on the RapidMiner as the main application that processes the training model with pre-processing on text mining and the application of the model for sentiment analysis, after that for the final result, the Tableau application is used as the process of making a graph dashboard.

In detail, the experiment process carried out is divided into several focus processes, starting from the collection of research and datasets regarding hotel reviews that will be used as training and testing datasets, after which the model is built based on training carried out using training datasets commonly known as supervised learning, training models are carried out using the Naïve Bayes algorithm which was chosen because it has a level of classification accuracy for predictions that precision. In addition, the advantages of the Naïve Bayes algorithm are that it has an efficient computational training time, can deal with noise in the data, and operates from estimates with a low-order probability in the training data so that it is easy to update (Webb, 2016). In short, the algorithm is a simple but fast, accurate, and reliable algorithm to be used for many purposes, which is especially very well used in natural language processing (NLP) problems (Chauhan, 2022).

After the model is formed on the dataset, the text clustering process is carried out using the k-means algorithm which aims to group words on the dataset that have similarities, k-means will group unsupervised learning by selecting many clusters or commonly referred to as the K variable, the k-means algorithm works iteratively to assign each data point to one of the K groups based on the features provided. Data points are grouped based on the similarity of features (Trevino, 2016), The centroid of cluster K can be used to label new data, the grouping makes it possible to find and analyze the groups that have been formed. In this study, K, or the number of groups was determined into three groups or clusters selected from the case orientation in the sentences in the hotel review dataset, namely orientation regarding accommodation facilitation, services, and locations regarding reviewed hotels.

The dataset used in this study is a hotel review dataset obtained from the Kaggle website, the dataset is used for experiments such as training on sentiment analysis models and other hotel review datasets are used as datasets for testing. The dataset used for the training model is a combination of datasets shared by Datafiniti named "datafiniti hotel reviews" (Datafiniti, 2019) sourced from the datafiniti business database and a dataset by

* Author Corresponding

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11564
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Jiashen Liu named "515k hotel reviews data in Europe" (Jiashen Liu, 2017) sourced from scrapping results www.booking.com. The data for training consists of 11,500 positive review data and 11,500 review data that are negative which are labeled sentiment so that the model can read and learn each word or sentence and its sentiment.

The two test datasets are also made from several combinations of review collections separated from each file having a fully positive and negative sentiment that aims to test and see if the model built is successful and can predict correctly. For the application of the sentiment analysis model, there are 3 datasets used, namely the hotel reviews dataset Named "trip advisor hotel reviews" which contains 20,491 data sourced from www.tripadvisor.com used also in the literature (Nature, Ryu, & Lee, 2016), then another dataset shared by Harmanpreet Singh Named "labeled hotel reviews" (Harmanpreet Singh, 2017), the last test dataset used was the "515k hotel reviews data in Europe" dataset which had been transformed so that it had 1,031,476 data and excluded data that had been used for training.

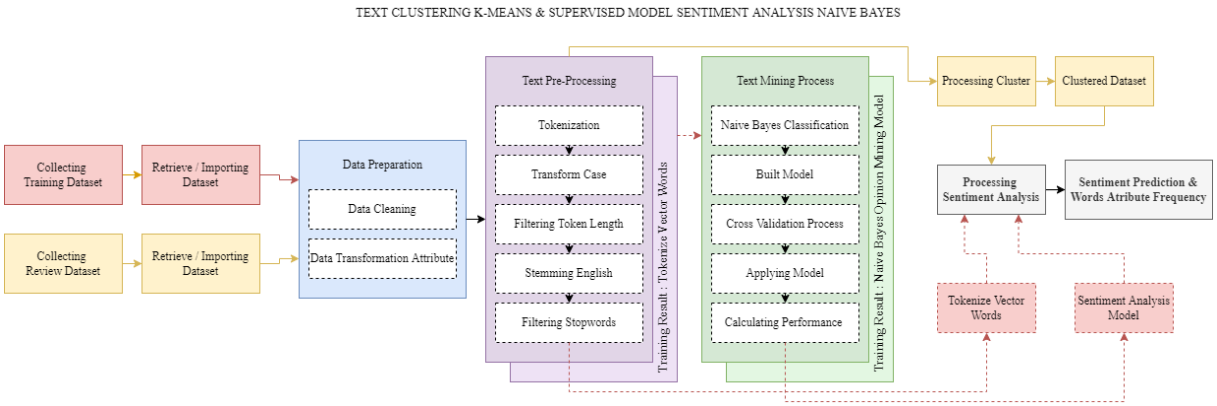

Fig. 1. Text Clustering & Sentiment Analysis Terminologies

The training model is a process of sentiment analysis model built on a text-shaped dataset that has previously been pre-processed, the model is trained using a Naïve Bayes algorithm which aims to find a sentence or word patterns in the dataset text with the sentiment contained in it based on the training dataset used. The training dataset begins with text pre-processing with the TF-IDF approach so that it can measure statistics, frequency of appearance, and evaluation of whether each word formed is relevant to the word in the document or training dataset.

The dataset that has been pre-processed then enters the main process, namely the text mining process using a cross-validation operator which is set 20 times using the Naïve Bayes classification algorithm so that it can find out the pattern of word frequency and classify the sentiment contained in it. After the classification is completed, cross-validation will also test the performance of the prediction accuracy of the model that has been formed from the results of text mining training. The results of text mining will form a model that can be used to read and predict sentiment analysis on other hotel review datasets.

The text clustering process is carried out on the hotel review dataset to form a pattern of grouping words so that the words that are most often mentioned as the main topic in each cluster group are discussed by the entire data in the dataset. Text clustering is done using the k-means cluster algorithm chosen because it is a reasonably simple algorithm and has a process with good results. Before clustering each dataset, the data preparation and text pre-processing process is carried out with the same process in the previous training model. The data is processed text clustering using the k-means algorithm operator with a total of three K or three clusters, three clusters are taken from the focus orientation of the hotel review namely location orientation, accommodation facilities, and services. after the clustering is completed, it will provide the final result in the form of a dataset that is already clustered. The sentiment analysis or opinion mining process is the last stage by combining all the results of the training model and text clustering process, the process begins with applying the sentiment analysis Naïve Bayes model that has been built and tokenizes vector word from the results of text pre-processing in the hotel review dataset that has been carried out text clustering. The stage of applying the sentiment analysis model to the RapidMiner begins with placing the retrieve model operator who has previously been trained using the Naïve Bayes algorithm and retrieving tokenized vector word as a result of the pre-processing text in the training data, then the three hotel review datasets are first carried out text pre-processing with the same pre-processing specifications as was done during training and also text clustering, after which the dataset begins to be connected to the application model operator for the start of the sentiment analysis process using a pre-built model.

* Author Corresponding

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11564
e-ISSN : 2541-2019
p-ISSN : 2541-044X

RESULT

The results of the experiment process at the sentiment analysis model training stage using the Naïve Bayes classification algorithm with 23,000 training data provide 90.90% accuracy results, this accuracy is the result of accuracy obtained from the results of cross-validation with the TF-IDF approach which is folded 20 times, details about the accuracy performance and precision of the built model can be seen in Table 1. Then from the results of the pre-processing text in the training process, a tokenized vector word was also obtained with a total of 371 variable attributes that will be used as word references for text mining.

Table 1. Naïve Bayes Model Performance

	True Negative	True Positive	Class Precision
Pred. Negative	10576	1170	90.04%
Pred. Positive	924	10330	91.79%
Class Recall	91.97%	89.83%	

After the model is completed, to ensure that the model can function and be used properly, sentiment analysis testing is carried out on the testing dataset, namely two datasets consisting of positive and negative hotel review datasets by providing prediction results with the intended goal of reading and finding different sentiments in the dataset, the results can be seen in Fig. 2.

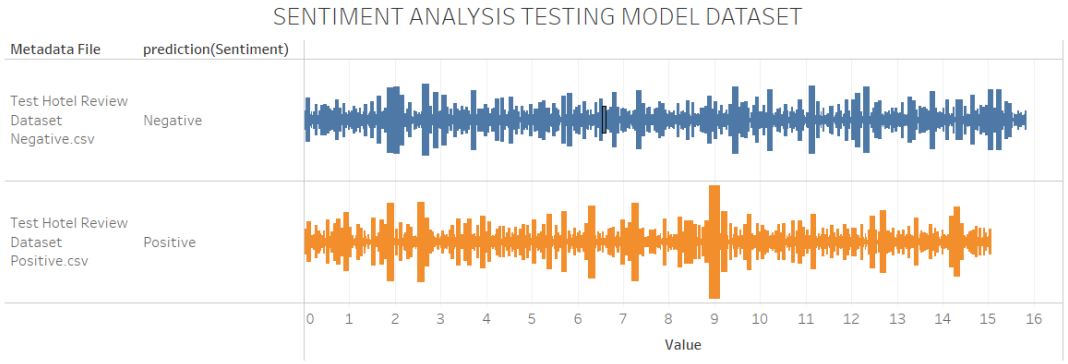

Fig. 2. Sentiment Analysis Testing Dataset Result

Furthermore, it carried out a text clustering process on all three hotel review datasets with the aim of grouping word variables and finding the frequency of mentions so that the main topic of each cluster was obtained. Clustering is carried out using the K-means algorithm in each dataset that has been pre-processed. Because there are quite a lot of word variables, which are 371, therefore the cluster results are limited to only the most word variables displayed. The following is the application of the text mining process in the form of text clustering and opinion mining in the three-dataset hotel review can be seen in Fig. 3. to Fig. 5.

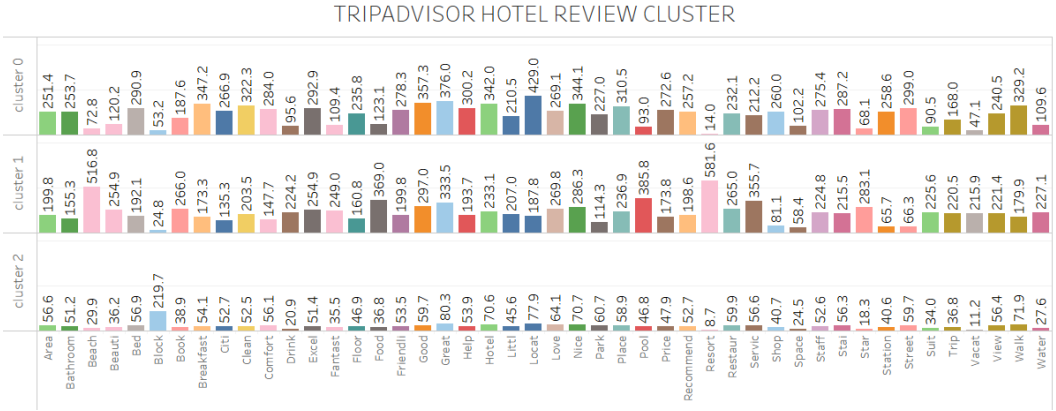

Fig. 3. TripAdvisor Review Cluster Graph

* Author Corresponding

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11564
e-ISSN : 2541-2019
p-ISSN : 2541-044X

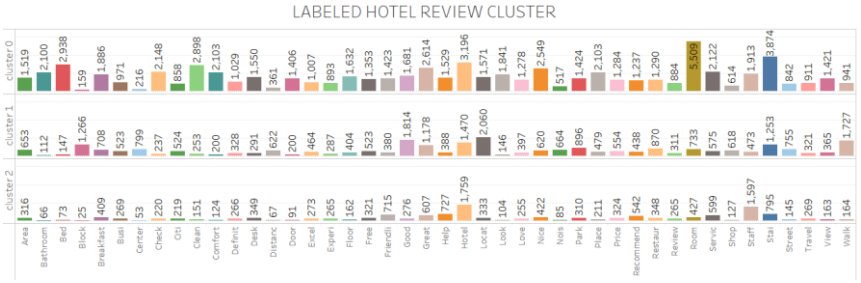

Fig. 4. Labeled Hotel Review Cluster Graph

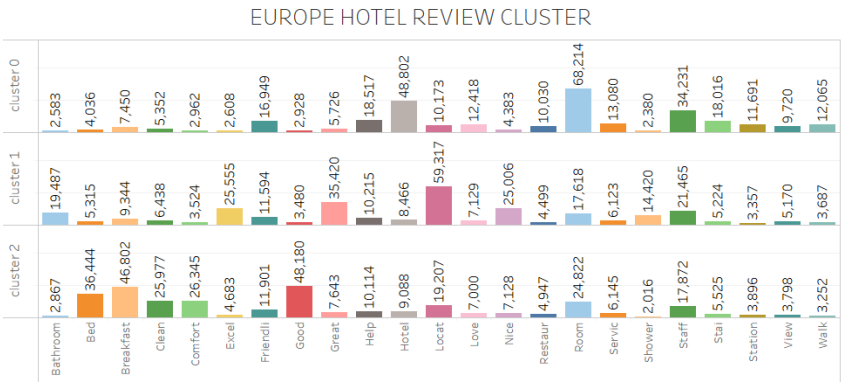

Fig. 5. Hotel in Europe Review Cluster Graph

In the three clusters of each dataset that has been displayed in the form of a graph using the Tableau application the average of the most frequently mentioned words or commonly referred to as word frequency and the main topic in each group is obtained, from this data the orientation of the customer or reviewer can be known by looking at the majority of the most mentioned word properties oriented to which group in each cluster. since all that is needed is the words that are most often mentioned, only those words that have a high appearance are displayed on the chart. More details about the averages of the most frequently mentioned words in each hotel review dataset can be found also in Table 2 to Table 4.

Table 2. Tripadvisor Review Average Impression Word Cluster

TripAdvisor Average Mentioned Word Impression Variables					
Cluster 0		**Cluster 1**		**Cluster 2**	
Location	429.0	Resort	581.6	Block	219.7
Great	376.0	Beach	516.8	Space	164.7
Good	357.3	Pool	385.8	Great	80.3
Breakfast	347.2	Food	369.0	Walk	71.9
Nice	344.1	Service	355.7	Nice	70.7
Hotel	342.0	Great	333.5	Hotel	70.6
Walk	329.2	Good	297.0	Love	64.1
Clean	322.3	Nice	286.3	Park	60.7
Place	310.5	Star	283.1	Restaurant	59.9
Help	300.2	Love	269.8	Street	59.7
Street	299.0	Book	266.0	Comfort	56.1
Excellent	292.0	Restaurant	265.0	Breakfast	54.1
Bed	290.9	Beauty	254.9	Help	53.9
Comfort	284.0	Excellent	254.9	Friendly	53.5
Friendly	278.3	Fantastic	249.0	Recommend	52.7
Staff	275.4	Place	236.9	City	52.7
Price	272.6	Hotel	233.1	Staff	52.6
Love	269.1	Water	227.1	Clean	52.5
City	266.9	Suite	225.6	Excellent	51.4
Shop	260.0	Staff	224.8	Bathroom	51.2
Station	258.6	Drink	224.2	Price	47.9
Recommend	257.2	View	221.4	Floor	46.9

* Author Corresponding

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11564
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Bathroom	253.7	Trip	220.5	Pool	46.8
Area	251.4	Vacation	215.9	Little	45.6
View	240.5	Stay	215.5		44.5
Orientation Cluster :		**Orientation Cluster :**		**Orientation Cluster :**	
Location		**Services**		**Accommodation Facilities**	

Table 3. Labeled Hotel Review Average Impression Word Cluster

Labeled Hotel Average Mentioned Word Impression Variables					
Cluster 0		**Cluster 1**		**Cluster 2**	
Room	5.509	Locate	2.060	Hotel	1.759
Stay	3.874	Good	1.814	Staff	1.597
Hotel	3.196	Walk	1.727	Stay	795
Bed	2.938	Hotel	1.470	Help	727
Clean	2.898	Block	1.266	Friendly	715
Great	2.614	Stay	1.253	Great	607
Nice	2.549	Great	1.178	Service	599
Check	2.148	Park	896	Recommend	542
Service	2.122	Restaurant	870	Room	427
Place	2.103	Centre	812	Nice	422
Comfort	2.103	Street	799	Breakfast	409
Bathroom	2.100	Room	733	Restaurant	348
Staff	1.913	Breakfast	708	Location	333
Breakfast	1.886	Noise	664	Price	324
Looks	1.841	Area	653	Free	321
Good	1.681	Distance	622	Area	316
Floor	1.632	Nice	620	Park	310
Locate	1.571	Shop	618	Good	276
Desk	1.550	Service	575	Excellent	273
Help	1.529	Price	554	Busy	269
Area	1.519	City	524	Travel	269
Park	1.424	Free	523	Definite	266
Friendly	1.423	Busy	523	Experience	265
View	1.421	Place	479	Review	265
Door	1.406	Staff	473	Love	255
Orientation Cluster :		**Orientation Cluster :**		**Orientation Cluster :**	
Accommodation Facilities		**Location**		**Services**	

Table 4. Hotel Review in Europe Average Impression Word Cluster

Hotel in Europe Average Mentioned Word Impression Variables		
Cluster 0	**Cluster 1**	**Cluster 2**

* Author Corresponding

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 3, July 2022
DOI : https://doi.org/10.33395/sinkron.v7i3.11564
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Room	68.214	Location	59.317	Good	48.180
Hotel	48.802	Great	35.420	Breakfast	46.802
Staff	34.231	Excellent	25.555	Bed	36.444
Help	18.517	Nice	25.006	Comfort	26.345
Stay	18.016	Staff	21.465	Clean	25.977
Friendly	16.949	Bathroom	19.487	Room	24.822
Service	13.080	Room	17.618	Location	19.207
Love	12.418	Shower	14.420	Staff	17.872
Walk	12.065	Friendly	11.594	Friendly	11.901
Station	11.691	Help	10.215	Help	10.114
Location	10.173	Breakfast	9.344	Hotel	9.088
Restaurant	10.030	Hotel	8.466	Great	7.643
View	9.720	Love	7.129	Nice	7.128
Breakfast	7.450	Clean	6.438	Love	7.000
Great	5.726	Service	6.123	Service	6.145
Clean	5.352	Bed	5.315	Stay	5.525
Nice	4.383	Stay	5.224	Restaurant	4.947
Bed	4.036	View	5.170	Excellent	4.683
Comfort	2.962	Restaurant	4.942	Station	3.896
Good	2.928	Walk	4.499	View	3.798
Excellent	2.608	Comfort	3.524	Walk	3.252
Bathroom	2.583	Good	3.480	Bathroom	2.867
Shower	2.380	Station	3.357	Shower	2.016
Orientation Cluster : Accommodation Facilities		**Orientation Cluster : Location**		**Orientation Cluster : Services**	

After searching for the frequency of the most frequently appearing words and their orientation, sentiment checks can then be carried out on each cluster using Naïve Bayes' sentiment analysis model. With the results obtained, the three clusters formed in each hotel review dataset have an overall positive sentiment. The results can be seen in Fig. 6.

Row No.	metadata_file	label	prediction(Sentiment)	confidence(Negative)	confidence(Positive)
1	TripAdvisor Cluster 0.csv	Trip Advisor Sentiment	Positive	0.000	1.000
2	TripAdvisor Cluster 1.csv	Trip Advisor Sentiment	Positive	0.000	1.000
3	TripAdvisor Cluster 2.csv	Trip Advisor Sentiment	Positive	0.000	1.000

Row No.	metadata_file	label	prediction(Sentiment)	confidence(Negative)	confidence(Positive)
1	Labeled HR_Cluster 0.csv	Labeled HR Sentiment	Positive	0	1
2	Labeled HR_Cluster 1.csv	Labeled HR Sentiment	Positive	0	1
3	Labeled HR_Cluster 2.csv	Labeled HR Sentiment	Positive	0	1

Row No.	metadata_file	label	prediction(Sentiment)	confidence(Negative)	confidence(Positive)
1	Hotel in Europe Review_Cluster 0.csv	Hotel Europe Sentiment	Positive	0	1
2	Hotel in Europe Review_Cluster 1.csv	Hotel Europe Sentiment	Positive	0	1
3	Hotel in Europe Review_Cluster 2.csv	Hotel Europe Sentiment	Positive	0	1

Fig. 6. Sentiment Analysis Result Clustered Dataset

DISCUSSIONS

The overall results of the experiments that have been carried out, provide a clearer picture of the sentiment contained and also know the main topics discussed in the three hotel review datasets discussed. From these results, it can be seen that each cluster in each dataset has a word frequency that is often mentioned there, also the main orientation of the discussion can be taken. For example, the TripAdvisor dataset talks about the main topics of location in the first cluster, services in the second cluster, and accommodation facilities in the third cluster with the entire cluster having a positive sentiment. Likewise, the Labeled Hotel Review and Hotel in

* Author Corresponding

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Europe dataset talk about the main topics regarding accommodation facilities in the first cluster, locations in the second cluster, and services in the third cluster with the entire cluster having positive sentiments.

CONCLUSION

The application of opinion mining by building a model with the Naïve Bayes classification algorithm and K-means text clustering in the hotel review dataset will provide the advantage that not only an accurate sentiment reading of 90.90% but information on the main topic is also obtained, moreover, the frequency of words grouped into clusters will help to provide more explanation of the content of the conversation discussed in the dataset. In addition, the training dataset is an important thing that needs to be prepared properly because the supervised model and tokenized vector word that is formed will always adjust on the training dataset used. For further studies, it is possible to compare the performance obtained from the combination of classification algorithms and other clustering or even improvements in the model that can be done by processing experiments on RapidMiner with the addition of other operators.

REFERENCES

Alam, Md. H., Ryu, W.-J., & Lee, S. (2016). Joint multi-grain topic sentiment: modeling semantic aspects for online reviews. *Information Sciences*, *339*, 206–223. doi:10.1016/j.ins.2016.01.013

Chauhan, N. S. (2022, April 8). Naïve Bayes Algorithm: Everything You Need to Know - KDnuggets. Retrieved July 6, 2022, from https://www.kdnuggets.com/2020/06/naive-bayes-algorithm-everything.html

Datafiniti. (2019). Hotel Reviews | Kaggle. Retrieved July 7, 2022, from https://www.kaggle.com/datasets/datafiniti/hotel-reviews

Farisi, A. A., Sibaroni, Y., & Faraby, S. al. (2019). Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier. *Journal of Physics: Conference Series*, *1192*, 012024. doi:10.1088/1742-6596/1192/1/012024

Ghorpade, T., & Ragha, L. (2012). *Featured-based sentiment classification for hotel reviews using NLP and Bayesian classification*. In *2012 International Conference on Communication, Information & Computing Technology (ICCICT)* (pp. 1–5). IEEE. doi:10.1109/ICCICT.2012.6398136

Harmanpreet Singh. (2017). Hotel Reviews | Kaggle. Retrieved July 7, 2022, from https://www.kaggle.com/datasets/harmanpreet93/hotelreviews

Hu, Y.-H., Chen, Y.-L., & Chou, H.-L. (2017). Opinion mining from online hotel reviews – A text summarization approach. *Information Processing & Management*, *53*(2), 436–449. doi:10.1016/j.ipm.2016.12.002

Jiashen Liu. (2017). 515K Hotel Reviews Data in Europe | Kaggle. Retrieved July 7, 2022, from https://www.kaggle.com/datasets/jiashenliu/515k-hotel-reviews-data-in-europe/metadata

Križo, P., Madzík, P., Vilgová, Z., & Sirotiaková, M. (2018). Evaluation of the Most Frequented Forms of Customer Feedback Acquisition and Analysis (pp. 562–573). doi:10.1007/978-3-319-95204-8_47

Mergel, I., Edelmann, N., & Haug, N. (2019). Defining digital transformation: Results from expert interviews. *Government Information Quarterly*, *36*(4). doi:10.1016/J.GIQ.2019.06.002

Redhu, S., Srivastava, S., Bansal, B., & Gupta, G. (2018). Sentiment Analysis Using Text Mining: A Review. *International Journal on Data Science and Technology*, *4*(2), 49. doi:10.11648/j.ijdst.20180402.12

Shi, H.-X., & Li, X.-J. (2011). *A sentiment analysis model for hotel reviews based on supervised learning*. In *2011 International Conference on Machine Learning and Cybernetics* (pp. 950–954). IEEE. doi:10.1109/ICMLC.2011.6016866

Trevino, A. (2016, December 6). Introduction to K-means Clustering. Retrieved July 6, 2022, from https://blogs.oracle.com/ai-and-datascience/post/introduction-to-k-means-clustering

Webb, G. I. (2016). Naïve Bayes. In *Encyclopedia of Machine Learning and Data Mining* (pp. 1–2). Boston, MA: Springer US. doi:10.1007/978-1-4899-7502-7_581-1

What is Digital Transformation - Salesforce. (n.d.). Retrieved July 4, 2022, from https://www.salesforce.com/ap/products/platform/what-is-digital-transformation/

Zhang, X., & Yu, Q. (2017). *Hotel reviews sentiment analysis based on word vector clustering*. In *2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA)* (pp. 260–264). IEEE. doi:10.1109/CIAPP.2017.8167219

Zvarevashe, K., & Olugbara, O. O. (2018). *A framework for sentiment analysis with opinion mining of hotel reviews*. In *2018 Conference on Information Communications Technology and Society (ICTAS)* (pp. 1–4). IEEE. doi:10.1109/ICTAS.2018.8368746

* Author Corresponding