❒ 1536

SentiHotel: a sentiment analysis application of hotel services using an optimized neural network

Dyah Apriliani[1], Taufiq Abidin[2], Edhy Sutanta[3], Amir Hamzah[4], Oman Somantri[5]
[1,2]Department of Informatics Engineering, Politeknik Harapan Bersama, Indonesia
[3,4]Department of Informatics Engineering, Institut Sains & Teknologi AKPRIND Yogyakarta, Indonesia
[5]Department of Informatics Engineering, Politeknik Negeri Cilacap, Indonesia

Article Info

ABSTRACT

An assessed hotel service is necessary for tourists and everyone who is traveling, however currently it is still difficult to find recommended hotel information. The solution provided in this research is to propose a smart application that has been developed by implementing machine learning in it. The purpose is to build a sentiment review smart application by applying the sentiment analysis hybrid model of the best neural network (NN) algorithm model that has been optimized using genetic algorithms. To get the right model, the research method was carried out with experiments starting from the initial stages of conducting data preprocessing, tokenization, weighting, modeling experiments, and conducting the system evaluation stage to determine the success of the proposed model. The progress of the application development system is by using the prototyping model. SentiHotel is a sentiment application that was successfully built to provide a solution for tourists in assessing a hotel service. The software validation test is carried out using the blackbox method and the results show that the SentHotel application is in accordance with the expected result; all system functions can run properly.

Corresponding Author:

Oman Somantri
Department of Informatics Engineering
Politeknik Negeri Cilacap
Jln. Dr. Soetomo No.01 Sidakaya Cilacap, Indonesia
Email: oman_mantri@yahoo.com

1. INTRODUCTION

The service level quality of a company engaged in the service sector is part of the criteria for evaluating a service user. Services that are beyond expectation will provide an increased level of trust in these service providers and will be worth the price, the more satisfying the service, the higher the service fees that must be paid by users. The hotel and lodging service sector currently provides enormous opportunities for these service providers, especially in the tourism sector. Assessment of the level of a hotel service satisfaction can come from various sources, one of which is a review comments of tourists or travelers who have used the lodging services [1].

The development of artificial intelligence (AI), especially machine learning, makes everything easier to apply. In accordance with its advantages, one part of ML is a supervised learning algorithm, which can be used not only for control and monitoring systems [2], but also in data science where it can be implemented in a data classification that applies in various fields, such as health [3], livestock and agriculture [4], [5], economy and industry [6], transportation [7], education, health [8] and others [9].

The rise of social media which currently attract more and more user, provides a huge potential, especially for many companies. They no longer need to expose their products in a paid service advertisement. This benefit of social media also provides an advantage for data seekers, especially data that comes from comments from social media users which are used as a decision support. Data derived from these comments is used as a dataset to be processed using text mining technology. This technology allow data in the form of written text in a review and comments to become data that can be classified or predicted [10]. The use of sentiment analysis that comes from social media has been used by several previous researchers [11], [12]. It was used for the sentiment assessment of the relocation of the Indonesia capital city [13], as well as the use of real time social media bigdata for sentiment [14]. The sentiment analysis is also used to assess film reviews [15]. Another study is the application of sentiment by optimizing the accuracy of the support vector machine model to overcome hyper-parameters [16].

Building a trust from the service sector business is currently a top priority for these company owners. The consumers trust that are assessed from a service has a very significant effect on business continuity in this service sector. The lodging service business of a hotel is currently growing rapidly and is unique because it is almost said that the level of service determines the success or failure of the hotel. The more people feel that they are getting services that are worth the money paid, the more likely that person will have an impact in the future, since it is almost certain that the experiences they have will be shared with those around them either directly or through social media. Marketing through a review comment written by hotel customers will indirectly affect the failure of the hotel accommodation service business especially negative review, because prospective customer will see the review results from comments of hotel users; if the results of the review are positive sentiments, then it will affect their decision to stay at the hotel, and vice versa.

Research on hotel and restaurant sentiment has actually been carried out by several previous researchers using various models and algorithms [17]-[20], however, almost all of these studies use English as input data as well as the data sets. Previous research that has been carried out in the search for a hotel service sentiment review model using the decision tree model produces the best accuracy rate of 88.54% [21], but it is still not optimal. The next research was carried out by applying a neural network model that had optimized parameter values and resulted in a classification accuracy rate of 88.99% [22]. Based on a model with a known level of accuracy, this study develops and builds a sentiment analysis-based application to classify hotel service ratings based on customer reviews using Indonesian text, slightly different from [23]. The basic contribution of this research is that it offers an intelligent sentiment analysis application SentiHotel which is developed by integrating a neural network model that has been optimized for parameter values using a genetic algorithm. This application was developed to make it easier for tourists and everyone who wants to find the best hotel recommendations with the best satisfying service. The contribution of this research is to develop a new application based on machine learning that has a better level of accuracy and can provide better benefits since the application proposed and developed can be accessed anywhere by everyone. The constructed classification in the system proposed consists of two sentiments; positive and negative sentiment.

2.   RESEARCH METHOD
2.1. Dataset

To build a system and applications, sentiment data for hotel review assessments obtained from the site google maps [24] with the keyword hotels in the Central Java region is used. The data used is in the form of text data which is processed in the data preprocessing process and then processed into a model applied to the application. An overview of the dataset used in the search for the built-in application model is as shown in Figure 1. The data used is Indonesian text data taken in 2019 which consists of a collection of negative assessment comments with a 1-2 stars rating and positive reviews with a rating between 4-5 stars. Overall, the number of review comments for this hotel rating is 414 reviews [25].

5/5 2 tahun lalu di G Google

Hotel baru yang menarik di kota tegal. Mempunyai fasilitas yang lengkap. Menu breakfast lengkap dan ada menu jajanan gorengan khas tegal/slawi yaitu tahu aci. Selain menu standar yang biasa ada,seperti tradisional dan western

1/5 2 tahun lalu di G Google

mahal, karyawan tidak ramah, kamar sempit, makanan tidak istimewa, and not rocemmended

(a)   (b)

Figure 1. Sample dataset for the proposed model, (a) Positive comment review, (b) Negatif comment review

2.2. Proposed model in system

The development of a hotel service sentiment analysis application in this study applies a previously obtained model, namely a model based on the neural network algorithm with predetermined parameters. The depiction of the model of the neural network that is applied is as shown in Figure 2 where the expected data output has two neurons. This means that the expected result is a classification of two different labels, namely positive sentiment or negative sentiment. The process of tokenization, filtering text, and weighting of text data is carried out during data processing stage. The depiction of the applied model adopts the research that has been done before, which is shown in Figure 2. Two classifications of review sentiment have been assigned in the model obtained, namely positive sentiment and negative sentiment. To get the accuracy of the model, data validation is used using k-fold cross validation with predefined parameters and produces the expected level of accuracy.
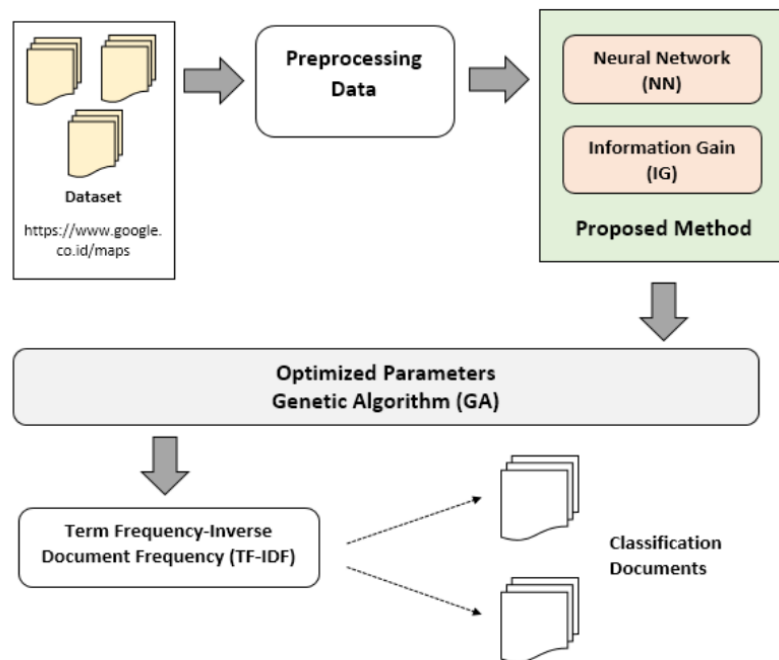

Figure 2. Framework of SentiHotel model used [21]

2.3. Stages of the system development

To build a system in the intelligent hotel sentiment analysis application, each stage is made using the Prototyping system development model. The stages carried out in this model are the system requirements analysis stage, building prototyping, protoype evaluation, system coding, system testing, system evaluation and system implementation. Each stage that is carried out is adjusted to the circumstances of the situation and conditions in the field, however it still follow the concept and is built according to planning. The depiction of the activities carried out in the development of this system in outline stages is shown in Figure 3.
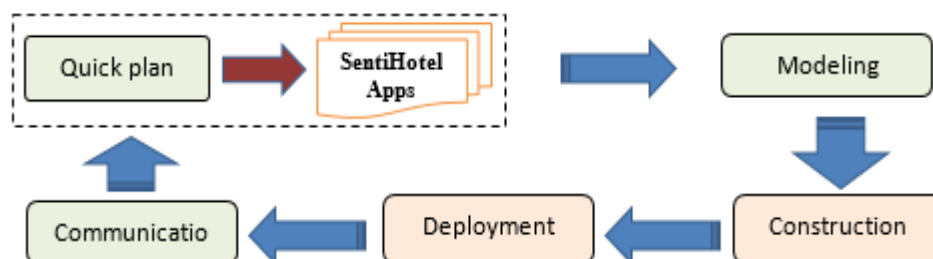

Figure 3. Model prototyping system development method [26]

3.   RESULTS AND DISCUSSION
3.1. Proposed model
In this article, the proposed model is a neural network-based model that has been optimized by applying the genetic algorithm. The specified parameters are the values obtained after parameter optimization by using a genetic algorithm and selecting the best weight using information gain (IG). The NN model obtained in Figure 4 is an NN model that uses the sigmoid activation function used (1).

$$fx = \frac{1}{1+e^{-x}} \qquad (1)$$

The model obtained consists of 10 input neurons, 7 hidden neurons, and 2 output neurons. In this model, there is 1 bias neuron in the input layer and 1 bias neuron in the hidden layer. The proposed model was obtained using the Rapid Miner Studio software analysis, Windows 10 operating system, Core i7 processor, and 8 GB RAM.
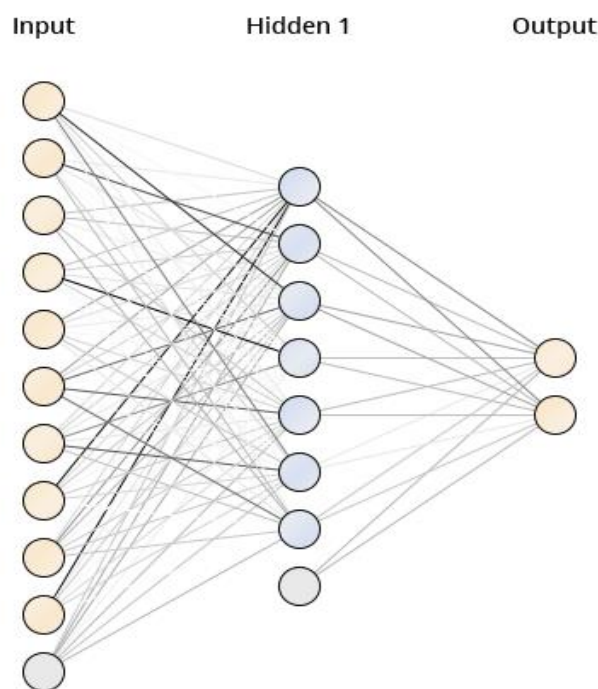

Figure 4. Neural network model applied to the application

The best parameters that have been obtained to produce this model are using the tournament selection type, training cycles=94, learning rate=0.8589114570.858911457, and momentum=0.586464677. The expected level of accuracy is obtained by using (2), (3) and (4).

$$Accuracy = \frac{True\ Positive\ (TP)+True\ Negative\ (TN)}{True\ Positive\ (TP)+False\ Negative\ (FN)+\ True\ Negative\ (TN)+False\ Positive\ (FP)} \qquad (2)$$

$$precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Postive\ (FP)} \qquad (3)$$

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Negative\ (FN)} \qquad (4)$$

Based on the experimental results in this model, the best level of accuracy for the classification of hotel service ratings is 88.99%. This level of accuracy is obtained based on the predetermined parameter combination settings. It has produced various combinations of accuracy level values, and the experimental results which are shown in Table 1 and Figure 5.
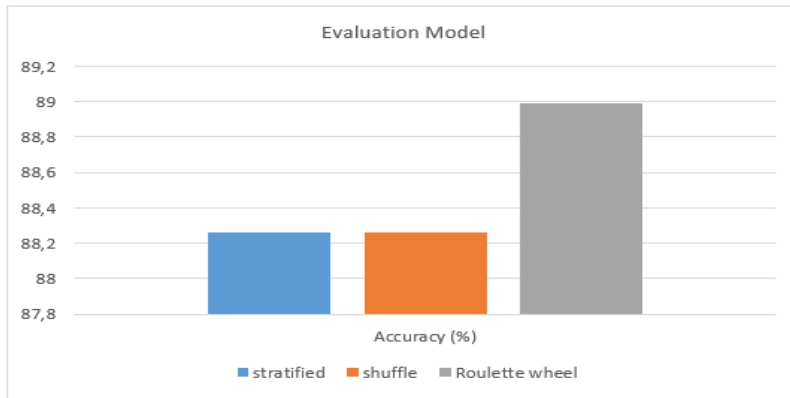
Figure 5. Evaluation of the best NN model

Table 1. The evaluation of the best model uses several sampling methods

Sampling	Accuracy (%)
stratified	88,26
shuffle	88,26
Roulette wheel	88,99

3.2. System development design

In the development of smart applications that are made, the modeling system is shown in the use case figure. In the figure, it shows that the user consists of two actors, namely admin and end user. These actors have their respective functions. The admin is given access to analyze text data that will be input including carrying out the processes needed to produce sentiment classifications starting from data preprocessing to producing sentiment classifications which will be forwarded into the hotel recommendations system with the best service. The end-user are slightly different. They have limited access where they are only capable of seeing hotel recommendations and hotel details. The use case depiction of the designed system is as shown in Figure 6.
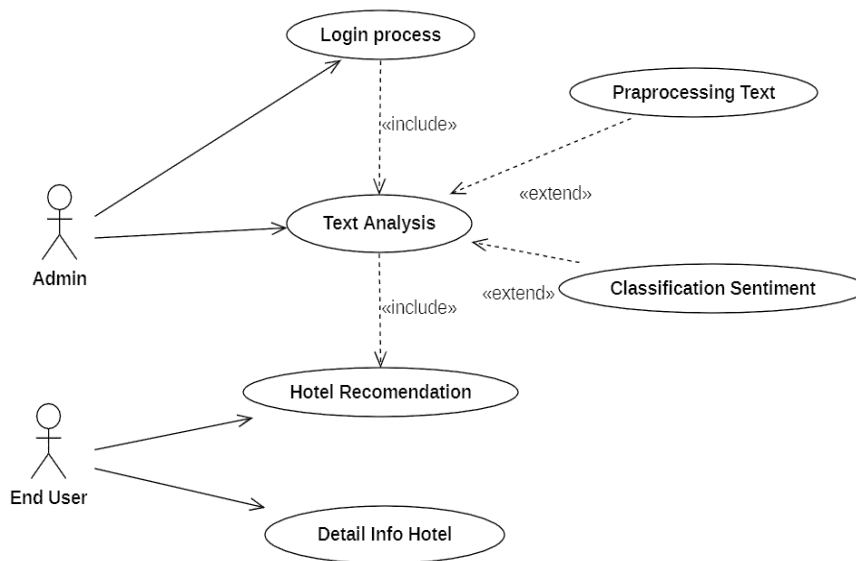

Figure 6. Use case diagram of smart application of hotel service sentiment SentiHotel

3.3. System implementation

The login page is a feature used specifically for admins who will enter the application. This login system can be accessed only by registered admin users, therefore not all users can enter the system. On this interface, there are two data input, username and password, which will be used as a user validation system if you want to enter the system. The image shown in Figure 7 shows the main page of the admin user. There are several features that are displayed on this main page, including accapitulations regarding the number of hotels reviewed, the number of visitor reviews, and user data. The main page access of the visitor section is also displayed in this feature.
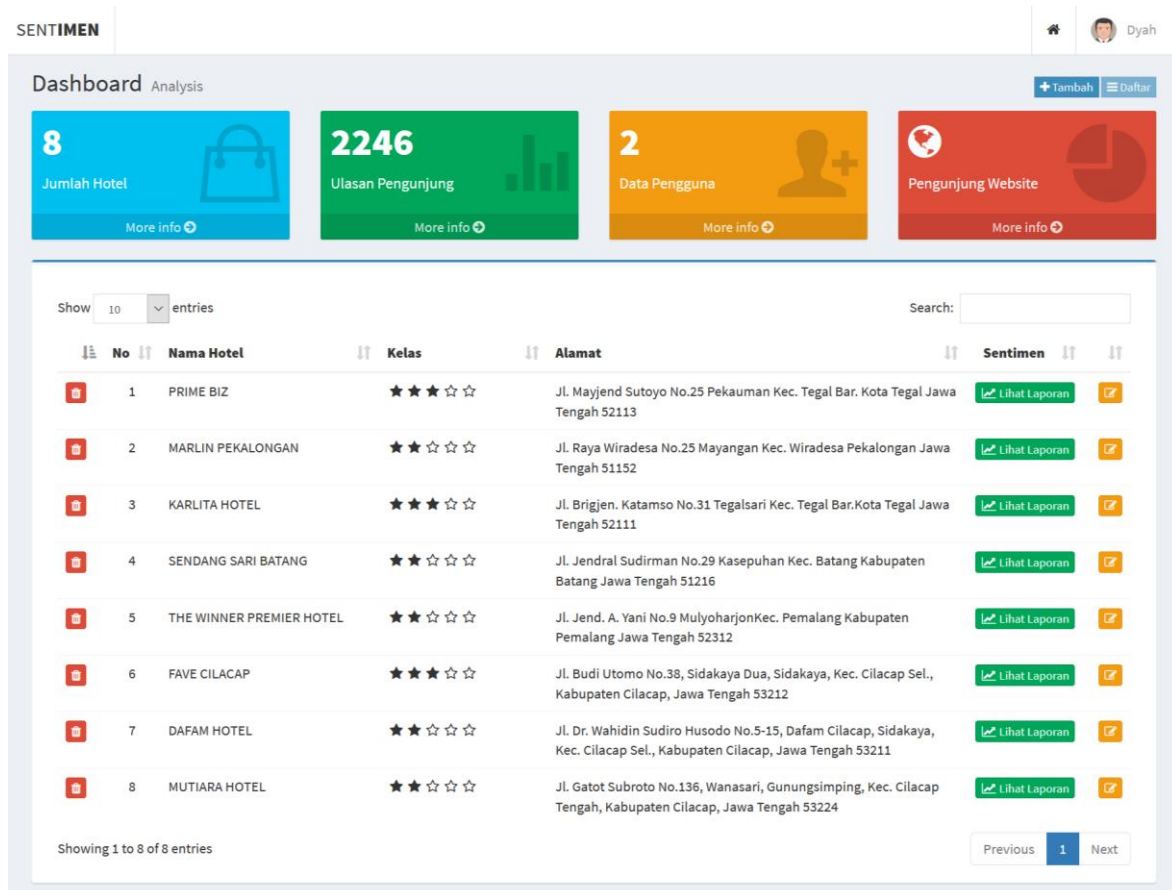
Figure 7. Main admin page interface

On this admin page, a recapitulation and a list of the names of the hotels being reviewed are also displayed, along with a "class" rating which contains an asterisk rating totaling 5 (five) stars. In the data list, the hotel address and the results of sentiment analysis are also displayed as well as an edit feature for other analysis processes. There are facilities to change the hotel data in this feature if there are any changes. Data change will adjust to the new data input. To provide an information obtained by the end-user as the user of this sentiment analysis, the main display containing information related to the hotel review sentiment is created. This page contains data regarding this application information as well as some information that can be displayed about a hotel service. This information is in the form of hotel data that can be selected according to the details of the data that have been input. To see details, the user just has to select "more info". The interface features of the main end-user page are shown in Figure 8.

On the feature page of Figure 8, list of hotel names is compiled based on incoming data and the results of the analysis that have been carried out on the previous admin page. End-users can only access it without logging in to the system in this feature. If you want to know detailed information about the sentiment results from the hotel, then in Figure 8 you can see a detailed feature menu display that provides it. Figure 9 shows detailed information about the hotel based on the results of the sentiment analysis that has been carried out. In this display, the information obtained is detailed hotel information as well as the address equipped with a map of the selected hotel address. In this feature, we can see the review sentiment display of this hotel which can be displayed graphically and can be seen on a monthly basis, as well as the comments that come in for this hotel review.

3.4. System evaluation

System evaluation is carried out to provide an idea that the applications that have been made are aligned properly. This testing process using the Black Box method that focused on system functions that are in process and have been made. The test results are shown in Table 2, Table 3, Table 4, Table 5, and Table 6. Table 2 shows the system testing carried out on the funnctional system login done by the admin. The system is tested using two functions alternately in this test. From the results, it is found that the functional login system runs as it should.
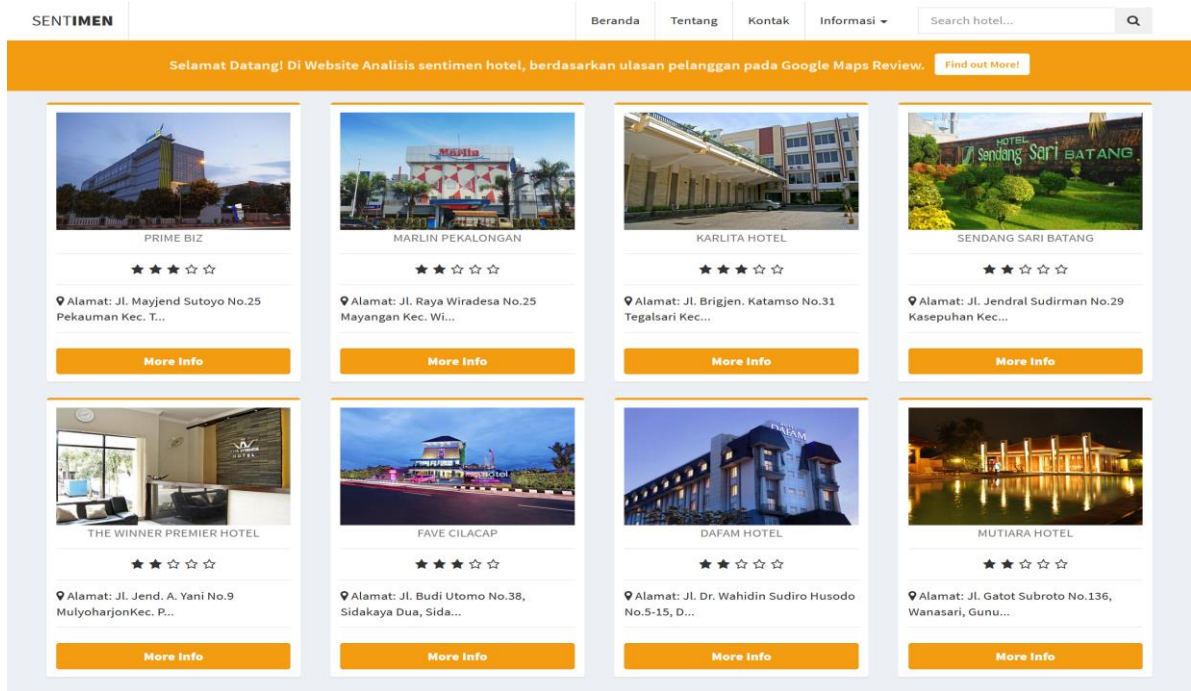
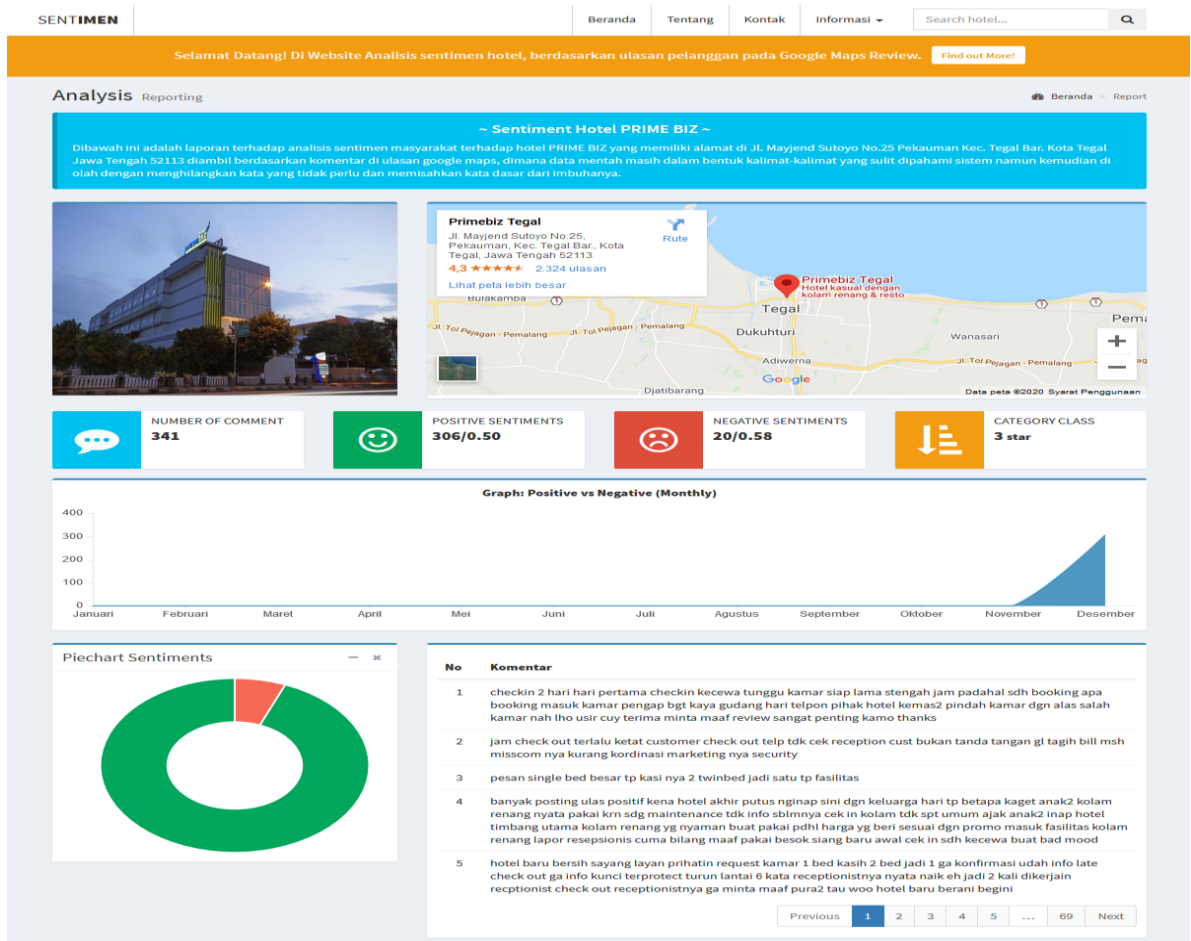Figure 8. The end-user main page interface

Figure 9. Interface of end-user analysis detail page

Table 2. Testing the login page (login)

Field	Input	Output	Annotation
Username & Password	Data != DB	Login Failed	Denied
Username & Password	Data == DB	Login Success	Granted

Table 3. Data upload test (data validation)

Field	Input	Output	Annotation
Rawdata (.csv)	Is Null	Upload Pending	Denied
Rawdata (.csv)	Not Null	Upload Success	Granted

Table 4. Stemming testing (sentence correction)

Field	Input	Output	Annotation
Netizen Comment (array)	Array[i] == Words with affix	Stemming	Processed
Netizen Comment (array)	Array[i] != Words with affix	Not Stemming	Not Processed

Table 5. Deletion test of the connecting word (stopword)

Field	Input	Output	Annotation
Netizen Comment (array stemmed)	Array[i] == Words with affix	Processed	Deleted
Netizen Comment (array stemmed)	Array[i] != Words with affix	Not Processed	Not deleted

The data upload test carried out on the system is done by performing the data upload functional test. The system is functional as expected, the results are shown in Table 3. The next test is carried out by testing the stemming and stopword data preprocessing analysis, in which the results are in accordance with the results as shown in Table 4 and Table 5. The next step is to conduct a test on the function of database access. Table 6 shows the results obtained. The result of functional database access on the system is as it should be. Likewise, the test done on the functional database access of the end-user report page shows that the system language is running properly, as it is shown in Table 7.

Table 6. Testing the home page (home)

Field	Input	Output	Annotation
Hotel data from database	Array[i] == Exist	Data Shown	Displays hotel data
Hotel data from database	Array[i] != Exist	Data not shown	Do not display hotel data

Table 7. Testing report pages (report)

Field	Input	Output	Annotation
Data analysis from the database	Array[i] == Exist	Displays hotel data	Show hotel analysis
Data analysis from the database	Array[i] != Exist	Do not display hotel data	Do not show hotel analysis

3.5. Limitation and future research

SentiHotel has shortcoming in the applied model (neural network), especially in the not yet optimal and significantly high level of accuracy. This increase in accuracy can be answered by conducting advanced experiments to get the terabit model with the best classification accuracy rate. SentiHotel has not been given a feasibility study assessment by end users, namely tourists who need assessment information and require hotel service information. The assessment has not yet been carried out and requires further experiments, especially in terms of analyzing the use of system acceptance by users. The usability system method that can be carried out for further research can be done using, the TAM, UTAUT, and other methods [27] so that SentiHotel is ready to be used and implemented to provide information to users regarding the assessment of a hotel service.

4. CONCLUSION

An application called SentiHotel is proposed and developed to apply a model obtained with the best performance. The applied neural network optimization model has enabled the system to functionally have different capabilities, namely being able to carry out the sentiment process of hotel service review assessments which has an accuracy rate of 88.99%. This study provides new contributions based on other previous studies where the development of a new application of sentiment analysis based on the neural network model is proposed thus it provides benefits for anyone who wants to get a hotel service recommendation. The sentiment application analysis of hotel services functionally makes it easy for users, especially tourists who are looking for the best hotel reference. For further research, this application requires a continuous experimental effort that will lead to the may be even better results. The difficulty of implementing a model which contain an algorithm, provides its own level of difficulty; therefore collaboration efforts with programmers are very important.

ACKNOWLEDGEMENTS

REFERENCES

[1]   R. Nunkoo, V. Teeroovengadum, C. M. Ringle, and V. Sunnassee, "Service quality and customer satisfaction: The moderating effects of hotel star rating," *Int. J. Hosp. Manag.*, vol. 91, p. 102414, Oct. 2020, doi: 10.1016/j.ijhm.2019.102414.

[2]   A. Gupta, C. Chakraborty, and B. Gupta, "Monitoring of Epileptical Patients Using Cloud-Enabled Health-IoT System," *Trait. du Signal*, vol. 36, no. 5, pp. 425–431, Nov. 2019, doi: 10.18280/ts.360507.

[3]   R. Dwivedi, S. Dey, C. Chakraborty, and S. Tiwari, "Grape Disease Detection Network based on Multi-task Learning and Attention Features," *IEEE Sens. J.*, 2021, doi: 10.1109/JSEN.2021.3064060.

[4]   R. García, J. Aguilar, M. Toro, A. Pinto, and P. Rodríguez, "A systematic literature review on the use of machine learning in precision livestock farming," *Computers and Electronics in Agriculture*, vol. 179. Elsevier B.V., p. 105826, Dec. 01, 2020, doi: 10.1016/j.compag.2020.105826.

[5]   T. van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Computers and Electronics in Agriculture*, vol. 177. Elsevier B.V., p. 105709, Oct. 01, 2020, doi: 10.1016/j.compag.2020.105709.

[6]   Z. Kang, C. Catal, and B. Tekinerdogan, "Machine learning applications in production lines: A systematic literature review," *Computers and Industrial Engineering*, vol. 149. Elsevier Ltd, p. 106773, Nov. 01, 2020, doi: 10.1016/j.cie.2020.106773.

[7]   D. Truong, "Using causal machine learning for predicting the risk of flight delays in air transportation," *J. Air Transp. Manag.*, vol. 91, p. 101993, Mar. 2021, doi: 10.1016/j.jairtraman.2020.101993.

[8]   A. W. Muhammad, G. W. Sasmito, and I. Riadi, "Colorectal Polyp Detection Using Feedforward Neural Network with Image Feature Selection," in *Proceeding - 2018 International Symposium on Advanced Intelligent Informatics: Revolutionize Intelligent Informatics Spectrum for Humanity, SAIN 2018*, Mar. 2019, pp. 26–31, doi: 10.1109/SAIN.2018.8673371.

[9]   J. Wen, S. Li, Z. Lin, Y. Hu, and C. Huang, "Systematic literature review of machine learning based software development effort estimation models," *Inf. Softw. Technol.*, vol. 54, no. 1, pp. 41–59, 2012, doi: 10.1016/j.infsof.2011.09.002.

[10] D. M. E. D. M. Hussein, "A survey on sentiment analysis challenges," *J. King Saud Univ. - Eng. Sci.*, vol. 30, no. 4, pp. 330–338, 2018, doi: 10.1016/j.jksues.2016.04.002

[11] F. A. Pozzi, E. Fersini, E. Messina, and B. Liu, "Sentiment Analysis in Social Networks," *Elsevier,* 2016, pp. 284.

[12] B. Liu and L. Zhang, "A Survey of Opinion Mining and Sentiment Analysis," in *Mining Text Data*, Boston, MA: Springer US, 2012, pp. 415–463.

[13] E. Sutoyo and A. Almaarif, "Twitter sentiment analysis of the relocation of Indonesia's capital city," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 9, no. 4, pp. 1620–1630, Aug. 2020, doi: 10.11591/eei.v9i4.2352.

[14] N. D. Zaki, N. Y. Hashim, Y. M. Mohialden, M. A. Mohammed, T. Sutikno, and A. H. Ali, "A real-time big data sentiment analysis for iraqi tweets using spark streaming," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 9, no. 4, pp. 1411–1419, Aug. 2020, doi: 10.11591/eei.v9i4.1897.

[15] D. H. Abd, A. R. Abbas, and A. T. Sadiq, "Analyzing sentiment system to specify polarity by lexicon-based," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 10, no. 1, pp. 283–289, Feb. 2021, doi: 10.11591/eei.v10i1.2471.

[16] L. K. Ramasamy, S. Kadry, and S. Lim, "Selection of optimal hyper-parameter values of support vector machine for sentiment analysis tasks using nature-inspired optimization methods," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 10, no. 1, pp. 290–298, Feb. 2021, doi: 10.11591/eei.v10i1.2098.

[17] C. Bucur, "Using Opinion Mining Techniques in Tourism," *Procedia Econ. Financ.*, vol. 23, no. October 2014, pp. 1666–1673, 2015, doi: 10.1016/s2212-5671(15)00471-2.

[18] F. Ali, K. S. Kwak, and Y. G. Kim, "Opinion mining based on fuzzy domain ontology and Support Vector Machine: A proposal to automate online review classification," *Appl. Soft Comput. J.*, vol. 47, pp. 235–250, 2016, doi: 10.1016/j.asoc.2016.06.003.

[19] O. Somantri, D. A. Kurnia, D. Sudrajat, N. Rahaningsih, O. Nurdiawan, and L. Perdana Wanti, "A Hybrid Method Based on Particle Swarm Optimization for Restaurant Culinary Food Reviews," in *Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019*, Oct. 2019, doi: 10.1109/ICIC47613.2019.8985842.

[20] Y. C. Chang, C. H. Ku, and C. H. Chen, "Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor," *Int. J. Inf. Manage.*, vol. 48, pp. 263–279, 2019, doi: 10.1016/j.ijinfomgt.2017.11.001.

[21] D. Apriliani, T. Abidin, E. Sutanta, A. Hamzah, and O. Somantri, "Sentiment analysis for assessment of hotel services review using feature selection approach based-on decision tree," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 4, pp. 240–245, Jun. 2020, doi: 10.14569/IJACSA.2020.0110432.

[22] D. Apriliani, T. Abidin, E. Sutanta, A. Hamzah, and O. Somantri, "Sentiment analysis for Indonesia hotel services

review using optimized neural network," in *Journal of Physics: Conference Series*, Jun. 2020, vol. 1538, no. 1, p. 12060, doi: 10.1088/1742-6596/1538/1/012060.

[23] S. Ali Alsaidi, A. T. Sadeq, and H. S. Abdullah, "English poems categorization using text mining and rough set theory," *Bull. Electr. Eng. Informatics*, vol. 9, no. 4, pp. 1701–1710, Aug. 2020, doi: 10.11591/eei.v9i4.1898.

[24] "Google Maps." [Online]. Available: https://www.google.com/maps/ .

[25] O. Somantri and D. Apriliani, "Dataset: Sentiment Analysis Application of Hotel Services in Indonesia." doi: 10.5281/zenodo.4635253.

[26] R. S. Pressman, *Software Engineering A Practitioner's Approach 7th*. McGraw-Hill Education, 2010.

[27] G. W. Sasmito, L. O. M. Zulfiqar, and M. Nishom, "Usability Testing based on System Usability Scale and Net Promoter Score," in *2019 2nd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2019*, Dec. 2019, pp. 540–545, doi: 10.1109/ISRITI48646.2019.9034666.

BIOGRAPHIES OF AUTHORS

Dyah Apriliani, received Bachelor-degree in Information Technology at the Universitas Ahmad Dahlan (UAD) Indonesia. She received his Master/s degree in Information System from Universitas Dipenogoro, Indonesia. The area of his research interest lies in information system, decision support system, business intelligence, and intelligent system.

Taufiq Abidin, received Bachelor Degree in Education from Universitas Negeri Semarang Indonesia and later he received his Master/s degree in Informatics from Universitas Dian Nuswantoro Indonesia. His research areas of interest are informatics, software engineering and database.

Edhy Sutanta, received Bachelor Degree of Informatics Management & Computer Engineering form IST AKPRIND Yogyakarta, Indonesia in 1996, received Master of Computer Science from Universitas Gadjah Mada, Yogyakarta, Indonesia in 2006, and completed his Doctorate Degree at Computer Science from Universitas Gadjah Mada, Yogyakarta, Indonesia in 2017. Currently he is an associate profesor at Department of Informatics Engineering in IST AKPRIND Yogyakarta, Indonesia. His research areas of interest are database system, database analysis & design, information system, schema matching, and e-Government.

Amir Hamzah, he received his Bachelor Degree in statistics form Institut Pertanian Bogor Indonesia, received his Master/s Degree in electrical engineering from Universitas Gadjah Mada Yogyakarta Indonesia in 1998, and completed his Doctorate Degree at electrical engineering from Universitas Gadjah Mada Indonesia in 2009. His research area of interest are emotion mining, text mining, information retrieval, and opinion mining.

Oman Somantri, received Bachelor Degree in Informatics from STMIK Sumedang Indonesia and later he received his Master/s degree in Informatics from Universitas Dian Nuswantoro Indonesia. His research areas of interest are data mining, text mining, decision support system, information retrieval and intelligent system