Sentiment Analysis for Assessment of Hotel Services Review using Feature Selection Approach based-on Decision Tree

Dyah Apriliani[1], Taufiq Abidin[2], Edhy Sutanta[3], Amir Hamzah[4], Oman Somantri[5]

Department of Informatics Engineering, Polytechnic of Harapan Bersama, Tegal, Indonesia[1, 2]
Department of Informatics Engineering, Institut Sains & Teknologi AKPRIND Yogyakarta, Yogyakarta, Indonesia[3, 4]
Department of Informatics Engineering, State Polytechnic of Cilacap, Cilacap, Indonesia[5]

Abstract—To get the best hotel accommodation equipped with great services is all what a tourist want. Hotel reviews found in social media sometimes become a reference to book a hotel room. The problem is there is sometimes inaccuracy in understanding the reviewer's sentiment; therefore sentiment analysis approach is used in this study. The sentiment analysis approach use three algorithms within this article; Naïve Bayes, Support vector machines, and decision tree. The result of the experiment is that decision tree is the best algorithm, however the accuracy level still become a focus since it is not optimal. The purpose of this study is to find a hybrid sentiment analysis model of an intelligent application that can be used as a decision support for hotel service assessment recommendations problem. In this paper, we proposed a model which was developed using the feature selection (FS) approach, whereas the improvement of model accuracy was done using information gain (IG). In this study, the experiment was carried out through five stages, namely taking the research dataset in the form of hotel service assessment texts, data pre-processing, weighting, experimental models, and evaluation. Experiments were conducted to get the best accuracy on the proposed model, while the evaluations were carried out to determine the accuracy of the model. Based on the experimental results, the best accuracy level in the model is 88.54%.

Keywords—Sentiment analysis; feature selection; decision tree; support vector machines; Naïve Bayes; hotel services

I. INTRODUCTION

People who are traveling generally require the best hotel recommendations that suit their needs. Some people believe that expensive hotel rents do not necessarily guarantee good services that meet their needs, in this regard what is needed by them is the best hotel service at an appropriate price, even though they differ in class.

The existence of social media currently has a broad impact. Nowadays social media is also used as a reference to get information and news [1], including recommendations and ratings of hotel services. Opinions and comments about hotel services are widely posted by people who have experienced and stayed at the hotel. Opinions or comments that appear on social media can be indicated as positive, negative, or mediocre. But the problem is that the opinions or comments contained in the media are still difficult to interpret whether they are positive, negative or mediocre. In this case, text

mining based computerized sentiment analysis (SA) is proposed because of its capabilities [2][3].

Application of SA technology is textual; therefore it allows to get the inclusive sentiment information. The advantages possessed by the SA will produce a decision support, whether in the form of sentiment classification, review of an object, detection of a spam, and others [4]. The methods often used in SA include support vector machines (SVM), Naïve Bayes, neural networks, decision trees (DT), Bayesian networks, and maximum entropy. These methods can be chosen according to the problem that will be used as the object of research [5]. Too many attributes in the model have caused problems in the poor classification results in the text mining model [6]. Another problem is the difficulty of determining the optimal parameters therefore the accuracy obtained is still low.

The novelty of this study lies in the proposed model, namely the optimization of the increase in accuracy produced on the DT model by using the IG in the classification of hotel services so as to produce a better level of accuracy. Another advantage of the proposed model is that when it is applied to a sentiment analysis application system, it can be used and utilized by tourists to get an overview of hotel ratings with the best service since the model used has a high accuracy level.

II. RELATED WORKS

Several previous studies have been conducted related to the application of SA, one that was done by Y. H. Hu & K. Chen [7] who examined the predictions of hotel reviews, including visibility and interaction reviews between star hotels and rating reviews. The results of this study stated that the tree model (M5P) was superior to linear regression. The tree model also supports SVM, this is because it is better in modeling the interaction effect. Research conducted by D. Gräbnera & M. Zankerb [8] examined the classification of consumer review based on sentiment analysis on hotel review classifications. In this study the Lexicon model is proposed for review classification and produces a better model with an accuracy rate of 90%.

Research conducted by S. Nadali et al. [9] examined a sentiment classification for consumer reviews using fuzzy logic, in contrast to what was done by A. Reyes & P. Rosso [10] who examined the sentiment of an objective decision support derived from subjective data and focused on detecting

irony in customer reviews. Research by Y. C. Chang et al. [11] analyzed social media as a source of text data extracted and visualized at the Hilton hotel, and reviews from Tripadvior. Other studies on opinion classification using maximum entropy and K-Means clustering were conducted by A. Hamzah & N. Widyastuti [12] where the results stated that K-Means are better than maximum entropy with an average precision of 3%.

In addition, research on opinion mining towards the review using the optimization model was carried out by [13], this research was conducted by proposing a Naïve Bayes model based on feature selection on customer reviews of culinary foods. Various attempts were made to improve the accuracy of text mining, one of which was with feature selection (FS). Efforts to improve accuracy with FS have been done to improve accuracy in a classification model, such as sentiment analysis and text categorization [14], [15]–[18].

In contrast to previous studies, this paper discusses a proposed DT-based model of sentiment analysis for the classification of hotel service ratings. Based on the advantages obtained during feature selection process [19]-[21], sentiment analysis was developed based on FS and IG as the best method to improve the accuracy of the model.

III. RESEARCH METHOD

A. Proposed Method

At this stage the dataset is included into the model. Experiments are carried out on several models and predefined parameter combinations. The model proposed in this study is FS using DT-based IG. At this stage DT is used as a classical model that is optimized using FS therefore it is expected to get the results with the best classification accuracy. A description of the stages in the proposed model is shown in Fig. 1. Data validation in this study uses Cross Validation with the accuracy obtained is the best value. Evaluation of the model is done to find out whether the model obtained is in accordance with what is desired. Model evaluation is done by comparing

the level of accuracy obtained by the proposed model with other relevant models, namely SVM and DT.

B. Dataset and Tools

To get the best sentiment analysis model in hotel service valuation, the research was conducted in five stages, namely taking the research dataset in the form of hotel service assessment text, data pre-processing, weighting, model experimentation, and evaluation. Experiments carried out to get the expected best model. The research dataset used came from the website https://www.google.com/maps, where there are text data in the form of a collection of opinions and comments that have been posted by the public for hotel services located in the Central Java province of Indonesia. The research dataset used is Indonesian-language text data between 2015 and 2018. An example of the text dataset used is shown in Fig. 2. In Fig. 2, the text data used contains user reviews of hotel services. The amount of text included into the model is determined at the time of preprocessing data. To get good analysis results, RapidMiner Studio 9.2 software tools and Windows 7 operating system are used in the study, while the hardware specifications used are Intel Corei5 processor and 4Gb memory.

Fig. 1. Examples of Research Datasets used.
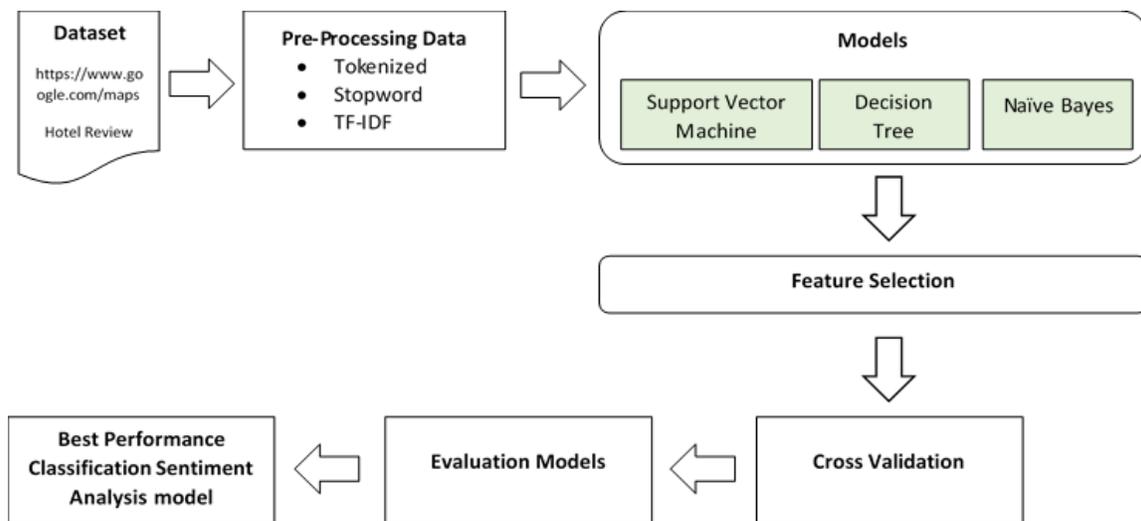
Fig. 2. Proposed Model Framework.

C. Preprocessing Data

Preprocessing data is done before the dataset is applied into the model for experimentation with the aim of getting the best model. The steps taken include the process of tokenization, namely the separation of each word in the sentence contained in the text, filter text, stopword, and the weighting process using the term frequency inverse document frequency (TF-IDF) method [21] - [23]. At this stage, the training data and testing data are determined. Training data is done by separating the overall results of the dataset obtained, namely 70% training data and 30% testing data. The data is further classified into two labels, namely "positive" and "negative".

D. Decision Tree and Information Gain (IG)

DT is a model that represents an alternative class of a classification algorithm. DT structure resembles a tree, where every node it has is a test attribute, each branch represents the result of the test, and the leaf node represents the class [24] [25]. The DT method can be calculated using equation (1), and to calculate the entropy value equation (2) is used.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si) \qquad (1)$$

$$Entropy(S) - \sum_{i=1}^{n} - Pi * \log_2 Pi \qquad (2)$$

IG is one of the best algorithms that can be used for feature selection [26]. In this study, the IG calculation is done using equations (3) and (4), whereas to measure the effectiveness of an attribute in the data classification it is calculated by equation (5).

$$Info(D) = -\sum_{i=1}^{c} p_i \log_2(p_i) \qquad (3)$$

where

c: the number of values in the target attribute

pi: total sample of class i

$$Info_A(D) = \sum_{j=1}^{V} \left(\frac{Dj}{D}\right) x\ Info(D_j) \qquad (4)$$

$$Gain(A) = |\ Info(D) - Info_A(D)\ | \qquad (5)$$

IV. RESULTS AND DISCUSSION

The experiments in this study were conducted on three classic models of sentiment analysis reviews for hotel service assessments, namely DT, SVM, and Naïve Bayes. In the experiment using the SVM method, K-Fold settings are performed with the aim to get the best accuracy, and 3 parameter criterion to get the best results. The k-Fold values used are 10, 8, 6, and 4 since the number of k-Fold used will affect the resulting accuracy therefore the combination of k-Fold parameter determination is done. Variation of values in the criterion parameters used are dot, radial, polynomial.

A. Results of Experiments on the Classical Model

The results of sentiment analysis experiments using three classical models are shown in Table I for the DT method, Table II for the SVM method, and Table III for the Naïve Bayes method.

TABLE I. THE RESULTS OF THE SENTIMENT ANALYSIS EXPERIMENT WITH THE DECISION TREE METHOD

Sampling Type	Criterion	Validation			
		k-Fold=10 (%)	k-Fold=8 (%)	k-Fold=6 (%)	k-Fold=4 (%)
Stratified	gain_ratio	87.44	87.63	87.35	87.35
	information_gain	86.53	**88.26**	86.72	86.35
	gini_index	86.89	87.90	87.17	87.17
Shuffled	gain_ratio	87.54	87.44	87.35	87.26
	information_gain	86.99	88.17	87.08	86.35
	gini_index	86.99	86.99	87.17	86.62
Linear	gain_ratio	87.00	85.59	85.43	84.18
	information_gain	85.27	83.41	83.43	83.54
	gini_index	86.89	83.50	83.70	83.27

TABLE II. THE RESULTS OF SENTIMENT ANALYSIS EXPERIMENTS WITH THE SUPPORT VECTOR MACHINE METHOD

Sampling Type	Criterion	Validation			
		k-Fold=10 (%)	k-Fold=8 (%)	k-Fold=6 (%)	k-Fold=4 (%)
Stratified	dot	87.81	87.90	87.72	87.90
	radial	87.90	87.90	87.90	88.08
	polynomial	87.72	87.72	87.72	87.72
Shuffled	dot	87.81	87.90	87.90	87.99
	radial	**88.26**	87.72	88.26	86.63
	polynomial	87.72	87.72	87.72	87.72
Linear	dot	87.55	87.32	87.25	87.27
	radial	87.91	87.14	87.16	86.18
	polynomial	87.73	87.68	87.70	87.73%

TABLE III.   RESULTS OF SENTIMENT ANALYSIS EXPERIMENTS WITH THE NAÏVE BAYES METHOD

K-Fold	Sampling Type		
	Stratified (%)	*Shuffled (%)*	*Linear (%)*
10	52.14	56.23	52.14
9	51.50	56.87	51.50
8	56.33	56.77	47.30
7	55.87	56.96	48.13
6	56.96	56.23	49.50
5	57.14	56.78	51.14
4	56.50	57.51	52.23
3	57.51	57.05	54.87
2	57.32	57.33	**61.59**

Based on experimental results on the DT, SVM, and Naïve Bayes method, the highest value of accuracy using the decision tree method is 88.26%, and support vector machine is 88.26%, while for the highest value using Naïve Bayes is 61.59%. Based on experiments, these values indicate that the level of accuracy produced using each method has different result therefore the selection of the model used cannot guarantee that all models have the same results.

B. Experiment Results on Optimized Models

Model optimization is done to improve the accuracy of the sentiment analysis classification results in hotel services, so we get the best model. FS optimization is done by applying IG to the DT and SVM methods. Fig. 3 is decision tree model after optimization; moreover Table IV shows the results of sentiment optimization analysis on the DT model using IG, while the results on the SVM method are shown in Table V and Fig. 4.

Based on Table IV, it is known that the highest level of accuracy in DT models using IG is 88.54%. The highest accuracy is obtained in the DT model and the application of IG with criterion = gini_index and k-Fold = 8. Based on the experimental results on the SVM model using IG in Table V, the highest accuracy level obtained was 88.45% in the polynomial kernel.

C. Evaluation of Models

Evaluation of the sentiment analysis model for assessment of hotel services review using the DT-based approach based on DT is done by comparing the accuracy values obtained from the results of experiments that have been conducted. In summary, a comparison of the accuracy levels between the models obtained is shown in Table VI. The highest accuracy is obtained in the DT model that is optimized using the FS method, which is 88.54%. The FS-based DT model has a better level of accuracy, this is after an optimization using IG with its ability to provide the best weight value in the DT model therefore it leads to an increase in accuracy. The resulting increased accuracy between DT + FS and SVM + FS has a difference that is not too significant, this is because the two models have almost the same level of ability in the modeling process, but both models produce different levels of accuracy but not too far away.

The level of accuracy generated is very much influenced by the parameter settings in the model used, such as the k-Fold value in the model validation process, and the determination of sampling type. Based on the experimental results obtained in the proposed model, there is a difference with the model proposed in other previous studies that have been conducted.
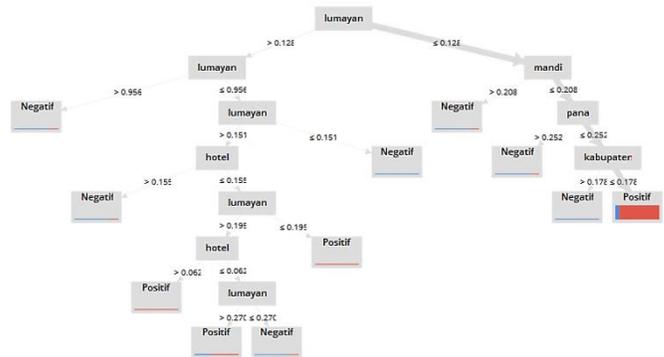
Fig. 3.   Proposed Model with Decision Tree Feature Selection Approach.

TABLE IV.   THE RESULTS OF SENTIMENT ANALYSIS EXPERIMENTS ON THE DT MODEL USING IG

Validation	Sampling Type		
	Stratified (%)	*Shuffled (%)*	*Linear (%)*
k-Fold=10	87.71	87.72	86.72
k-Fold=8	87.63	87.63	85.32
k-Fold=6	87.72	87.81	85.70
k-Fold=4	87.26	87.54	85.82
k-Fold=10	87.72	87.72	87.36
k-Fold=8	88.17	87.81	85.87
k-Fold=6	87.90	87.35	85.61
k-Fold=4	87.44	87.72	85.82
k-Fold=10	87.72	87.99	86.54
k-Fold=8	**88.54**	87.63	85.50
k-Fold=6	87.35	87.54	85.98
k-Fold=4	87.71	87.08	85.54
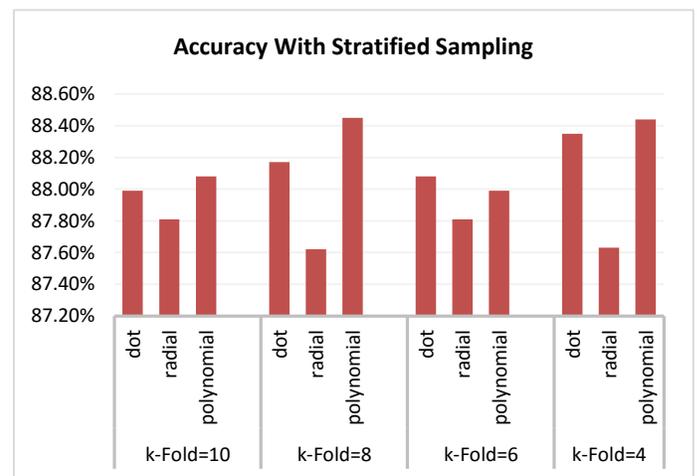
Fig. 4.   Perfomance Accuracy SVM + IG with Stratified Sampling.

TABLE V. THE RESULTS OF SENTIMENT ANALYSIS EXPERIMENTS ON THE SVM MODEL USING IG

Validation	Kernel type	Sampling Type		
		Stratified (%)	Shuffled (%)	Linear (%)
k-Fold=10	dot	87.99	87.90	86.91
	radial	87.81	83.81	87.73
	polynomial	88.08	87.90	87.18
k-Fold=8	dot	88.17	87.44	85.77
	radial	87.62	87.81	86.68
	polynomial	**88.45**	88.26	86.77
k-Fold=6	dot	88.08	88.08	86.07
	radial	87.81	87.63	86.89
	polynomial	87.99	88.26	86.80
k-Fold=4	dot	88.35	88.36	86.18
	radial	87.63	87.17	86.91
	polynomial	88.44	87.99	86.91

TABLE VI. COMPARISON OF ACCURACY BETWEEN MODELS

No	Model	Accuracy (%)
1	Naïve Bayes	61.59
2	Support Vector Machine	88.26
3	Decision Tree	88.26
4	Support Vector Machine + Feature Selection	88.45
5	Decision Tree + Feature Selection (Proposed Method)	**88.54**

However, the process of finding the best model in the sentiment analysis for assessment of hotel services review classification has a combination of models that can be applied according to specified parameters. The process of finding the best parameters in modeling becomes a thing to do therefore the model obtained is the best model with the best level of accuration.

V. CONCLUSION

The use of the FS method to optimize sentiment analysis for the assessment of hotel services review using the FS approach based on DT in this study is to improve and provide the best accuracy. The accuracy value obtained is influenced by the parameters of sampling type, criterion, and determination of the k-Fold value in the DT model. This research will then conduct a model optimization experiment by doing parameter value optimization on the proposed model to improve the accuracy of the desired model classification for the better result.

The focus of the next study should be the selection of parameter values. Other optimization algorithms can be used to find the value of these parameters. In addition, the combination of the selection settings for the parameters would be optimized; therefore it greatly affects the accuracy value of the resulted model.

ACKNOWLEDGMENT

REFERENCES

[1] W. Fan and M. D. Gordon, "Unveiling the Power of Social Media Analytics," Commun. ACM, vol. 12, no. JUNE 2014, pp. 1–26, 2013.
[2] B. Liu, Sentiment Analysis and Subjectivity, 2nd ed. Handbook of natural language processing, 2010.
[3] L. Di Caro and M. Grella, "Sentiment analysis via dependency parsing," Comput. Stand. Interfaces, vol. 35, no. 5, pp. 442–453, 2013.
[4] K. Ravi and V. Ravi, A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, vol. 89, no. June 2015. Elsevier B.V., 2015.
[5] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of Sentimental Reviews Using Machine Learning Techniques," Procedia Comput. Sci., vol. 57, pp. 821–829, 2015.
[6] P. Koncz and J. Paralic, "An approach to feature selection for sentiment analysis," in 2011 15th IEEE International Conference on Intelligent Engineering Systems, 2011, pp. 357–362.
[7] Y. H. Hu and K. Chen, "Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings," Int. J. Inf. Manage., vol. 36, no. 6, pp. 929–944, 2016.
[8] D. Gräbnera and M. Zankerb, "Classification of customer reviews based on sentiment analysis.," Technol. Tour. , p. 12, 2012.
[9] S. Nadali, M. A. A. Murad, and R. A. Kadir, "Sentiment classification of customer reviews based on fuzzy logic," Proc. 2010 Int. Symp. Inf. Technol. - Eng. Technol. ITSim'10, vol. 2, pp. 1037–1044, 2010.
[10] A. Reyes and P. Rosso, "Making objective decisions from subjective data: Detecting irony in customer reviews," Decis. Support Syst., vol. 53, no. 4, pp. 754–760, 2012.
[11] 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., no. September, pp. 0–1, 2017.
[12] A. Hamzah and N. Widyastuti, "Opinion classification using Maximum Entropy and K-Means Clustering," in 2016 International Conference on Information & Communication Technology and Systems (ICTS), 2016, pp. 162–166.
[13] O. Somantri and D. Apriliani, "Opinion mining on culinary food customer satisfaction using naïve bayes based-on hybrid feature selection," Indones. J. Electr. Eng. Comput. Sci., vol. 15, no. 1, pp. 468–475, 2019.
[14] E. M. Mashhour, E. M. F. El Houby, K. T. Wassif, and A. I. Salah, "Feature Selection Approach based on Firefly Algorithm and Chi-square," Int. J. Electr. Comput. Eng., vol. 8, no. 4, p. 2338, Aug. 2018.
[15] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," Procedia Comput. Sci., vol. 17, pp. 26–32, 2013.
[16] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN," Expert Syst. Appl., vol. 40, no. 2, pp. 621–633, Feb. 2013.
[17] S. P. Rajamohana and K. Umamaheswari, "Hybrid approach of improved binary particle swarm optimization and shuffled frog leaping for feature selection," Comput. Electr. Eng., vol. 67, pp. 497–508, 2018.
[18] A. K. Uysal and S. Gunal, "A novel probabilistic feature selection method for text classification," Knowledge-Based Syst., vol. 36, pp. 226–235, 2012.
[19] A. K. Uysal, "An improved global feature selection scheme for text classification," Expert Syst. Appl., vol. 43, pp. 82–92, 2016.
[20] S. Lei, "A Feature Selection Method Based on Information Gain and Genetic Algorithm," 2012 Int. Conf. Comput. Sci. Electron. Eng., pp. 355–358, 2012.
[21] K. Chen, Z. Zhang, J. Long, and H. Zhang, "Turning from TF-IDF to TF-IGM for term weighting in text classification," Expert Syst. Appl., vol. 66, pp. 1339–1351, 2016.
[22] D. Kim, D. Seo, S. Cho, and P. Kang, "Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec," Inf. Sci. (Ny)., vol. 477, pp. 15–29, Mar. 2019.

[23] T. Zhang and S. S. Ge, "An Improved TF-IDF Algorithm Based on Class Discriminative Strength for Text Categorization on Desensitized Data," in Proceedings of the 2019 3rd International Conference on Innovation in Artificial Intelligence - ICIAI 2019, 2019, pp. 39–44.

[24] B. Tidke and R. Mehta, "A Comprehensive Review and Open Challenges of Stream Big Data," in In Soft Computing: Theories and Applications, Singapore: Springer, 2018, pp. 89–99.

[25] K. W. DeGregory et al., "A review of machine learning in obesity," Obes. Rev., vol. 19, no. 5, pp. 668–685, May 2018.

[26] X. Deng, Y. Li, J. Weng, and J. Zhang, "Feature selection for text classification: A review," Multimed. Tools Appl., vol. 78, no. 3, pp. 3797–3816, Feb. 2019.