International Journal of Innovative Computing, Information and Control
Volume **20**, Number **2**, April **2024**

ICIC International ©2024 ISSN 1349-4198
pp. 359–372

SOFT VOTING CLASSIFIER WITH OPTIMIZED WEIGHT USING PARTICLE SWARM OPTIMIZATION ON SENTIMENT ANALYSIS FOR ONLINE CREDIT AND LOAN APPLICATION REVIEWS

Amil Ahmad Ilham*, Elly Warni and Pahrul

Department of Informatics
Universitas Hasanuddin
Jl. Poros Malino Km. 6, Bontomarannu, Gowa, Sulawesi Selatan 92171, Indonesia
elly@unhas.ac.id; pahrul19d@student.unhas.ac.id
*Corresponding author: amil@unhas.ac.id

Received June 2023; revised October 2023

Abstract. *Online loan applications have significantly facilitated the process of obtaining loans for individuals. However, like other financial services, online loans also come with risks. Some of the possible risks include irresponsible use, fraud, high fees, and getting stuck in a debt cycle. Therefore, users should understand the risks and consider them when making the decision to take an online loan from a company. A system that can be developed to provide recommendations for online loan applications is through sentiment analysis by utilizing reviews or ratings from the customers against the application. In this research, a sentiment analysis with the ensemble Soft Voting Classifier was developed and optimized using the Particle Swarm Optimization algorithm. The system implemented three individual classification algorithms, namely Random Forest, XGBoost, and Support Vector Machine, all of which were trained using the same dataset. The results show that the Soft Voting Classifier with optimized weights using the Particle Swarm Optimization algorithm has the highest classification accuracy rate compared to the other three individual classification models. The Soft Voting Classifier achieved an accuracy rate of 87.3%, while the Random Forest, XGBoost, and Support Vector Machine achieved accuracy rates of 80.9%, 80.7%, and 79.9%, respectively.*
Keywords: Online credit and loan applications, Random Forest, XGBoost, Support Vector Machine, Sentiment analysis, Soft Voting Classifier

1. **Introduction.** In today's era, smartphones facilitate access to a variety of services, including voice communications, data storage, wireless connectivity, and data storage [1]. One of the services that can be utilized via smartphone is online transaction services. The impact of fintech on society at large is significant, as it provides easier access to financial products, allowing transactions to be done practically and effectively through smartphones, e-Money, and even investments. Problems in transactions such as difficulty in finding items at shopping malls, difficulty in transferring funds to banks or ATMs, or reluctance to visit certain places due to unpleasant services, can be minimized with fintech so that everything can be done easily [2]. One of the financial technology or fintech products in Indonesia is the online loan in the form of an application. Online loans are a result of the progress in financial technology, which involves using technology in the financial system to create innovative products, business models, and services. This can impact the stability, efficiency, reliability, and security of the monetary and payment

DOI: 10.24507/ijicic.20.02.359

systems [3]. An online loan is a loan that can be applied for via a mobile app without requiring an in-person meeting [4].

The practice of online lending affects various aspects of the financial system, including the stability of monetary and financial systems, as well as the efficiency, smoothness, security, and reliability of payment systems [5]. The legal framework governing online loans is addressed by the Financial Services Authority Regulation No. 77/POJK.01/2016, which specifically pertains to information technology-based lending and borrowing services [6-9]. The use of online loan applications has several risks, such as irresponsible use, fraud, overpriced interest rates, and getting stuck in a cycle of debt. Therefore, it is important for users to understand these risks and conduct a risk assessment before deciding to take out an online loan. In Indonesia, there are several online applications that provide credit and loans that are well known to the public such as Akulaku, Kredivo, Indodana, and AdaKami. However, the many reviews or opinions that exist regarding these online loan applications make potential consumers feel hesitant and unsure in choosing which application is easy, safe, and convenient to use, as well as those that suit their needs and can provide satisfaction. Therefore, a system that is able to provide advice to the public about the right online loan application is needed. One system that can be developed is through sentiment analysis by utilizing reviews or community assessments of the online loan application. The main goal of sentiment analysis is to analyze thoughts, define feelings, and determine polarity [10]. The process of sentiment analysis involves analyzing a piece of text to determine whether its overall tone is positive, negative, or neutral. The main goal of this analysis is to understand people's attitudes and preferences, and how this information can help businesses to grow and expand [11,12]. Sentiment analysis utilizes various techniques such as Natural Language Processing (NLP), information retrieval, data mining, and knowledge management to extract subjective information from huge amounts of unstructured data. Its aim is to identify and analyze the sentiment or opinion expressed in the text [13]. Sentiment analysis can be performed using classification methods both supervised and unsupervised [14].

The researchers had utilized text mining techniques, including sentiment analysis, to analyze user opinions expressed through social media [15]. Sentiment analysis can be built using machine learning algorithms to classify text into positive, negative, or neutral sentiment categories based on the words and sentences used in the text. Özçift [16] built a sentiment analysis model using the ensemble Soft Voting Classifier method. Soft Voting Classifier was also used by Kibria et al. [17] and Manconi et al. [18] to conduct research with classification cases and managed to get good and accurate results. Another research was also conducted by Al Amrani et al. [19], compared to Random Forest and Support Vector Machine algorithms. Li et al. [20] also conducted sentiment analysis research using the XGBoost classification algorithm. Hayatin et al. [21] had conducted sentiment analysis research using the Naïve Bayes algorithm and the Particle Swarm Optimization (PSO) algorithm. They found that PSO had succeeded in increasing accuracy. Other researchers had also used PSO in sentiment analysis cases. In 2020, Machová et al. [22] conducted research with classification cases using PSO and also managed to get high accuracy.

Our research uses the PSO metaheuristic algorithm to optimize the weights of each algorithm used in Soft Voting Classifier, thereby significantly improving the accuracy of sentiment analysis. This study offers a valuable contribution to the field of text mining or NLP by presenting an approach that has been shown to improve the accuracy of Soft Voting Classifiers compared to the respective individual algorithms, making it an innovation worthy of attention in sentiment analysis. As outlined, Section 2 sets out the research method. This section includes details of the research flow, dataset used, and models used. Section 3 presents results and discussion, and Section 4 contains our conclusions.

2. Materials and Method.

This research aims to develop a sentiment analysis system using Soft Voting Classifier with three classification algorithms, namely Random Forest, XGBoost, and Support Vector Machine (SVM). To increase accuracy, the Soft Voting Classifier then uses the PSO algorithm to find optimal weights for each individual algorithm. This system was implemented and tested using a dataset of online credit and loan applications.

2.1. Research process.

Figure 1 shows the design of the research system from the data collection process to the model evaluation results.

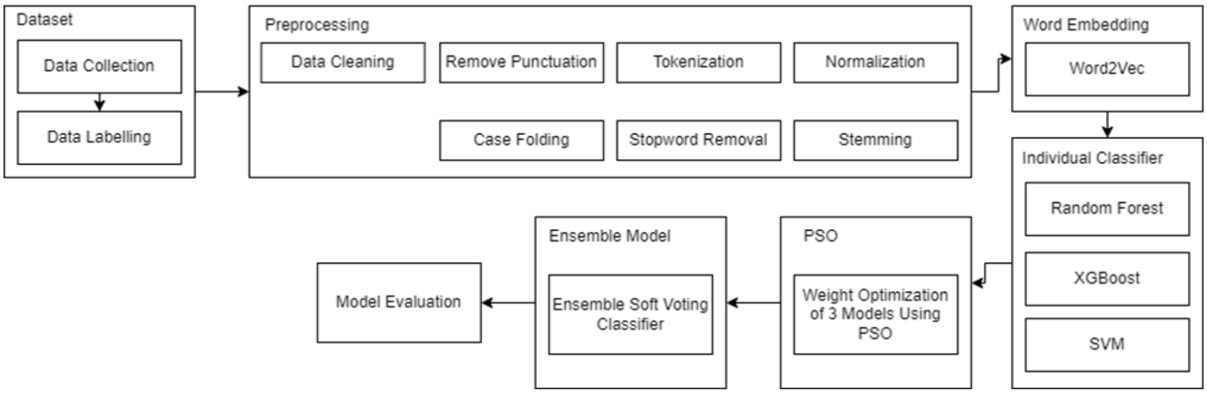


FIGURE 1. Research process

2.2. Data collections and data labelling.

During the data collection phase, the Google-play-scraper Python library is utilized to perform data scraping on user reviews and comments of online loan applications found on the Google Play Store. After the data has been collected, the next step is to manually label the data by giving a sentiment label to each review. Sentiment labels are created to classify each review into 3 categories, namely negative, neutral, and positive. Reviews that have negative sentiments will be labeled 1, reviews that have neutral sentiments will be labeled 0, and reviews that have positive sentiments will be labeled 2.

2.3. Data preprocessing.

Data preprocessing is a process to clean data from interference and transform raw data into a form that is ready for further processing. In general, data preprocessing aims to produce clean reviews that are then ready to be processed. Here are some of the preprocessing steps used in this research.

2.3.1. *Data cleaning.* The procedure of data cleaning involves the removal of erroneous, repetitive, incongruous, and inadequate data from a dataset. It is also aimed at eliminating data that can potentially create noise while performing analysis [23].

2.3.2. *Remove punctuation.* Removing symbols or punctuation marks such as periods, question marks and commas in text is known as remove punctuation [24].

2.3.3. *Case folding.* Case folding is the process of converting each letter character in a document to lowercase. The aim is to make the text data consistent during processing for the purpose of performing sentiment classification [25].

2.3.4. *Tokenization.* Tokenization refers to the process of splitting an input string into tokens [26]. Tokenizing is the process of splitting a sentence into words or tokens.

2.3.5. *Stopword removal.* The elimination of words that lack significance is referred to as stopword removal [27]. This research uses a dictionary that contains a list of stop-words in Bahasa such as the words “ke”, “lagi”, “yang” and many more.

2.3.6. *Normalization.* The process of normalization is employed to enhance the quality of words by converting non-standard words to standard words. In this study, a slang dictionary is utilized which includes non-standard words and their corresponding standard words. For instance, the non-standard word in Bahasa “udah” is replaced with the standard word “sudah”.

2.3.7. *Stemming.* The purpose of the stemming process is to decrease or eliminate prefixes and suffixes from a word [28].

2.4. **Word embedding (Word2Vec).** This research uses Word2Vec to generate word embedding with Continuous Bag-of-Words (CBOW) technique. CBOW utilizes neighboring words to predict the co-occurrence of each individual word, which in turn generates a word vector. These word vectors are then utilized as classification features [29-31].

2.5. **Classification model.** The individual classification models used in this research are Random Forest, XGBoost and SVM.

2.5.1. *Random Forest.* Random Forest is a type of supervised machine learning approach that can be applied to tasks involving both regression and classification [32]. Random Forest is an ensemble classifier consisting of several decision trees, where the output is determined by the class that is predicted most frequently by the individual trees [33,34].

2.5.2. *XGBoost.* XGBoost is a machine learning technique that uses an ensemble learning approach that is more complex than gradient boosting. XGBoost combines tree-based learning algorithms and solvers for linear models [35]. XGBoost can be preferred for two reasons, namely that it is faster in executing tasks and produces more accurate models [36].

2.5.3. *Support Vector Machine.* SVM is based on the concept of creating a hyperplane, which is also referred to as a decision boundary or optimal boundary, to maximize the distance between the nearest data points (support vectors) and effectively separate different classes [37]. SVM has been utilized effectively in various studies on text classification due to its major advantages, including its robustness in high-dimensional spaces, suitability of each function, strength in handling sparsely sampled datasets, and ability to solve most linear text classification problems [38].

2.6. **Optimized weight using PSO.** The inspiration for developing the PSO algorithm came from studying the behavior of birds flying in flocks and fish swimming in schools within their natural habitats [39]. In this algorithm, the objective is to optimize the problem by manipulating particles or potential solutions within the problem space using specific functions to determine the particles’ position and velocity. A particle’s movement is impacted by the best solution discovered by that particle, as well as the overall best solution generated by other particles [40].

The process of obtaining the optimal weights will be executed through a well-structured methodology depicted in Figure 2, which is a flowchart that describes the steps involved in the PSO algorithm.

The process begins by determining the fitness function, which serves as a measure of evaluating the quality of the solution in the context of the problem being optimized. After that, a population of particles is initialized, where each particle represents a potential solution. Evaluation is done by calculating the fitness function value for each particle.

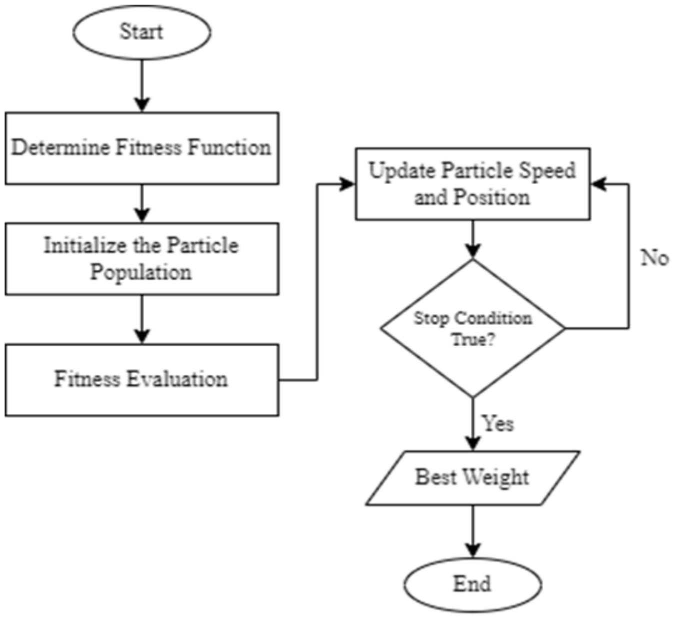


FIGURE 2. PSO flowchart for finding optimal weight

Next, the particles update their velocities and positions based on the influence of the best particle ever found by the particle itself (pbest) and the influence of the best particle ever found by the entire population (gbest). This process repeats over multiple iterations, with stopping criteria that can be determined based on the maximum number of iterations or the achievement of a desired result. Once the iterations are complete, the best weight that produces the highest fitness value is determined as the optimal solution found by the PSO algorithm. This weight is then implemented into the Soft Voting Classifier model.

2.7. **Soft Voting Classifier.** Soft Voting is an ensemble method that is considered more complex, because it takes account of the probability value of each prediction made by each model in the classification process [41]. The Soft Voting Classifier employs the mean probability of each class predicted by each model to determine the class with the highest average of prediction probabilities [42]. The Soft Voting technique allows models to harness the strengths and advantages of various models participating in ensemble learning, leading to more accurate and reliable predictions. Soft Voting is often used in cases where different classification models are employed to address complex classification problems. This study will utilize the class prediction probability of each model and the weight of each model to arrive at a final decision using the Soft Voting Classifier.

$$\hat{y} = \arg\max_i \sum_{j=1}^{m} w_j p_{ij} \tag{1}$$

In Equation (1), w refers to the weight generated by the PSO algorithm. PSO is used to calculate the optimal weights for each individual algorithm used in the classification process. These weights reflect the relative contribution of each algorithm in the final decision-making process. The weight represents the level of confidence or relative importance of each algorithm in the final decision. Additionally, p refers to the probability generated by each algorithm for each prediction class.

Figure 3 illustrates the working principle of the Soft Voting Classifier, which involves combining the probability outcomes from each algorithm for all classes, while considering the optimal weights calculated earlier through the PSO optimization algorithm. As such,

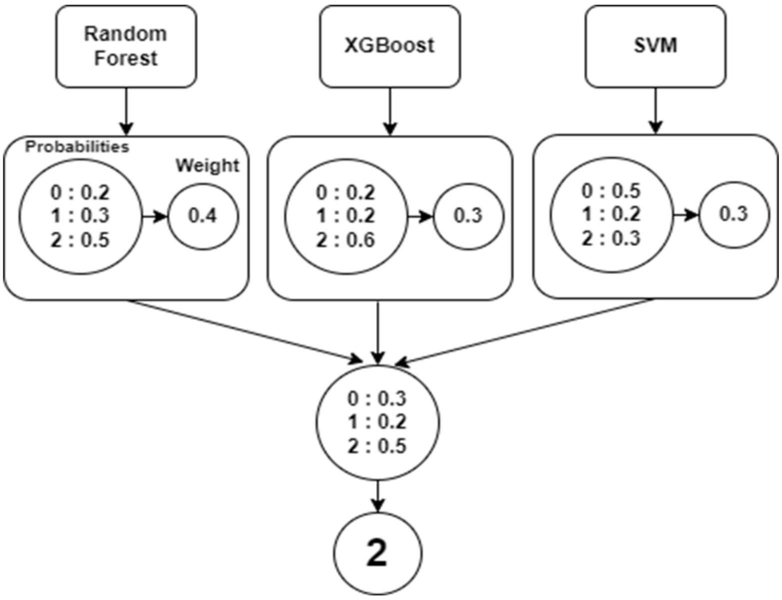


FIGURE 3. Soft Voting Classifier example

the model not only takes account of the probabilities from each algorithm but also considers the respective weights associated with the contributions of each algorithm in the final decision-making process. This enables the model to provide more accurate and reliable results by leveraging information from all the algorithms involved in the classification process.

2.8. **Evaluation method.** To evaluate the performance of the model, this study uses a confusion matrix to calculate Accuracy, Recall, Precision, and F1 Scores. The confusion matrix is a widely used method for evaluating the accuracy and effectiveness of a classification algorithm by determining its level of correctness. By analyzing the confusion matrix, the True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) predictions of the model can be determined, and these values can then be used to calculate the Accuracy, Recall, Precision, and F1 Score, which provides a comprehensive evaluation of model performance [43].

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \times 100 \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

3. **Result and Discussion.**

3.1. **Data collections and data labeling.** The dataset that was collected came from a number of user reviews of online credit and loan applications, namely the Kredivo, Akulaku, Indodana, and AdaKami. After duplicate data was removed, the remaining number of reviews was 17,640 reviews. Table 1 shows examples of collected reviews and their

Table 1. Examples of collected reviews and their manual labeling

Review (in Bahasa)	Label
Penagih tidak sopan	1
Aplikasi Busuk!!!	1
Pinjaman Uang Disini Mantap Bro, Gue Udah Setuju! Dan Sudah Cair Ke Rekening Mantap!!! Saya sangat puas dengan dengan adanya pinjol.	2
Saya menunggu di terima.	0
Sangat bagus aplikasi nya mudah dan cepat pinjam uang	2
Tadi saya baru install aplikasi ini	0
Mohon Bantuan Nya Untuk Menghapus Data Pribadi Saya Jika Masih Ada Di Sini Terima Kasih	0
Pinjaman Uang Sangat Ramah	2
Pengalaman saya di pinjol ini sangat puas langsung cair beberapa menit saja.	2

corresponding manual labels. When creating the dataset, each review was manually labeled "0", "1", or "2". Label "0" indicates neutral sentiment, label "1" indicates negative sentiment, and label "2" indicates positive sentiment.

3.2. **Data preprocessing.** The data preprocessing stage includes several steps, namely data cleaning, removing punctuation marks, case folding, tokenizing, stopword removal, normalization, and stemming.

Table 2 shows examples of reviews before and after undergoing each preprocessing stage. The first preprocessing is data cleaning. As can be seen in Table 2, "Review 1" and "Review 2" are duplicates so through the data cleaning process, one of them is deleted, so the duplicate review is only counted as one review. The next stage is removing punctuation where all punctuation marks in the review are removed, where in this example the punctuation marks ",", "!", and "." are removed. At the case folding stage, all the words in review are converted into lowercase letters, which are then separated into independent words through a tokenization process. Stopword removal is the process of removing conjunctions, where in this example the conjunctions in Bahasa such as "dan", "ke", and "dengan" are removed. The next stage is normalization, namely changing non-standard words into standard words, where in this example the words in Bahasa such as "gue" and "udah" are changed to standard form to become "saya" and "sudah", respectively. The final stage is stemming, namely removing prefixes and suffixes in a word, where in this example language words such as "pinjaman" are changed to "pinjam" and "adanya" becomes "ada".

3.3. **Word representation.** After undergoing a series of preprocessing steps, such as eliminating irrelevant data and obtaining clean text, the review data was prepared and poised for conversion into numerical form. For this purpose, we opted to utilize the Word2Vec technique, specifically the Continuous Bag-of-Words (CBOW) method. This method facilitates the numerical representation of words in vector form. Table 3 exemplifies the word vectorization outcomes attained through the Continuous Bag-of-Words technique.

In this study, 200 vectors were employed to portray each review, indicating that each review was transformed into a 200-dimensional vector. These vectors were generated by considering the context of the words and phrases within the reviews, enabling the representation of the semantic meaning of the words.

TABLE 2. Examples of preprocessing reviews

Review (in Bahasa)	
Review 1	Pinjaman Uang Disini Mantap Bro, Gue Udah Setuju! Dan Sudah Cair Ke Rekening Mantap!!! Saya sangat puas dengan dengan adanya pinjol.
Review 2 (duplicate)	Pinjaman Uang Disini Mantap Bro, Gue Udah Setuju! Dan Sudah Cair Ke Rekening Mantap!!! Saya sangat puas dengan dengan adanya pinjol.
Preprocessing stages	
Data cleaning	Pinjaman Uang Disini Mantap Bro, Gue Udah Setuju! Dan Sudah Cair Ke Rekening Mantap!!! Saya sangat puas dengan dengan adanya pinjol.
Remove punctuation	Pinjaman Uang Disini Mantap Bro Gue Udah Setuju Dan Sudah Cair Ke Rekening Mantap Saya sangat puas dengan dengan adanya pinjol
Case folding	pinjaman uang disini mantap bro gue udah setuju dan sudah cair ke rekening mantap saya sangat puas dengan dengan adanya pinjol
Tokenization	"pinjaman", "uang", "disini", "mantap", "bro", "gue", "udah", "setuju", "dan", "sudah", "cair", "ke", "rekening", "mantap", "saya", "sangat", "puas", "dengan", "adanya", "pinjol"
Stopword removal	"pinjaman", "uang", "disini", "mantap", "bro", "gue", "udah", "setuju", "sudah", "cair", "rekening", "mantap", "saya", "sangat", "puas", "adanya", "pinjol"
Normalization	"pinjaman", "uang", "disini", "mantap", "bro", "saya", "sudah", "setuju", "sudah", "cair", "rekening", "mantap", "saya", "sangat", "puas", "adanya", "pinjol"
Stemming	"pinjam", "uang", "disini", "mantap", "bro", "saya", "sudah", "setuju", "sudah", "cair", "rekening", "mantap", "saya", "sangat", "puas", "ada", "pinjol"

TABLE 3. Examples of numeric representation using Word2Vec

Clean review (in Bahasa)	dim0	dim1	dim2	dim3	dim4	...	dim199
mantap	−1.1360	0.1874	0.9969	−1.7700	1.0744	...	−0.9382
sangat bagus dan bermanfaat	−0.4157	−0.3638	−1.5438	1.0637	0.4084	...	−0.8425
aplikasi busuk sangat meresahkan	−0.1388	1.8679	0.2977	−1.1115	−0.5153	...	2.2784
tolong hapus data saya	−0.4931	0.5117	−1.6121	0.4395	−0.1547	...	0.3575
buat orang sengsara pinjaman online	1.1336	0.3137	−0.1239	0.3524	−0.2673	...	−1.2793
pinjaman uang sangat ramah	−2.9299	−1.5099	−1.3463	1.9393	−0.2241	...	−1.3378
hadeh aplikasi apan sih ga jelas	0.7595	0.6279	−0.2691	−0.3750	0.6832	...	0.0572

3.4. **Model evaluation.** After completing the process of converting words into vectors, the next step is to divide the dataset into two important parts in developing an evaluation of the sentiment analysis model. The dataset is dinamically splitted into training data and testing data to ensure the accurate testing and evaluation of the model. This dataset is divided using a standard ratio, where 80% of the data is used for training, leaving the remaining 20% for testing.

The first evaluation carried out was to measure the level of classification accuracy individually for each model used, namely Random Forest, XGBoost, and SVM, using the same training data and testing data. Next, the level of classification accuracy was measured on the Soft Voting Classifier which was built from these three models and optimized using the PSO algorithm. PSO is used to determine the optimal weight for each individual model, which is then taken into consideration by the Soft Voting Classifier in determining the analysis sentiment classification. Table 4 shows the optimal weights produced by PSO for each model.

Table 4. The optimal weight generated by PSO

Model	Optimal weight
Random Forest	0.39957691
XGBoost	0.3892255
SVM	0.21119759

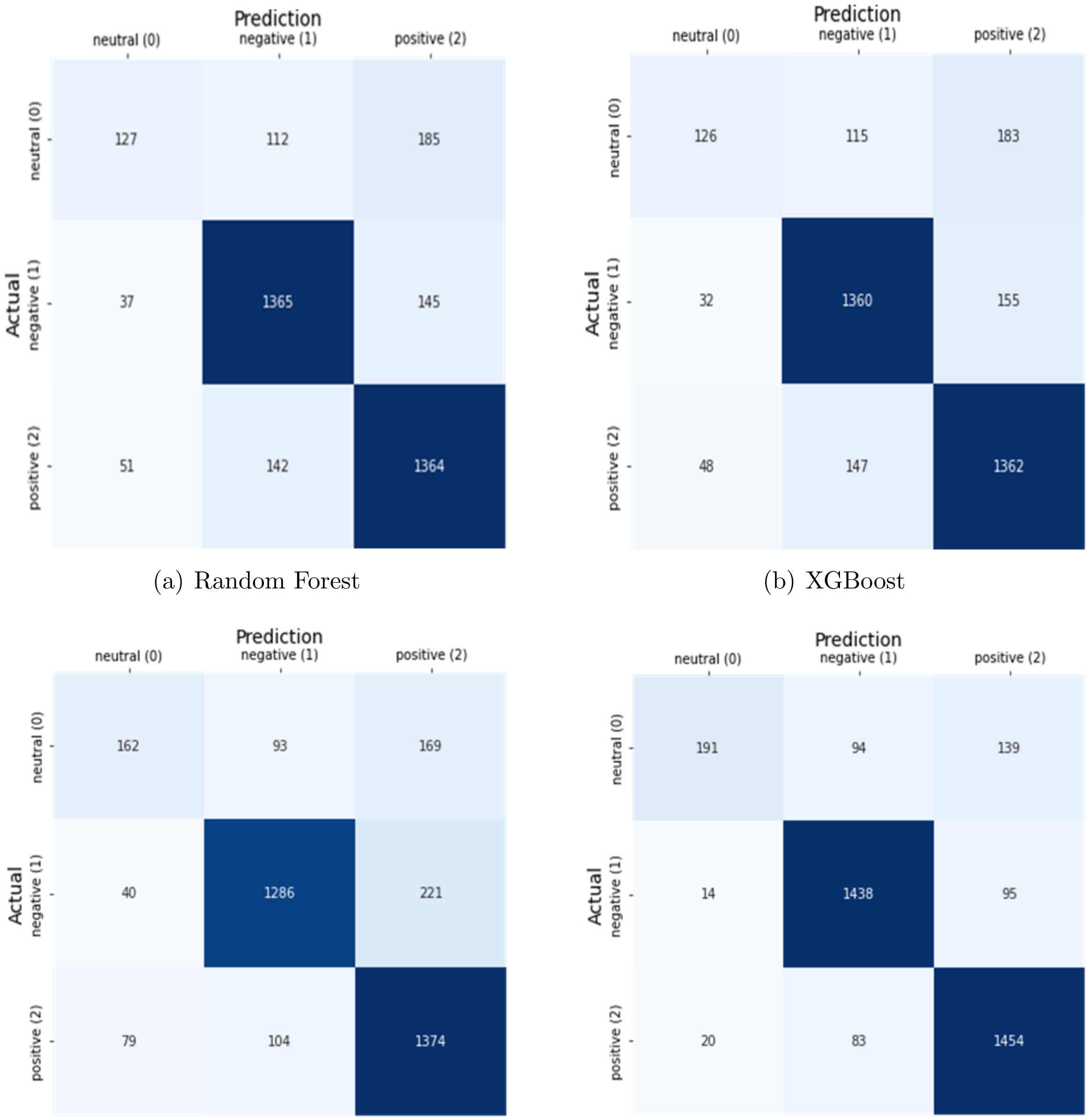
The weight of each model reflects the extent to which each model is considered reliable or has better performance in correctly predicting sentiment analysis classes. In the Soft Voting Classifier, the weight function is to provide a greater influence on the predictions of a model that is considered more reliable or has better performance, while also providing a smaller influence on models that are less reliable. By considering these weights, the Soft Voting Classifier can combine predictions from multiple models in a more adaptive way and can improve the overall accuracy and performance of the ensemble.

This research uses a confusion matrix table to assess the performance of each individual model and the Soft Voting Classifier, as shown in Figure 4.

The confusion matrix is used to calculate performance metrics, namely Accuracy, Precision, Recall, and F1 Score for each model, providing an evaluation of its overall performance. Figure 4 shows that the largest number of TPs was achieved by the Soft Voting Classifier, namely 3,083 TPs, followed by the Random Forest model with 2,856 TPs, the XGBoost model with 2,848 TPs, and the SVM model with 2,822 TPs. TP indicates the match between predicted and actual class. For multiclass classification, the TN value is equal to zero, so a high TP value indicates a high classification accuracy value (Equation (2)).

Another important classification metric is the F1 Score (Equation (5)), which is calculated based on the Recall (Equation (3)) and Precision values (Equation (4)). As seen in Figure 4, the Random Forest model produces 88 FNs and 297 FPs in the neutral class, 254 FNs and 182 FPs in the negative class, 330 FNs and 193 FPs in the positive class. The XGBoost model produces 80 FNs and 298 FPs in the neutral class, 262 FNs and 187 FPs in the negative class, 338 FNs and 195 FPs in the positive class. The SVM model produces 119 FNs and 262 FPs in the neutral class, 197 FNs and 261 FPs in the negative class, 390 FNs and 183 FPs in the positive class. The Soft Voting Classifier produces 34 FNs and 233 FPs in the neutral class, 177 FNs and 109 FPs in the negative class, 234 FNs and 103 FPs in the positive class. FN in the neutral class indicates a mismatch between the predicted class and the actual class, where the predicted class is a neutral class, but its actual class is not a neutral class. FP in the neutral class indicates a mismatch between the predicted class and the actual class, where the predicted class is the other class, but its actual class is a neutral class. This definition of FN and FP also applies to other classes, namely the negative class and the positive class.

As shown in Figure 5, Random Forest achieved Precision of 79.6%, Recall of 69%, F1 Score of 79.6%, and Accuracy of 80.9%. XGBoost achieved Precision of 79.4%, Recall of



(a) Random Forest (b) XGBoost

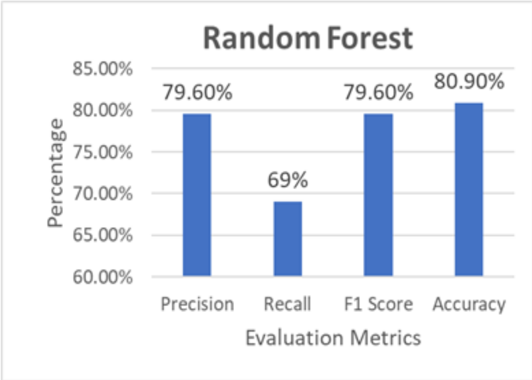
(c) SVM (d) Soft Voting Classifier

Figure 4. Confusion matrix

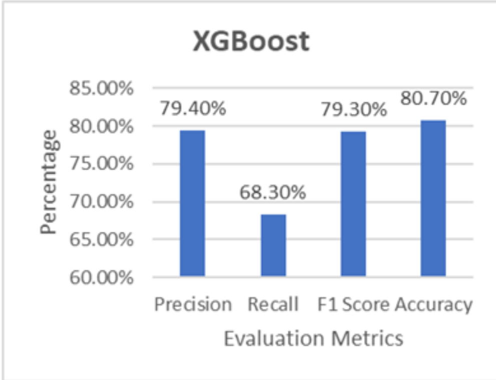
68.3%, F1 Score of 79.3%, and Accuracy of 80.7%. SVM achieved Precision of 79.3%, Recall of 70%, F1 Score of 79.2%, and Accuracy of 79.9%. Soft Voting Classifier achieved Precision of 87.2%, Recall of 77%, F1 Score of 86.5%, and Accuracy of 87.3%.

Of the four models, the highest accuracy value was obtained by the Soft Voting Classifier, namely 87.3%. This high accuracy value corresponds to the number of TPs obtained by the Soft Voting Classifier, namely 3,083 TPs, which is the highest among other models. The Soft Voting Classifier's F1 Score value also shows the highest among other models, namely 86.5%. This shows that the Soft Voting Classifier is able to classify reviews through sentiment analysis more precisely, compared to the other three individual models.

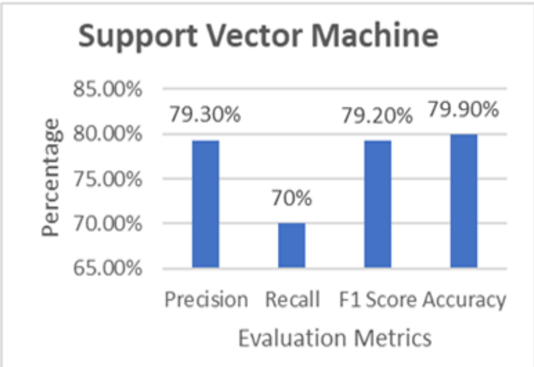
Figure 6 provides empirical evidence regarding the performance of classification models in sentiment analysis in the context of online credit and loan applications. These results provide empirical evidence regarding the ability of the Soft Voting Classifier model to recognize and classify sentiment with a high level of accuracy. The research results show



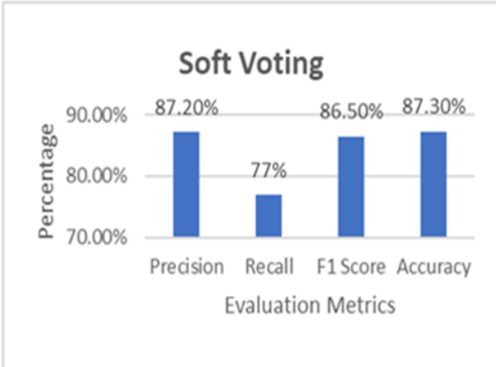
(a) Random Forest



(b) XGBoost



(c) SVM



(d) Soft Voting Classifier

FIGURE 5. Performance of individual model and Soft Voting Classifier

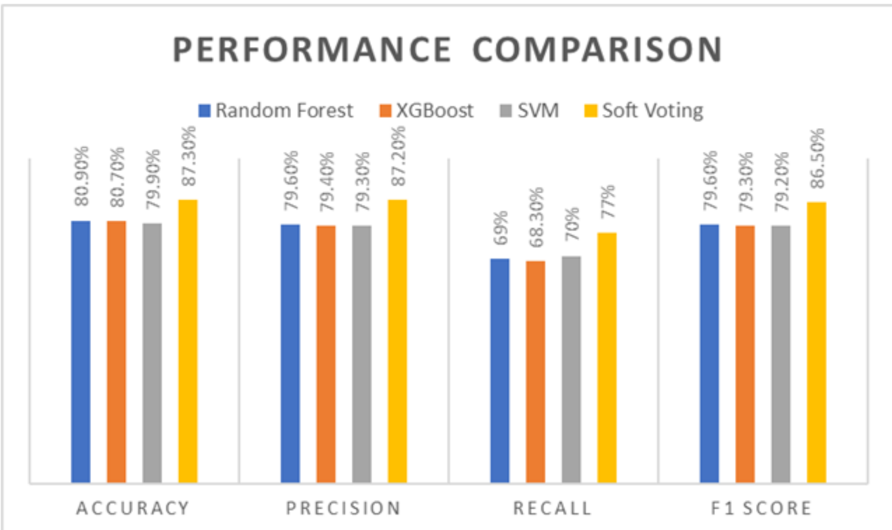


FIGURE 6. Model performance comparison

that the Soft Voting Classifier stands out as the model with the most superior performance when compared to the three other individual classification models used in this research. Although other models, namely Random Forest, XGBoost, and SVM also show quite good performance in sentiment analysis, Soft Voting Classifier consistently shows superiority in various aspects of performance.

4. **Conclusions.** This research aims to develop the Soft Voting Classifier as an ensemble model in sentiment analysis on online credit and loan application datasets. The PSO algorithm is implemented to optimize and generate optimal weights for the classifier. The classifier is composed of three individual classification models: Random Forest, XGBoost, and SVM. The results showed that the optimized Soft Voting Classifier improved the classification accuracy rate compared to the other three individual classification algorithms. The Soft Voting Classifier achieved an accuracy rate of 87.3%, while the Random Forest, XGBoost, and SVM achieved accuracy rates of 80.9%, 80.7%, and 79.9%, respectively.

For further research, it is recommended to try other metaheuristic algorithms to find the optimal weight of each individual classification algorithm used in the Soft Voting Classifier such as the firefly, bee and ant algorithms to increase the accuracy of sentiment analysis results.

Acknowledgment. The researchers express gratitude to research members in Information System and Cloud Computing Laboratory, the Department of Informatics, Faculty of Engineering, Universitas Hasanuddin.

REFERENCES

[1] O. AbuAlghanam, L. Albdour and O. Adwan, Multimodal biometric fusion online handwritten signature verification using neural network and support vector machine, *International Journal of Innovative Computing, Information and Control*, vol.17, no.5, pp.1691-1703, 2021.

[2] I. Kamil, D. Anggraini and H. Prihanto, Socialization of financial technology (fintech): Smart in choosing and deciding the use of online fund loan applications, *Jurnal Pustaka Dianmas*, vol.2, no.2, pp.1-6, 2022.

[3] Jasman, N. Y. Arifin, A. L. Setyabudhi and O. Veza, Online loans during the COVID-19 pandemic for the batam community, *International Journal Economic and Business Management (EABMIJ)*, vol.4, no.2, pp.119-124, 2022.

[4] D. T. Subagiyo, L. R. Gestor and S. Sulistiyo, Characteristic of illegal online loans in Indonesia, *Indonesia Private Law Review*, vol.3, pp.69-84, 2022.

[5] W. Arista and R. Rusmini, Law enforcement of online loan application services during the pandemic period in Palembang City as an effort to protect consumers, *Lex Librum: Jurnal Ilmu Hukum*, vol.9, no.1, pp.67-74, 2022.

[6] I. P. A. S. Sinaga, The legal protection for customers using online loan services, *Jurnal Daulat Hukum*, vol.5, pp.211-226, 2022.

[7] P. R. Hidayat, Legal protection for users of illegal online loan services, *Jurnal Justiciabelen*, vol.5, pp.16-25, 2022.

[8] I. K. B. Kurnianta, I. M. I. Sastrawan and I. B. G. A. Mahaputra, Legal protection of borrower's personal data in online loan application services, *ICBLT 2022*, pp.497-504, 2023.

[9] I. E. Khaq and A. Hidayat, The law enforcement against an illegal online loans platform, *Journal of Law Theory and Law Enforcement*, vol.1, pp.85-98, 2022.

[10] V. Sukhavasi, A comprehensive survey of sentiment analysis: Techniques, applications and open problems, *International Conference on Technological Emerging Challenges in Computer and Engineering*, Department of CSE, Tirumala Engineering College, pp.273-284, 2021.

[11] N. V. Babu and E. G. M. Kanaga, Sentiment analysis in social media data for depression detection using artificial intelligence: A review, *SN Computer Science*, vol.3, 74, 2021.

[12] X. Wang, Y. Sheng, H. Deng and Z. Zhao, CharCNN-SVM for Chinese text datasets sentiment classification with data augmentation, *International Journal of Innovative Computing, Information and Control*, vol.15, no.1, pp.227-246, 2019.

[13] A. Alsayat, Improving sentiment analysis for social media applications using an ensemble deep learning language model, *Arabian Journal for Science and Engineering*, vol.47, pp.2499-2511, 2022.

[14] M. S. Vohra and J. Teraiya, Applications and challenges for sentiment analysis: A survey, *International Journal of Engineering Research*, vol.2, 2013.

[15] B. Andrian, T. Simanungkalit, I. Budi and A. F. Wicaksono, Sentiment analysis on customer satisfaction of digital banking in Indonesia, *International Journal of Advanced Computer Science and Applications*, vol.13, no.3, 2022.

[16] A. Özçift, Medical sentiment analysis based on soft voting ensemble algorithm, *Yönetim Bilioim Sistemleri Dergisi*, vol.6, no.1, pp.42-50, 2020.

[17] H. B. Kibria, M. Nahiduzzaman, M. O. F. Goni, M. Ahsan and J. Haider, An ensemble approach for the prediction of diabetes mellitus using a soft voting classifier with an explainable AI, *Sensors*, vol.22, 7268, 2022.

[18] A. Manconi, G. Armano, M. Gnocchi and L. Milanesi, A soft-voting ensemble classifier for detecting patients affected by COVID-19, *Applied Sciences*, vol.12, 7554, 2022.

[19] Y. Al Amrani, M. Lazaar and K. E. El Kadiri, Random forest and support vector machine based hybrid approach to sentiment analysis, *Procedia Computer Science*, vol.127, pp.511-520, 2018.

[20] Z. Li, Q. Zhang, Y. Wang and S. Wang, Social media rumor refuter feature analysis and crowd identification based on XGBoost and NLP, *Applied Sciences*, vol.10, 4711, 2020.

[21] N. Hayatin, G. I. Marthasari and L. Nuarini, Optimization of sentiment analysis for Indonesian presidential election using Naïve Bayes and particle swarm optimization, *Jurnal Online Informatika*, vol.5, no.1, pp.81-88, 2020.

[22] K. Machová, M. Mikula, X. Gao and M. Mach, Lexicon-based sentiment analysis using the particle swarm optimization, *Electronics*, vol.9, 1317, 2020.

[23] H. T. Ismet, T. Mustaqim and D. Purwitasari, Aspect based sentiment analysis of product review using memory network, *Scientific Journal of Informatics*, vol.9, no.1, pp.73-83, 2020.

[24] R. Meifitrah, I. Darmawan and O. N. Pratiwi, Sentiment analysis of tokopedia application review to service product recommender system using neural collaborative filtering for marketplace in Indonesia, *IOP Conference Series: Materials Science and Engineering*, vol.909, 012071, 2020.

[25] M. A. Nurrohmat and A. SN, Sentiment analysis of novel review using long short-term memory method, *Indonesian Journal of Computing and Cybernetics Systems*, vol.13, no.3, pp.209-218, 2019.

[26] U. Lestari, T. Romadhani, S. Suraya and E. Fatkhiyah, Sentiment analysis for extracting student opinion data on higher education services using the Naïve Bayes classifier and support vector machine methods (case study Akprind Institute of Science and Technology Yogyakarta), *Jurnal TAM (Technology Acceptance Model)*, vol.13, no.1, pp.51-56, 2022.

[27] A. Pradana and M. Hayati, The effect of stemming and removal of stopwords on the accuracy of sentiment analysis on Indonesian-language text, *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, vol.4, no.4, pp.375-380, 2019.

[28] A. F. Hidayatullah, The influence of stemming on Indonesian tweet sentiment analysis, *Proc. of International Conference on Electrical Engineering Computer Science and Informatics*, 2015.

[29] M. A. Fauzi, Word2Vec model for sentiment analysis of product reviews in Indonesian language, *International Journal of Electrical and Computer Engineering (IJECE)*, vol.9, no.1, pp.525-530, 2019.

[30] A. Iriani, Hendry, D. H. F. Manongga and R.-C. Chen, Mining public opinion on radicalism in social media via sentiment analysis, *International Journal of Innovative Computing, Information and Control*, vol.16, no.5, pp.1787-1800, 2020.

[31] T. Mikolov, K. Chen, G. Corrado and J. Dean, Efficient estimation of word representations in vector space, *arXiv Preprint*, arXiv: 1301.3781, 2013.

[32] A. P. Rodrigues, R. Fernandes, A. Aakash, B. Abhishek, A. Shetty, K. Atul, K. Lakshmanna and R. M. Shafi, Real-time Twitter spam detection and sentiment analysis using machine learning and deep learning techniques, *Computational Intelligence and Neuroscience*, e5211949, 2022.

[33] J. Asian, M. D. Rosita and T. Mantoro, Sentiment analysis for the Brazilian anesthesiologist using multi-layer perceptron classifier and random forest methods, *Jurnal Online Informatika*, vol.7, no.1, pp.132-141, 2022.

[34] A. Özçift, Random forests ensemble classifier trained with data resampling strategy to improve cardiac arrhythmia diagnosis, *Computers in Biology and Medicine*, vol.41, pp.265-271, 2011.

[35] A. Samih, A. Ghadi and A. Fennan, Enhanced sentiment analysis based on improved word embeddings and XGboost, *International Journal of Electrical and Computer Engineering*, vol.13, no.2, pp.1827-1836, 2023.

[36] K. Shakhovska, N. Shakhovska and P. Veselý, The sentiment analysis model of services providers' feedback, *Electronics*, vol.9, 1922, 2020.

[37] T. Adugna, W. Xu and J. Fan, Comparison of random forest and support vector machine classifiers for regional land cover mapping using coarse resolution FY-3C images, *Remote Sensing*, vol.14, 574, 2022.

[38] D. A. Kristiyanti, D. A. Putri, E. Indrayuni, A. Nurhadi and A. H. Umam, E-wallet sentiment analysis using Naïve Bayes and support vector machine algorithm, *Journal of Physics: Conference Series*, vol.1641, 012079, 2020.
[39] T. M. Shami, A. A. El-Saleh, M. Alswaitti, Q. Al-Tashi, M. A. Summakieh and S. Mirjalili, Particle swarm optimization: A comprehensive survey, *IEEE Access*, vol.10, pp.10031-10061, 2022.
[40] M. Noor, W. Gata, Risnandar, Fakhrudin and A. Novitarani, Optimization of sentiment analysis of Program Sembako (BPNT) based on Twitter, *Journal of Applied Engineering and Technological Science (JAETS)*, vol.4, pp.223-234, 2022.
[41] N. Peppes, E. Daskalakis, T. Alexakis, E. damopoulou and K. Demestichas, Performance of machine learning-based multi-model voting ensemble methods for network threat detection in Agriculture 4.0, *Sensors (Basel, Switzerland)*, vol.21, 7475, 2021.
[42] G. P. de. Oliveira, A. Fonseca and P. C. Rodrigues, Diabetes diagnosis based on hard and soft voting classifiers combining statistical learning models, *Brazilian Journal of Biometrics*, vol.40, pp.415-427, 2022.
[43] N. T. S. Saptadi, A. Suyuti, A. A. Ilham and I. Nurtanio, Modeling of organic waste classification as raw materials for briquettes using machine learning approach, *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol.14, no.3, pp.577-585, 2023.

Author Biography

Amil Ahmad Ilham obtained his Bachelor Degree of Electrical Engineering from Universitas Hasanuddin, Indonesia in 1997; Master of Information Technology from The University of Newcastle, Australia in 2003 and Doctor of Informatics from Kyushu University, Japan in 2011.

Dr. Ilham is currently an Associate Professor at Department of Informatics, Faculty of Engineering, Universitas Hasanuddin, Indonesia. His research interests include information systems, data analytics and cloud computing. He is an active member of IEEE since 2016. He is the Head of Information System and Cloud Computing Laboratory, Department of Informatics, Faculty of Engineering, Universitas Hasanuddin.

Elly Warni received her Bachelor Degree of Electrical Engineering from Universitas Hasanuddin, Indonesia in 2003 and Master of Engineering in Informatics from Universitas Hasanuddin, Indonesia in 2008.

Ms. Warni is currently an Assistant Professor at Department of Informatics, Faculty of Engineering, Universitas Hasanuddin, Indonesia. Her research interests include data mining, computer vision and knowledge discovery. She is a research member in Information System and Cloud Computing Laboratory, Department of Informatics, Faculty of Engineering, Universitas Hasanuddin.

Pahrul accomplished his Bachelor Degree of Informatics at Department of Informatics, Faculty of Engineering, Universitas Hasanuddin, Indonesia in 2023.

Mr. Pahrul is currently a Research Assistant at Department of Informatics, Faculty of Engineering, Universitas Hasanuddin. His research interests include data analytics and machine learning. He is a research member in Information System and Cloud Computing Laboratory, Department of Informatics, Faculty of Engineering, Universitas Hasanuddin.