JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i2.5640

CONTENT-BASED FILTERING CULINARY RECOMMENDATION SYSTEM USING DEEP CONVOLUTIONAL NEURAL NETWORK ON TWITTER (X)

Zahwa Dewi Artika[1*]; Erwin Budi Setiawan[2]

Informatics[1,2]
Telkom University, Bandung, Indonesia[1,2]
https://telkomuniversity.ac.id/[1,2]
wawalesmana@student.telkomuniversity.ac.id[1*], erwinbudisetiawan@telkomuniversity.ac.id[2]

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract*—Along with the development of technology, social media has become integral to everyday life, especially for sharing content like culinary reviews. Social media platform X (formerly Twitter) is often used for sharing culinary recommendations, but the abundance of information makes it difficult for users to find relevant suggestions. In order to improve rating prediction performance, this study suggests a recommendation system model that is more thoroughly created utilizing Content-Based Filtering (CBF) combined with Deep Convolutional Neural Network (CNN) and optimised with Particle Swarm Optimization (PSO). Data was collected from PergiKuliner and Twitter, totaling 2644 reviews and 200 cuisines. The preprocessing involved text processing, translation, and polarity assessment. Post-labeling, 7438 data were labeled with 0 and 1562 with 1. Label 0 means not recommended while label 1 means recommended. The imbalance is handled by applying the SMOTE method after observing that the fraction of data labeled 0 and 1 is 65.2%. CBF employed TF-IDF feature extraction and FastText word embedding, while Deep CNN handled classification. PSO optimisation was applied to enhance the accuracy of culinary rating predictions. The results showed an initial accuracy of 76.32% with the baseline Deep CNN model, which increased to 86.06% after Nadam optimisation with the best learning rate, and further reached 86.18% after PSO optimisation on dense units. The 9.86% accuracy improvement from the baseline model demonstrates the effectiveness of the combined methods.*

Keywords*: content-based filtering, deep CNN, fasttext, particle swarm optimization, TF-IDF.*

Intisari—*Seiring dengan perkembangan teknologi, media sosial telah menjadi bagian tak terpisahkan dari kehidupan sehari-hari. Pengguna media sosial kini memiliki akses mudah untuk berbagi konten, termasuk ulasan kuliner, yang menjadi topik populer karena terkait dengan kebutuhan pokok manusia, yaitu pangan. Platform media sosial X (sebelumnya Twitter) adalah salah satu media yang digunakan untuk berbagi rekomendasi kuliner. Namun, banyaknya informasi membuat pengguna kesulitan menemukan rekomendasi yang sesuai dengan preferensi mereka. Untuk meningkatkan performa prediksi rating, penelitian ini menyarankan sebuah model sistem rekomendasi yang dikembangkan lebih lengkap dengan menggunakan Content-Based Filtering (CBF) yang dikombinasikan dengan Deep Convolutional Neural Network (CNN) dan dioptimasi dengan Particle Swarm Optimization (PSO). Data dikumpulkan dari situs PergiKuliner dan Twitter, dengan total 2644 review dan 200 kuliner. Proses preprocessing meliputi text processing, translasi, dan penilaian polaritas. Setelah melalui tahap labelling, didapatkan 7438 data bernilai 0 dan 1562 data bernilai 1. Label 0 berarti tidak direkomendasikan sedangkan label 1 berarti direkomendasikan. Melihat rasio perbandingan antara jumlah data berlabel 0 dan 1 yang memiliki persentase 65.2%, maka diterapkan metode SMOTE untuk menangani ketidakseimbangan tersebut. CBF menggunakan ekstraksi fitur TF-IDF dan word embedding FastText, sementara Deep CNN digunakan untuk klasifikasi. Optimasi PSO diterapkan untuk meningkatkan akurasi prediksi rating kuliner. Hasil penelitian menunjukkan akurasi 76.32% dari model Deep CNN baseline, yang meningkat menjadi 86.06% setelah optimasi Nadam dengan learning rate terbaik, dan* *mencapai 86.18% setelah optimasi PSO pada dense units. Peningkatan akurasi sebesar 9.86% dari model baseline menunjukkan efektivitas kombinasi metode yang digunakan.*

***Kata Kunci**: pemfilteran berbasis konten, deep CNN, fasttext, particle swarm optimization, TF-IDF.*

INTRODUCTION

In this contemporary era of rapid technological development, social media has become an indispensable component of the everyday life of individuals worldwide. Social media has facilitated more accessible access to various forms of content, one related to gastronomy or culinary arts. The study of culinary arts will continue to be relevant in our modern society due to its intrinsic link with the basic biological needs of humans. The act of uploading photos and reviews of food and beverages consumed by social media users, as well as the indirect inspiration of their audience to try these culinary delights, exemplifies the profound impact of social media on the culinary arts. One of the social media platforms that has been utilized in this regard is X, which was previously known as Twitter.

X is an online social networking platform launched in July 2006. This social media post is made by tweeting and can contain up to 140 characters [1]. Approximately 316 million active users on platform X can generate an estimated 500 million tweets daily [2]. This situation may result in users' difficulties assimilating and managing extensive volumes of information and maintaining interactions with a large number of individuals. In the context of culinary-related topics, the impact can also lead to the difficulty of users finding culinary recommendations that suit their preferences and tastes from the many information intensities that exist. Based on these problems, a recommendation system is needed to help users find culinary items that fit their preferences.

A recommendation system is an system that evaluates user conduct or viewpoints, after which it processes the data to offer users recommendations based on their preferences [3]. Recommendation systems have been used in various works of literature. Due to the basic working principles of recommendation systems that depend on data, some previous studies have utilized data from platform X to improve the performance of recommendation systems [4]. One widely used method in recommendation systems is content-based filtering (CBF). CBF algorithms provide recommendations to users based on item descriptions and users' personal preferences. Furthermore, CBF algorithms only use ratings or profile information for active users, which makes them effective in situations when there is a dearth of user data but an abundance of information on the things [5].

In the study of Ortakci, O.U.Y. and Albayati, A.N.K. [4], machine learning-based NLP techniques are blended with CBF. The accuracy value of the developed recommendation system is 0.8624, or 86.24%. These findings demonstrate the possibility of developing recommendation outcomes from the system created by attempting to use deep learning-based techniques, which are actually more capable than machine learning.

CBF has been combined with several deep learning-based classification methods, including deep convolutional neural networks (CNN) [6]. By using fewer parameters than other deep learning models, CNN models can classify text and achieve good prediction accuracy using less computational power. [7]. The architecture of deep CNN [8] itself consists of a series of layers, including convolutional, pooling, and fully connected layers. Each layer extracts unique features, which undergo sub-sampling before being classified [9].

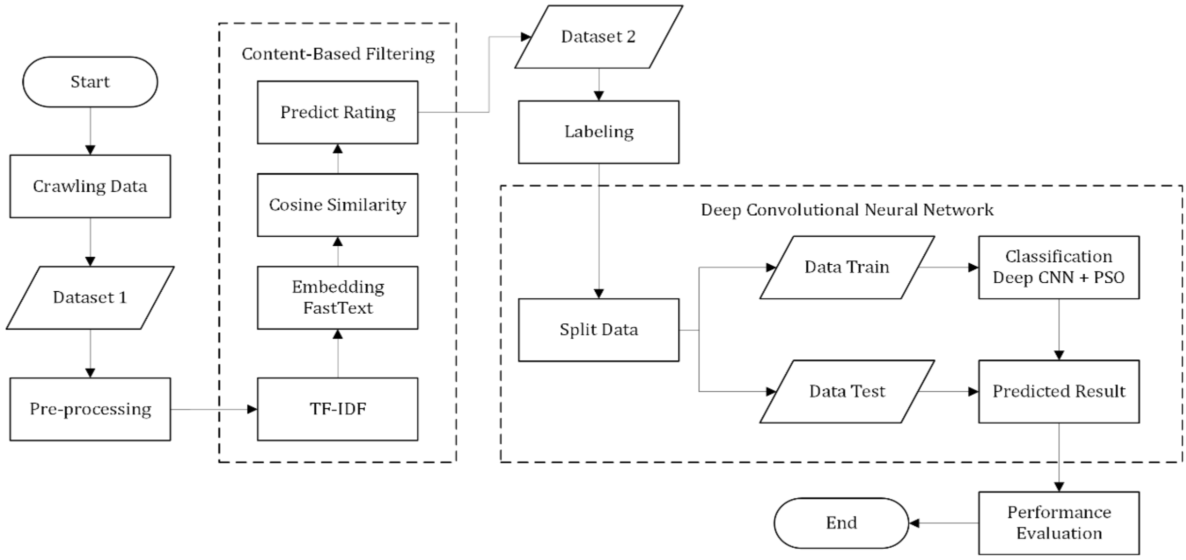
Besides being combined with classification methods, some studies also add optimization methods for better accuracy results. One example is the particle swarm optimization (PSO) method [10], [11]. PSO is an optimization method that mimics the behavior of a swarm or group of particles to find the best solution in a search space [12]. The addition of this optimization method succeeded in providing a higher fitness value than the comparative recommendation results [11]. Combining this method with classification algorithms, such as K-Means, has also resulted in better accuracy [10].

This research is based on various references from previous studies on recommendation systems. In the research by Suvarna, B. and Balakrishna, S. [6], deep CNN was used for a product recommendation, and its performance was evaluated using a fashion product dataset. The proposed model has results that outperform existing models using deep neural networks in terms of accuracy and precision values. The accuracy value generated from this model is 89.02%. However, this research has not applied the optimization method.

Another study by Aguerchi, K., Jabrane, Y., Habba, M., & El Hassani, A. H. [13] employed the Particle Swarm Optimization (PSO) technique to optimize hyperparameters within CNN

VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i2.5640

architectures for mammography image classification. The proposed PSO-CNN model achieved substantial performance gains, recording accuracy rates of 98.23% and 97.98% on the DDSM and MIAS datasets, respectively. Compared to traditional optimization approaches, the findings indicate that PSO-CNN presents a robust solution for CNN architecture optimization by significantly reducing the manual configuration required while enhancing model accuracy.



Source: (Research Results, 2024)

Figure 1. Flowchart Design System

This research proposes a recommendation system model that is more fully developed using CBF and deep CNN classification and optimized using the PSO method to obtain a better performance value in rating prediction. To the best of my knowledge, no research has used a combination of methods similar to the one that will be used in this research. In the process, feature extraction and word embedding are also utilized to improve the results.

MATERIALS AND METHODS

This research uses the overall system illustrated in Figure 1. The designed recommendation system implements the processes of data collection, preprocessing, content-based filtering, deep CNN classification, PSO optimization, and performance evaluation. The research process was conducted using Google Colab tools with the Python programming language.

1. Crawling Data

Two types of datasets were used in this research. The first dataset is data containing a list of culinary in the Bandung area from the PergiKuliner website, which contains information such as culinary places, culinary types, addresses, price ranges, ratings, and descriptions whose brief descriptions can be seen in Table 2. The total amount of data in this dataset is 200 culinary data. The second dataset contains reviews of culinary places obtained from Twitter or X. The dataset crawling process from X uses the Application Program Interface (API). The dataset obtained contains user usernames, culinary place names, and tweets containing culinary review text, as seen in Table 1. The amount of data in this dataset is 2644 review data.

Table 1. Culinary Reviews Dataset

culinary name	text	...	username
de.u Coffee	@juaramageran Bandung has many. I'll give you this...	...	aarrddyyee_95
de.u Coffee	@moejaaaa What kind of coffee shop are you looking for? Y...	...	aarrddyyee_95
...	...	...	...
Kartika Sari	?? Want to ask sorry because I'm not from Bandung, kal...	...	tanyakanrl
Golden Lamian	?? guys do you prefer marugame udon or ...	...	tanyakanrl

Source: (Research Results, 2024)

Table 2. Culinary Dataset

culinary name	type	...	description
150 Coffee and Garden	Kafe	...	150 Coffee and Garden is a café that is located...

culinary name	type	...	description
Ambrogio Patisserie	Kafe	...	Ambrogio Patisserie is a bakery...
...	...	...	...
Xing Fu Tang	Bubble Tea	...	Xing Fu Tang is a contemporary teahouse that...
Yoshinoya	Jepang	...	Yoshinoya is the perfect choice for you...

Source: (Research Results, 2024)

2. Preprocessing

In the preprocessing stage, datasets containing reviews from Twitter or X are processed in three main stages: text processing, translating into English, and polarity scoring. The text processing stage itself consists of the following processes.

A. Case Folding: Converting upper case text data into lower case.
B. Data Cleaning: Cleaning hashtags, URLs, punctuation, emoji, mentions, usernames, and irrelevant elements from tweets.
C. Tokenization: Breaking down sentence text into individual words.
D. Stopword Removal: Removing words that do not add value. This process is done with the help of the nltk.corpus library.
E. Stemming: Reducing words to their primary form by removing all suffixes, prefixes, confixes, and infixes. This process is done with the help of the Sastrawi Python library.

After being processed, the text review will be translated from Indonesian into English using the GoogleTranslator library. Next, the translated text will calculate the polarity score on a scale of 0 to 5 to evaluate the review and identify whether it falls into the positive or negative category.

3. Content-based Filtering

The recommendation process with content-based filtering utilizes content-based filtering techniques that rely on user profiles and features extracted from previously evaluated items [14].

In CBF, an item's features and characteristics are compared to identify items with similarities. Similar items are then recommended to users who have previously shown interest in or used similar items. To improve the recommendation results, CBF uses TF-IDF feature extraction and FastText word embedding in this research.

3.1. Feature Extraction TF-IDF

Term Frequency Inverse Document Frequency (TF-IDF) method has become widely used in the information retrieval and text mining domains to assess the relatedness of individual words in a document corpus. Term Frequency (TF) refers to the frequency of occurrence of a word in a dataset [15], whereas Inverse Document Frequency (IDF) quantifies the word's importance in the dataset, with words that occur frequently in a large number of documents having a low weight and words that occur in a small number of documents having a higher weight [16].

This research uses TF-IDF as a vectorizer to convert text features into numerical vector form. The TF-IDF formula used is as follows [17].

$$TF * IDF = TF(w_i, d) * IDF\ (w_i) \quad (1)$$

$$IDF(w_i) = \log\left(\frac{N}{DF(w_i)}\right) \quad (2)$$

N is the total number of documents, with w_i representing the weight of the ith term in the document. Meanwhile, $DF(w_i)$ is the number of documents containing the word (w_i).

3.2. Word Embedding FastText

FastText is a word representation library that contains 2 million common crawl words with 300 dimensions and provides 600 billion word vectors. FastText's embedding method uses morphological features to identify complex words, thus making it an optimal choice for representing vectors. This capability also enhances its generalisability. FastText generates vectors using n-grams, which facilitates handling unknown words [18].

4. Labeling

During the labeling process, the value of the dataset generated by the CBF process is transformed into a binary representation, with values of 0 and 1 denoting the absence or presence of a particular characteristic. The absence of a rating value exceeding 0.5 denotes a culinary that has been labeled 0. Conversely, the presence of a rating value above 0.5 indicates a culinary that has been labeled 1. This binary classification is employed to identify culinary items included in the recommended category and those not.

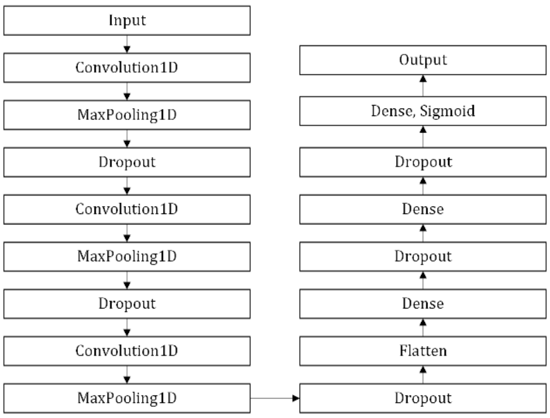
5. Deep Convolutional Neural Network

Deep CNN generally contain more layers than traditional CNN. This increased depth enables Deep CNN to model complex functions and capture hierarchical features within input data [19]. The model constructed in this study was created using the TensorFlow Keras library, whose layer structure consists of input, 1D convolution, 1D max-pooling, dropout, flatten, dense, and output layers.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i2.5640

The convolution and max-pooling layers employ a one-dimensional (1D) approach to accommodate the data type employed in this research, namely text. The convolution layer is responsible for classifying the text and extracting its features, detecting important patterns, and producing a feature map. The pooling layer, in turn, takes the maximum value of the pooling window to reduce the data dimension. Subsequently, the data passes through the dropout layer, which prevents overfitting by silencing a random number of neurons during the training process. A flatten layer is also added for smoothing by converting the features into a vector. Finally, the dense layer is added to minimize the risk of overfitting.



Source: (Research Results, 2024)

Figure 2. Deep CNN Model Architecture

6. Particle Swarm Optimization

PSO is a stochastic, swarm-based algorithm inspired by the collective foraging behavior of bird flocks and fish schools. In PSO, each solution is represented as a particle navigating through a multidimensional search space, incorporating historical best positions and interactions with neighboring particles to guide its trajectory. Through iterative movement, these particles collectively converge toward optimal solutions, demonstrating PSO's effectiveness across diverse application domains, facilitated by a relatively straightforward parameter tuning process [20][21][22].

7. Evaluation

Performance measurement in recommendation systems can be done with classification accuracy, as the main parameter, using the confusion matrix [6]. The metrics used to evaluate the proposed method are accuracy [4][6][23]. The formula for calculating the accuracy value based on the confusion matrix is shown in Table 3. In addition, Root Mean Square Error (RMSE) [24] and Mean ABsolute Error (MAE) is also used as a measurement metric for rank prediction.

Table 3. Confusion Matrix

		Predicted Class	
		Positive	*Negative*
Actual class	*Positive*	*True Positive* (TP)	*False Negative* (FN)
	Negative	*False Positive* (FP)	*True Negative* (TN)

Source: (Research Results, 2024)

The accuracy formula refers to how many samples are correctly identified out of all samples [6]. The accuracy formula is shown in formula (3).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Then to calculate the RMSE, the formula used is formula (4).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y_\iota})^2} \quad (4)$$

Furthermore, the formula used to calculate MAE is formula (5).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \widehat{Y_\iota}| \quad (5)$$

From the formula, it is known that n is the total number of predictions, Y_i is the observation value, and Y_i ^ is the predicted value. The performance value is considered better if the RMSE [24] and MAE value is lower.

RESULTS AND DISCUSSION

1. Data Preparation Result

In the process of data preparation, data is preprocessed. Previously, the process of crawling review data from Twitter or X was accomplished with the Tweet Harvest tool. The obtained data is then subjected to the preprocessing procedure, creating a rating dataset, as depicted in Table 4. This rating dataset is merged with culinary reviews obtained from various authenticated websites that have been previously collected, with the grabfood site serving as a representative example. The total number of users included in the rating dataset is 44.

Table 4. Ratings Dataset

Account	Culinary Name	Rating
BaseBDG	150 Coffee and Garden	3.366071
BaseBDG	Ayam Geprek Pangeran	3.000000
...	...	...
grabfood	Xing Fu Tang	4.800000
grabfood	Yoshinoya	4.800000

Source: (Research Results, 2024)

VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i2.5640

2. Content-based Filtering Result

The first dataset containing the culinary list is utilized in the CBF stage. The columns "culinary name," "type," and "description" are combined in the column "content" as a feature that describes the culinary. Subsequently, a TF-IDF vectorizer is applied to the data to fill empty values with empty strings. Subsequently, a function is defined to obtain a sentence embedding using the FastText model. This function takes the average of the embedded words in the sentence or returns a null vector if no words are found. The resulting FastText embedding is extracted into an array and combined with the TF-IDF matrix.

Subsequently, the cosine similarity value between all culinary items is calculated to ascertain the degree of similarity between each culinary item and the others. The prediction result values from the CBF process, which were obtained by applying TF-IDF and FastText, are presented in Table 5. The evaluation of the results indicates that the tested model exhibits a low error rate, with a Mean Absolute Error (MAE) value of 0.062 and a Root Mean Square Error (RMSE) of 0.094.

Table 5. CBF TF-IDF and FastText Results

Culinary Name	BaseBDG	...	undipmenfess
150 Coffee and Garden	0.674	...	0.06445
Ambrogio Patisserie	0.092417	...	0.062731
...	...	...	...
Xing Fu Tang	0.092266	...	0.063128
Yoshinoya	0.596	...	0.063049

Source: (Research Results, 2024)

3. Classification Result

The classification procedure is executed by utilizing the deep convolution neural network (CNN) approach, utilizing the CBF result dataset, which has undergone a labeling process resulting in the generation of binary labels, specifically 0 and 1. This labeling process has produced a dataset of 7438 instances with label 0 and 1562 instances with label 1. Considering the imbalanced data distribution between the labels 0 and 1, the synthetic minority oversampling technique (SMOTE) will be applied. Previous research by Shujuan Wang et al. [25] has applied the SMOTE method to overcome the imbalance of data distribution and successfully achieved better accuracy value results in the classification process.

In this experiment, classification was conducted in three scenarios. The initial experimental scenario entails classification with the baseline Deep CNN model and SMOTE at four distinct split ratios. In the subsequent experimental scenario, a range of optimization algorithms is applied to the Deep CNN model utilizing the split ratio that yielded the most favorable outcomes in the preceding scenario. The optimal result from one of the tested optimization algorithms is then incorporated into the optimal learning rate value. Finally, the third experimental scenario applies the particle swarm optimization (PSO) method to the Deep CNN model resulting from the second scenario. The objective is to identify the optimal dense unit value in the two existing dense layers, to improve performance.

3.1 Deep CNN Baseline SMOTE Model

Experiments were conducted on the Deep CNN baseline, utilizing a dropout rate of 0.5, a batch size of 64, and 50 epochs. The activation function employed in the final dense layer was sigmoid. The split ratio employed comprised four categories, including 90:10, 80:20, 70:30, and 60:40. The results of this experiment are presented in Table 6. Based on these results, the optimal accuracy value is achieved at a split ratio of 80:20, 76.32%. This result indicates that the split ratio 80:20 will be employed in the subsequent experimental scenario.

Table 6. Baseline Model Results

Split Ratio	Accuracy (%)
90:10	75.01%
80:20	**76.32%**
70:30	74.12%
60:40	75.63%

Source: (Research Results, 2024)

3.2 Deep CNN Model Optimization

In the second scenario, seven optimization algorithms were implemented: Adam, Nadam, Adamax, Adagrad, Adadelta, SGD, and RMSprop. This optimization aimed to enhance model performance metrics and minimize losses during the classification process. The epoch value employed in this study was identical to that used in the first scenario, namely 50. The results of this experiment are presented in Table 7.

Table 7. Deep CNN Model with Optimization Results

Optimizer	Accuracy (%)
Adam	76.41% (+0.09)
Nadam	**76.72% (+0.40)**
Adamax	67.86% (-8.46)
Adagrad	54.96% (-21.36)
Adadelta	53.74% (-22.58)
SGD	60.04% (-16.28)
RMSprop	76.10% (-0.22)

Source: (Research Results, 2024)

Table 7 illustrates that Nadam optimization yields the most accurate results, with an accuracy of 76.72%, an increase of approximately 0.40% compared to the experimental outcomes in the

VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i2.5640

initial scenario. Furthermore, by utilizing the Nadam optimization algorithm with the optimal learning rate, an accuracy of 86.06% was attained, representing a notable enhancement of 9.34% compared to the experimental outcomes before implementing the optimal learning rate. The outcomes of utilizing the optimal learning rate in Nadam optimization are presented in Table 8.

Table 8. Best Learning Rate for Nadam Result

Learning Rate	Accuracy (%)
0.007196856730011514	86.06% (+9.34)

Source: (Research Results, 2024)

3.3 Optimization of Dense Layer in Deep CNN Model Using Particle Swarm Optimization (PSO)

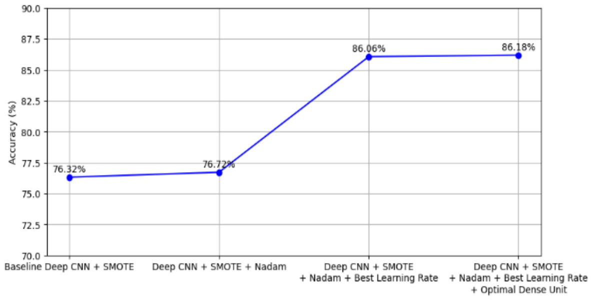
In the third experimental scenario, the PSO method optimizes the dense unit value parameter for two dense layers in the Deep CNN model. This scenario was conducted to identify the optimal dense unit value to enhance the performance of the Deep CNN model. This experiment set upper and lower bound values for the dense units parameter at 32 and 256, respectively. Some settings were also performed, including a batch size of 64, several particles of 20, and an iteration count of 20. The results of this experiment are presented in Table 9. Table 9 illustrates that the optimal dense unit 1 and dense unit 2 values derived from the PSO results improve accuracy by 0.12% compared to the third scenario, reaching 86.18%.

Table 9. Deep CNN Model with Best Dense Unit Result

Dense Unit 1	Dense Unit 2	Accuracy (%)
95.07996645684426	141.53066208357933	86.18% (+0.12)

Source: (Research Results, 2024)

The three scenarios demonstrate an enhancement in model performance, as evidenced by the increase in accuracy value observed in each experimental scenario. This improvement in accuracy is depicted in Figure 3.



Source: (Research Results, 2024)

Figure 3. Accuracy Improvement Graph

Table 10 shows the significant test results, with a significant change from scenario 1 to scenario 2. Changes from scenario 2 to scenario 3 were also experienced, although not as significant as those in scenario 1 to scenario 2. Adding an optimization algorithm with the best learning rate value can significantly improve model performance. Using optimal dense unit values is also proven to improve model performance.

Table 10. Significant Test Result Between All Test Scenarios

	S1 → S2	S2 → S3
Z-Value	225.91815877461508	2.8431137401092674
P-Value	0.0	0.0044675124172863 79
Significant ?	TRUE	TRUE

Source: (Research Results, 2024)

CONCLUSION

This research developed a recommendation system using the CBF and Deep CNN classification methods. The dataset used consisted of 2644 ratings, 200 of which were culinary and 44 of which were users. TF-IDF feature extraction and FastText word embedding were employed in the CBF stage. The rating prediction resulted in an MAE value of 0.062 and an RMSE value of 0.094. Subsequently, the classification process was implemented to enhance the performance metrics and accuracy of the system in providing culinary recommendations. In the classification stage, the Deep CNN model was integrated with the Nadam optimization algorithm and a learning rate of 0.007196856730011514, resulting in an accuracy value of 86.06%. The optimal value of dense units from the PSO optimization results on the Deep CNN model yielded higher results than those from the Deep CNN model with default dense units. The accuracy obtained at the end was 86.18%, an increase of 9.86% from the baseline model.

These results illustrate the satisfactory performance of the proposed recommendation system. The outcomes have important practical ramifications, especially for the creation of tailored recommendation systems in the food industry. Businesses can improve customer happiness and engagement by offering more precise and personalized suggestions by combining CBF with deep learning techniques and optimization algorithms like PSO. Furthermore, the research's techniques and conclusions can be applied to other fields where personalization is essential, like e-commerce, media streaming, and travel advice.

However, this study does have limitations, particularly concerning the small size of the dataset and the fact that the data is exclusively from the X domain. These limitations may affect the generalizability of the results to other contexts or larger datasets. It is anticipated that this research will be further developed in the future by utilizing a larger quantity of data, employing alternative filtering techniques such as collaborative filtering or hybrid filtering, or optimizing additional parameters on the Deep CNN model utilizing PSO to achieve a more optimal performance value than that observed in this research.

REFERENCE

[1] D. Murthy, "Sociology of Twitter/X: Trends, Challenges, and Future Research Directions," *Annu. Rev. Sociol.*, vol. 50, no. 1, pp. 169–190, Aug. 2024, doi: 10.1146/annurev-soc-031021-035658.

[2] D. S. Krishna, G. Srinivas, and P. V. G. D. Prasad Reddy, "A Deep Parallel Hybrid Fusion Model for disaster tweet classification on Twitter data," *Decis. Anal. J.*, vol. 11, p. 100453, Jun. 2024, doi: 10.1016/j.dajour.2024.100453.

[3] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, and R. Kashef, "Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities," *Appl. Sci.*, vol. 10, no. 21, p. 7748, Nov. 2020, doi: 10.3390/app10217748.

[4] A. N. K. Albayati and O. U. Y. Ortakci, "Recommendation Systems on Twitter Data for Marketing Purposes using Content-Based Filtering," in *HORA 2022 - 4th International Congress on Human-Computer Interaction, Optimization and Robotic Applications, Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/HORA55278.2022.9799989.

[5] R. Widayanti, M. H. R. Chakim, C. Lukita, U. Rahardja, and N. Lutfiani, "Improving Recommender Systems using Hybrid Techniques of Collaborative Filtering and Content-Based Filtering," *J. Appl. Data Sci.*, vol. 4, no. 3, pp. 289–302, 2023, doi: 10.47738/jads.v4i3.115.

[6] B. Suvarna and S. Balakrishna, "An Efficient Fashion Recommendation System using a Deep CNN Model," in *International Conference on Automation, Computing and Renewable Systems, ICACRS 2022 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 1179–1183. doi: 10.1109/ICACRS55517.2022.10029063.

[7] M. Amjad, A. Gelbukh, I. Voronkov, and A. Saenko, "Comparison of Text Classification Methods Using Deep Learning Neural Networks," 2023, pp. 438–450. doi: 10.1007/978-3-031-24340-0_33.

[8] S. Ouhame, Y. Hadi, and A. Ullah, "An efficient forecasting approach for resource utilization in cloud data center using CNN-LSTM model," *Neural Comput. Appl.*, vol. 33, no. 16, pp. 10043–10055, Aug. 2021, doi: 10.1007/s00521-021-05770-9.

[9] A. S. Sabyasachi, B. M. Sahoo, and A. Ranganath, "Deep CNN and LSTM Approaches for Efficient Workload Prediction in Cloud Environment," *Procedia Comput. Sci.*, vol. 235, pp. 2651–2661, 2024, doi: 10.1016/j.procs.2024.04.250.

[10] A. B. Bramantyo and A. T. Wibowo, "Collaborative Filtering Movie Recommendation System using K-means Clustering with Particle Swarm Optimization," in *2023 International Conference on Data Science and Its Applications, ICoDSA 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 242–247. doi: 10.1109/ICoDSA58501.2023.10276506.

[11] V. Narendra, J. Christopher, and V. Arunachalam, "Particle Swarm Optimization for dietary recommendations," in *2023 IEEE IAS Global Conference on Emerging Technologies, GlobConET 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/GlobConET56651.2023.10149930.

[12] R. A. Rudiyanto and E. B. Setiawan, "Sentiment Analysis Using Convolutional Neural Network (CNN) and Particle Swarm Optimization on Twitter," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 9, no. 2, pp. 188–195, Feb. 2024, doi: 10.33480/jitk.v9i2.5201.

[13] K. Aguerchi, Y. Jabrane, M. Habba, and A. H. El Hassani, "A CNN Hyperparameters Optimization Based on Particle Swarm Optimization for Mammography Breast Cancer Classification," *J. Imaging*, vol. 10, no. 2, p. 30, Jan. 2024, doi: 10.3390/jimaging10020030.

[14] Y. Afoudi, M. Lazaar, and M. Al Achhab, "Hybrid recommendation system combined content-based filtering and collaborative

VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i2.5640

prediction using artificial neural network," *Simul. Model. Pract. Theory*, vol. 113, Dec. 2021, doi: 10.1016/j.simpat.2021.102375.

[15] F. Firmansyah *et al.*, "Comparing Sentiment Analysis of Indonesian Presidential Election 2019 with Support Vector Machine and K-Nearest Neighbor Algorithm," in *2020 6th International Conference on Computing Engineering and Design (ICCED)*, IEEE, Oct. 2020, pp. 1–6. doi: 10.1109/ICCED51276.2020.9415767.

[16] S. Ranjan and S. Mishra, "Comparative Sentiment Analysis of App Reviews," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, Jul. 2020, pp. 1–7. doi: 10.1109/ICCCNT49239.2020.9225348.

[17] A. Boluki, J. Pourmostafa, R. Sharami, and D. Shterionov, "Evaluating the Effectiveness of Pre-trained Language Models in Predicting the Helpfulness," *Proc. 2023 Intell. Syst. Conf.*, vol. 4, pp. 1–21, 2023, doi: https://doi.org/10.48550/arXiv.2302.10199.

[18] M. Umer *et al.*, "Impact of convolutional neural network and FastText embedding on text classification," *Multimedia Tools and Applications*, vol. 82, no. 4, pp. 5569–5585, 2023.

[19] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5455–5516, Dec. 2020, doi: 10.1007/s10462-020-09825-6.

[20] 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, doi: 10.1109/ACCESS.2022.3142859.

[21] E. H. Houssein, A. G. Gad, K. Hussain, and P. N. Suganthan, "Major Advances in Particle Swarm Optimization: Theory, Analysis, and Application," *Swarm Evol. Comput.*, vol. 63, p. 100868, Jun. 2021, doi: 10.1016/j.swevo.2021.100868.

[22] A. G. Gad, "Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review," *Arch. Comput. Methods Eng.*, vol. 29, no. 5, pp. 2531–2561, Aug. 2022, doi: 10.1007/s11831-021-09694-4.

[23] N. Yarahmadi Gharaei, C. Dadkhah, and L. Daryoush, "Content-based Clothing Recommender System using Deep Neural Network," in *26th International Computer Conference, Computer Society of Iran, CSICC 2021*, Institute of Electrical and Electronics Engineers Inc., Mar. 2021. doi: 10.1109/CSICC52343.2021.9420544.

[24] Y. Tai, Z. Sun, and Z. Yao, "Content-Based Recommendation Using Machine Learning," in *IEEE International Workshop on Machine Learning for Signal Processing, MLSP*, IEEE Computer Society, 2021. doi: 10.1109/MLSP52302.2021.9596525.

[25] S. Wang, Y. Dai, J. Shen, and J. Xuan, "Research on expansion and classification of imbalanced data based on SMOTE algorithm," *Sci. Rep.*, vol. 11, no. 1, p. 24039, Dec. 2021, doi: 10.1038/s41598-021-03430-5.