RESTAURANT RECOMMENDATION SYSTEM USING ADVANCED COLLABORATIVE FILTERING AND REVIEW TEXT CONTENT APPROACH

DANIEL RIANDY[1] , AMALIA ZAHRA[2]

[1]Bina Nusantara University, Computer Science, Jakarta, Indonesia

[2] Bina Nusantara University, Computer Science, Jakarta, Indonesia

E-mail: [1]daniel.riandy@binus.ac.id, [2]amalia.zahra@binus.edu

ABSTRACT

Recommendation system can provide recommendations to consumers on which restaurants might be liked by these consumers. This method plays the major role in food industries. The approach taken is to use collaborative filtering for the final touch in providing restaurant recommendations, but before this stage, it will go through several stages, including Natural Language Processing (NLP) aiming to create a new scoring which is the result of sentiment analysis combined with previous ratings (review text content). Then, clustering will be carried out on customers, to accelerate and optimize the recommendation process. From these two approaches, collaborative filtering is then carried out, so that it can provide some recommendations that are most suitable for consumers. However, in the recommendation system, there is a problem called the Cold Start Problem, this is a condition where the consumer is new so we cannot give recommendations. To overcome this problem, Location Based Filtering will be used, in this method, the consumer must fill in the location where he is located, so that we can provide recommendations for the best restaurant close to the location where the consumer is located. To measure model performance Root Mean Squared Error (RMSE) is used, the results obtained for the RMSE of the proposed method are 0.55, better than the baseline RMSE at 0.62. The time needed to provide recommendations on the proposed method is 247 ms, faster than the baseline model with a speed of 1.83 s. To overcome the Cold-Start Problem, Location Based Filtering is used, in this method the consumer must fill in the location where he is, so that he can provide the best restaurant recommendations that are close to the location where the consumer is.

Keywords: *Recommendation system, Collaborative filtering, Natural Language Processing, Machine Learning, Clustering, Cold Start Problem, Review Text Content*

1. INTRODUCTION

Food is a primary human need, therefore the culinary business is one business that will always be sought and needed by humans. However, it also does not rule out the possibility that there will be many business rivals in this culinary business.

The problem of competition in the culinary business world also does not stop with the many existing competitors, but the Covid-19 pandemic has become a new problem for business people who are struggling in the culinary sector. The COVID-19 pandemic has had a huge impact on the business world. Manufacturing industry, trade, hotels and restaurants, the mining sector, the transportation and communications sector, and the construction sector. This decrease can be seen from the decline in the Weighted Net Balance (WNB) of business activities in the 1st quarter of 2020, which was -5.56%, which was down from 7.79% in the 4th quarter of the previous year. This pattern is recorded for the first time because from 2016 – 2019 there has always been an increase. In particular, in the trade, hotel, and restaurant sectors, its SBT fell by 5.8% from 2.76% in the previous quarter to -3.04%.

In this research a recommendation system based on collaborative filtering will be built, to provide the right recommendations to users. However, in order to produce an appropriate and computationally fast recommendation system, several approaches will be taken. First, to help produce more accurate recommendations, it will utilize the review column and rating score, to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

determine a new score [17], where the new score will be the rating for the collaborative filtering process. Then, it will use a clustering approach to improve recommendations as previously done by Chee Hoon Ha [19] but added to the extraction carried out so that the rating used is not the default rating on the data, but a new scoring that has been aggregated in such a way by combining reviews given by customers with previous ratings. As well as simplifying the computational process, so that the recommendation process can run quickly because the data has been grouped based on their respective clusters which are obtained from the results of grouping customers based on the type of restaurant they prefer [10].

2. METHOD

The following are some of the theories and methods used in this research.

2.1 Collaborative Filtering

Collaborative filtering is a recommendation system that provides recommendations to users which are the result of matching the preferences of a user with other users who have the same preferences [4]. There are already many sites or applications that use this technique to make the user experience when using the application can be personalized well according to the preferences of the user, in other words, the appearance of one user to another will not be the same because it has been displayed based on the preferences of each user. Proven that several large companies have used this technique and have been shown to increase sales significantly. Amazon managed to increase sales by 29%, Netflix increased sales by 60% [13], and many other examples. The first step is to calculate the similarity of all data using the Euclidean distance as in Equation 1:

$$sim(i,j) = \sqrt{\sum_{k=1}^{n}(R_{k,i} - R_{k,j})^2} \quad (1)$$

Where i and j are the i-th data and j-th data, respectively, k represents the k-th item, while R represents the rating. After that, the rating of the user will be predicted using Equation 2:

$$P_{i,j} = \frac{\sum_{k=1}^{n} R_{k,i} \times sim(i,j)}{\sum_{k=1}^{n} sim(i,j)} \quad (2)$$

Where i and j are the i-th data and j-th data, respectively, k represents the k-th item, while R represents the rating.

2.2 K-Means Clustering

K-means clustering is the most commonly used and simplest unsupervised learning algorithm. The purpose of K-means clustering itself is to partition the existing data and then group them into k clusters, where each cluster is expected to have the same tendency. First of all, the value of k must be determined, then the training process will be carried out iteratively to find the most optimal number of clusters. In this process, the distance from each data point will be calculated to get the optimal cluster, the distance calculation will use the Euclidean distance in Equation (1). This approach aiming to simplify the recommendation process so that the algorithm might generate the recommendation faster [5].

2.3 Word2vec

Word2vec is used to represent word vectors and has two modeling architectures, namely continuous bag-of-word (CBOW) and Skip-gram [21]. In this study, the author uses the Skip-gram architecture. Skip-gram architecture can recognize three layers, the first layer is the input layer, the second layer is the hidden layer, and the last layer is the output layer [25]. In the input layer in the hidden layer, there is a weight matrix (W) derived from random values, this layer is used to activate the hidden layer. The calculation of the hidden layer with the weight'(W') matrix is generated from the Output layer [24]. The steps taken by word2vec to vectorize are as follows. First, the words sourced from the text are converted to vectors using one-hot encoding, obtaining vectors and dimensions. Next, the vector w(t) resulting from the encoding becomes the input vector and the output target. W(t) is input and becomes output. Second, the input vector of the term will be the hidden layer of the sum |v| which means the number of features used. This hidden layer is the product of the weight vector W[|v|, N] and the input vector w(t). Third, the Hidden layer has no activation function, because H[1,k] will be directly passed to the output layer. Fourth, the output layer will be calculated in multiplication between H[1, N] and W'[N, |v|] then the result is vector U. Fifth, Calculate the probability of the vector with the softmax function with the equation.

$$p(w_{c,j} = wO_c | wl) = \frac{e^{u_{c,j}}}{\sum_{j=1}^{v} e^{u_j}} \quad (3)$$

Where w_(c,j) is the jth word predicted at the c-th context position, wO〗_c is the actual word that is in the c-th context position, wl is the input word, and u_(c,j) is the jth value in the U vector to predict the word in the c-th context position.

2.4 Random Forest

Random forest one of machine learning method [16] that is a development of the Decision

Journal of Theoretical and Applied Information Technology
28th February 2023. Vol.101. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

tree method, by applying bootstrap aggregating and random feature selection methods to produce more optimal predictions [22]. First, we will create a subset of the data from the overall dataset using bootstrap. Next from that subset of data, a tree will be created, which this process is done repeatedly, until the n-th tree, where the value of n has been determined. For each tree creation, we use entropy as a splitting determinant for each node, then the entropy will be searched for each attribute, the entropy formula can be seen in Equation 4.

$$Entropy(Y) = -\sum_i p(c|Y) \,.\, log_2 \,.\, p(c|Y) \quad (4)$$

Where Y is the set of cases and p(c|Y) is the probability of occurrence of a value in Y against class c. Then after getting the entropy value, the information gain value can be calculated using the formula in Equation 5, where information gain can be used to measure the effectiveness of an attribute in classifying data. The higher the information gain value, the better the splitting is.

$$IG(Y,a) = Entropy(Y) - \sum Values(a) \frac{|Y_v|}{|Y_a|} Entropy(Y_v) \quad (5)$$

Where Value(a) is all possible values in the case set a, Y_v is a subclass of Y with class v corresponding to class a and Y_a all values corresponding to a. This method approach itself aiming to predict the sentiment of review[11][17].

2.5 Sentiment-Based Text Rating

To utilize sentiment information into a new score, the results of the modeling in the previous process using random forest can be used to convert the review data in the main dataset [3], so that it can be averaged with the old rating so that it becomes a new score, with the following formula [18]:

$$TextRating = \left[\frac{P}{P+N} \times 4\right] + 1 \quad (6)$$

2.6 Proposed Methods

The proposed method in order to contribute to this research lies in making a new scoring which we can see more detail in Figure 1. This needs to be done to avoid randomness that might occur when the customer fills in the rating value so that there is no synchronization between the rating and the review text using Equation 4 [18]. So, we will first combine the rating and the results of the sentiment analysis from the review text into a new score. Then, it is added to grouping users based on certain attributes, so that it is hoped that the scope of the collaborative filtering process can be reduced further so that recommendations can be more focused on the same group and running time will also be faster because there are fewer candidates for recommendations [10]. Then also prepared a Location-Based Filtering approach to overcome the Cold start problem. The output that will be given by this research is in the form of the 10 best restaurants that will be given as recommendations to users.

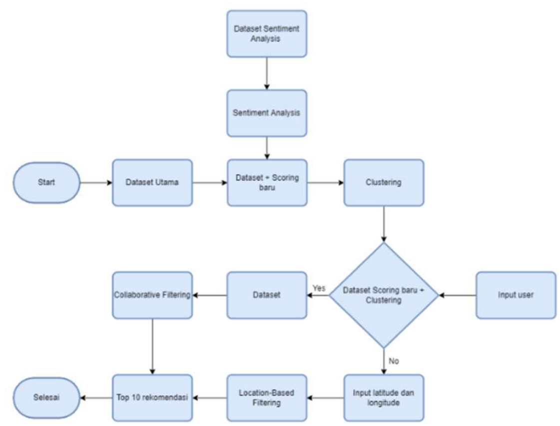

Figure 1. Flowchart for proposed methods

3. RELATED WORKS

This research aiming to contribute to supporting the culinary business, especially restaurants, plus, during the pandemic, more and more people tend to avoid trying new restaurants. With this recommendation system, it is hoped that it will make it easier for new users to be more aware of the restaurants around them (with the help of location-based filtering) and then they will be recommended accordingly using collaborative filtering or content-based filtering. Of course, it also helps the restaurant so that their restaurant is more likely to be reached by users.

Previously, Mr. Achmad Arif Munaji and Mr. Andi Wahyu Rahardjo conducted research using restaurant datasets from the kaggle site which aimed to recommend restaurants using collaborative filtering[9]. So, from the previous user data, it will be seen which restaurants the user has visited. After that, the ID of the restaurant will be taken and then matched with other users who have also visited the restaurant, because if the rating at the previous restaurant is relatively high, then it can be assumed that the two users have food tastes that tend to be the same. The process of equating correlations or seeing the same trend in this study utilizes Pearson Correlation. If the correlation that has been calculated shows a significant number, then that will be the option to be recommended to the user.

Next is the journal written by Mr. Ashish Gandhi. In this journal, a recommendation system is built using machine learning to predict the rating that will be given by the customer, if the prediction results are positive then it will become a recommendation for the user or negative will not be included in the recommendation[6]. What is considered as a recommendation is the rating from the customer. The dataset used is taken from the Yelp application. The predictions are made using 2 algorithms, namely Logistic Regression and Support Vector Machine (Linear and RBF Kernel). The features used also vary, from user-level features, to restaurants-level features, and there are also several new columns created from existing columns. . The maximum predictive results are linear SVM with an accuracy rate of data testing at 69.89%. The more data training for the model, will reduce the tendency of overfitting. The author's suggestion for future research is to understand more about the available features, because there may be features that have a high enough influence on ratings (positive/negative) because the features that the researcher chose before may not be optimal.

This research was written by JinHyun Jooa, SangWon Bangb and GeunDuk Parka[7]. The researcher aims to provide recommendations to users of a business communication tool (with NFC) or places that can be visited by users using association rules and collaborative filtering. In this study, datasets from users will be collected from NFC on their smartphones. More or less data will be processed using clustering using k-means to reduce the scope to several clusters, so that the recommendation process can be processed in a relatively faster time. Then collaborative filtering will be carried out to recommend places that can be visited to users based on the preferences of users who have similar historical data, so it is assumed that they have the same interest.

The following research was written by Arora, G., Kumar, A., Devre, G. S., & Ghumare, A[2]. This research was conducted to create a recommendation system for films, because there have been many recommendation systems for e-commerce users, this time it will be made to recommend films, for Android users. The approach taken is to find the similarity of each user using city block distance and euclidean distance to find the weight of similarity between users, which will then be found by users who have the same tendency as other users, and then recommendations will be given according to the high level of similarity. It is said that, researchers use a Hybrid recommendation engine, where researchers combine content-based filtering and collaborative filtering to produce the most optimal recommendation results for users.

This research was written by Abdulla, G. M., & Borar, S[1]. This research was conducted to help provide recommendations to the user/customer which size is suitable for them, without having to bother to measure it yourself. Datasets are divided into 2 types, namely observable features and latent features. The latent feature here utilizes the word2vec-based skip gram model. Then to ensure whether the size ordered by the customer is right, to assist the recommendation process, a classification will be used using the Gradient Boosting Classifier to classify, if the customer exchanges the product, it means that the size does not fit, and positive if the customer does not exchange the product. Then predictions are made using 3 models, namely observable features model, latent features modal and a combination of both. Evaluation matrix used is accuracy, precision and AUC ROC. Then at the end, clustering will be carried out using K-means after getting sizing from the prediction model results to be able to group customers and products/SKUs into several categories so that it will be easier to recommend. The resulting output is in the form of clustering from the SKU table containing the dimensions and sizes of each brand, containing 6 clusters. Then for the user table, we get 5 clusters, which contain the customer body type.

This journal was written by Mr. N. Varatharajan, J Guruprasad and K. Mathumitha. In this journal, what is considered as a recommendation is the user's food preferences and the ratings given by other users. The recommendation system is carried out based on content-based filtering and collaborative filtering. The dataset is taken from the yelp application [12]. The content-based filtering is calculated based on "Similarity", where it is seen from various aspects, the favorite food of 1 user with the profile of the restaurant, cuisine, and others. Meanwhile, collaborative filtering uses the Funk Matrix factorization and SVD++ techniques. After that the user can log in, so that they can recommend restaurants that have good menus and ratings. In this case, the researcher says that the recommendation system has been successfully created and has provided suitable recommendations for users.

This research was written by Celestine Iwendi [25]. One of the techniques used in item recommendation is known as item-based recommendation system or item–item collaborative filtering. Presently, item recommendation is based completely on ratings like 1–5, which is not included in the comment section. In this context, users or customers express their feelings and thoughts about

Journal of Theoretical and Applied Information Technology
28th February 2023. Vol.101. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

products or services. This paper proposes a machine learning model system where 0, 2, 4 are used to rate products. 0 is negative, 2 is neutral, 4 is positive. This will be in addition to the existing review system that takes care of the users' reviews and comments, without disrupting it. Actually, they use the same dataset from Yelps, along with a mean absolute error (MAE) of 21%, precision at 79%, recall at 80% and F1-Score at 79%. Our model shows scalability advantage and how organizations can revolutionize their recommender systems to attract possible customers and increase patronage. Also, the proposed similarity algorithm was compared to conventional algorithms to estimate its performance and accuracy in terms of its root mean square error (RMSE), precision and recall. Results of this experiment indicate that the similarity recommendation algorithm performs better than the conventional algorithm and enhances recommendation accuracy.

This research was written by Phongsavanh Phorasim and Lasheng Yu[10]. In this study, the researchers made recommendations for a film, using a collborative filtering approach to recommend films, but first used k-means clustering. Clustering is done with the hope that it can help reduce the scope of recommendations, so that recommendations are more accurate because users will be recommended to the same cluster, so they are likely to have the same characteristics. Therefore, clustering is carried out first and then the similarity is sought (in this study using Euclidean distance) and then recommendations are given to the user. This research helps problems that often occur in collaborative filtering which has a long running time due to the large number of similarities that must be sought between users and the accuracy that tends to be low, the solution offered by the researcher is to combine clustering and collaborative filtering approaches to optimize recommendations.

Next is the research from Chee Hoon Ha regarding the Advanced recommendation system[19]. The research was conducted for the Yelp dataset, aiming to recommend restaurants, using several approaches with Collaborative Filtering as a baseline that will be compared with several other approaches. First, it will be compared with the K-nearest neighbors (KNN) algorithm, then this algorithm is combined with clustering and SVD. For the baseline, it produces RMSE 1.474 then for the best results there is the KNN and Clustering combining algorithm with RMSE 1.123.

4. RESULT AND DISCUSSION

The first thing to do is to conduct training on text data, to form a model to analyze the level of sentiment in the text reviews given by customers [19]. The accuracy obtained for this sentiment analysis model using the Random forest approach is 87% on the train data and 86% on the test data, there is still quite high overfitting so that the difference in accuracy is also quite significant. Then the resulting confusion matrix can be seen as shown in the Figure 2.

		Prediction	
		Standard	Substandard
Aktual	Standard	37	13
	Substandard	19	31

Figure 2. Confusion Matrix

Right after the model obtained, it will be implemented in our main data, so that a new score will be obtained as planned at the beginning. Preprocessing data in such a way on the main data, by doing cleaning, feature engineering and combining several data into 1 main data, then the recommendation process can be carried out.

4.1. Clustering

In the collaborative filtering approach, clustering will be carried out on the restaurant category favored by customer n. By looking at the silhouette score in this k-means model [23], we can see from Figure 3 that the most significant results are obtained when there are 6 clusters.

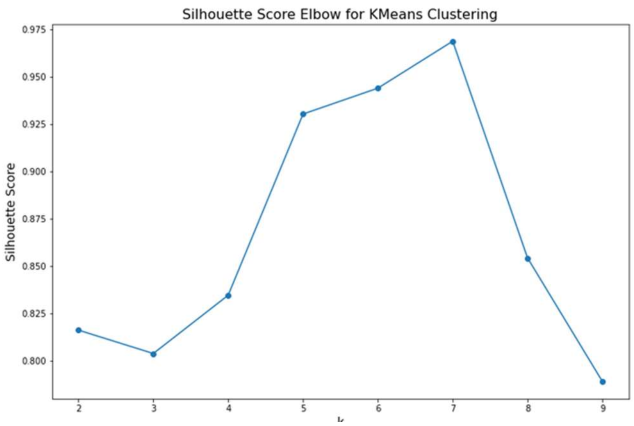

Figure 3. Silhouette score on user-level clustering

So that we will divide the user-level into 6 groups. So that it can be compared to the collaborative filtering approach by utilizing this clustering (Advanced Collaborative Filtering) with ordinary collaborative filtering which does not utilize clustering in the process.

The second one is we do the clustering to the restaurant-level, based on the location where the restaurant belong. According to the location of the restaurant from the map below, we can simply say that the restaurant is spread into 7 clusters since there are six states in the Data.

The clustering process will still be carried out so that the results obtained are more certain by referring to the silhouette score, because after plotting using the elbow method, the resulting angle is not significant so that the optimal cluster cannot be determined [14]. At the first level, the results of the silhouette in Figure 4 show that the optimal number of clusters is 7.

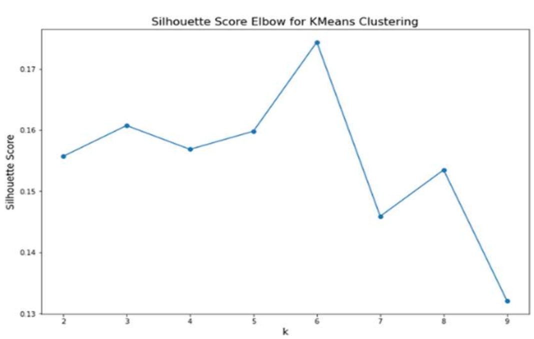

Figure 4. Silhouette score on state-level clustering

The clustering process will still be for each state, clustering will be carried out again for the location level, where after implementing k-means clustering and calculating the various silhouette scores for each state.

4.2. Collaborative filtering

The last process to be carried out is to build a collaborative filtering model. The results obtained are as the Table 1 below.

Algorithm	**Provide Recommen-dation**	**Training Time**	**RMSE**
CF	1.83 s	16h 51m 52s	2.77
Advanced CF + Review Text Content	247 ms	1h 11m 33s	3.20

Table 1. Performance summary

As can be seen that the RMSE obtained for Advanced collaborative plus Review text content review shows higher error results compared to ordinary Collaborative Filtering, but it is not significant. The more crucial thing is that the training time is very different where Collaborative filtering to conduct training on the data so that the RMSE value is obtained it takes 16 times longer than the proposed method. Meanwhile, to provide regular Collaborative Filtering recommendations also takes 7.4 times longer than the proposed method.

4.3 Location-Based Filtering

Location-based filtering approach utilize the location of this new customer, to find the top 10 restaurants surround his/her area, the restaurant that will recommend to this customer based on the location and the score. So instead of just picking the restaurant nearest to the customer we provide the nearest restaurant and this restaurant received good rating, rated by other customer, which shown in Figure 5.

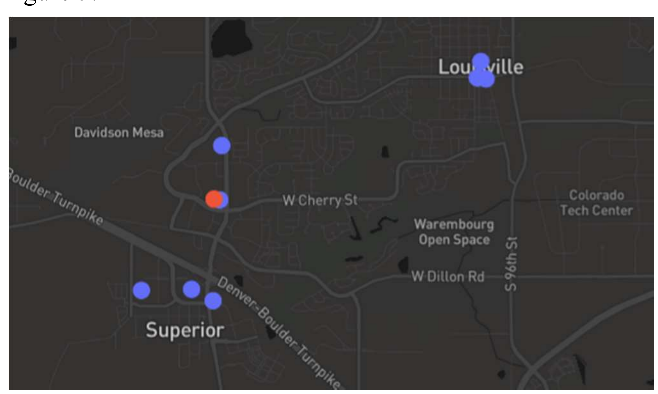

Figure 5. Top-10 nearest restaurants

We can see from the figure 6 above, the use case where the customer is at Trinity United

Methodist Church and here we provide him/her top 10 restaurants in our list and let him/her choose which one maybe he/she will like the most.

In the future, it can be added to do modeling on the NLP model, the small amount of data when building the model is the main influence on the poor accuracy results so that it also affects the RMSE in the collaborative filtering approach, if the model results are good then the RMSE will also decrease. So that in the future it can be added data,

5. CONCLUSION AND FUTURE WORKS

In the results of this study, two things were carried out aimed at maximizing the recommendation system. The first is Review text-content, which is a technique that uses Natural Language Processing and then implements a formula to extract sentiment value from a review, where the results of this extraction will change the previous rating value and create a new score. This approach has proven to be successful, by providing a more minimal RMSE value. In a previous study it was found that the minimum RMSE value was at 1.1 [19], but with this approach, the minimum RMSE value was 0.5. Second, utilizing clustering to maximize the time to provide recommendations, has also been successfully carried out. It is proven by the faster time in providing recommendations on the proposed method, which is only 2.47 ms, while using ordinary collaborative filtering, it takes 1.83s.

With all this findings, small error rate and fast to generate recommendation, this particular solution will help to give an accurate recommendation that can beneficial for both customer site and restaurant site, along with that the time is also the part that always be concerned, but we generate it faster in this research, so this recommendation system can cover that problem.

In the next research, if you want to continue this research, you can do several things that make it possible to optimize the time to generate predictions and minimize the RMSE generated by the model, so that recommendations can be made more accurate and in a shorter time. In the text content review section, we can examine newer approaches and the latest state of the art machine learning models, so as to maximize machine learning models, and produce good sentiment analysis. In this study, only the baseline model from random forest was used for sentiment analysis, but the results provided were quite optimal. Likewise in the clustering process, you can use methods that are the latest state of the art in clustering cases, so that they can provide more optimal results when providing recommendations, and even have the opportunity to help provide more hyperparameter tuning on the NLP model and also using a good model or maybe vectorization, maybe you can use the latest state of the art so that the results obtained on model accuracy can be more significant, which causes RMSE results on collaborative filtering to also be much better.Also we can add some method like trusted-based collaborative filtering and/or association rules that might process the data so we can generate better recommendation system [8].
accurate predictions so that the RMSE obtained is more minimal.

6. ACKNOWLEDGMENT

I would like to express my special thanks of gratitude to all my lecturer who gave me this golden opportunity to do this wonderful project regarding recommendation system. Which also helped me a lot during this project. I came to know about so many new things I am really thankful to them. Also, I would like to thank my family and family who always support me to finish this project within the limited time frame.

REFERENCES:

[1] Abdulla, G. M., & Borar, S. (2017). Size recommendation system for fashion e-commerce. In KDD Workshop on Machine Learning Meets Fashion.

[2] Arora, G., Kumar, A., Devre, G. S., & Ghumare, A. (2014). Movie recommendation system based on users' similarity. International journal of computer science and mobile computing, 3(4), 765-770.

[3] Chinsha, T. C., & Joseph, S. (2014). Aspect based opinion mining from restaurant reviews. International Journal of Computer Applications, 975, 8887.

[4] Dastidar, S. G. (2017). Restaurant Recommendation System. International Journal of Business Analytics and Intelligence, 5(2), 22.

[5] Eyjolfsdottir, E. A., Tilak, G., & Li, N. (2010). Moviegen: A movie recommendation system. *UC Santa Barbara: Technical Report*.

[6] Gandhe, A. (2015). Restaurant recommendation system. cs229. stanford. edu.

[7] Jiang, L., Cheng, Y., Yang, L., Li, J., Yan, H., & Wang, X. (2019). A trust-based collaborative filtering algorithm for E-commerce recommendation system. Journal of Ambient Intelligence and Humanized Computing, 10(8), 3023-3034.

[8] Jooa, J., Bangb, S., & Parka, G. (2016). Implementation of a recommendation system

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

using association rules and collaborative filtering. *Procedia Computer Science*, *91*, 944-952.

[9] Munaji, A. A., & Emanuel, A. W. R. (2021). Recommendation System Based on User Ratings with Collaborative Filtering. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1077, No. 1, p. 012026). IOP Publishing.

[10] Phorasim, P., & Yu, L. (2017). Movies recommendation system using collaborative filtering and k-means. *International Journal of Advanced Computer Research*, *7*(29), 52.

[11] Krishna, A., Akhilesh, V., Aich, A., & Hegde, C. (2019). Sentiment analysis of restaurant reviews using machine learning techniques. In *Emerging Research in Electronics, Computer Science and Technology* (pp. 687-696). Springer, Singapore.

[12] Varatharajan, M. N., Guruprasad, J., & Mathumitha, K. (2020). Restaurant Recommendation System Using Machine Learning. International Educational Applied Research Journal, 4(3), 1-4.

[13] Koren, Y. (2009). Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 447-456).

[14] K. Mahendru (2019). How to Determine the Optimal K for K-Means.

[15] Melville, P., Sindhwani, V., Sammut, C., & Webb, G. I. (2010). Encyclopedia of machine learning.

[16] Taboada, M. (2016). Sentiment analysis: an overview from linguistics. Annual Review of Linguistics, 2, 325-347.

[17] Ganu, G., Elhadad, N., & Marian, A. (2009, June). Beyond the stars: improving rating predictions using review text content. In *WebDB* (Vol. 9, pp. 1-6).

[18] Beigi, G., Hu, X., Maciejewski, R., & Liu, H. (2016). An overview of sentiment analysis in social media and its applications in disaster relief. *Sentiment analysis and ontology engineering*, 313-340.

[19] Ha, C. H., & Univerisy, S. (2021). Yelp Recommendation System Using Advanced Collaborative Filtering.

[20] Acosta, J., Lamaute, N., Luo, M., Finkelstein, E., & Andreea, C. (2017). Sentiment analysis of twitter messages using word2vec. *Proceedings of student-faculty research day, CSIS, Pace University*, *7*, 1-7.

[21] Jotheeswaran, J., & Koteeswaran, S. (2016). Feature selection using random forest method for sentiment analysis. *Indian Journal of Science and Technology*, *9*(3), 1-7.

[22] Ogbuabor, G., & Ugwoke, F. N. (2018). Clustering algorithm for a healthcare dataset using silhouette score value. *International Journal of Computer Science & Information Technology*, *102*(2018), 27-37.

[23] Zhou, Y., Wilkinson, D., Schreiber, R., & Pan, R. (2008, June). Large-scale parallel collaborative filtering for the netflix prize. In *International conference on algorithmic applications in management* (pp. 337-348). Springer, Berlin, Heidelberg.

[24] Lilleberg, Joseph, and Yun Zhu. "Support Vector Machines and Word2vec for Text Classification with Semantic Features." Cognitive Informatics & Cognitive Computing (2015): n. pag. Print.

[25] Iwendi, C., Ibeke, E., Eggoni, H., Velagala, S., & Srivastava, G. (2022). Pointer-based item-to-item collaborative filtering recommendation system using a machine learning model. *International Journal of Information Technology & Decision Making*, *21*(01), 463-484.