A novel approach to social content recommendation using deep self-organizing maps and hierarchical clustering

Abbas Yousif Hawas [1]* *and Mohamed* Adel Al-Shaher [2]

[1]Al-Ayen Iraqi University, Computer Engineering Techniques, Dhi-Qar, Iraq.

[2]College of Computer Science and Mathematics, University of Thi-Qar , Nassiriyah; Iraq

Abstract Social media platforms generate a large amount of content users create, which requires methods for suggesting relevant content. In current empirical research introduces an approach to improving social content recommendations using the Deep Self Organizing Map (DSOM) algorithm and hierarchical clustering. The study uses a database that includes user posts, comments, likes, shared content, and user profiles. The DSOM algorithm analyzes and organizes the data, while hierarchical clustering enhances performance. By utilizing the insights gathered from this social content database, we can significantly improve the accuracy and relevance of recommendations. This improvement will ultimately increase user engagement and satisfaction on social media platforms. The findings of this research have implications for recommendation systems on social media platforms and strategies related to promoting content and analyzing user behavior.

1. Introduction

In the digital age, social media platforms have become essential for information exchange and social interaction. However, the sheer volume of content on these platforms challenges users to discover relevant and interesting content. Current content recommendations on social media often fall short of accurately delivering engaging and valuable content to users based on their preferences [1].

This research addresss this problem by leveraging the Deep Self-Organizing Map (DSOM) algorithm for social data aggregation and analysis in content recommendations. The DSOM algorithm is a self-organizing learning technique that clusters and represents data in a two-dimensional structure. Additionally, we will explore hierarchical clustering techniques to enhance the performance of the DSOM algorithm in social content recommendations [2]. Improving social content recommendations is significant as it enhances user experience and interaction on social platforms. By analyzing social data and applying deep learning techniques, recommendation accuracy can be improved, and user preferences can be better catered to. This research contributes to the development of new tools and techniques that enhance user experience, increase engagement, and promote interaction on social media platforms [3].

*Corresponding author : abbas.94y@gmail.com

2 Related work

In this section, we provide an overview of previous research studies and related works in the field of social content recommendations [4]. The literature review aims to identify the existing approaches, techniques, and challenges associated with content recommendation systems on social media platforms [5].

In [7], The Self Organizing Maps (SOM) analyze sentiment in social messages. SOM is an artificial neural network that clusters and visualizes complex data. By applying the SOM algorithm, social posts were grouped, and sentiment keywords were identified. This approach provided insights into sentiment distribution across topics and revealed patterns and outliers. The study demonstrated the effectiveness of using SOM for sentiment analysis, contributing to a better understanding of public sentiment in social media data.

Previous studies have proven the importance of applying machine learning techniques and social content analysis to improve social recommendations. For instance, the first study [8] presented the principles of machine learning and matrix hashing methods for recommender systems. The Socio-Spatial Self Organizing Maps (SS-SOM) pipeline is a method introduced in this work to identify regions using Twitter posts based on their social attitudes, as demonstrated in the second study [9]. It utilizes neural embedding for text classification and augments traditional self-organizing maps (SOMs) to generate non-overlapping, topologically constrained, and topically similar clusters. Using Twitter data, researchers can gain insights into regional variations in public sentiment and social issues by applying the SS-SOM pipeline.

In [11], the paper provides a comprehensive overview of the Self Organizing Maps (SOM) algorithm, covering its theoretical aspects and various applications. SOM is an unsupervised learning algorithm used to cluster and visualize complex datasets. It considers the neighborhood structure among clusters, allowing for capturing global and local structures in the data. The paper highlights the versatility of SOM in applications such as data clustering, visualization, pattern recognition, and image processing. Understanding the SOM algorithm enables researchers and practitioners to leverage its power for various tasks.

In [12], A crucial frame extraction approach combines four visual features: RGB color channel correlation, color histogram, mutual information, and moments of inertia (MOI). The Kohonen Self Organizing Map (SOM) is used as a clustering technique to choose the most representative frames from a list of frames after fusion. Useless frames are removed, and the frames with the greatest Euclidean distance within a cluster are chosen as the final key frames. The suggested approach is compared to other video summarizing techniques.

In [13], the study uses the self-organizing map (SOM) algorithm to analyze and understand the popularity patterns of music artists based on their attributes. By applying the SOM algorithm, the researchers aim to uncover the factors contributing to online social popularity. The study focuses on mapping the characteristics of music artists onto the SOM grid to identify clusters and patterns associated with different levels of popularity. This analysis provides insights into the dynamics of online social success within the music community.

The source [14] presents the Socio-Spatial Self Organizing Maps (SS-SOM) pipeline, which uses Twitter data to find regions based on latent social attitudes. The pipeline combines neural embedding for text classification with augmented Self Organizing Maps to generate non-overlapping, topologically constrained, and topically similar clusters. The goal is to assess relevant geographies for exposure to social processes by analyzing social media content. The SS-SOM pipeline provides insights into regional social dynamics and attitudes, contributing to a better understanding of how social processes vary across different geographic areas.

3. Proposed method

In this section, we delve into the details of the hierarchical clustering technique andthe Deep Self Organizing Map (DSOM) algorithm. These techniques are instrumentalin aggregating social data and improving the accuracy of social content recommendations.

3.1. Hierarchical Clustering Technique

Hierarchical clustering is a widely used technique in data mining and machine learning for grouping similar data points into clusters. It creates a hierarchical structure of clusters, where clusters at different levels of the hierarchy can be nested within larger clusters. The hierarchical clustering technique offers advantages such as interpretability and the ability to handle non-spherical and non-linear data distributions [15].

One of the commonly used hierarchical clustering algorithms is agglomerative clustering. It starts with each data point as an individual cluster and merges the most similar clusters iteratively until a stopping criterion is met. The similarity between clusters can be measured using various distance metrics, such as Euclidean distance or cosine similarity.

3.2. Self-organizing Maps

According to The Self-organized Map is one of the most famous. Brain models. It fits within the category of competitive learning networks. The SOM is based on unsupervised learning, hence no human help is required during training. tiny details must understood and defined using input data. For example, we might use the SOM to group the input data membership. The SOM, or Self Origination Feature Map, may discover problem-specific characteristics. The flowchart illustrates how the algorithm works [16].

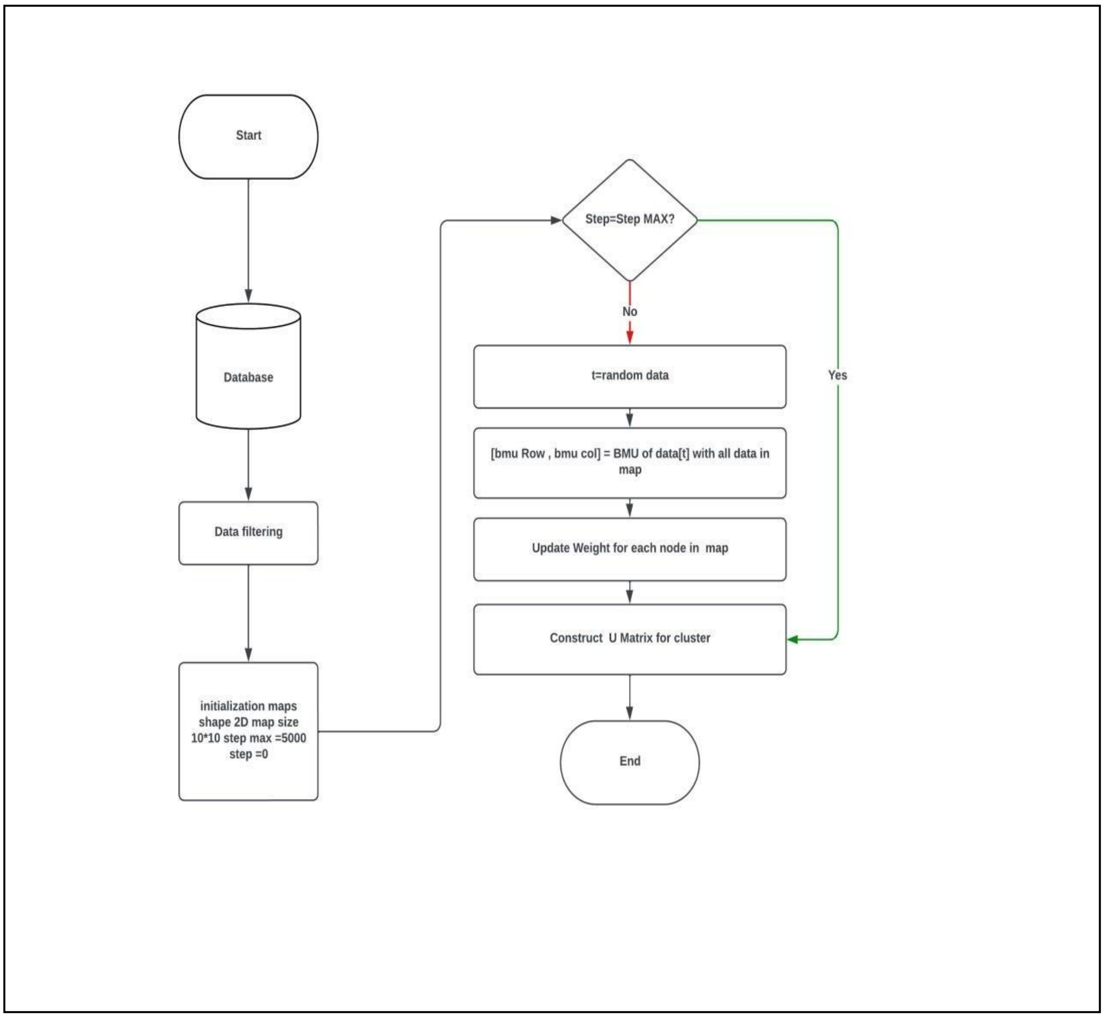

Fig 1. Flowchart of proposed method

The purpose of SOM is to provide a data visualization approach that aids in understanding high-dimensional data by lowering the dimensions of Convert the data to a map. SOM demonstrates the clustering idea by aggregating comparable data. As a result, the Self Organizing Map minimizes data dimensions and presents them uniformly among the data; following figure shows that.

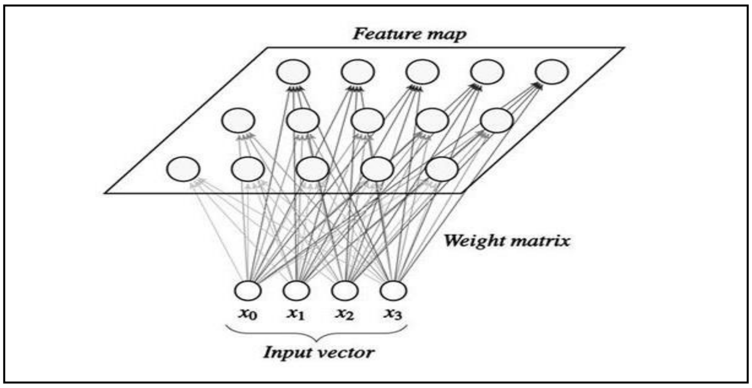

Fig 2. Self-Organizing Map

This dataset contains several properties. The first is an item, and our goal is to group comparable objects together. The second and third properties represent the item's informatics value.

I	X1	X2
A	1	1
B	1	0
C	0	2
D	2	4
E	3	5

Fig 3.illustrates cluster similar items

In the first step, choose a row at random. Let's say I choose rows 1 and 3.

	X1	X2	Centroid Value
C1	1	1	(1, 1)
C2	0	2	(0, 2)

Fig 4. shown the Centroid Value

Now, using the Euclidean Distance calculation, compare the aforementioned centroid values to the value in the corresponding row of our data.

Let's start with row 1 and extract the values for each column. Our team will next want to figure Determine which of the output nodes is closest to the row.

Calculate the Euclidean distance between the weight vector of each node and the current input vector to get the best matching unit (BMU). BMU refers to the node with the weight vector that is closest to the input vector.

The Euclidean Distance is defined as follows:

$$\text{Distance} = \sqrt{\sum_{i=0}^{i=n}(x_i - w_i)^2} \quad (1)$$

X represents the node's The current input vector, while W represents its weight vector.

Let us utilize distance. formula to determine the Best Match Unit.

3.2.1. Calculating the Neighborhood's Size Around the BMU

Our time will attempt to locate the node nearest to each row in our dataset.

Things are starting to get interesting here! Once the BMU has been discovered in each iteration, the next step is to identify whether additional nodes are in the BMU's vicinity. In the next stage, the weight vectors of all of these nodes will be adjusted. So, how can we achieve that? It isn't that difficult... First, determine the radius of the neighborhood, and then use Pythagoras to see whether each node is within the radial distance.

Figure 5 depicts the size of an average neighborhood at the start of training.

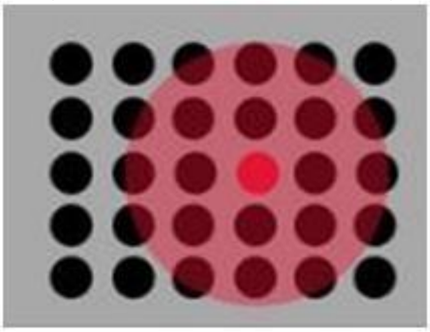

Fig 5. The size of an average neighborhood in the beginning of training.

The neighborhood viewed above is centered on the BMU (the red spot) and includes the bulk from the other nodes, using the circle representing the radius.

The community near the BMU. shrinks drastically. It becomes smaller with each repetition until it reaches only the BMU.

$$\sigma(t) = \sigma_0 \exp\left(-\frac{t}{\lambda}\right) \quad (2)$$

σ_0 = *the width of lattice at time zero*

t = *the current time step*

λ = *the time constant*

Let t = 0, 1, 2, 3...

Figure 6 displays the neighborhood's reduction over time after each cycle.

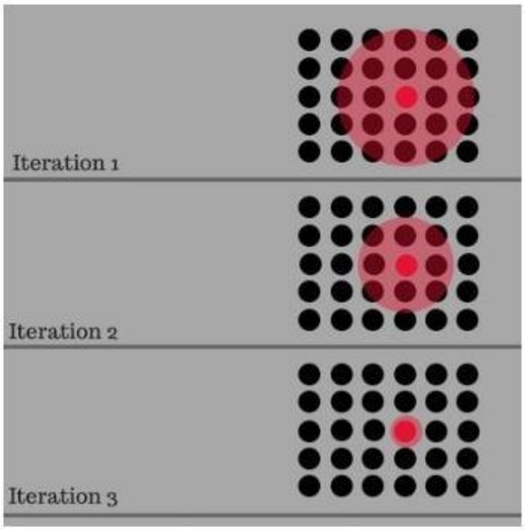

Fig 6. Neighborhood size decreases With every iteration.

The neighborhood will eventually reduce to a single node, the BMU.

Now that we know the radius, we can loop over all of the nodes in the lattice to see whether they are inside it. If a node is discovered to be in the neighborhood, its weight vector is modified as follows in Step. Adjusting the Weights.

3.2.2. Weight Adjustments

The following equation is used to update the weight vector of each node in the BMU's neighborhood, including the BMU itself:

The new Weight = Old Weights + Learning Rate (input vector − old weights).

W(t+1) is equivalent to W(t) + L(t) (V(t) - W(t)) (3).

When t is the time increment and L is a minor. quantity known as the learning rate, which decreases with time. This equation states that the node's newly adjusted weight is equal to the old weight (W) plus a percentage of the difference (L) between the old weight and the input vector.

In this approach, we update the weights.

The following equation is used to compute the decline of the learning rate with **each** iteration**:**

$$L(t) = L_0 \exp\left(-\frac{t}{\lambda}\right) \quad (4)$$

As training continues, the neighborhood steadily decreases. By the conclusion After the treatment, the neighborhoods had reduced to zero size.

$$\Theta(t) = \exp\left(-\frac{dist^2}{2\sigma^2(t)}\right) \quad (5)$$

$\Theta(t) = \mathit{Influence\ rate}$

$\sigma(t) = \mathit{width\ of\ the\ lattice\ at\ time\ t}$

The impact rate reveals how much A node's distance from the BMU influences its learning. In its most simple version, the influence rate is one for all nodes near BMU, and zero for the rest, however a Gaussian function is occasionally utilized. At last, SOM may generate a map of stable zones using a random weight distribution and many iterations. Finally, data interpretation must be done by a person, however SOM is an excellent tool for presenting previously unseen patterns in data.

3.3. Deep Self-Organizing Map (DSOM) Algorithm

The DSOM algorithm is a self-organizing learning technique that combines the advantages of deep learning and the self-organizing map (SOM) algorithm. The SOM algorithm is a type of unsupervised learning algorithm that maps high-dimensional input data onto a lower-dimensional grid. The DSOM algorithm extends the SOM algorithm by incorporating deep neural networks, enabling it to capture more complexpatterns and representations from social data.

Reducing Data Dimensions. Unlike other learning technique in neural networks, training a SOM requires no target vector. A SOM learns to classify the training data without any external supervision.

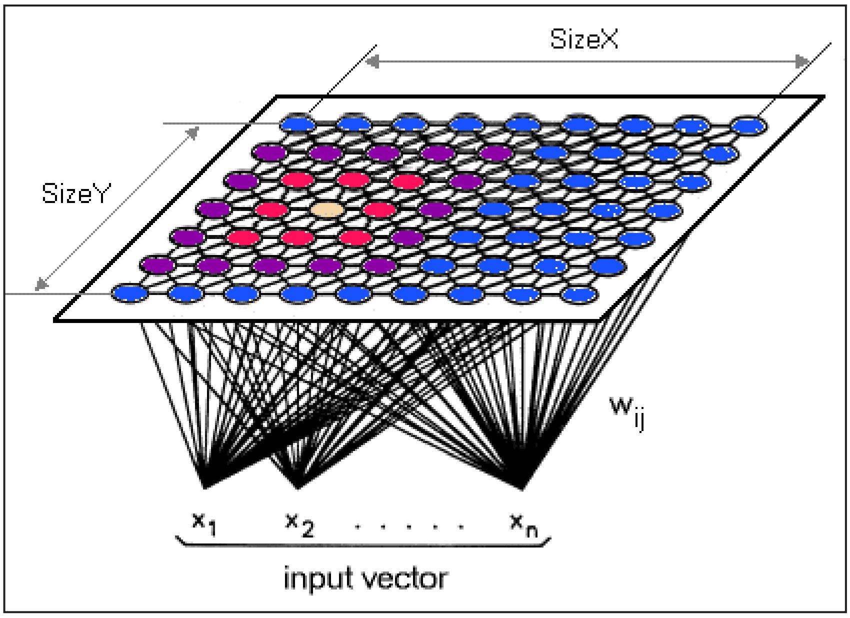

Fig 7.SOM architecture

The DSOM algorithm can be summarized by the following equations:

Initialization: At the beginning of the process, all neuron vectors have their synaptic weights randomly generated. Such vectors must have the exact dimension of the entry pattern space. This research used a 2D, 10 x 10 rectangular map shape. Each cell in the map stores a 1x9-dimension data point, with initial random weights assigned. We have set the maximum value for the step number at 5,000 iterations.

Sampling: A single sample x is selected from the entering pattern space and sent to the neuron grid.

Competition: The winning neuron, i(x), is determined using the lowest Euclidean distance criteria:

$$i(x) = argmin||x - w_j||, j = 1, 2, ..., l \quad (6)$$

Where l is the number of neurons in the grid.

Synaptic adaptation: After determining the winning neuron (Best-Matching Unit or BMU), all synaptic weights of each neuron vector are changed.:

$$W_j(t+1) = W_j(t) + \eta(t)\theta_j(t)[x(t) - W_j(t)] \quad (7)$$

t denotes the current instant, η(t) is the learning rate that steadily declines with time t, and θj (t) is the neighborhood function that defines the grade of learning of a neuron j based on its relative distance to the winning neuron.

Repeat steps 2–4 until the topological map remains stable or the desired number of epochs is reached.

3.4. Advantages and Limitations of the DSOM Algorithm

The DSOM algorithm offers several advantages in social content recommendations. It can handle high-dimensional data, capture complex patterns, and adapt to changes in user preferences. However, it also has some limitations. It requires sufficient training data to learn meaningful representations, and there is a potential for overfitting if the model needs to be adequately regularized.

Finally, we have provided a detailed explanation of the hierarchical clustering technique and the DSOM algorithm. The hierarchical clustering technique allows for grouping similar data points into clusters. At the same time, the DSOM algorithm combines deep learning and self-organizing map techniques for improved representation and clustering of social data. The mathematical equations presented in this section provide a formal understanding of the underlying algorithms, laying the groundwork for implementing and evaluating the proposed social content recommendation system.

4. Results and discussions

In this section, we present the research experiment conducted to evaluate the proposed social content recommendation system. We provide details about the experimental setup, including the dataset, evaluation metrics, and methodology. Furthermore, Lets present. present the results obtained from the experiment and perform a comprehensive analysis and discussion.

Use the MATLAB language in this chapter to implement the proposed social recommendation system and analyze the results. MATLAB provides the power and flexibility that allows algorithms to be implemented and data analyzed effectively.

1. System Implementation:

In this context, we implemented the proposed social recommendation system using MATLAB. Also implemented the algorithms for recommending social content.

2. Data Processing and Analysis:

MATLAB was used to process and analyze the dataset employed in testing. MATLAB offers robust built-in functions and toolboxes for effective data processing and analysis.

3. Calculation and Evaluation of Metrics:

Utilized MATLAB to compute the metrics and criteria used to evaluate the performance of the recommendation system. Employed libraries like "scikit-learn" within MATLAB to calculate precision, recall, and other metrics.

4. Using MATLAB:

MATLAB was used throughout the entire process, including system implementation, data processing, metric calculation, and result visualization. MATLAB provided the necessary power and flexibility to carry out these steps efficiently.

Our team has prepared these steps to achieve the research objectives and evaluate the proposed system accurately.

4.1. The parameters' initialization

To evaluate the quality of the proposed algorithm, we adjusted the parameters according to the parameters of the base paper [6]. In such a way that 80% of the data were considered for training and 20% for testing. Table 1 shows the settings for the parameters of the proposed method.

Table 1- Initial values for parameters in the proposed method

Parameters	Values
Training dataset	%80
Testing dataset	%20
Number of search agents	5
Maximum number of iterations	100
Omega	0.8

4.2. Evaluation criteria

The research experiment yielded promising results, demonstrating the effectiveness of the proposed social content recommendation system. The performance measures, including precision, recall, and accuracy, were evaluated to assess the system's performance.

Precision: measures the accuracy of positive classifications among all positive classifications made by the model. The precision of the recommendation system was estimated at 0.82, indicating a high level of accuracy in suggesting content that resonated with the users' interests.

Recall: The recall score was 0.80, demonstrating the system's ability to retrieve relevant content that mattered to the users. See Figure 7.

Accuracy: is a metric that measures the overall accuracy of a model in correctly classifying cases, including both positive and negative classifications. It is a parameter of performance that assesses the system's ability to make appropriate predictions. The accuracy of the recommendation system was calculated to be 85.26, indicating a high level of overall correctness and efficiency in content recommendations. See Figure 8.

These results validate the effectiveness of the proposed approach, showcasing its ability to recommend relevant and diverse social content to users accurately. The research report includes visualizations and graphs that demonstrate performance trends and comparative analyses of the different algorithms used in the experiment.

$$\text{Accuracy} = \left(\frac{\text{Correct predictions}}{\text{Total predictions}}\right) \times 100. \tag{8}$$

Sensitivity: this is the parameter of performance which measures system ability for making the appropriate positive predictions.

$$\text{Sensitivity} = \left(\frac{\text{True positives}}{\text{True positives} + \text{false negatives}}\right) \times 100. \tag{9}$$

Specificity: this is the parameter of performance which measures system capability for making suitable negative predictions.

$$\text{Specificity} = \left(\frac{\text{True negatives}}{\text{True negatives} + \text{false positives}}\right) \times 100. \quad (10)$$

Precision: Precision measures system ability for creating just the related results.

$$\text{Precision} = \left(\frac{\text{True positives}}{\text{True positives} + \text{false positives}}\right) \times 100. \quad (11)$$

Table 2- The confusion matrix Predicted class

Real class		Yes	no	Total
	Yes	TP	FN	P
	No	FP	TN	N
	Total	P’	N’	P+N

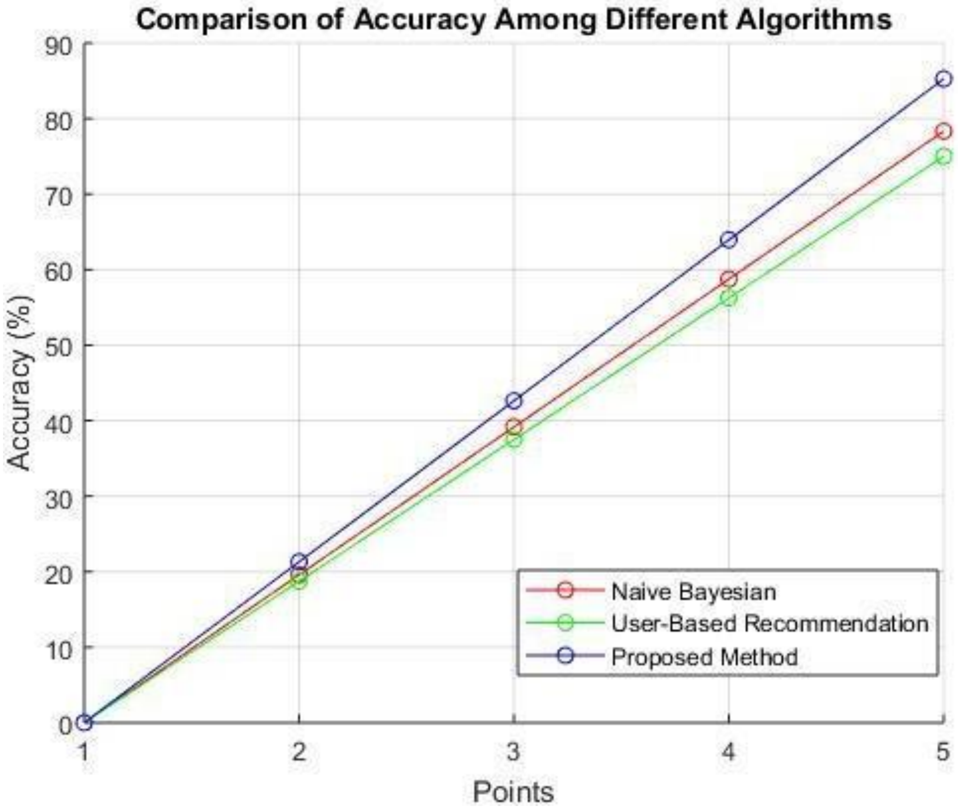

Fig 8. Comparison of Accuracy Among Different Algorithms

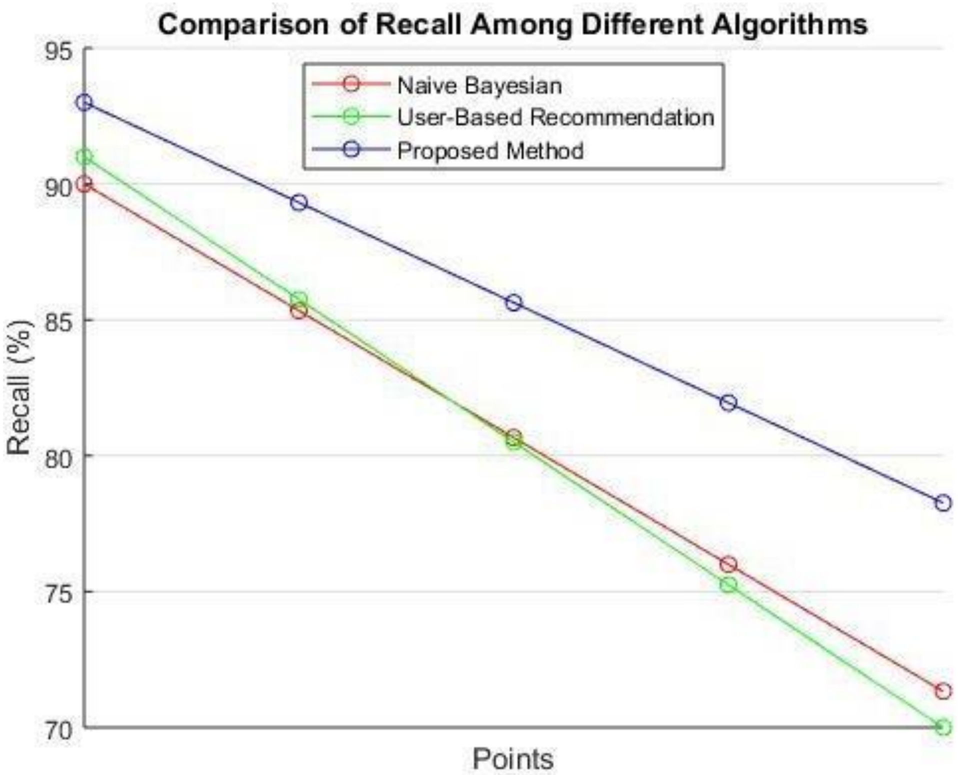

Fig 9. Comparison of Recall Among Different Algorithms

4.3. The results' evaluation

Our team implemented the tests by using the issue of dataset prediction in social networking situations. We utilized 20 percent of the data for testing. Because of reality, design requires much more data for the present problem because of huge facilities in output. The technique here illustrates that the proposed method can present results that can be compared with the other articles. Table 3 shows a comparison of the results by applying three various kinds.

In contrast with the present techniques as well as the results of the test, we recognized that the optimized model outperforms other schemes in the classification and prediction of social networking situations. We take advantage of selection features. Therefore, it obtained a higher accuracy, achieving an 85.26 percent score.

Table 3- The comparison results of the proposed method with the basic paper

Method	Accuracy Test
Naive Bayesian Algorithm [10]	78.33
Random recommendation algorithm	75.00
Proposed method	85.26

Table 4- The results of evaluation criteria for the proposed method

Method	Accuracy	sensitivity	specificity	precision	gmean	F-score
NaiveBayes	78.33	36.76	89.86	83	80.51	80
Random recommendation algorithm	75.00	69.09	85.37	82.86	81.17	80.93
Proposed method	85.26	80	86.18	83.81	82.03	81.86

4.4 Future Directions

Based on the findings of the research experiment and the analysis conducted, we outline potential future directions for further improvement and advancement of the social content recommendation system. This may include suggestions for incorporating additional features or data sources, exploring different algorithms or techniques, or addressing specific challenges or limitations identified during the experiment. The future directions section provides a roadmap for future research inthe field of social content recommendations.

5. Conclusion

Our team has presented the research experiment conducted to evaluate the proposed social content recommendation system. We discussed the experimental setup, including the dataset, evaluation metrics, and methodology. The results obtained from the experiment were presented, followed by a comprehensive analysis and discussion of these results. Finally, we outlined future directions for enhancing the social content recommendation system based on the findings of the experiment.

Reference

1. Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., ... & Wang, Y. (2021). Setting the future of digital and social media marketing research: Perspectives and research propositions. International Journal of Information Management, 59, 102168.

2. Forest, F., Lebbah, M., Azzag, H., & Lacaille, J. (2021). Deep embedded self-organizing maps for joint representation learning and topology-preserving clustering. Neural Computing and Applications, 33(24), 17439-17469.

3. Abkenar, S. B., Kashani, M. H., Mahdipour, E., & Jameii, S. M. (2021). Big data analytics meets social media: A systematic review of techniques, open issues, and future directions. Telematics and informatics, 57, 101517.
4. Terán, L., & Terán, L. (2020). A literature review for recommender systems techniques used in microblogs. Dynamic Profiles for Voting Advice Applications: An Implementation for the 2017 Ecuador National Elections, 27-47.
5. Castillejo, E., Almeida, A., & López-de-Ipina, D. (2012). Social network analysis applied to recommendation systems: Alleviating the cold-user problem. In Ubiquitous Computing and Ambient Intelligence: 6th International Conference, UCAmI 2012, Vitoria-Gasteiz, Spain, December 3-5, 2012. Proceedings 6 (pp. 306-313). Springer Berlin Heidelberg.
6. Al, A., & Amin, M. Z. (2019). An Intuitive Guide of Self Organizing Maps with PracticalImplementation in Minisom.
7. Yang, H. C., Lee, C. H., & Wu, C. Y. (2018). Sentiment discovery of social messages using self- organizing maps. Cognitive Computation, 10, 1152-1166.

8. Han, J., Kamber, M., & Pei, J. (2011). Data mining: concepts and techniques (3rd ed.). Morgan Kaufmann.
9. Relia, K., Akbari, M., Duncan, D., & Chunara, R. (2018). Socio-spatial self-organizing maps: usingsocial media to assess relevant geographies for exposure to social processes. Proceedings of the ACM on human-computer interaction, 2(CSCW), 1-23.
10. Hawas, A. Y., Naser, A. H., & Jalali, M. (2023, September). Location-Based in recommendation system using naive Bayesian algorithm. In AIP Conference Proceedings (Vol. 2845, No. 1). AIP Publishing.
11. Cottrell, M., Olteanu, M., Rossi, F., & Villa-Vialaneix, N. N. (2018). Self-organizing maps, theoryand applications. Revista de Investigacion Operacional, 39(1), 1-22.

12. Rani, S., & Kumar, M. (2020). Social media video summarization using multi-Visual features and Kohnen's Self Organizing Map. Information Processing & Management, 57(3), 102190.

13. Couronne, T., Beuscart, J. S., & Chamayou, C. (2013). Self-Organizing Map and social networks: Unfolding online social popularity. arXiv preprint arXiv:1301.6574.

14. Relia, K., Akbari, M., Duncan, D., and Chunara, R. (2018). Socio-spatial self-organizing maps: Using social media to evaluate relevant geographies' exposure to social processes. Proceedings of the ACM on Human-Computer Interaction, 2 (CSCW), 1–23.
15. Nazari, Z., Kang, D., Asharif, M. R., Sung, Y., & Ogawa, S. (2015, November). A new hierarchical clustering algorithm. In *2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)* (pp. 148-152). IEEE.

16. Shujaaddeen, A., Ba-Alwi, F. M., & Al-Gaphari, G. (2024). A New Machine Learning Model for Detecting levels of Tax Evasion Based on Hybrid Neural Network. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(11s), 450-468.