Vipin Kumar Singh, 2024, 12:5
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

# Enhancing Friend Recommendations in Social Media Networks

**M. Tech Scholar Vipin Kumar Singh, HoD Nagendra Patel**
Department of Computer Science Engineering
Rewa Institute of Technology, Rewa, India

**Abstract- Friend recommendation is a key feature of social networking platforms, designed to connect users with similar or familiar individuals. This concept, popularized by social networks like Twitter and Facebook, often employs a "friends-of-friends" approach, where users are introduced to connections within their extended social circles. Users generally don't connect with random individuals; rather, they form connections within their friends' networks. However, existing recommendation methods are limited in scope and lack efficiency. To address these shortcomings, we propose a new friend recommendation model that enhances accuracy and breadth by utilizing collaborative filtering. This approach allows us to analyze similarities and differences in users' preferences, activities, and interests to deliver more relevant user-to-user suggestions. Additionally, location-based friend recommendation systems are gaining popularity as they bridge the digital and physical worlds, providing a deeper understanding of user interests and preferences. Our model aims to broaden recommendation options by connecting users with shared interests and nearby locations, creating a more meaningful social experience.**

**Keywords- Friend recommendation, collaborative filtering, social network, Recommendation system.**

## I. INTRODUCTION

Friend recommendation is a fundamental service in location-based social network (LBSN) platforms, helping users connect with familiar or like-minded individuals. Approximately 71% of internet users engage with online social networks, a number expected to grow significantly as social networking remains one of the most popular online activities, driven by high interaction rates and mobile advancements. With the rapid increase in smartphone usage, mobile social networks have expanded, providing new features that enhance user engagement. Facebook leads the market in user reach and engagement, with billions of active monthly users [1].

Recent developments in localization techniques have further enhanced social networks by enabling users to share locations and location-related content, giving rise to location-based social networks (LBSNs). LBSNs offer friend recommendations based on users' historical location data, suggesting connections between individuals with similar location histories. Traditional friend recommender systems rely on user profiles, social connections, and interactions, but location data can make recommendations more relevant, as users with overlapping location histories are more likely to have shared interests [2].

While conventional friend recommendation systems exist, few algorithms leverage LBSN data. Early models often relied on GPS data, but check-in information, which is more context-specific, offers richer insights into user behavior. Check-in data, commonly collected by LBSNs, provides a more nuanced view than simple GPS coordinates. Our proposed recommendation model incorporates user profiles, interests, and check-in location histories, employing collaborative filtering to broaden and improve recommendation accuracy [3].

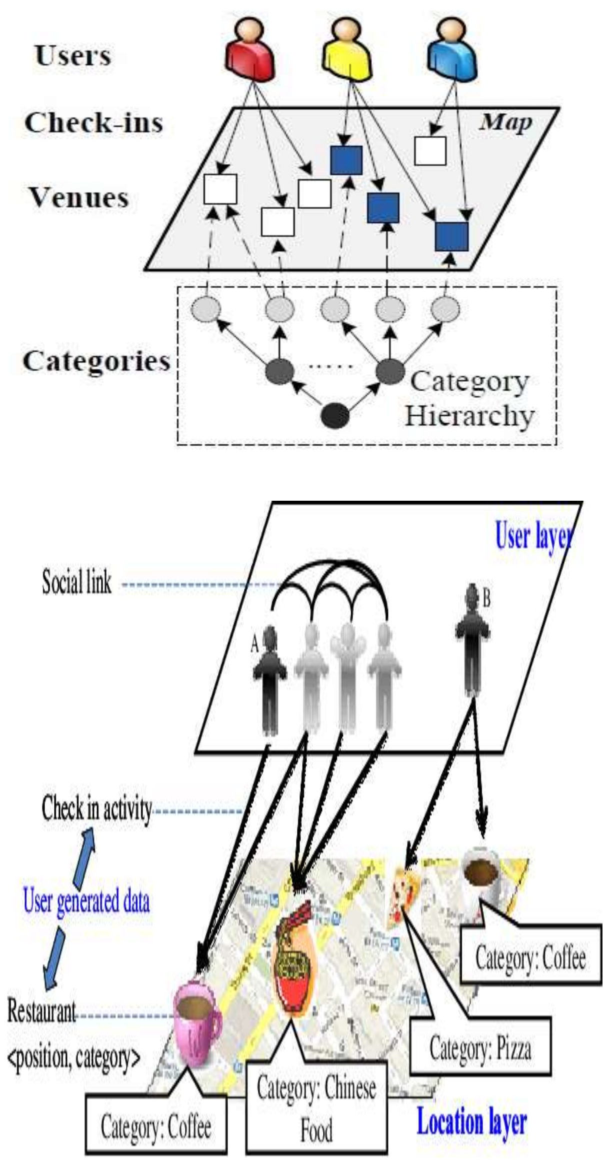

Figure 1: Location Based Social Network

LBSNs are not merely about appending location data to traditional social networks; they introduce a new social structure based on real-world location interdependencies. Individuals connect through shared location-tagged content, such as text, images, and videos, creating unique social relationships shaped by physical proximity and location histories [4]. In this framework, an individual's location history over time, along with their interests and activities inferred from location data, becomes a significant factor in building meaningful social connections within LBSNs [5].

LBSN is consists of a G <U, C> and social network G <U, E >. In G <U, E> U is the set of users and E is the set of edges which connects or indicates a social connection between different users in LBSN. In G <U, C> Check_in ‘c’ belongs to set C and shows user ‘u’ belongs to set U has a check in activity at location l at time t [2].

## II. RELATED WORK

Temporal, spatial and social correlation are three main attributes of any LBSN. However, the situation which include these three features cannot be solved in previous algorithms. There is no method which utilizes all information properly A new approach of friend recommendation is proposed, which aims to recommend friends with similar location preference for LBSN's users. This approach first, use the method of local random walk based on Markov chain to calculate the user's friendship similarity on social network. Second, it calculates the user’s location preference similarity in the real world based on check-in data and finally recommend friends to users by building a mixed user preferences model [6].

A new friend recommendation model (FE-ELM), is proposed where friend recommendation is regarded as a binary classification problem. In this model first feature extraction is done by using different strategies and then in training process ELM is selected as classifier to learn the the spatial-temporal feature, social feature, and textual feature, finally experiments are performed on real datasets for better efficiency and accuracy [7].

The new properties and challenges that location brings to recommender systems for LBSNs are discussed in this paper. First, author has categorized the recommender systems by the objective of the recommendation, which include locations, users, activities, or social media. Second, they categorize the by the methodologies employed, including content-based, link analysis-based, and collaborative filtering. Then finally, classify the systems by the data sources used, including user profiles, user online histories, and user location histories. For each category, the goals and contributions of each system are summarised and highlights the representative research effort. It introduces the concepts, unique properties, challenges, evaluation methods and future work for recommender systems in LBSNs [8].

Hierarchical-graph-based similarity measurement (HGSM) framework is proposed here, which models people's location histories and determines the similarity between users. In this framework, 3 factors sequence property of users' movements, Hierarchy property of geographic spaces, Popularity of different locations are considered. Using HGSM to estimate the similarity between users, a collaborative filtering-based method is also employed in our system to find an individual's interest in unvisited geospatial regions [9].

A friend recommendation algorithm is proposed which is known as Random walk-based context-aware friend recommendation algorithm (RWCFR). This model uses an undirected un-weighted graph that represents users, locations, and their relationships. RWCFR constructs a sub-graph according to the user's present environment. Popular users and famous places in region are added to this sub- graph. After constructing the sub-graph, this sub-graph is given as input to algorithm, and it calculates the recommendation possibilities of users for suggesting becoming potential friend. A list of potential friends is generated according to output of the random walk algorithm [10].

Recommendation system make use of user profile, friend description and past behaviour for recommendation but no attention has been given to personalization based explicitly on social networks. Author has used information such as social graph among users, tracks & tags from last.fm social network which effectively incorporates bonds of friendship. We have done number of experiments between the Random Walk with Restarts model and user-based collaborative filtering model. The results prove that the graph model gains from the additional information implanted in social knowledge [11].

The paper analyzes the main challenges of the collaborative filtering algorithm and provides several solutions. To solve cold start problem for the new user, we could replenish user's profile in different ways, the general approach is to require user provide their profile while login the social account and for the new friend, we could combine the collaborative filtering and content-based recommender algorithm. There are few solutions for the sparsity problem. The first one uses filling or decreasing the dimension to decrease the sparsity of the matrix. Another solution improves the efficiency of the algorithms without changing the sparsity of the matrix. [12].

Recent advancements in friend recommendation systems have focused on improving accuracy and relevancy by leveraging various approaches in social media. Kung et al. (2024) introduced an embedding-based retrieval method using Approximate Nearest Neighbor (ANN) query expansion to enhance friend recommendations through refined data retrieval strategies [14]. Similarly, Su (2024) explored product recommendations within social circles, addressing challenges related to inadequate labeling and the social context of user networks [15]. Ding et al. (2024) proposed a meta-path-aware dynamic graph learning model to integrate user mobility, enriching friend suggestions by analyzing temporal connections in dynamic networks [16]. Ramakrishna et al. (2023) introduced HCoF, a hybrid collaborative filtering technique that combines social and semantic elements to boost recommendation accuracy [17]. Furthermore, Alshammari and Alshammari (2023) implemented a collaborative filtering engine specifically tailored for Facebook, demonstrating how personalized recommendations can be fine-tuned within established social networks. These studies illustrate the potential of combining multiple data sources and advanced modeling techniques to refine friend recommendation systems in social media [18].

## III. PROBLEM DOMAIN

The traditional collaborative filtering recommendation algorithm is having lack of accuracy and efficiency as this uses formal method of filtering which makes it inefficient to use at alone. In terms of recommendation made by the collaborative filtering algorithm it may be concluded that the algorithm needs many more improvements.

By implementing traditional collaborative filtering recommendation algorithm, we get less accuracy which makes it typical to use and inefficient to apply on huge datasets i.e. Big Data. Dealing with big data the less accuracy makes it inappropriate and less accurate. As applying this algorithm on huge amount of data in real world applications the less accuracy will not be efficient for making recommendations to users. The numbers of attributes which are available are totally considered for extracting information to recommend friends to users which makes the collaborative filtering recommendation algorithm inefficient. Also, the higher the number of attributes used to make recommendations, results in higher computing time and higher number of comparisons to be made. The overall dimensions included for making recommendation should be removed as per the requirement.

Apart from this the k-means clustering applied previously with the collaborative filtering algorithm can be replaced by different clustering technique. There are some drawbacks that can be seen in the k-means clustering technique which may be overcome by replacing this clustering technique with the newer one. In the k-means clustering the numbers of the clusters that should be made need to be defined at the start of the algorithm which makes it inefficient to use if the numbers of the clusters are not properly defined.

One more thing to be noted, that is the dimensionality of the given dataset should be less in number to lower the comparisons that will be made at the time of execution. The more the number of the dimensions to evaluate the results, makes the accuracy lesser and requires more time to make recommendations to the user. Hence to reduce the number of attribute or the dimensionality of the dataset is major task.

## IV. PROPOSED METHOD

The problem observed in the previous algorithm can be removed by replacing the existing techniques by newer techniques. As in the previous, the algorithm combines the K-means clustering technique with the PCA as dimensionality reduction technique. Combining both this techniques in the collaborative filtering algorithm was a solution proposed earlier by the authors.

Here we have proposed a better clustering technique as compared to the k-means clustering, while keeping the PCA as earlier it was used. The k-means clustering can be replaced by the hierarchical clustering as it is better clustering technique to work on. The PCA will be used as the dimensionality reduction technique to decrease the dimensionality of the data.

The Hierarchical clustering will provide better results in comparison to the k-means clustering, as stated that in hierarchical clustering there is no need to define the number of clusters at the beginning of the clustering. Defining the required number of clusters after applying the hierarchical clustering will make it feasible to break the clusters as per the dataset. But before applying the clustering technique on the dataset the dataset should be improved. If the Input to the algorithm will be accurate then the obtained output will be more efficient. So, to improve the input dataset the dimensionality reduction should be done and to do this the PCA have to be applied on the dataset.

In final words we are going to apply the PCA on the dataset before giving it as input and after getting the principal components this are given as input to the hierarchical clustering. The collaborative filtering algorithm will firstly perform the PCA and after that the hierarchical clustering is applied and the final recommendations are made. Hence in this way the collaborative filtering algorithm can be improved and the recommendations can be made accurate.

### 1. Algorithm For Proposed Approach

The proposed algorithm using both the techniques, the first one is the PCA which will help in reducing the dimensions of the given dataset and the second one is the clustering technique which is the hierarchical clustering. Here in our algorithm we are applying the PCA at first because it will reduce the dimensions of data and after that the hierarchical clustering will be performed on the obtained principal components. The working algorithm is as follows:

- Step 1: Data collection - collect the friend related data like name, rating etc. in the form of csv file.
- Step 2: Data pre-processing - perform manual data analysis and eliminate the feature which is less correlate to another feature.
- Step 3: Perform PCA (principal component analysis) on the data and save the data in to csv file.
- Step 4: Define hierarchical clustering (agglomerative) model.
- Step 5: Train the hierarchical clustering (agglomerative) model on the data.
- Step 6: Take the one user input and apply PCA on that.
- Step 7: Perform the prediction in the input it gives the cluster id.
- Step 8: Fetch all the friend detail which belong to this cluster id and make the list of it.
- (This list is recommended friend list)

### 2. Flowchart of the Proposed Approach

Below we have given the flowchart for the proposed approach which will help in understanding the flow of the steps performed:

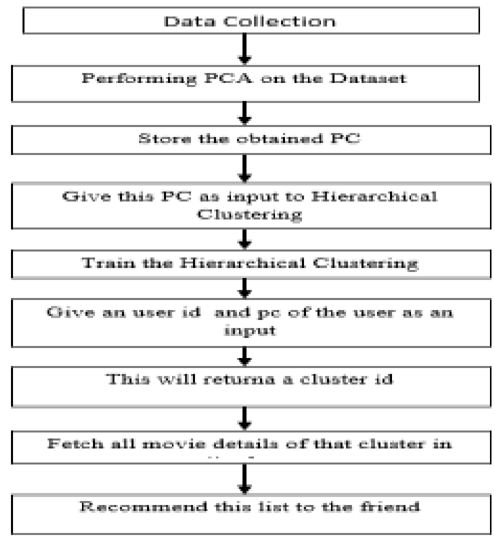


Figure 2: Flowchart for the proposed system

### 3. Pearson Correlation Coefficient

The Pearson correlation coefficient is a formula that measures the strength between variables and relationships. It is very helpful statistical formula is often referred to as the 'Pearson R' test. Whenever we want to find how strong relationship is between two variables, it is a good idea to apply a Pearson correlation coefficient test.

### Formula

In order to see how strong, the relationship is between 2 variables, a formula must be followed to produce what is referred to as the coefficient value. The coefficient value varies between -1.00 and 1.00. If the coefficient value is – ve, then it means the relationship between the variables is negatively correlated, and if the value is + ve, then it shows variables are positively correlated, or both values varies together either increase or decrease.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

**Pearson Correlation Coefficient Formula**

Note: The above examples only use data for 3 people, but the ideal sample size to calculate a

Pearson correlation coefficient should be more than 10 people.

A comparision between different similarity calculation techniques is also discussed here which suggest why we have chosen pearsons correlation.

Suppose we have 2 vectors x & y and we want to measure the similarity or degree of closeness between them. A basic similarity function is the inner product

$$\text{Inner}(x, y) = \sum_i x_i y_i = \langle x, y\rangle$$

If x tends to be high where y is also high, and low where y is low, Higher the inner product vectors are more similar. The inner product is unbounded. A way to make it bounded between -1 and 1 is to divide by the vector's L2 norms which results in giving the cosine similarity.

$$\text{CosSim}(x, y) = \sum_i x_i y_i / \sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2} = \langle x, y\rangle / \|x\|\,\|y\|$$

This is bounded between 0 and 1 if x and y are non-negative. Cosine similarity is not invariant to shifts/change. If x was shifted to x+1, the cosine similarity would change. Pearson correlation is invariant. Let X and Y be the respective means:

$$\text{Corr}(x, y) = \sum_i (x_i - X)(y_i - Y) / \sqrt{\sum (x_i - X)^2}\,\sqrt{\sum (y_i - Y)^2}$$

Correlation is the cosine similarity between centered value of x and y i.e mean value, it is also bounded between -1 and 1. People generally think about cosine similarity in terms of vector angles, but it can be not be used as a correlation, if you think of the vectors as paired samples then correlation is invariant to both scale & location changes of x and y.

## V. EXPERIMENTAL RESULTS AND EVALUATION

### 1. Data Set Processing and Experimental Result

In this section, we implemented set of experiments that show for evaluating the impact of proposed system on recommendation. We have done different experiments on the Friend data set. In currently, we have a tendency to perform experiments on move choice knowledge collected from the friend recommendation web-based recommender system. The information set contained 600,000 choices from 824 users and one, 50 friends, with every user choice a minimum of twenty things on more details table 1 and figure 3.

Table 1 Data Set Attributes

| FriendID | Sex | Age | Location | Category | FriendChoice | QualityIndex | FriendshipMode | Discussion |
|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 24 | 5 | 4 | 4 | 5 | 2 | 10 |
| 2 | 1 | 28 | 4 | 1 | 3 | 7 | 1 | 15 |
| 21 | 1 | 29 | 5 | 8 | 8 | 7 | 0 | 20 |
| 4 | 0 | 27 | 8 | 0 | 8 | 6 | 5 | 23 |
| 5 | 0 | 38 | 5 | 7 | 8 | 7 | 1 | 5 |
| 6 | 1 | 30 | 4 | 3 | 3 | 5 | 2 | 20 |
| 7 | 1 | 33 | 8 | 2 | 6 | 6 | 2 | 50 |
| 8 | 1 | 48 | 7 | 6 | 2 | 7 | 0 | 35 |
| 29 | 1 | 45 | 8 | 1 | 8 | 6 | 1 | 20 |
| 10 | 1 | 21 | 2 | 3 | 8 | 7 | 3 | 20 |
| 11 | 0 | 24 | 7 | 8 | 6 | 4 | 3 | 13 |
| 12 | 1 | 34 | 8 | 4 | 8 | 6 | 0 | 20 |
| 13 | 1 | 18 | 3 | 9 | 7 | 7 | 3 | 20 |
| 24 | 0 | 37 | 5 | 6 | 6 | 6 | 2 | 10 |
| 15 | 1 | 27 | 8 | 5 | 7 | 9 | 2 | 20 |
| 16 | 1 | 19 | 7 | 0 | 3 | 5 | 3 | 40 |
| 17 | 1 | 38 | 4 | 7 | 4 | 6 | 2 | 15 |
| 26 | 0 | 40 | 3 | 6 | 6 | 5 | 6 | 10 |
| 19 | 1 | 43 | 0 | 5 | 8 | 7 | 0 | 10 |

```
FriendID
Sex
     0- male
     1- female
Age
Location  0 - zONE0
          1 - zONE1
          2 - zONE2
          3 - zONE3
          4 - zONE4
          5 - zONE5
          6 - zONE6
          7 - zONE7
          8 - zONE8
Category  0 - A++
          1 - A+
          2 - A
          3 - B++
          4 - B+
          5 - B
          6 - C++
          7 - C+
          8 - C
FriendChoice - Range (1-10)
QualityIndex- Range (1-10)
FriendshipMode
               0 - family
               1 - family friend to friend
               2 - school frend
               3 - other friend
```

Figure 3 Cleaning of Friend Dataset

In this figure 3 cleaning of friend dataset. During cleaning we have clean all attributes like sex (0 to male and 1 to female); locations are dividing into zone wise 0 to 8, friend category is dividing into 0 to 8 and etc.

```
C:\Users\ths\Desktop\Final 1>python main.py
main.py:5: DeprecationWarning: time.clock has been deprecated in Python 3.3 and will be removed from Python 3.8: use ti
e.perf_counter or time.process_time instead
  start_time = time.clock()
  FriendID      Sex      Age  Location  Category  FriendChoice  QualityIndex  FriendshipMode  Discussion
0 -0.670433 -1.358585 -0.536815 -0.281032 -0.114765     -1.565208     -1.247735        0.401634   -0.942108
```

Figure 4 Display Data Set Attribute and it's Calculate Exaction Time

In this figure 4 display all attribute on given data set like FriendID, sex, age, location, category, friend choice, quality index, payment mode and discount. Total exaction time taken 7.48 seconds.

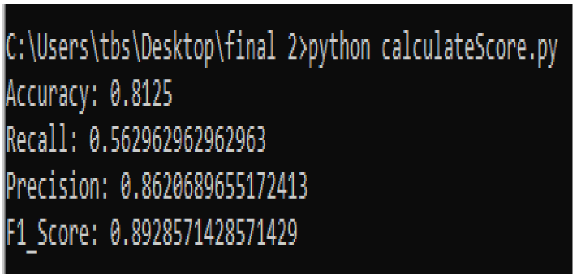


Figure 5 Display Accuracy, Recall, Precision and F1-Score on given Data Set

In this figure 5 displays performance on given data set. Accuracy, recall, precision and f1-score are 81.25 %, 90.90%, 86.95 % and 86.95%.

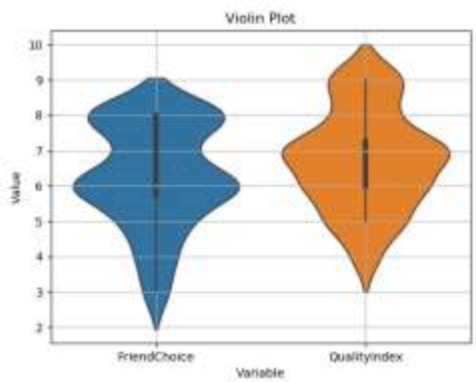


Figure 6 Violin plot between Friend Choice and Quality Index

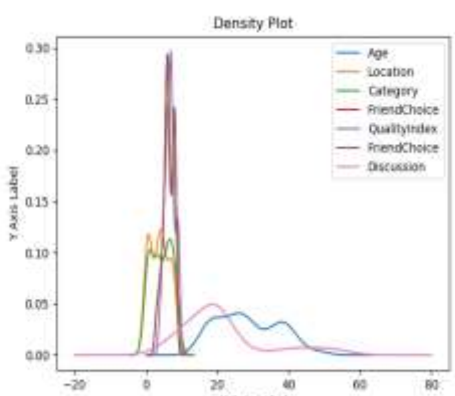


Figure 7 Density Plot in all Attributers on Given Data Set

In this figure 7 displays density plot in all attributers on given data set attributes like age, location, category, friend choice, quality index, payment mode and discount.

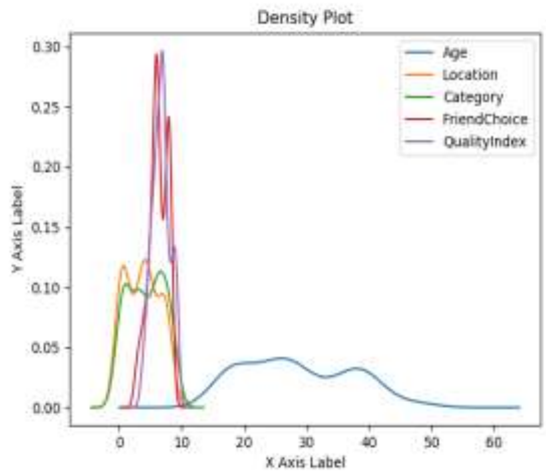


Figure 8 Density Plot in Some Attributers on Given Data Set

In this figure 8 displays Density Plot in Some Attributers on Given Data Set attributes like age, location, category, friend choice and quality index.

Developing a solution is an approach proving mechanism but to prove its results is a complicated task because it measures each and every step of the solution and let it compare with the existing mechanisms. So as to do that effectively this chapter gives a detailed result analysis to prove effectiveness of the suggested mechanism.

For making the analysis of the proposed approach we have used the Kaggle dataset the data about friends is taken from the Kaggle dataset and the friend_likes pattern and user details are combined from the Kaggle dataset. The experiment was carried out to evaluate the accuracy of the recommendations produced by the algorithm we have proposed in our paper. The accuracy term is calculated in this experiment by which the comparison between the proposed and the existing algorithm can be made.

We are applying this data on the previous collaborative algorithm with pca and k-means and the results are obtained, so the accuracy of the previous algorithm is calculated.

Accuracy = ({Relevant Document} intersection {Retrieved Document} / {Relevant Document}) *100
Now the proposed algorithm with hierarchical clustering is taken for analysis. The collaborative filtering algorithm along with pca and hierarchical clustering is analyzed over the same data. This algorithm's accuracy is compared with the existing algorithm.

The experiment clearly results in an increase in the accuracy of the recommendations made by our proposed algorithm. The results are compared between both the algorithms using k-means clustering with pca and hierarchical clustering with pca in terms of accuracy are shown in the following graph:

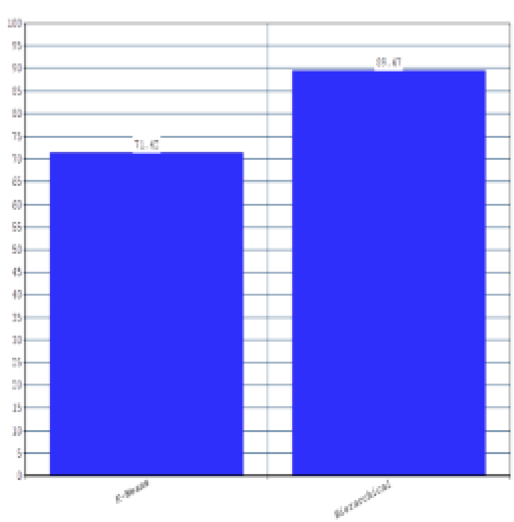


Figure 9: Accuracy results for both the algorithms

The Fig.9 clearly concludes that the proposed hierarchical clustering works much better as compared to the previously used k-means clustering. The results in terms of accuracy of the proposed algorithm is higher than the earlier clustering technique. So it is better to use the Hierarchical clustering with pca on the collaborative filtering algorithm as compared to earlier one.

## VI. CONCLUSION AND FUTURE WORK

The proposed research investigates friend recommendations generated by the system, utilizing a hierarchical clustering approach combined with Principal Component Analysis (PCA) to enhance accuracy assessment. System accuracy is evaluated by examining the overlap between recommended friends and previous friend "likes" made by the user, providing a measure of how well the recommendations align with the user's historical preferences. Experimental results demonstrate improved performance over previous algorithms, highlighting the effectiveness of the hierarchical clustering and PCA combination.

Future research could extend these findings by employing additional datasets, testing metrics beyond accuracy, and exploring alternative clustering techniques to further refine the recommendation algorithm. This approach could lead to a more versatile and robust system capable of adapting to varied user data and recommendation contexts.

## REFERENCES

1. Haruna K, Ismail MA, Damiasih D, Sutopo J, Herawan T. A collaborative approach for research paper recommender system. Plos One. 2017, 12(10):e0184516.https://doi.org/10.1371/journal.pone. 0184516 PMID: 28981512
2. Rojas G, Garrido I. Toward a rapid development of social network-based recommender systems. IEEE Latin America Transactions. 2017, 15(4):753–759.
3. Huang S, Zhang J, Wang L, Hua XS. Social Friend Recommendation Based on Multiple Network Correlation. IEEE Transactions on Multimedia. 2016, 18(2):287–299.
4. Corbellini A, Mateos C, Godoy D, Zunino A, Schiaffino S. An architecture and platform for developing distributed recommendation algorithms on large-scale social networks. Journal of Information Science. 2015, 41(5):686–704.

5. Fields B, Jacobson K, Rhodes C, Inverno M, Sanler M, Casey M. Analysis and Exploitation of Musician Social Networks for Recommendation and Discovery. IEEE Transactions on Multimedia. 2011, 13 (4):674–686
6. Chamoso P, Rivas A, Rodrı ´guez Sara, Bajo J. Relationship recommender system in a business and employment-oriented social network. Information Sciences. 2018, s 433–434:204–220.
7. Guo G, Zhang J, Zhu F, Wang X. Factored similarity models with social trust for top-N friend recommendation. Knowledge-Based Systems. 2017, 122:17–25.
8. Zhang Z, Liu H. Social recommendation model combining trust propagation and sequential behaviors. Applied Intelligence. 2015, 43(3):695–706.
9. Maier C, Laumer S, Eckhardt A, Weitzel T. Giving too much social support: social overload on social networking sites. European Journal of Information Systems. 2015, 24(5):447–464.
10. Lee S, Koubek RJ. The effects of usability and web design attributes on user preference for e-commerce web sites. Computers in Industry. 2010, 61(4):329–341.
11. AndreasenT, Jensen PA, Nilsson JF, Paggio P, Pedersen BS, Thomsen HE. Content-based text querying with ontological descriptors. Data & Knowledge Engineering. 2004, 48(2):199–219
12. Huang S, Zhang J, Wang L, Hua XS. Social Friend Recommendation Based on Multiple Network Correlation. IEEE Transactions on Multimedia. 2016, 18(2):287–299.
13. Kung, P. P. H., Fan, Z., Zhao, T., Liu, Y., Lai, Z., Shi, J.,... & Venkataraman, G. (2024, July). Improving Embedding-Based Retrieval in Friend Recommendation with ANN Query Expansion. In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 2930-2934).
14. Su, N. (2024). Promoting by Looking into Friend Circles: An Inadequately-labeled and Socially-aware Financial Technique for Products Recommendation. IEEE Access.
15. Ding, D., Yi, J., Xie, J., & Chen, Z. (2024). Meta-path aware dynamic graph learning for friend recommendation with user mobility. Information Sciences, 666, 120448.
16. Ramakrishna, M. T., Venkatesan, V. K., Bhardwaj, R., Bhatia, S., Rahmani, M. K. I., Lashari, S. A., & Alabdali, A. M. (2023). HCoF: Hybrid Collaborative Filtering Using Social and Semantic Suggestions for Friend Recommendation. Electronics, 12(6), 1365.
17. Alshammari, M., & Alshammari, A. (2023). Friend recommendation engine for Facebook users via collaborative filtering. International Journal of Computers Communications & Control, 18(2).