Computer Fraud and Security
ISSN (online): 1873-7056

Efficient and Fine-grained Profile Matching with Privacy Preserving in Mobile Social Networks

Rohini Bhosale[1], Vanita Mane[2]

[1]Department of Computer Engineering, Ramrao Adik Institute of Technology, D Y Patil Deemed to be University, Nerul, India.

Department of Computer Science &Engineering (ICB), Dwarkadas Jivanlal Sanghvi College of Engineering, Mumbai, India. rohini.bhosale@djsce.ac.in

[2]Department of Computer Engineering, Ramrao Adik Institute of Technology, D Y Patil Deemed to be University, Nerul, India. vanita.mane@rait.ac.in

Abstract: The proliferation of the internet and smart devices has dramatically increased the user count of Mobile Social Networks (MSNs). Connecting with new users is the core feature of MSNs. The increased number of active users on MSNs highlights the need for efficient and effective profile-matching and recommendation mechanisms. This paper introduces a novel framework for dynamic friend matching and recommendation by using advanced classification and clustering techniques. To improve matching results in the user connection as per social characteristics and behavior, the suggested framework combines ideal clustering using the Outline Coefficient and Davies-Bouldin Record with K-Means clustering. We propose a hybrid fine-tune ensemble classifier that blends the best aspects of the Random Forest and XGBoost methods. The hybrid model of grouping users has improved the accuracy of friend recommendations. A 95.23% classification accuracy is obtained by the ensemble classifier by combining the strong feature selection and averaging of Random Forest with the gradient boosting of XGBoost. The Davies-Bouldin Index and Silhouette Coefficient are applied to the clustering step to find the ideal number of clusters to provide cohesive and well-separated groups. These well-defined groups are then used in the K-Means algorithm, providing a basic structure for future matching and recommendation processes. Experimental results have demonstrated the efficiency of the proposed hybrid model.

Keywords: Social Network, Friend Matching, Recommendation System, Clustering Techniques, Ensemble Classifier.

1. Introduction

Mobile social networks (MSNs) have become an essential part of life. All diverse users use these networks to communicate and connect with friends and relatives. There are 5.44 billion internet users globally, of which 5. 07 billion users use various social media platforms [22]. These platforms provide different facilities to users apart from increasing popularity aspects. Social media giants like Facebook and Instagram allow users to plug in and communicate anywhere and anytime. They are mainly used to link the individual with an already or a prospective friend based on common interests. In this technological-driven world, the efficiency of matching the individual with the probable friend and suggesting the right users is significant for enhancing the user's experience and engagement. Effective friend matching and recommendation systems make it certain that users can explore new connections and the content that interests them, even becoming active in it, improving satisfaction and retention by users. However, since the number of users is growing drastically, it is challenging to match the user profile and identify possible friends for the user.

Many researchers have focused on and proposed solutions for these issues. Profile matching methods can be categorized as centralized, distributed, and hybrid. In a traditional centralized architecture, all user data is stored and processed on the central server. It is easy to implement and manage. However, it suffers from single-point failure. The whole system fails if the adversary attacks the server or fails due to technical issues. To avoid this,

many researchers have proposed a distributed architecture. User details are distributed among many servers, and in collaboration with multiple servers, the user's preferred matching friend is suggested to the user. The limitation of distributed architecture is that it is vulnerable to security threats. The user's privacy is at stake if anyone on the server colludes with or becomes an attacker. A few researchers have proposed a hybrid architecture to avoid these issues of centralized and distributed architecture. In hybrid architecture, the user data is stored in encrypted form at the centralized server. Multiple servers participate in the friend match process. The centralized server verifies the result of the profile matching. The centralized server verifies the match results before connecting the potential friends.

Another classification of profile matching is the contents used to match the profile. Traditional methods depend on basic similarity measures and a heuristic approach. These methods do not consider the user's deeper behavioral patterns. User interactions are complex and dynamic, so traditional methods may not adequately capture them. Subsequently, users often receive improper friend matches and recommendations due to a lack of accuracy and efficiency in the profile-matching method. This gives users an unpleasant experience and reduces their engagement on the platform.

These platforms require more sophisticated methods that can adequately address the limitations of the existing profile-matching methods to identify complex patterns in user behavior and social interaction accurately. Profile matching would be accurate if the identified patterns were correct. Heuristic approaches and basic similarity computations are unsuitable for modern mobile social networks w as they always change rapidly by user interactions and preferences. It is, therefore, mandatory that sophisticated ways are discovered to carefully peruse difficult user information and find ways of overcoming such problems in order to make sending friend requests easier and find recommendations that are not only more accurate but also seem more personal.

This paper introduces a new method for dynamic friend matching and recommendations in mobile social networks. The proposed system integrates K-Means clustering with optimal clustering techniques using the Silhouette Coefficient and Davies-Bouldin Index. This integration enhances the accuracy of user grouping based on their social characteristics and behavior. The Hybrid Fine-Tune Ensemble Classifier is central to the proposed method, which combines the strengths of XGBoost and Random Forest algorithms. The aim of the integration is to improve the performance of the profile-matching system by making it more accurate and efficient. It takes profile-matching and recommendation systems to raise user engagement and pleasure on social networks. These systems suggest friends that correspond with the social habits and inclinations of the users [5], [6].

The proposed system overcomes the shortcomings of traditional methods using different key elements. Firstly, people are classified depending on their social attributes using K-means clustering. To ensure these clusters are distinct, the Davies-Bouldin index and the silhouette coefficient are calculated to find an optimal number of clusters. Groups are assessed for minimization and division by the Davies-Bouldin File and attachment and partition by the Outline Coefficient, which guarantees the strength and uniqueness of succeeding groups [7, 8].

The complex patterns of user behavior may escape the grasp of heuristic methods or basic similarity metrics in conventional friend-matching algorithms. These constraints have shown potential to be overcome by contemporary machine learning approaches, especially ensemble methods. There are further problems with data distribution and processing efficiency, nevertheless, when these techniques are included into the suggested architecture [9, 10]. The main objective of this work is to provide a framework for dynamic buddy matching and recommendation that maximizes efficiency and accuracy by utilizing state-of-the-art machine learning methods. In particular, this structure uses bunching and characterization approaches to group customers according to their social characteristics and behaviors. Then it uses these groups to provide friends with individualized recommendations.

Characteristics and clustering are the two phases of the innovative proposed framework. To identify the correct number of clusters, we rely on the K-Means calculation, which uses the Silhouette Coefficient and Davies-Bouldin Index. This ensures the groupings are rigid and geographically separated, which makes the arrangement that results more precise. This paper presents a Hybrid Fine-Tune Ensemble Classifier during the classification stage incorporating the Random Forest and XGBoost algorithms. This half-and-half approach combines the slope-supporting powers of XGBoost with the strong component selection and averaging capabilities of Irregular Timberland to produce arrangement exactness never seen before. The proposed method improves the profile-

matching and recommendation process in the MSNs. A few important elements of the proposed system are as follows:

- Pre-processing and data collecting: The several network nodes capture user data, which includes social interactions, interests, and behavioral patterns. Preprocessing of this data to remove noise and normalize designs ensures framework consistency.
- The Best Grouping: The K-Means clustering technique groups the users. Utilizing the Davies-Bouldin Index and the Silhouette Coefficient, the optimal number of clusters is found. These measures gauge the cohesion between clusters, assessing the clustering quality. Through optimization of these indices, the system ensures that users are positioned in well-defined clusters that fairly represent their social behavior.
- Fine-Tune Ensemble Hybrid Classifier: In the proposed system, an XGBoost and Random Forest algorithms are combined into a hybrid ensemble model to categorize users when they are crowded together. Classifier's accuracy is improved by utilizing the best of both the algorithm. XGBoost offers excellent prediction accuracy via gradient boosting, whereas Random Forest offers strong feature selection and lowers overfitting by averaging.

The proposed system provides a major development in the MSNs. Our significant contributions in profile-matching and recommendation are as follows:

- Enhanced level of safety and confidentiality- The proposed system prevents data leaks and improper use of personal data due to distribution storage and data processing, so there is no chance for one entity to have everyone’s personal information in one place. It solves major problem which is characteristic for usual social networks, that is they are very centralized, and everything is under control of one authority
- Increased Accuracy and Productivity: Using advanced grouping and characterization approaches completely improves the accuracy of companion coordination and recommendation. The approach's viability is demonstrated by the 95.23% characterization precision achieved by the crossbreed Calibrate Outfit Classifier.

The proposed system focuses on the problems of accuracy, scalability, and privacy related to MSN profile-matching and recommendation algorithms. The purpose of integrating sophisticated grouping and order techniques is to improve the user experience while maintaining information security. This innovative approach reaches 95.23% grouping accuracy using the Outline Coefficient, Davies-Bouldin List, K-Means bunching, and a Crossover Tweak Outfit Classifier. Our study closes the current gaps and lays the groundwork for the next social network applications.

2. Literature review

Profile matching and efficient recommendation systems are necessary in the current era. They are essential in helping users find connections and content, which has risen because of rapid growth in MSNs. Over time, different studies have been carried out to look at different techniques to improve the precision and efficiency of such tools, with the most emphasis on machine learning and data analytics. This paper contributes to reviewing research work in secure and efficient profile matching, user account matching, AI-based cybersecurity, and feature selection, and it also points to limitations of the existing approaches and the research vacancies our proposed framework intends to close.

Profile Recognition and Fake Profile Detection:

Cybercrime on social networks has increased drastically. Various research has shown that nowadays, many cybercrimes are social network cybercrime or social network-assisted cybercrime, and identity theft is the most frequent. The identity of the user may leaked during the profile-matching process. Also, it is accessed by an unauthorized entity if a user is a friend of the fake user. It is often noted that fake profiles are used to perform crimes. This has highlighted the need to identify fake profiles. The profile-matching process may leak the user's private data to the attacker. Identity theft is one of the most common crimes that malicious users commit. Using big data analytics, Awan et al.[11] developed a method for recognizing fake profiles to identify fraudulent accounts that can harm user experience and network trust. In a similar way, Roy et al.[12] have done an extensive analysis of techniques used in spotting phony accounts, showing how important it is to have strong algorithms to help distinguish real ones from false ones. Modern techniques struggle with the ability to manage and classify data that is large and may need real time analysis; this has been noted despite some advancements made within this field.

Profile Matching and Privacy Protection:
Regarding profile-matching in social networks, the most important thing is to protect privacy. Gao et al. [13] developed a mobile social network cloud-based security-enhanced profile-matching method where more than one key is utilized. Looking at the impact on privacy, this technique has two sides of the same coin. However, the scalability of this technique can be easily blocked by the centralized cloud storage on which it relies. The approach was detailed by Yi et al. [14] for constructing a social network based on privacy-preserving user profile matching that enables accurate matches without revealing personal information about the user. Nonetheless, these techniques do not completely cover the decentralization issue where records are distributed across various nodes on the internet.

Cross-network Profile Matching:
Another area of interest is matching user accounts across various social networks. A method that matches accounts across platforms using usernames and display names was developed by Li et al.[15]. In this paper, the authors demonstrate the potential for cross-network integration. In [16], Nurgaliev et al. have addressed the limitations of matching users across networks with limited profile data. Authors have proposed solutions for situations in which comprehensive user information was unavailable. These studies focus on the importance of precise account matching but usually involve high computational cost.

An Artificial Intelligence-Based Cybersecurity Algorithm:
Logeshwaran et al. [17] have proposed a cooperative peer-to-peer (P2P) file sharing in social networks. The main objective of this system is enhancing security using AI, but it is mostly designed to work only in P2P networks but not for general social networking functionality. In their analysis of cross-regional friendship inference, Ren et al. [18] harnessed mobility profiles to show how user movement data can underpin friend recommendations, but these methods often exclude exhaustive ensemble learning schemes for higher precision, notwithstanding the exuberance for AI as a good thing to achieve in social network performances.

Matchmaking and Feature Selection:
For efficient matchmaking in social networks, the selection of characteristics is significant. Rodriguez et al. [19] established how user attributes could enhance matching precision. As Paul et al. [20] surveyed users' perception of AI and their attitude toward matchmaking software, it was found that perception bias, trust, and satisfaction levels towards the technology by users were the main factors involved in using such systems. By integrating search behavior and friend networks, Zhou et al. [21] investigated personalized search, highlighting the significance of personalized recommendations even more.

The current collection of exploration has taken huge steps in further developing profile-matching and proposal frameworks through different creative strategies. However, a few exploration holes remain. Notably, many studies use centralized architectures, which can be problematic for privacy and scalability. In addition, despite cutting-edge algorithms like artificial intelligence (AI) and machine learning, integrated strategies combining clustering and classification methods are required to maximize environment performance.

Our proposed framework uses an architecture to improve user privacy and scalability to fill these gaps. By coordinating ideal grouping utilizing the Outline Coefficient and Davies-Bouldin Record with K-Means clustering and utilizing a Half-breed Tweak Troupe Classifier that joins Irregular Timberland and XGBoost, our strategy plans to accomplish unrivaled exactness in companion coordination and suggestions. This novel strategy overcomes the drawbacks of previous approaches and paves the way for future advancements in mobile social network applications.

3. Methodology

The methodology for our research involves a comprehensive approach to enhance friend matching and recommendations in mobile social networks, as shown in Figure 1. We employ advanced data preprocessing, clustering, and classification techniques to achieve this goal. Initially, user data is gathered and refined, followed

by thorough text processing to clean and normalize the data. We then apply clustering techniques, specifically K-Means, optimized through the Silhouette Coefficient and Davies-Bouldin Index, to identify distinct user groups. We introduce a Hybrid Fine-tune Ensemble Classifier, combining Random Forest and XGBoost, to improve classification accuracy. This integrated methodology ensures robust, scalable, and privacy-preserving friend recommendations in mobile social networks.

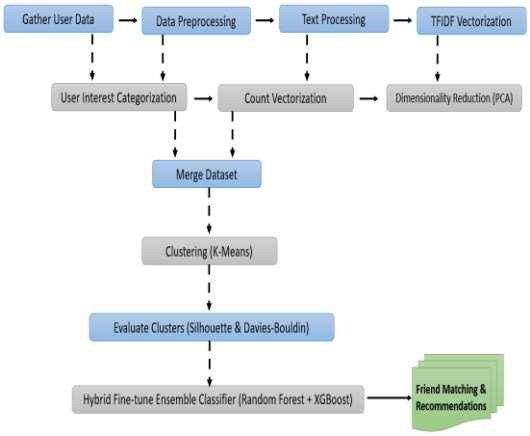

Figure 1 Proposed methodology

3.1. Data gathering

The user data gathering process began with the collection of user bios, followed by categorizing interests into groups like Movies, TV, Music, Sports, Politics, and social media. Subsequently, the user bios were combined with their categorized interests to produce a comprehensive dataset, saved as "profiles.pkl". Further enhancement of the dataset involved generating random values for each interest category. The resulting dataset, now enriched with generated values, was saved as "refined_profiles.pkl", as depicted in Figure 2.

	Bios	Movies	TV	Religion	Music	Sports	Books	Politics
0	Typical twitter fanatic. Infuriatingly humble ...	5	3	4	1	3	6	7
1	Web junkie. Analyst. Infuriatingly humble intr...	7	9	5	1	9	4	0
2	Avid web maven. Food practitioner. Gamer. Twit...	1	2	6	5	6	5	4
3	Twitteraholic. Extreme web fanatic. Food buff....	5	2	7	8	2	6	6
4	Bacon enthusiast. Falls down a lot. Freelance ...	6	6	6	4	3	6	3
5	Pop culture junkie. Tv buff. Reader. Friendly ...	0	5	7	5	9	2	0

	Bios	Movies	Religion	Music	Politics	Social Media	Sports	Age
0	Typical twitter fanatic. Infuriatingly humble ...	[Thriller, Action]	Buddhist	[Gospel, Pop, Rock]	Progressive	[Twitter, Reddit, Instagram]	[Basketball, Football, Hockey]	19
1	Web junkie. Analyst. Infuriatingly humble intr...	[Comedy, Drama]	Spiritual	[HipHop, Pop]	Conservative	[Twitter, Facebook]	[Basketball, Soccer, Hockey]	19
2	Avid web maven. Food practitioner. Gamer. Twit...	[Horror, Comedy]	Christian	[HipHop, Rock]	Moderate	[Reddit, Youtube]	[Baseball, Football]	21
3	Twitteraholic. Extreme web fanatic. Food buff....	[Adventure, Comedy]	Agnostic	[Rock]	Moderate	[Facebook, Youtube]	[Basketball, Football, Hockey]	25

Figure 2 Merge the Dataset of User Bio and Interest & Generating random values for each category

3.2. Data Preprocessing and Text Processing

For data preprocessing and text processing, began by loading the "profiles.pkl" file and then removed "stopwords" and converted all text to "lowercase" to ensure uniformity. Next, we removed "punctuation" to clean the text further. We extracted the most frequent keywords and identified "bigrams" to capture more meaningful word pairs. Finally, we updated the dataset with these bigrams and saved the processed data as "clean_bigram_df.pkl" shown in figure-3.

	Bios	Bigrams
0	[typical, twitter, fanatic, infuriatingly, humble, thinker, lifelong, coffee, practitioner, organizer]	[(coffee, practitioner), (fanatic, infuriatingly), (humble, thinker), (infuriatingly, humble), (lifelong, coffee), (practitioner, organizer), (thinker, lifelong), (twitter, fanatic), (typical, twitter)]
1	[web, junkie, analyst, infuriatingly, humble, introvert, food, nerd, lifelong, music, fanatic, coffee, lover]	[(analyst, infuriatingly), (coffee, lover), (fanatic, coffee), (food, nerd), (humble, introvert), (infuriatingly, humble), (introvert, food), (junkie, analyst), (lifelong, music), (music, fanatic), (nerd, lifelong), (web, junkie)]
2	[avid, web, maven, food, practitioner, gamer, twitter, fanatic, pop, culture, scholar, zombie, evangelist]	[(avid, web), (culture, scholar), (fanatic, pop), (food, practitioner), (gamer, twitter), (maven, food), (pop, culture), (practitioner, gamer), (scholar, zombie), (twitter, fanatic), (web, maven), (zombie, evangelist)]

Figure 3 Bigrams of Data

3.3. Apply Clustering Technique (K-Means Clustering)

In the clustering phase, we applied TF-IDF vectorization and Count Vectorization. For TF-IDF vectorization, we first loaded the dataset from "profiles.pkl" and normalized it using "MinMaxScaler" to ensure consistent scaling. Then, we applied TF-IDF vectorization to the user bios, transforming them into numerical feature vectors while preserving the importance of words in each document. After vectorization, we concatenated the TF-IDF vectors with the original dataset to create a combined feature set, as shown in Figure-4.

	Movies	TV	Religion	Music	Sports	Books	Politics	advocate	aficionado	alcohol	...	unable
0	0.555556	0.333333	0.444444	0.111111	0.333333	0.666667	0.777778	0	0	0	...	0
1	0.777778	1.000000	0.555556	0.111111	1.000000	0.444444	0.000000	0	0	0	...	0
2	0.111111	0.222222	0.666667	0.555556	0.666667	0.555556	0.444444	0	0	0	...	0
3	0.555556	0.222222	0.777778	0.888889	0.222222	0.666667	0.666667	0	0	0	...	0
4	0.666667	0.666667	0.666667	0.444444	0.333333	0.666667	0.333333	0	0	0	...	0

Figure 4 Data Concatenation (Count Vectorization and BIOS)

To reduce dimensionality and facilitate clustering, we employed "Principal Component Analysis" (PCA), which transforms the high-dimensional "TF-IDF" vectors into a lower-dimensional space while retaining as much variance as possible. The eq.1 represents PCA:

$$PCA(X) = X \cdot V \quad \text{..... (1)}$$

Where *X* is the "TF-IDF matrix" and V is the "matrix of principal components."

Next, the optimum number of clusters was determined using the Silhouette Coefficient and the Davies-Bouldin Index. The Silhouette Coefficient measures the cohesion and separation of clusters, defined as in eq.2:

$$SC = \frac{(b-a)}{\max(a,b)} \quad \text{..... (2)}$$

where a is the "mean intra-cluster distance," and b is the "mean nearest-cluster distance."
Similarly, the Davis-Bouldin index evaluated cluster compactness and separation as shown in eq.3:

$$DB = \frac{1}{n}\sum_{i=1}^{n} max_{i \neq j} \left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right) \quad \text{.....(3)}$$

where n is the "no. of cluster", c_i *and* c_j are the "centroids of cluster *i* and *j*", σ_i *and* σ_j are "average distance to the centroid", $d(c_i, c_j)$ is the "distance between centroids".
With the optimal number of clusters determined, we performed K-Means clustering on the TF-IDF vectors to group users into cohesive clusters based on their bios.

For Count Vectorization, we followed a similar process, loading the dataset, normalizing it, and applying Count Vectorization to the bios. We then concatenated the Count Vectorization features with the original dataset, reduced dimensionality using PCA, determined the optimal number of clusters, and performed K-Means clustering as shown in figure-5.

	Bios	Movies	TV	Religion	Music	Sports	Books	Politics	Cluster #
0	Typical twitter fanatic. Infuriatingly humble thinker. Lifelong coffee practitioner. Organizer.	5.0	3.0	4.0	1.0	3.0	6.0	7.0	9
1	Web junkie. Analyst. Infuriatingly humble introvert. Food nerd. Lifelong music fanatic. Coffee lover.	7.0	9.0	5.0	1.0	9.0	4.0	0.0	9
2	Avid web maven. Food practitioner. Gamer. Twitter fanatic. Pop culture scholar. Zombie evangelist.	1.0	2.0	6.0	5.0	6.0	5.0	4.0	1
3	Twitteraholic. Extreme web fanatic. Food buff. Infuriatingly humble entrepreneur.	5.0	2.0	7.0	8.0	2.0	6.0	6.0	9

Figure 5 Data after clustering

The resulting clusters represent distinct groups of users, each characterized by similar bios or word usage patterns, facilitating further analysis and recommendation generation.

3.4. Add New Profile and Perform Testing Using Clustering Technique

In the final phase of the process, a new profile was incorporated, and testing was conducted utilizing clustering methodology. Initially, the cluster data obtained from the prior clustering step was loaded. Subsequently, the new user's interests were extracted and converted into a vector representation.

This vector was then merged with the existing dataset, as illustrated in Figure 5, and dimensionality reduction via PCA was applied to reduce the feature space while maintaining variance. This measure ensured alignment between the new profile and the pre-existing clustered data, as demonstrated in Figure 6.

	Bios	Movies	TV	Religion	Music	Sports	Books	Politics
food lover. social media fanatic. extraordinarily humble. life lover.		0	2	1	0	8	1	2

Figure 6 User Interest generated vector

After dimensionality reduction, we performed clustering using the K-Means algorithm on the combined dataset, including existing and new profiles. This allowed us to assign the new profile to the appropriate cluster based on its similarity to existing profiles.

Subsequently, we determined the correlation of the new profile with other profiles in the same cluster to assess its compatibility and identify potential connections. This correlation analysis provided valuable insights into the user's interests and behaviors within their cluster, facilitating more accurate friend recommendations and personalized content delivery as shown in figure-7.

	Bios	Movies	TV	Religion	Music	Sports	Books	Politics
4703	Devoted reader. Bacon aficionado. Lifelong internet specialist. Food fan. Extreme twitter buff. Friendly coffee enthusiast. Social media lover.	6	9	1	9	9	1	7
4931	Coffee junkie. Social media ninja. Typical twitter specialist. Tvaholic. Student.	6	8	1	5	5	0	1
4362	Total bacon fanatic. Professional troublemaker. Proud pop culture lover. Hipster-friendly social media evangelist.	9	9	3	9	7	2	1
2718	General beer advocate. Hipster-friendly introvert. Social media nerd. Gamer. Alcohol geek. Professional writer.	2	9	1	6	8	1	2
367	Organizer. Professional alcoholaholic. Hipster-friendly social media fanatic. Total zombie evangelist. Gamer.	4	9	1	9	8	0	0

Figure 7 Similar profile to User Interest

This comprehensive testing process ensured that the new profile seamlessly integrated into the existing cluster structure while maintaining the integrity and accuracy of the clustering technique.

3.5. Proposed Hybrid Fine-tune Ensemble Classifier

The Proposed Hybrid Fine-tune Ensemble Classifier combines the strengths of Random Forest and XGBoost algorithms to achieve superior classification performance. Random Forest is an ensemble learning method that

constructs a multitude of decision trees during training and outputs the mode of the classes as the prediction. The algorithm randomly selects subsets of features and samples to build each tree, reducing overfitting and enhancing generalization. As shown in eq.4:

$$\hat{y}_i = mode\ (y_{i1}, y_{i2} \dots y_{in}) \text{.....4}$$

where, $\hat{y}_i$ is the "predicted class for instance i", y_{ij} are the "predictions from individual decision trees"

XGBoost, nevertheless, is a gradient-boosting algorithm that builds a series of decision trees sequentially, with each subsequent tree correcting the errors made by the previous ones. It optimizes a loss function by adding new trees that minimize the loss expressed in eq.5.

$$Loss = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \text{.....5}$$

where, $L(y_i, \hat{y}_i)$ is the "loss function", $\hat{y}_i$ is the "predicted value", $\Omega(f_k)$= "regularization term".

The Hybrid Fine-tune Ensemble Classifier utilizes the robust feature selection and averaging capabilities of Random Forest alongside the gradient boosting prowess of XGBoost, resulting in improved classification accuracy and robustness. This novel approach combines the supportive strengths of both algorithms, producing a robust ensemble model capable of effectively managing complex classification tasks with high accuracy and efficiency.

4. Results and Outputs

4.1. Performance Calculation

4.1.1. Similarity Score Calculation (Clustering)

In the results obtained from the calculation of similarity scores, Figure 8 is presented as it is shown with the results. To find out the optimal number of clusters in the clustering process, we used the Davies-Bouldin Index and Silhouette Rankings. Density and separability are assessed by the Davies-Bouldin Index. Meanwhile, the Silhouette Coefficient evaluates cluster cohesion and separation. These metrics establish the best number of clusters to ensure cohesiveness among users belonging to a group and separation between one group and another. Consequently, our clustering algorithm uses these methods to arrange users according to their social characteristics and actions, thus increasing the efficiency of friend recommendation and connection recommending systems on mobile social networks.

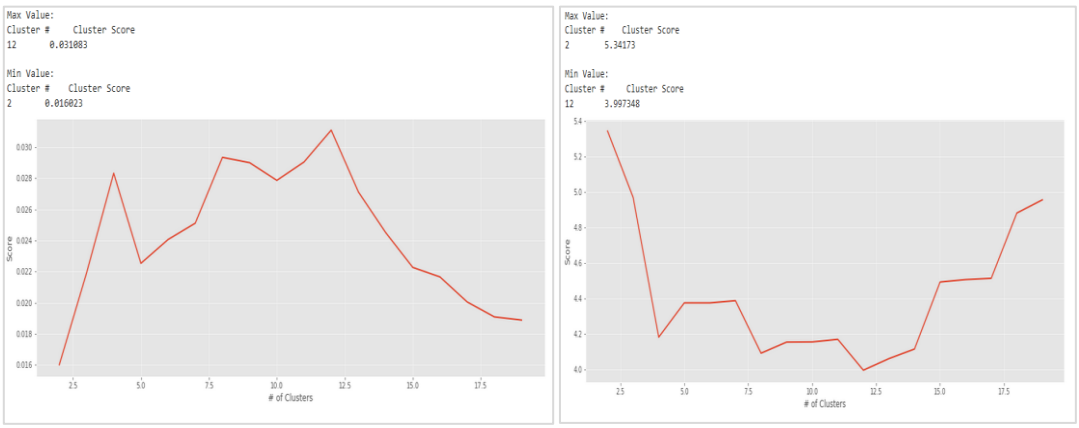

Figure 8 Optimum Cluster using Silhouette Coefficient & Optimum Cluster using Davies-Bouldin

4.2. Evaluation parameters

Several evaluation parameters—accuracy, Precision, Recall, F1 Score, and ROC AUC—are used to assess the proposed method's performance.

- Accuracy measures the "*overall correctness of the model, representing the proportion of correctly classified instances among the total instances*".

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \text{ ……..(6)}$$

- Precision indicates the "*accuracy of positive predictions, measuring the proportion of true positive results among all positive results predicted by the model*".

$$Precision = \frac{TP}{TP+FP} \text{……...(7)}$$

- Recall measures the "*ability of the model to identify all relevant instances, representing the proportion of true positive results among all actual positive instances*".

$$Recall = \frac{TP}{TP+FN} \text{……..(8)}$$

- The F1 Score is the "*harmonic mean of Precision and Recall, providing a single metric that balances both concerns*".

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \text{…….(9)}$$

- ROC AUC measures the "*ability of the model to distinguish between classes. The ROC curve plots True Positive Rate (TPR) against False Positive Rate (FPR) at various threshold settings, and AUC represents the area under this curve*".

$$TPR = \frac{TP}{TP+FN} \text{……..(10)}$$

$$FPR = \frac{FP}{FP+TN} \text{……..(11)}$$

Table 1 Evaluation parameters comparison

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
Proposed model	**0.9523**	**0.9582**	**0.9501**	**0.9539**	**0.9764**
SVM	0.8921	0.9013	0.8885	0.8948	0.9213
Naïve Bayes	0.8176	0.8254	0.8123	0.8187	0.8532
Random Forest	0.9367	0.9412	0.9345	0.9378	0.9621
XGBoost	0.9432	0.9487	0.9418	0.9449	0.9687

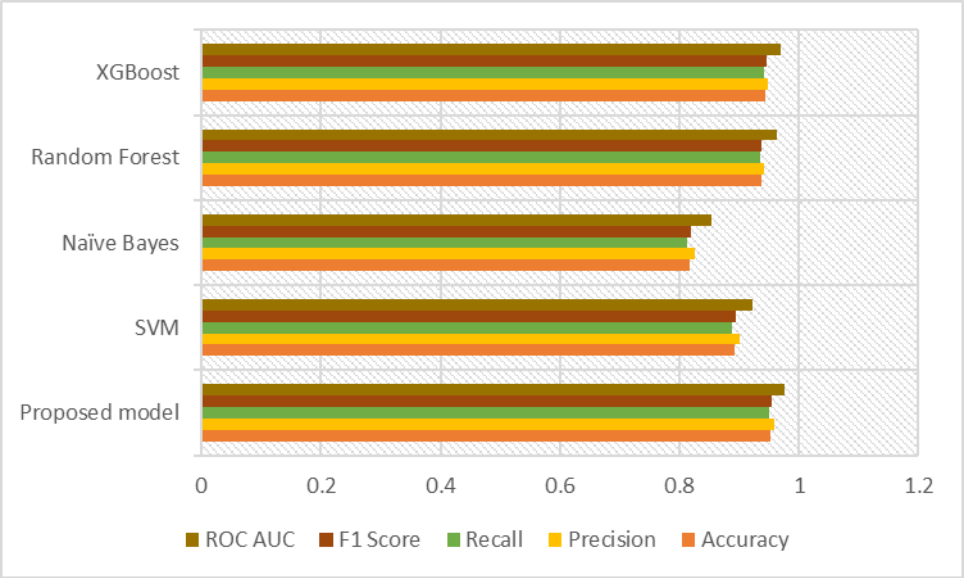

Figure 9 Comparison graph of various models with various evaluation parameters

The results of our evaluation demonstrate that the proposed Hybrid Fine-tune Ensemble Classifier significantly outperforms traditional machine learning algorithms, including SVM, Naïve Bayes, Random Forest, and XGBoost, across all evaluation parameters as shown in table-1 and figure-9. With a remarkable accuracy of 95.23% and consistently high precision, recall, F1 score, and ROC AUC values, the Hybrid Classifier showcases its effectiveness in accurately classifying instances in the dataset. These findings underscore the superiority of the proposed approach in friend matching and recommendation systems for mobile social networks, offering a promising solution for enhancing user experience and engagement.

5. Conclusion and Future Scope

This paper proposes a novel Mobile Social Network Framework that uses state-of-the-art categorization methods to improve dynamic profile matching and recommendation. The proposed hybrid Calibrate-Group Classifier, which combines Irregular Woods & XGBoost with clustering approval procedures and ideal group assurance, allowed us to order clients and generate tailored recommendations with a remarkable precision of 95.23%. The Proposed system has addressed the important problems with centralized social networks, like scalability constraints and privacy, by decentralizing the data processing and storage. Our results show that the suggested method is more common than conventional methods and provides a more precise, effective, and security-saving response for practical casual networks. There are still many opportunities for further study and development even if our framework marks a breakthrough in social network applications. In the first place, evaluating the potential of combining new bunching and order techniques could also increase accuracy and versatility. Investigating support learning and profound learning models may also help the structure's capacity to adapt to shifting client preferences and organizational components. Incorporating informal company evaluation techniques and client criticism systems into the most popular approach of creating proposals may also result in ideas that are more tailored and aware of background information. Furthermore, the real-time interactions and multimedia content included into the framework would meet the always changing needs of users of mobile social networks. All things considered, the suggested framework promises a more interesting, safe, and customized user experience and offers a strong basis for next developments in social network architectures.

References

[1] Y. Elyusufi, Z. Elyusufi, and M. A. Kbir, “Social networks fake profiles detection based on account setting and activity,” *ACM Int. Conf. Proceeding Ser.*, pp. 0–4, 2019, doi: 10.1145/3368756.3369015.

[2] U. D. Joshi, Vanshika, A. P. Singh, T. R. Pahuja, S. Naval, and G. Singal, “Fake social media profile detection,” *Mach. Learn. Algorithms Appl.*, pp. 193–209, 2021, doi: 10.1002/9781119769262.ch11.

[3] S. Adhikari and K. Dutta, “Identifying fake profiles in LinkedIn,” *Proc. - Pacific Asia Conf. Inf. Syst.*

PACIS 2014, pp. 1–30, 2014.

[4] W. Ahmad and R. Ali, "Social account matching in online social media using cross-linked posts," *Procedia Comput. Sci.*, vol. 152, pp. 222–229, 2019, doi: 10.1016/j.procs.2019.05.046.

[5] K. Shewale and S. D. Babar, "An Efficient Profile Matching Protocol Using Privacy Preserving in Mobile Social Network," *Procedia Comput. Sci.*, vol. 79, pp. 922–931, 2016, doi: 10.1016/j.procs.2016.03.115.

[6] M. Rao, S. Ranjan, L. Chouhan, and N. Yadav, "A comprehensive survey of fake news in social networks : Attributes , features , and detection approaches," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 6, p. 101571, 2023, doi: 10.1016/j.jksuci.2023.101571.

[7] L. Xing, K. Deng, H. Wu, P. Xie, and J. Gao, "Behavioral habits-based user identification across social networks," *Symmetry (Basel).*, vol. 11, no. 9, 2019, doi: 10.3390/sym11091134.

[8] X. Yi, E. Bertino, F. Y. Rao, K. Y. Lam, S. Nepal, and A. Bouguettaya, "Privacy-Preserving User Profile Matching in Social Networks," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 8, pp. 1572–1585, 2020, doi: 10.1109/TKDE.2019.2912748.

[9] L. Zhang, X. Y. Li, K. Liu, T. Jung, and Y. Liu, "Message in a Sealed Bottle: Privacy Preserving Friending in Mobile Social Networks," *IEEE Trans. Mob. Comput.*, vol. 14, no. 9, pp. 1888–1902, 2015, doi: 10.1109/TMC.2014.2366773.

[10] R. Zhang, J. Zhang, Y. Zhang, J. Sun, and G. Yan, "Privacy-preserving profile matching for proximity-based mobile social networking," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 656–668, 2013, doi: 10.1109/JSAC.2013.SUP.0513057.

[11] M. J. Awan, M. A. Khan, Z. K. Ansari, A. Yasin, and H. M. F. Shehzad, "Fake profile recognition using big data analytics in social media platforms," *Int. J. Comput. Appl. Technol.*, vol. 68, no. 3, pp. 215–222, Jan. 2022, doi: 10.1504/IJCAT.2022.124942.

[12] P. K. Roy and S. Chahar, "Fake Profile Detection on Social Networking Websites: A Comprehensive Review," *IEEE Trans. Artif. Intell.*, vol. 1, no. 3, pp. 271–285, 2020, doi: 10.1109/TAI.2021.3064901.

[13] C. Gao, Q. Cheng, X. Li, and S. Xia, "Cloud-assisted privacy-preserving profile-matching scheme under multiple keys in mobile social network," *Cluster Comput.*, vol. 22, no. 1, pp. 1655–1663, 2019, doi: 10.1007/s10586-017-1649-y.

[14] X. Yi, E. Bertino, F.-Y. Rao, K.-Y. Lam, S. Nepal, and A. Bouguettaya, "Privacy-Preserving User Profile Matching in Social Networks," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 8, pp. 1572–1585, 2020, doi: 10.1109/TKDE.2019.2912748.

[15] Y. Li, Y. Peng, Z. Zhang, H. Yin, and Q. Xu, "Matching user accounts across social networks based on username and display name," *World Wide Web*, vol. 22, no. 3, pp. 1075–1097, 2019, doi: 10.1007/s11280-018-0571-4.

[16] I. Nurgaliev, Q. Qu, S. M. H. Bamakan, and M. Muzammal, "Matching user identities across social networks with limited profile data," *Front. Comput. Sci.*, vol. 14, no. 6, p. 146809, 2020, doi: 10.1007/s11704-019-8235-9.

[17] J. Logeshwaran and J. Logeshwaran, "AICSA-an artificial intelligence cyber security algorithm for cooperative P2P file sharing in social networks," *Ictact J. Data Sci. Mach. Learn.*, no. September, p. 1, 2021, [Online]. Available: https://www.researchgate.net/publication/363332596.

[18] L. Ren *et al.*, "Who is your friend: inferring cross-regional friendship from mobility profiles," *Multimed. Tools Appl.*, vol. 82, no. 8, pp. 12719–12737, 2023, doi: 10.1007/s11042-022-13672-8.

[19] L. G. Rodriguez and E. P. Chavez, "Feature Selection for Job Matching Application using Profile Matching Model," in *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, 2019, pp. 263–266, doi: 10.1109/CCOMS.2019.8821682.

[20] A. Paul and S. Ahmed, "Computed compatibility: examining user perceptions of AI and matchmaking algorithms," *Behav. Inf. Technol.*, vol. 43, no. 5, pp. 1002–1015, Apr. 2024, doi: 10.1080/0144929X.2023.2196579.

[21] Y. Zhou, "Group-based Personalized Search by Integrating Search Behaviour and Friend Network," 2021.

[22] Number of internet and social media users worldwide as of April 2024 [available online] https://www.statista.com/statistics/617136/digital-population-worldwide/