International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

Improving Social Network Link Prediction with an Ensemble of Machine Learning Techniques

Sachin U. Balvir[1], Mukesh M. Raghuwanshi[2] and Purushottam D. Shobhane[3]

[1] Research Scholar, Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur, India
[1] Assistant Professor, Department of Computer Science and Engineering, S. B. Jain Institute of Technology, Management and Research, Nagpur,India
[2]Deen Engineering, S. B. Jain Institute of Technology, Management and Research, Nagpur, India
[2]Assistant Professor, Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune, India

E-mail address: sachin_balvir@yahoo.com[1], deanengineering@sbjit.edu.in[2], purushottam.shobhane@sitnagpur.siu.edu.in[3]

Received ## Mon.20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: Finding missing connections in social networks are a crucial task that has generated a great deal of interest lately. To solve this issue, there have been several machine learning methods proposed, however the majority of them concentrate on a particular kind of feature or technique. The goal of link prediction is to identify pairs of nodes that will either form a link in the future or not. In order to accurately predict links in social networks, we provide a hybrid machine learning methodology in this research that integrates many features and techniques. In this research, we present an ensemble strategy to enhance the resilience and accuracy of link prediction by utilizing the advantages of many machine learning approaches. Specifically, we combine seven popular methods, namely, logistic regression, support vector machines, random forests, and decision tree, naïve Bayes, k-NN and gradient boosting, and employ them together in a unique ensemble framework. In addition, we employ principal component analysis to lower the feature space's dimensionality and boost the model's computational effectiveness. To evaluate the proposed method, we run experiments on real-world social network dataset facebook. The results demonstrate that our method outperforms the state-of-the-art approaches in terms of F1 Score, accuracy, precision, and recall. All things considered, our method offers a viable way to improve prediction.

Keywords: Social networks, Link prediction, Principal component analysis, Dimensionality reduction, Ensemble framework, Network analysis

1. Introduction

Link prediction in social networks, which aims to predict the formation of new connections or ties between individuals in a social network, is a crucial task in the field of social network analysis. Applications for social network link prediction are diverse, such as making recommendations for new connections, predicting the evolution of the network structure over time, and detecting potential fraud or anomalous behaviour. Social networks can take many forms, including OSN's, as well as offline social networks such as professional networks and communities. Link prediction research has benefited greatly from the explosion of social connection data that has resulted from the rise of OSN's in recent years. In social networks, predicting links is a difficult task due to the complexity of the relationships between individuals in a network. Relationships can be influenced by many factors, such as demographics, shared interests, and past interactions. Additionally, the structure of social networks is constantly changing, making it difficult to predict the formation of new connections [1][2][3]. Despite these challenges, link prediction has become an active area of research, with numerous methods and techniques developed to tackle the problem.

Link prediction has several uses; it can be used to suggest friends and business associates, find possible joint ventures, and improve network security, among other things. The foundation for link prediction has been established by conventional techniques like preferred attachment and common neighbors. These techniques, however, frequently fail to capture the intricate, nuanced character of social interactions. Recent work has concentrated on using machine learning approaches to advance the accuracy and robustness of link prediction in order to overcome these limitations. Large datasets can be efficiently mined by machine learning models, especially those built on supervised learning, which can also reveal hidden patterns that conventional approaches might miss. These models are highly suited for the complex dynamics of social networks because they make use of a range of characteristics obtained from network topology, node attributes, and edge attributes.

The use of ensemble approaches, which combine many models to get a single, better prediction, has shown promise in a variety of fields, such as financial forecasting, picture recognition. The purpose of this study is to enhance prediction accuracy, robustness, and generalizability across various social networks by utilizing ensemble learning in link prediction. It aims to provide insights and workable solutions for enhancing social network analysis and promoting stronger, more meaningful connections in the digital sphere through thorough experimentation and analysis.

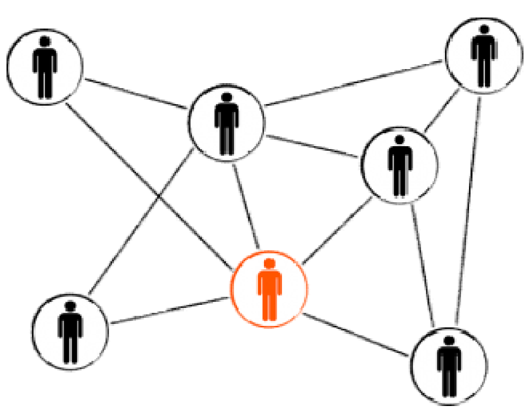


Figure 1: Link connected in social network

2. LITERATURE REVIEW

Links in social networks can be predicted with a range of techniques, which can be roughly divided into three categories: methods based on features, methods based on graphs, and hybrid approaches that blend feature- and graph-based techniques.

A. Graph-based Methods:

These techniques concentrate on the network's actual structure, considering the relationships between pairs of individuals in the network to make predictions. Some common graph based methods include the common neighbor's method, the resource allocation index, and the Adamic Adar index.

B. Feature-based Methods:

These methods focus on the attributes of the individuals in the network, such as their demographic information, interests, and past interactions [4]s[5][13]. These attributes are used to make predictions about the likelihood of new connections. One popular feature-based method is collaborative filtering, which uses a matrix factorization approach to make predictions based on the relationships between individuals.

C. Hybrid Methods:

These methods combine both graph-based and feature-based approaches to make predictions. For instance, when predicting something, a hybrid approach might take into account both the number of neighbors that two people have in common as well as their common interests. The development of deep learning-based link prediction techniques has gained popularity recently. These techniques take advantage of deep neural networks' strong representational capabilities to predict network topology and individual attribute values.

In the sections that follow, we'll look at a variety of social network linking techniques. The premise behind collaborative filtering is that people are more likely to engage if they have similar interests or behaviors. In order to find latent patterns, this method creates a matrix that represents the strength of interactions between nodes in the network. This matrix can be factorized using methods such as matrix factorization or singular value decomposition (SVD) [6]. Once factorization is complete, people with comparable latent patterns are found and used to predict future connections. The strength of these associations determines the ranking of the projected linkages. Because collaborative filtering may capture complex associations based on qualities and actions, it works well in a variety of social networks, including professional, academic, and internet networks [7]. Collaborative filtering works well, but it has drawbacks as well. For example, it might be sensitive to missing or noisy data, and it can be computationally expensive for big networks.

The premise behind the similar Neighbors approach, a graph-based method for social network connection prediction, is that two individuals have a higher chance of connecting if they have a large number of similar neighbors. This method involves calculating the number of common neighbors that exist in the network between every pair of individuals. Predictions can be rated based on which pairs have the highest likelihood of forming new connections—those with the most frequent neighbors [8][9][10]. Large networks can benefit from this method's key advantages, which are its computational efficiency and simplicity. Its drawbacks include its inability to identify intricate relationships based on common behaviors or interests and its reliance on the notion that the presence of shared neighbors is a strong predictor of the formation of new connections. In many link prediction applications, the Common Neighbors technique is well-liked and efficient despite these disadvantages. An approach to social network prediction based on graphs is the RA. It works on the premise that possible relationships can be inferred from the flow of "resources" (knowledge or influence) between people. Based on the quantity of shared neighbors and the quality of interpersonal ties, RA computes resource flow. It is projected that pairs with the highest resource flow will establish new connections; these predictions are ordered in order of likelihood [11][12]. Considering relationship strength and common neighbors, as well as being computationally efficient for vast networks, are some of RA's advantages. It can be susceptible to noisy or incomplete data, though, and it makes the assumption that resource flow correlates with connection likelihood, which may not always hold true. Notwithstanding these drawbacks, RA is a popsular option for link prediction since it is widely applied and efficient in a variety of networks, including academic, professional, and online social networks. One graph-based method for anticipating new links in social networks is the Adamic-Adar Index (AA). Assuming that ties to well-known people are more relevant, it calculates the inverse logarithm of each neighbor's degree to determine how similar two people are. In order to anticipate new connections, the AA approach ranks pairings according to their similarity scores, which are determined by taking into account the number and popularity of neighbors in common. For vast networks, the AA approach takes individual popularity into consideration and is computationally efficient. It can be susceptible to noisy or incomplete data, though, and it makes the assumption that popularity and connection likelihood are directly correlated—an assumption that may not always hold true.

In this study [15], the authors consider link prediction as a supervised learning task in an effort to incorporate many attributes as input data for classification. Authors have been using a select characteristics method to increase forecast accuracy. There have been two co-authorship data sets used in the experiments. This research [16] attempts to identify the characteristics that primarily determine the sign value using data from social media networks. The algorithm can accurately forecast the values of the linkages based on information retrieved from the users' neighbor relationships and self-statuses. Through the analysis of the models built across various datasets, study identifies the shared determining characteristics for link prediction. A model of link prediction for social multiplex networks (LPSMN) is proposed [17]. Author concatenate on these attributes as link representations after extracting explicit, latent, and graph structural aspects. Subsequently, an attention method is utilized to build a multiplex and improved forecasting model with the aid of external data from an established platform. The link prediction problem is sometimes handled as a binary classification problem. In order to enhance performance, this approach makes use of three distinct feature types. Finally, author use two social multiplex networks along with five artificial networks with varying degree distributions to generate an experimental scenario for further evaluation.

Multiple strategies are combined in hybrid methods for social network link prediction, increasing prediction accuracy and robustness. These strategies take advantage of the advantages of several ways to get around the drawbacks of specific methods. The author suggests [18] a novel hybrid method for spotting linkages in social networks. The method creates a hybrid system by combining supervised machine learning technique with an approach inspired by the particle swarm algorithm. The suggested solutions are compared against alternative strategies in an experimental research conducted using real-world data sets in the publication. Based on network communities, the author developed two hybrid link prediction algorithms [19]. In order to assess the efficacy of the suggested hybrid algorithms, author carried out an extensive computational campaign utilizing both artificial and real-world data sets. Experiments demonstrate that the prediction accuracy is improved by including details about communities and pertinent links. A hybrid link prediction model that integrates eight structure-based prediction techniques and automatically adjusts the weights assigned to each included approach is presented by the author, who tackles the link prediction problem as a time-series problem [20].

Using both network structure and node characteristics to improve prediction accuracy, hybrid approaches have become a potent tool for social network link prediction. Conventional techniques like Jaccard Coefficient, Adamic/Adar, and Common Neighbors depend solely on network structure, whereas attribute-based techniques make advantage of node attributes like user profiles and hobbies. These strategies are combined in hybrid methods, which incorporate node attributes and structural information to produce a prediction model that is more complete. In feature-based machine learning models, user attributes are combined with measures such as node degree. Adjacency matrices and node characteristics are both included in matrix factorization approaches like collaborative filtering and graph-regularized matrix factorization. Graph embedding techniques such as attributed network embeddings, Node2Vec, and DeepWalk use structural and attribute information to learn low-dimensional representations. Both elements are naturally integrated by deep learning techniques, particularly Graph Neural Networks (GNNs) and Variational Graph Autoencoders (VGAEs), which provide strong performance. In terms of accuracy and precision, hybrid approaches typically beat single-approach models; nevertheless, scalability and data sparsity present difficulties. It is still difficult to address privacy issues and dynamic networks, which emphasizes the necessity for cutting-edge methods in these fields. Hybrid approaches present a promising path toward more precise and thorough social network connection prediction in spite of these difficulties.

3. METHEDOLOGY

A. Dataset

The selection of a suitable dataset is a critical component in social network link prediction research. This entails taking into account variables including the kind of data required, the system's handling capability, the characteristics of the dataset and the dataset's cleanliness and quality. This guarantees the dataset satisfies the requirements of the study and is compatible with the capabilities of the system. Because they were pertinent to our goals, we concentrated on unweighted and undirected datasets for this study. Because interactions are frequently mutual, unweighted datasets minimize computer complexity by considering all edges equally, while undirected datasets make analysis easier.

We chose the Facebook dataset from the Stanford Network Analysis Project (SNAP) based on these requirements. With 1,446 nodes and 59,589 edges, this dataset is ideal for our purposes because it offers a thorough understanding of social interactions. Its unweighted and undirected characteristics are ideal for our research needs, allowing for the creation and testing of our hybrid link prediction techniques without introducing further complexity. Its reasonable size also allows for effective data processing and analysis.

B. Algorithm

Step 1: Read the dataset which is an edge list(graph)

Step 2: Data Preparation

- Retrieve not connected node pairs – Negative Samples
- Remove 20% Links from connected node pairs – Positive Samples. Remove link such that all nodes in the graph should remain connected.
- Prepare new dataset which will be the combination of positive & negative samples.

Step 3: Calculate prediction scores for each edge on new prepared dataset using following ten link prediction methods (Common Neighbors, Preferential Attachment, Jaccard Coefficient, Index, Adamic Adar Index, Resource Allocation Index, Sorensen Similarity Index, Salton Cosine Similarity Index, Hub Depressed Index, Hub Promoted Index, Leicht Home Index)

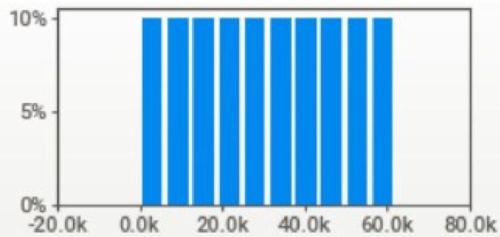
Step 4: Prediction scores for above methods is considered as Features. Examine the new feature set using a variety of machine learning algorithms. Logistic regression, Support Vector Machine, Random Forest, Decision Tree, k-NN, Naïve Bays, and Gradient Boosting are some of the machine learning techniques utilized.

C. *Feature Description*

The statistical analysis of the feature data set that we created for the application is displayed in Table 1 below.

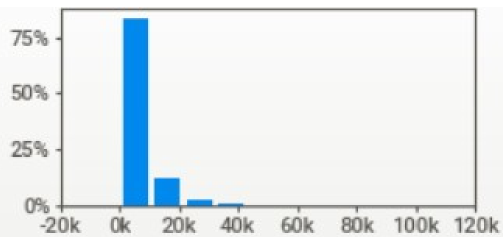
- Common_Neighbor's

MAX	62,011	RANGE	62,011
95%	58,910	IQR	31,006
Q3	46,508	STD	17,901
MEDIAN	31,006	VAR	320.5M
AVG	31,006		
Q1	15,503	KURT.	-1.20
5%	3,101	SKEW	0.00
MIN	0	SUM	1.9B



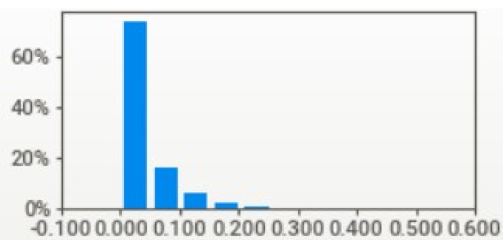
- Preferential attachment

MAX	107k	RANGE	107k
95%	20k	IQR	6,570
Q3	8k	STD	6,976
AVG	6k	VAR	48.7M
MEDIAN	4k		
Q1	1k	KURT.	11.2
5%	0k	SKEW	2.65
MIN	0k	SUM	371.2M



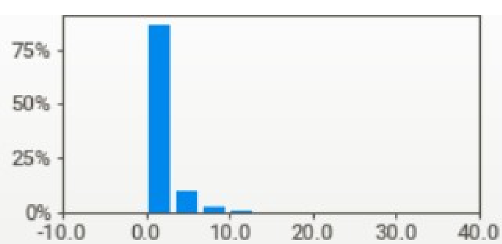
- Jaccard

MAX	0.511	RANGE	0.511
95%	0.136	IQR	0.052
Q3	0.052	STD	0.048
AVG	0.037	VAR	0.002
MEDIAN	0.019		
Q1	0.000	KURT.	7.00
5%	0.000	SKEW	2.23
MIN	0.000	SUM	2,290



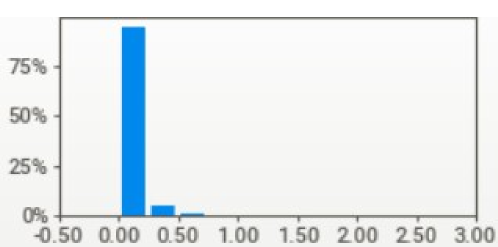
- Adamic

MAX	32.8	RANGE	32.8
95%	5.9	IQR	1.79
Q3	1.8	STD	2.23
AVG	1.4	VAR	4.97
MEDIAN	0.6		
Q1	0.0	KURT.	14.3
5%	0.0	SKEW	3.09
MIN	0.0	SUM	87,522



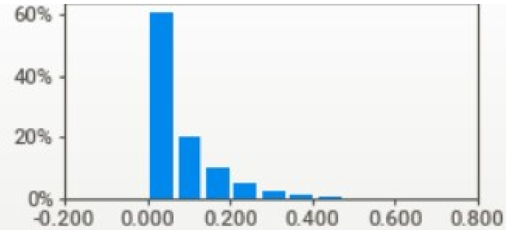
- Rallocation

MAX	2.48	RANGE	2.48
95%	0.25	IQR	0.075
Q3	0.07	STD	0.099
AVG	0.06	VAR	0.010
MEDIAN	0.02		
Q1	0.00	KURT.	29.5
5%	0.00	SKEW	3.79
MIN	0.00	SUM	3,731



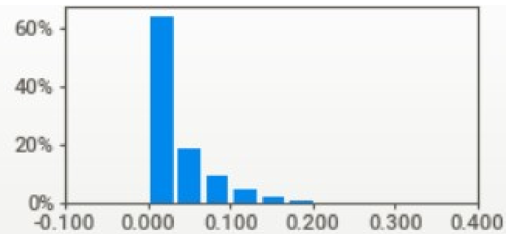
- Salton

MAX	0.680	RANGE	0.680
95%	0.251	IQR	0.110
Q3	0.110	STD	0.086
AVG	0.074	VAR	0.007
MEDIAN	0.045		
Q1	0.000	KURT.	3.22
5%	0.000	SKEW	1.67
MIN	0.000	SUM	4,573



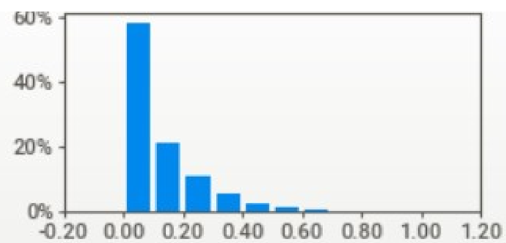
- Sorensen

MAX	0.338	RANGE	0.338
95%	0.119	IQR	0.050
Q3	0.050	STD	0.041
AVG	0.034	VAR	0.002
MEDIAN	0.019		
Q1	0.000	KURT.	3.84
5%	0.000	SKEW	1.80
MIN	0.000	SUM	2,091



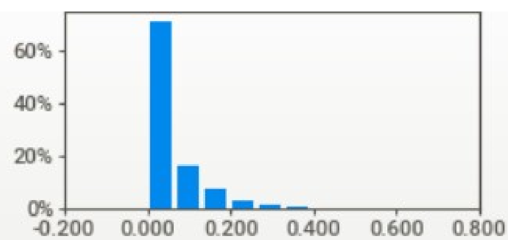
- Hub Promoted

MAX	1.00	RANGE	1.00
95%	0.37	IQR	0.167
Q3	0.17	STD	0.131
AVG	0.11	VAR	0.017
MEDIAN	0.07		
Q1	0.00	KURT.	5.14
5%	0.00	SKEW	1.85
MIN	0.00	SUM	6,969



- Hub Depressed

MAX	0.656	RANGE	0.656
95%	0.198	IQR	0.077
Q3	0.077	STD	0.069
AVG	0.053	VAR	0.005
MEDIAN	0.027		
Q1	0.000	KURT.	5.48
5%	0.000	SKEW	2.07
MIN	0.000	SUM	3,314



- LHN

MAX	0.143	RANGE	0.143
95%	0.004	IQR	0.001
Q3	0.001	STD	0.002
AVG	0.001	VAR	3.71e-6
MEDIAN	0.001		
Q1	0.000	KURT.	971
5%	0.000	SKEW	19.5
MIN	0.000	SUM	67.5

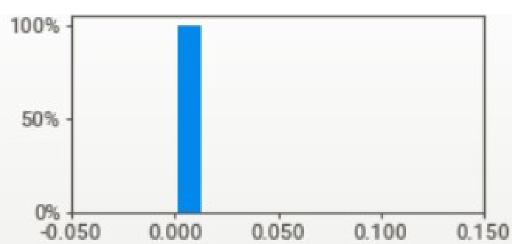


Figure 2: statistical analysis of feature data set

D. Proposed Model

An example of the suggested ensemble architecture for link prediction is shown in Figure 2. Preprocessing the incoming data to make sure it is clean and appropriate for analysis is the first step in the process. In this step, missing value handling, data normalization, and the conversion of categorical variables into numerical formats could all be involved. Following preprocessing, the data is split into two subsets: a testing set and a training set. The testing set is used to assess the performance of the machine learning models. The architecture's central process is the development and evaluation of seven different machine learning models. These models are chosen on the basis of their applicability and possible efficacy in managing the particular features of the dataset. To ascertain each model's accuracy and predictive power, it is first trained on the training set and subsequently assessed on the testing set.

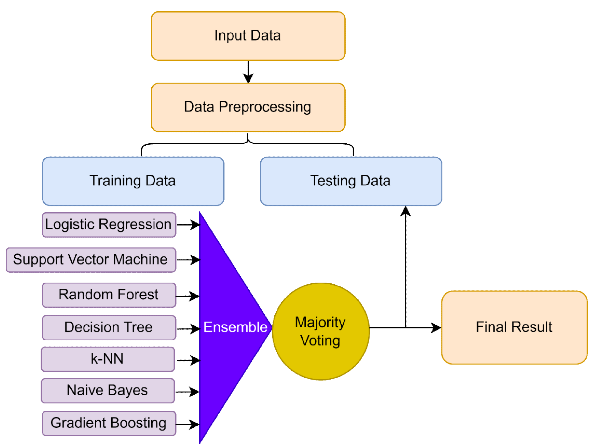


Figure 3: Proposed Ensemble Architecture

Following an independent evaluation of each of these seven models' performance, an ensemble approach is used to integrate their outputs. By utilizing each model's advantages, this technique generates predictions that are more reliable and accurate. Voting, averaging, and more complex strategies like stacking—in which the outputs of the base models are utilized as inputs for a higher-level model—are examples of ensemble approaches. Through the integration of several models' outputs, the ensemble technique seeks to mitigate the shortcomings of individual models while enhancing overall forecast accuracy and reliability.

E. Evaluating the effectiveness of link prediction methods

An important future topic for study is to develop methods that can combine link prediction with other techniques, like network centrality and community detection. Following are numerous metrics [14] that are normally used to calculate the effectiveness of link prediction methods, including:

- *Accuracy*: Accuracy is another important performance metric used in machine learning algorithms and is a reflection of how faithful the model is to the data. Promoting the percentage of accurately detected examples in regards to the amount of all occurrences in a dataset. It is the evaluation of the total accuracy of the model is done in the following way; it divide it by the number of times it successfully predicts outcomes which later turn out to be true.
- *Precision*: The percentage of true positive predictions a model makes among all positive predictions is known as precision. The ratio of true positives to the total of true positives and false positives is used to compute it.
- *Recall:* Recall, often referred to as sensitivity or true positive rate, is the proportion of real positive cases in a dataset that a model correctly finds out of all actual positive cases. The ratio of true positives to the total of false negatives and true positives is used to calculate it.
- *F1 score:* Recall and precision are combined into a single number, known as the F1 number, which provides a fair assessment of a model's performance. It is computed as the harmonic mean of recall and precision, with equal weight assigned to each metric.

4. RESULTS

We implemented seven classification machine learning model. However, the ensemble approach gave us the best accuracy. We achieve Ensemble Accuracy

E-mail:sachin_balvir@yahoo.com, deanengineering@sbjit.edu.in, purushottam.shobhane@sitnagpur.siu.edu.in

94.23%. The table 1 gives the complete snapshot of results of all the machine learning algorithms.

TABLE 1: EXPERIMENTAL RESULTS

SN	Machine Learning Model	Accuracy	Precision	Recall	F1-Score
1	Logistic Regression	93.38	92.69	93.38	92.76
2	SVM	90.46	87.11	90.46	86.31
3	Random Forest	93.07	92.45	93.07	92.64
4	Decision Tree	93.07	92.5	93.02	92.7
5	Naïve Bayes	90.91	91.47	89.33	91.16
6	k-NN	91.3	89.57	91.25	89.73
7	Gradient Boosting	92.99	92.16	92.99	92.22
8	Ensemble Method	94.23	92.71	92.36	92.51

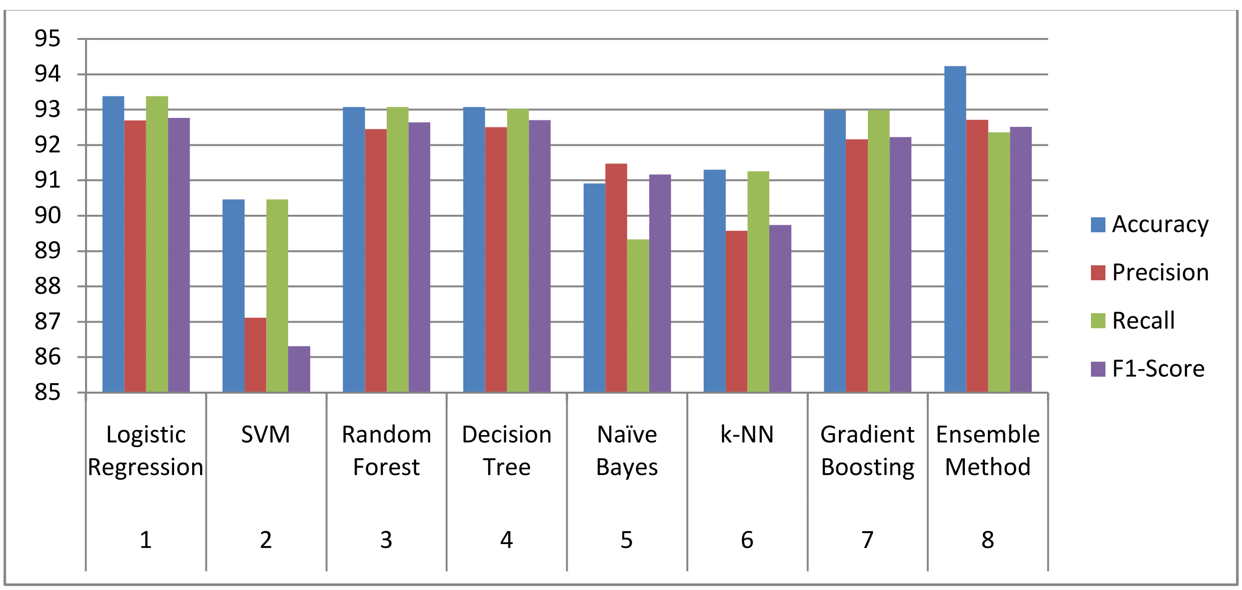


Figure 3: Graphical representation of results

The above table 2 compares the recall, accuracy, precision, and F1-Score of various machine learning models. Performance-wise, the Ensemble Method outperforms the others with the highest accuracy of 94.23%. Logistic Regression exhibits the highest level of recall (93.38% accuracy), while Decision Tree and Random Forest demonstrate excellent overall balance. Gradient Boosting delivers a well-rounded performance with an accuracy of 92.99%. While accuracy suffers, Naïve Bayes yields greater precision (91.47%). All indicators point to SVM as having the weakest performance. The ideal model depends on the specific application's requirements for accuracy, precision, recall, and balancing.

5. DISCUSSION

An ensemble machine learning model presents a new strategy to enhance the prediction accuracy of linkages in social networks. The major objective is to use a combination of diverse strategies to improve speed and circumvent the issues with typical link forecast methods. The first section of the study discusses the significance of link prediction for network analysis, connection discovery, and improving recommendation systems. The authors discuss the difficulty of forecasting broken linkages in large-scale social networks and offer an ensemble strategy to address these issues. The recommended ensemble approach is selecting and mixing various machine learning methods, including logistic regression, decision trees, support vector machines, and other machine learning approaches. The ensemble approach is more dependable and precise since every program contributes a distinct viewpoint to various aspects of link prediction.

The Ensemble Method outperforms all other models with the highest accuracy of 94.23%, demonstrating that it successfully integrates the benefits of multiple methods to generate superior overall performance. Logistic regression performs exceptionally well, yielding the best recall rate and accuracy rate of 93.38%. This makes it a valuable tool for identifying true positives. Random Forest and Decision Tree both perform well, with accuracy of 93.07%, precision of roughly 92.5%, and F1-Scores above 92.6%, indicating a solid balance between precision and recall. With well-rounded stats and an accuracy of 92.99%, gradient boosting is a strong choice for balanced performance as well. It comes in second.

Naive Bayes does rather well in precision, scoring 91.47%, indicating its ability to eliminate false positives, but having a little lower accuracy of 90.91%. k-NN performs admirably, with 91.3% accuracy and F1-Score with 89.73%; but, its precision and recall are slightly below the best models' levels. SVM performs the lowest of all the models, with an accuracy of 90.46% and an F1-Score of 86.31%, suggesting that it does a worse job of striking a balance between recall and precision. In the end, the model selection should be dependent on the specific requirements of the application, regardless of whether the application emphasizes recall, precision, overall accuracy, or a combination of these metrics.

6. Future Directions For Link Prediction Research

This field continues to evolve and advance, and there are many exciting future directions for research in this area. Some of these future directions include:

A. *Incorporating node attributes:* The majority of link prediction techniques used today solely take into account the network structure and ignore node characteristics like behavior, preferences, and demographic data. More precise and useful predictions may result from the incorporation of node properties into link prediction techniques.

B. *Dynamic networks:* Many social networks are dynamic in the actual world—that is, the interactions between nodes change over time. Future research should focus on creating link prediction techniques that can manage dynamic networks and anticipate changes in the connections over time.

C. *Multi-layer networks:* Many real-world networks consist of multiple levels, where a node-to-node relationship is represented by a different level. A social network might have both professional and friendship relationships, for instance.

D. *Scalability:* As networks become larger and more complex, it is imperative to develop link prediction systems that can handle these large and complicated networks. This involves creating new techniques especially tailored for large-scale networks as well as enhancing the effectiveness of currently used techniques.

E. *Heterogeneous networks:* Real-world networks typically consist of heterogeneous networks, or nodes with a variety of types and attributes. Future research should focus on creating link prediction techniques that can manage heterogeneous networks.

F. *Integration with other methods:* There are numerous techniques used to analyze and comprehend social networks; link prediction is but one facet of this process. Future studies should concentrate on developing methods that integrate community detection and network centrality with link prediction.

Future studies on predicting links in networks of people are anticipated to go in a plethora of fascinating ways, these being just a few. The growing significance of social networks in our daily lives and the demand for efficient techniques for analyzing and comprehending these networks are what are fueling the field's continuing expansion and development.

7. CONCLUSION

This research paper proposes an effective and new method for improving social network link prediction. The suggested approach improves on standard link prediction methods by using an ensemble of machine learning algorithms. Concerning precision, the Ensemble Method performs better than any one model alone, indicating that it leverages the advantages of several models to produce better overall results. It effectively finds a middle ground between preventing false positives and identifying true positives, as seen by its high precision, recall, and F1-Score. The benefits of combining different machine learning algorithms are demonstrated by the ensemble technique. The ensemble model may capture many network characteristics and topological traits since each approach provides unique insights on link prediction.

The study demonstrates that the ensemble model works better than link prediction techniques and individual methods. The accuracy, precision, and recall of the ensemble technique's prediction of missing social network linkages are enhanced. More network links and patterns are captured by the complementary algorithms. The ensemble approach is beneficial for recommendation systems, network analysis, and community identification. The adaptability of the ensemble model is advantageous to students and practitioners in social network analysis.

REFERENCES

[1] A. Papadimitriou, P. Symeonidis, and Y. Manolopoulos, “Fast and accurate link prediction in social networking systems,” *J. Syst. Softw.*, vol. 85, no. 9, pp. 2119–2132, 2012, doi: 10.1016/j.jss.2012.04.019.

[2] Y. L. Chen, C. H. Hsiao, and C. C. Wu, “An ensemble model for link prediction based on graph embedding,” *Decis. Support Syst.*, vol. 157, no. August 2021, p. 113753, 2022, doi: 10.1016/j.dss.2022.113753.

[3] S. S. Singh, D. Srivastva, A. Kumar, and V. Srivastava, “FLP-ID: Fuzzy-based link prediction in multiplex social networks using information diffusion perspective,” *Knowledge-Based Syst.*, vol. 248, p. 108821, 2022, doi: 10.1016/j.knosys.2022.108821.

[4] X. Liu, X. Li, G. Fiumara, and P. De Meo, “Link prediction approach combined graph neural network with capsule network,” *Expert Syst. Appl.*, vol. 212, no. August 2022, p. 118737, 2023, doi: 10.1016/j.eswa.2022.118737.

[5] N. N. Daud, S. H. Ab Hamid, M. Saadoon, F. Sahran, and N. B. Anuar, “Applications of link prediction in social networks: A review,” *J. Netw. Comput. Appl.*, vol. 166, no. February, 2020, doi: 10.1016/j.jnca.2020.102716.

[6] W. Peng, X. U. Baowen, W. U. Yurong, and & Zhou Xiaoyu, “011101:1-011101:38 State-of-the-Art,” *Sci China Inf Sci*, vol. 58, no. January, p. 11101, 2015, doi: 0.1007/s11432-014-5237-y.

[7] H. Wang, Z. Cui, R. Liu, L. Fang, and Y. Sha, “A Multi-type Transferable Method for Missing Link Prediction in Heterogeneous Social Networks,” *IEEE Trans. Knowl. Data Eng.*, pp. 1–13, 2023, doi: 10.1109/tkde.2022.3233481.

[8] S. Haghani and M. R. Keyvanpour, “A systemic analysis of link prediction in social network,” *Artif. Intell. Rev.*, vol. 52, no. 3, pp. 1961–1995, 2019, doi: 10.1007/s10462-017-9590-2.

[9] W. Li, T. Li, and K. Berahmand, “An effective link prediction method in multiplex social networks using local random walk towards dependable pathways,” *J. Comb. Optim.*, vol. 45, no. 1, pp. 1–27, 2023, doi: 10.1007/s10878-022-00961-z.

[10] Z. Gao and A. Rezaeipanah, “A Novel Link Prediction Model in Multilayer Online Social Networks Using the Development of Katz Similarity Metric,” *Neural Process. Lett.*, 2022, doi: 10.1007/s11063-022-11076-1.

[11] M. N. Abadeh and M. Mirzaie, “A differential machine learning approach for trust prediction in signed social networks,” *J. Supercomput.*, no. 0123456789, 2023, doi: 10.1007/s11227-023-05044-2.

[12] Y. Yang, Z. Ye, H. Zhao, and L. Meng, “A Novel Link Prediction Framework Based on Gravitational Field,” *Data Sci. Eng.*, vol. 8, no. 1, pp. 47–60, 2023, doi: 10.1007/s41019-022-00201-8.

[13] S. U. Balvir, M. M. Raghuwanshi and K. R. Singh, "A Comprehensive Survey on Learning Based Methods for Link Prediction Problem," 2023 6th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2023, pp. 1-7, doi: 10.1109/ISCON57294.2023.10112010

[14] Sahu, S.K.; Mokhade, A.; Bokde, N.D. “An Overview of Machine Learning, Deep Learning, and Reinforcement Learning-Based Techniques in Quantitative Finance: Recent Progress and Challenges”, Appl. Sci. 2023, 13, 1956. https://doi.org/10.3390/app13031956

[15] S. Aouay, S. Jamoussi and F. Gargouri, "Feature based link prediction," 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA), Doha, Qatar, 2014, pp. 523-527, doi: 10.1109/AICCSA.2014.7073243.

[16] F. Liu, B. Liu, X. Wang, M. Liu and B. Wang, "Features for link prediction in social networks: A comprehensive study," 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Seoul, Korea (South), 2012, pp. 1706-1711, doi: 10.1109/ICSMC.2012.6377983.

[17] Cao, J.; Lei, T.; Li, J.; Jiang, J. A Novel Link Prediction Method for Social Multiplex Networks Based on Deep Learning. Mathematics 2023, 11, 1705. https://doi.org/10.3390/math11071705

[18] S. Aouay, S. Jamoussi and F. Gargouri, "Hybrid Swarm Based Method for Link Prediction in Social Networks," 2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI), Vietri sul Mare, Italy, 2015, pp. 974-981, doi: 10.1109/ICTAI.2015.140.

[19] Kerkache, M. H., Sadeg-Belkacem, L., & Benbouzid-Si Tayeb, F. (2023). A hybrid approach for enhanced link prediction in social

networks based on community detection. International Journal of General Systems, 53(2), 154–183. https://doi.org/10.1080/03081079.2023.2265043

[20] F. Gao and K. Musial-Gabrys, "Hybrid structure-based link prediction model," 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, USA, 2016, pp. 1221-1228, doi: 10.1109/ASONAM.2016.7752394.

Sachin U. Balvir is a Ph.D. research scholar in Computer science and Technology in Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur, India. He did his M.Tech in Computer science and Engineering form Visvesvaraya National Institute of Technology (VNIT), Nagpur, India. He has more than ten research publication various international and national journals and conferences. His area of interest includes Web mining, social network analysis, link prediction. He can be contacted at email: sachin_balvir@yahoo.com

Dr. Mukesh M. Raghuwanshi received a Ph.D. in Computer Science from Visvesvaraya National Institute of Technology (VNIT), Nagpur, India. He did his M.Tech in Computer Science from Indian Institute of Technology (IIT), Kharagpur, India. He is currently working as Dean Engineering at S B Jain Institute of Technology, Management and Research, Nagpur. He is a highly ingenious mentor and passionate technocrat with more than 37 years of experience in academics, administration, software project development & management, and hardware installations & maintenance. He has a proven track record in establishing new institutes & departments. He has Expertise in the design and delivery of engineering curriculum and courses. He has published 2 books, 67 Journal papers, 54 conference papers, 15 book chapters, 5 copyrights and 2 patents. His areas of specialization are Genetic algorithm, multi-objective optimization, Algorithms, Data science. He can be contacted at email: deanengineering@sbjit.edu.in

Dr. Purushottam D. Shobhane obtained his Ph.D. degree in Mathematics from Gondwana University, Gadchiroli, India. He obtained his M.Sc. degree in Mathematics from RTM Nagpur University, Nagpur, India. He has been engaged in teaching for more than 30 years and in research for more than 10 years. At present he is working as Assistant Professor in CSE Department at Symbiosis Institute of Technology Nagpur, Symbiosis International (Deemed University) Pune, MH, India. He has authored 4 books on Mathematics and has presented more than 15 papers in International/National Journals/ Conferences. His research interests include General Theory of Relativity & Cosmology and Mathematical Modeling. He can be contacted at email: purushottam.shobhane@sitnagpur.siu.edu.in