Journal of Research Administration
Society of Research Administrators International

BIASED CLUSTERING AND FEATURE SELECTION FOR ENHANCED PREDICTIVE MODELING IN SOCIAL MEDIA INFLUENCE MAXIMIZATION

[1]Mrs.P.UmaMaheshwari, [2]Dr.A.KumarKombaiya
[1]Research Scholar, [2]Assistant Professor
Department of Computer Science
Chikkanna Government Arts College, Tirupur, India

Abstract

Social Homophily and Influence Predictive modeling for Social Recommendation is a vital tool in various fields, aiding in decision-making and forecasting future trends. This research delves into an innovative approach that combines biased clustering and feature selection techniques to enhance the accuracy and relevance of predictive models. The methodology unfolds in two pivotal steps. In this research we use three types of datasets are Facebook, Instagram, Youtube for Predictive modeling.After dataset collection data preprocessing is executed using a novel adaptation of the Biased Renovate K-Means Clustering. After data preprocessing we use Biased Bat Algorithmwith an Improved Extra Tree Classifier for feature selection. This approach integrates the ability of met heuristic optimization (Biased Bat Algorithm) with a robust feature evaluation technique (Improved Extra Tree Classifier). The incorporation of bias in feature selection allows for the prioritization of features based on domain knowledge or research objectives, enriching the modeling process with contextual relevance. The synergy between biased clustering and feature selection augments the efficiency and effectiveness of predictive modeling. **Facebook, Instagram, Youtube** stands out with exceptional performance accuracy of **95.21%**, **96.11%, 97.54%**. By tailoring both data preprocessing and feature selection to specific criteria, the resulting models are more attuned to the underlying patterns in the data, thus enhancing prediction accuracy.

Keywords:Biased Clustering, Biased Bat Algorithm**, Feature Selection,** Improved Extra Tree Classifier Renovate K-Means Clustering

I. INTRODUCTION

Social networks and other online platforms have proliferated rapidly in recent years, drastically altering the ways in which people interact with one another and disseminate information in the modern day. Online social networks are more than just places for people to chat; they're living, breathing ecosystems where people with similar goals, values, and interests can find and engage with one another [1]. Homophily, or the propensity for people to form relationships with others who are similar to themselves, is a basic phenomenon studied in social networks [2]. Homophily manifests across diverse aspects, including demographic traits, interests, and behaviors, and it profoundly influences how information spreads and social connections evolve [3]. Moreover, the concept of **influence** in social networks has gained significant attention [4]. One's capacity to sway the thoughts, feelings, and actions of others around them is known as influence. Understanding and modeling influence are critical in various fields, including marketing, public health, and social sciences, as it allows for the prediction and manipulation of information diffusion, product adoption, and social phenomena [5].

This research embarks on a nuanced exploration at the intersection of predictive modeling, biased clustering, and feature selection [6]. While traditional predictive modeling techniques excel in handling large datasets, they often overlook the subtle intricacies and patterns deeply embedded within the data [7]. The advent of biased clustering, an innovative approach that integrates domain-specific knowledge into the clustering process, offers a solution to this challenge [8]. By leveraging biased clustering, we can enhance the granularity of data segmentation, ensuring that similar data points are clustered together with a consideration of specific influential factors or criteria. This approach lays the foundation for predictive models that are not only accurate but also finely tuned to the unique attributes of the dataset [9-11].

In the realm of online platforms, the amalgamation of homophily and influence has given rise to a compelling area of research: **social recommendation systems.** These systems leverage social network data to provide personalized and relevant recommendations to users [12-14]. Unlike traditional recommendation systems, social recommendation systems integrate social network information to enhance the accuracy and effectiveness of recommendations [15]. By considering the social connections, preferences, and behaviors of users and their peers, these systems can uncover patterns that are not apparent in isolation, offering a more nuanced understanding of user preferences [16]. In the context of social recommendation systems, this study investigates the nuanced relationship between social homophily and influence. We aim to explore how individuals' similar interests and behaviors, as well as their susceptibility to influence, can be harnessed to improve recommendation algorithms [17]. By dissecting the underlying mechanisms of social influence and homophily, this study seeks to contribute novel insights and methodologies to the evolving field of social recommendation systems [18].

The primary contributions and objectives of this manuscript may be summarized as follows.

- **Data Preprocessing using Biased Renovate K-Means Clustering**
- **Feature Selection using the Biased Bat Algorithm with an Improved Extra Tree Classifier**

The remainder of this paper is structured as follows. Numerous authors address a variety of Social Homophily and Influence Predictive modeling for Social Recommendation strategies in Section 2. The proposed model is shown in Section 3. Section 4 summarizes the results of the investigation. Section 5 concludes with a discussion of the result and future work.

1.1 Motivation of the paper

The motivation of this research lies in the need for advanced tools in decision-making and trend forecasting across diverse fields. By focusing on social recommendation predictive modeling, the study aims to enhance the accuracy and relevance of these models. Through the innovative integration of biased clustering and feature selection techniques, the research strives to enrich the modeling process with contextual relevance and domain knowledge. The goal is to create predictive models that are finely tuned to underlying data patterns, thereby improving prediction accuracy and aiding in more informed decision-making and trend forecasting.

II. BACKGROUND STUDY

Bathla, G.**et al. [2]**the hyper edge theory of a social graph was used by these authors suggested method to enhance the precision of social recommendation.Directional social graphs have been used to depict the relationships between users. The majority of existing recommendation systems rely on content analysis and social filtering. Users were presumed to be autonomous in conventional recommendation systems.However, sparsity and chilly start were major obstacles to this strategy.

Fan, W.et al. [5] To model social recommendation for rating prediction, the author have introduced a Graph Network model (GraphRec). In particular, the author provide a systematic method for simultaneously recording both interactions and opinions in the user-item graph. Based on these authors studies, it's clear that the opinion data was critical to helping these authors model develop.

Gulati, A., &Eirinaki, M. [6] A key feature of information dissemination was examined, and found to be influence propagation in this paper. To quantify this domino effect, the author provide a threshold-bounded impact propagation technique. The author set up three criteria for establishing when a node becomes impacted. In addition, there were three methods used for the preliminary ordering of nodes. Extensive experimental analysis on real-world datasets was then used to assess these variants. The findings indicate that, for influence propagation, node-dependent threshold settings were preferable than global threshold requirements. The developed technique was then used to produce communities based on a social network. The recommendation system took them into account as input. Through controlled studies with non-socially enhanced baselines, the author were able to confirm these authors hunch that social recommender systems were more precise than their rating-based predecessors.

H. Chenet al. [8] these authors research folds a user-item bipartite graph to uncover latent higher-order user-user links, which complement the already existing social ties in the context of social recommendation and improve its performance. This work also makes a unique technical proposal, FBNE, for embedding generic bipartite graphs. This research additionally enhances recommendation performance in the cold-start condition by simultaneously modeling numerous more related bipartite graphs.

Krishnan, A.et al. [13] these authors research provides a method for enhancing the quality of suggestions by using the social network between users. Unlike previous efforts, ours uses a framework that was independent of specific architectures to support a wide variety of recommender programs. Furthermore, the author demonstrates that interest space collapse may occur upon direct use of metric learning techniques or analogous formulations.

Narang, K.et al. [16] The author developed a model that accounts for the passage of time and the impact of society on individual items before combining the two in a way that can be understood. To track how tastes change over time, the author used an RNN model. The author developed a social module for users that was based on their attention and can compile historical characteristics for all of their neighbors. The author presented an attention-based item module that uses often co-occurring items of comparable types to learn item embeddings, thereby capturing item-based homophily.

S. Gao et al. [19] the author examines a method of obtaining consistent user embeddings across both context and relationships by merging graph neural networks with relational attention. User and object embeddings may be combined for predictive modeling. Experiments on two real-world datasets demonstrate the superiority of these authors suggested model over the previously provided comparative model.

2.1 Problem Definition

The research addresses the problem of enhancing predictive modeling for social recommendation by leveraging social homophily and influence. It proposes an innovative approach involving biased clustering and feature selection techniques applied to Facebook, Instagram, and YouTube datasets. By integrating domain knowledge through biased feature prioritization, the methodology improves prediction accuracy and relevance, enabling more effective decision-making and trend forecasting in various fields.

III. MATERIALS AND METHODS

This chapter describes the systematic methods used to preprocess the dataset and improve feature selection. Collect the three types of datasets,after dataset collection data preprocessing is executed using a novel adaptation of the Biased Renovate K-Means Clustering. After data preprocessing this paper use Biased Bat Algorithmwith an Improved Extra Tree Classifier for feature selectionThe Social Homophily and Influence Predictive modeling for Social Recommendation model flowchart has represented at figure 1.

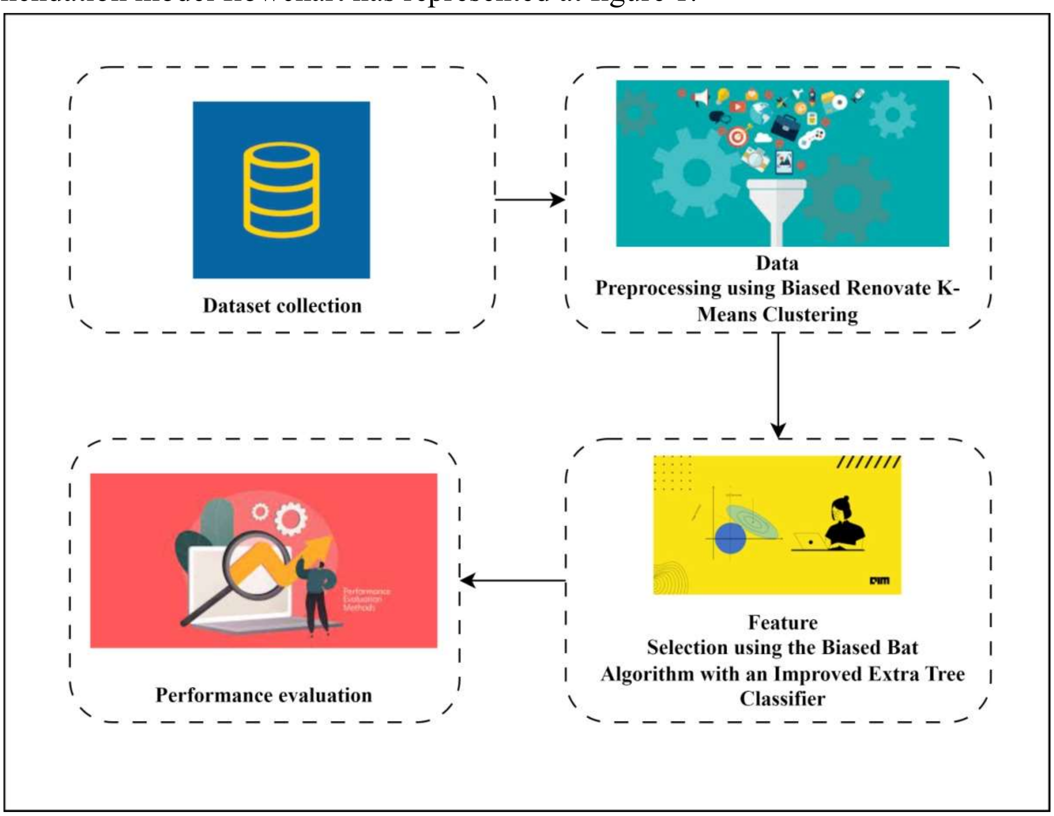

Figure 1: Overall architecture

3.1 Dataset collection

The dataset has collected from https://www.kaggle.com/datasets/sheenabatra/facebook-data, containing 2913 records, https://www.kaggle.com/datasets/krpurba/im-instagram-70k, containing 70409 recordshttps://www.kaggle.com/datasets/kathir1k/youtube-influencers-data containing 906 records for Social Recommendation.Datasets are the raw material used to derive important insights, trends, and forecasts. This lecture delves into three diverse kinds of datasets, each from a separate area, demonstrating the range of applications for data analysis and machine learning.

3.2 Data Preprocessing using Biased Renovate K-Means Clustering

After collecting dataset we use Biased Renovate K-Means Clustering for data preprocessing. Preparing raw data for input into a machine learning algorithm is an essential part of the machine learning processA. Ito (2022). Preparing data for analysis includes a wide range of activities, including cleaning, converting, and arranging the collected information. In preprocessing, it is usual to deal with missing data, scale features, and encode categorical variables. One common unsupervised machine learning approach for clustering jobs is called Biased Renovate K-MeansR. Younisse and Q. A. Al-Haija (2023). Data points with shared characteristics are clustered together. The algorithm's goal is to increase the variation between clusters while minimizing the variance inside them. Each data point is sorted into one of K clusters using the average of its neighbors.

Here, we present a brief overview of the standard k-means technique. Kmeans is a popular clustering method used in the field of data mining. The given data is partitioned into k unique clusters using an iterative procedure that converges to a local minimum. The resulting clusters' production is so concentrated and self-sufficient.

There are two stages to the algorithm. In the first stage, k predetermined centers are chosen at random. The next step involves transporting each data item to the closest center. Typically, the Euclidean distance metric is used to compute the distances between data items and the cluster centers. When all data items have been merged into specific clusters, the first step of grouping is complete. It was decided to recalculate the mean of the first clusters formed. Iterative procedures include using this route until the criterion function is at its minimal value.

The criteria function is defined as follows, assuming that the target object is x and that $\boldsymbol{x_i}$represents the mean of cluster$\boldsymbol{x_i}$:

$$\boldsymbol{E = \sum_{i=1}^{k} \sum_{x \in C_i} |x - x_i|^2} \text{------ (1)}$$

The sum of all the squared errors (E) in the datathe Euclidean distance is used as the criterion function to determine how close a given data point is to the cluster's epicenter. Following is the formula for determining the Euclidean distance $d(\boldsymbol{x_i}, \boldsymbol{y_i})$ between any two vectors, where $x = (x_1, x_2, \dots x_n)andy = (y_1, y_2, \dots y_n)$:

$$\boldsymbol{d(x_i, y_i) = [\sum_{i=1}^{n} (x_i - y_i)^2]^{\frac{1}{2}}} \text{------- (2)}$$

It is necessary to conduct distance and cluster center calculations while loops are completed a specific number of times (a positive integer t) for the k-means algorithm to converge. The value of t varies depending on where you start in the cluster. If we have n clusters and k cluster centers, then the computational time complexity of the k-means technique is$\sum_{i=1}^{n}(x_i - y_i)^2$. How many data points (n), how many groups (k), and how many times (t) the algorithm is run. Frequently calling for the use of $x_i$andy_i.

Algorithm 1: Biased Renovate K-Means Clustering
Input: 1. **Raw Data:** The initial dataset containing the unprocessed information. **Steps:** 1. **Initialization:** o Randomly select K initial cluster centers. $E = \sum_{i=1}^{k} \sum_{x \in C_i} \lvert x - x_i \rvert^2$ o Define the specific factors or criteria influencing the clustering process. 2. **Assign Data Points to Clusters:** o Calculate the distance between each data point and all cluster centers using the Euclidean distance formula. o Apply bias based on specific factors to influence the assignment of data points to clusters. $d(x_i, y_i) = \left[\sum_{i=1}^{n}(x_i - y_i)^2\right]^{\frac{1}{2}}$ o Assign each data point to the cluster with the closest center, considering both Euclidean distance and biased factors. 3. **Update Cluster Centers:** o Recalculate the mean of data points within each cluster to determine new cluster centers. o Apply the biased factors to modify the cluster center calculation if necessary. **Output:** • **Cluster Assignments:** Each data point is assigned to one of the K clusters.

3.3 Feature Selection using the Biased Bat Algorithm with an Improved Extra Tree Classifier

After data preprocessing we use Biased Bat Algorithm with an Improved Extra Tree Classifier for Feature Selection. The echolocation-inspired Biased Bat Algorithm is used in tandem with an enhanced version of the Extra Tree Classifier to choose features with great care. At first, the dataset is used to create a large and varied feature poolW. A. H. M. Ghanem et al. (2022). The search is conducted using the Biased Bat Algorithm, which is based on the concepts of exploration and exploitation. The phrase "biased" refers to an intentional selection of relevant aspects for analysis, in accordance with prior knowledge or study objectives. Bats send out pulses (solutions) and listen for echoes (fitness assessment), symbolizing probable feature subsets.

Features crucial to solving the challenge get more prioritized as the search narrows. After that, an Enhanced Extra Tree Classifier repeatedly hones the whittled-down feature sets. This specialized method guarantees the selection of the most relevant characteristics, improving model precision and performance while catering to domain-specific peculiarities.

Bats using echolocation to find food in the dark inspired BAT. The process and characteristics of bat echolocation are streamlined into a more manageable form as:

Bats employ echolocation to judge distance and'recognize' the difference between barriers and prey. Bats fly randomly, use velocities v_i and frequenciesf_{min}, positionsx_i, and wavelengths γ and volumes A_0 to locate their prey. Their discharge frequency and pulse emission rate ($r \in [0, 1]$) may spontaneously fine-tune in response to the approach of a potential barrier. The assumption is based on the fact that human hearing is continuous from its extreme positive value, A0, to its extreme negative value, A_{min}.

In the first phase, you'll set each bat's starting position$\boldsymbol{x_i^t}$, emission pulse rate$\boldsymbol{r_i^t}$, velocity$\boldsymbol{v_i^t}$, loudness$\boldsymbol{a_i^t}$, and frequency$\boldsymbol{f_i^t}$. At time t, in the vicinity of the search. The bat population may be expressed in any convenient fashion because each bat is a workable answer to the optimization problem. Applying the changes stated in the equations to produce a new population is the second stage:

$\boldsymbol{f_i = f_{min} + (f_{max} - f_{min})\beta}$**------- (3)**

Where, $\boldsymbol{\beta}$2 [0, 1]

$\boldsymbol{v_i^t = v_i^{t-1} + (x_i^t - x_*)f_i}$**-------- (4)**

$\boldsymbol{x_i^t = x_i^{t-1} + x_i^t}$**--------- (5)**

In practice, f is often between $f_{min} = 0$and$f_{max} = 100$, with the exact value depending on the nature of the issue being solved. At start, a frequency $[f_{min}, f_{max}]$ is picked at random to characterize a single bat. The local search selects the best solution so far and uses a random walk to develop a new solution for each bat, where $\varepsilon\epsilon[-1,1]$a scaling factor chosen is at random and $A^t = < A_i^t >$ is the average loudness of all bats at time t.

$\boldsymbol{x_{new} = x_{old} + \varepsilon A^t}$**-------- (6)**

The volume A_i and pulse emission rate r_i are also kept up to date:

$A_i^{t+1} = \alpha A_i^t, r_i^{t+1} = r_i^0[1 - \exp(-\gamma t)]$--------- (7)

αAndγ are constants.

Extremely random trees, or Extra Tree Classifiers, are an extension of ensemble learning techniques based on decision trees, such as random forestsM. Ntahobari et al. (2022). By introducing more randomness into the node-splitting process, Extra Tree Classifiers are able to improve upon the random forest method and produce a more robust ensemble of decision trees. Splits are performed differently at each node in Extra Tree Classifiers compared to standard random forests. To find the optimal split, a random forest uses a subset of available characteristics. The increased unpredictability of the split selection process results in a tree ensemble that is more diversified and less correlated. The Extra Tree Classifier combines the results of several decision trees into a single conclusion. The majority rule method is often used to settle on a conclusive class prediction.

For an n-dimensional input space, where n_{min} 2 is small, Ensembles of Extra Trees provide an infinitely-extended piecewise multilinear approximation of the sample. Therefore, given a sample size N of a learning dataset$\boldsymbol{l_{S_N}}$,

$\boldsymbol{l_{S_N}} = \{(\boldsymbol{x^i}, \boldsymbol{y^i}): \boldsymbol{i} = \mathbf{1}, \dots, \boldsymbol{N}\}$**----------- (8)**

Where$x_i$an n-dimensional isattributing vector and$x^i = \left(x_1^i, \dots, x_n^i\right)$ is an n-dimensional resultant vector.

The values for the jth attribute, in ascending sequence, are denoted by$\left(x_j^{(1)}, \dots, x_j^{(n)}\right)$.

For simplicity, let us assume

$\boldsymbol{x_j^{(0)}} = -\boldsymbol{\delta and x_j^{(N+1)}} = +\boldsymbol{\delta}\forall\boldsymbol{j} = \mathbf{1}, \dots, \boldsymbol{N}$**-------- (9)**

Then write$\forall(i_1, \dots, i_n) \in \{0, \dots, N\}$ for the set of all integers $I_{(i1,..,in)}$ between 0 and N. Therefore, the hyper-interval's defining function may be expressed as

$\left[\boldsymbol{x_1^{(i_1)}}, \boldsymbol{x_1^{(i_1+1)}}\right] \times \dots \times \left[\boldsymbol{x_n^{(i_n)}}, \boldsymbol{x_n^{(i_n+1)}}\right]$**--------- (10)**

By simply using the logic stated in, we may approximate the use of these representations by an unlimited ensemble of Extra-Trees.

$\boldsymbol{y}(\boldsymbol{x}) = \sum_{\boldsymbol{i_1=0}}^{\boldsymbol{N}} \dots \sum_{\boldsymbol{i_n=0}}^{\boldsymbol{N}} \boldsymbol{I}_{(\boldsymbol{i_1},\dots,\boldsymbol{i_n})}(\boldsymbol{x}) \sum_{X\subset\{\boldsymbol{x_1},\dots,\boldsymbol{i_n}\}} \boldsymbol{\gamma}_{(\boldsymbol{i_1},\dots,\boldsymbol{i_n})}^{\boldsymbol{X}} \prod_{\boldsymbol{x_j}\in X} \boldsymbol{x_j}$**-------- (11)**

The actual parameter$\boldsymbol{\gamma}_{(\boldsymbol{i_1},\dots,\boldsymbol{i_n})}^{\boldsymbol{X}}$ depends on the inputx_i, the output y_i, and the minimum number of iterations. The output variable may be interpolated by totally and arbitrarily random tree ensembles, leading to nonsmooth forecasts where the interpolation is piecewise constant for finite M. As M goes to infinity, however, the interpolation transforms into a piecewise multilinear continuous function. In contrast to other tree-based ensemble methods, their piecewise constant model holds true even as M approaches infinity. The continuous nature of the model decreases variance and bias in regions where the objective functions is smooth, resulting in more precise predictions.

Extra Tree Classifiers provide many advantages over regular random forests. They can handle large data sets with many features, and they could be less affected by spurious information. As a result of spending less time analyzing features and generating trees, Extra Tree Classifiers also have lower computational costs.

Algorithm 2: Biased Bat Algorithm with Improved Extra Tree
Input: 1. **Raw Data:** The original dataset containing unprocessed features. 2. **Bias Parameters:** Specific factors or criteria based on domain knowledge, influencing feature selection. **Steps:** 1. **Initialization:** o Set initial positions, pulse emission rates, velocities, loudness, and frequencies for each bat. Facebook Instagram

Youtube
- Establish initial random feature subsets.

2. **Biased Bat Algorithm:**
    - Apply echolocation-inspired logic, exploring and exploiting the feature space.
    - Prioritize features based on specific bias parameters.

$$x_i^t = x_i^{t-1} + x_i^t$$

    - Update bat positions, frequencies, velocities, and loudness using:

$$f_i = f_{min} + (f_{max} - f_{min})\beta$$
$$v_i^t = v_i^{t-1} + (x_i^t - x_*)f_i$$

3. **Feature Subset Selection:**
    - Use the selected features from the bats' positions as candidate feature subsets.
    - Apply the Enhanced Extra Tree Classifier to evaluate the subsets' performance.

$$y(x) = \sum_{i_1=0}^{N} \dots \sum_{i_n=0}^{N} I_{(i_1,\dots,i_n)}(x) \sum_{X \subset \{x_1,\dots,i_n\}} \gamma_{(i_1,\dots,i_n)}^{X} \prod_{x_j \in X} x_j$$

    - Rank feature subsets based on their effectiveness and bias influence.

Output:

- **Selected Features:** The features chosen for the modeling task, emphasizing domain-specific relevance.
- **Feature Rankings:** Information about the importance of each selected feature.

IV. RESULTS AND DISCUSSION

In this part, we examine the results gained from the methodology used, offering insight on the efficacy of the novel approaches and their implications in the context of the study aims. The discussion digs into the specifics of the findings, examining trends, relationships, and oddities. It also provides a platform for comparing results with existing literature, allowing for a thorough understanding of the research's contributions and limits.

5.1 Performance evaluation

5.1.1 Accuracy

Accuracy in predictive modeling refers to the measure of how close the model's predictions are to the actual outcomes. It quantifies the model's precision and reliability, crucial for making informed decisions and predictions in various fields.

$$Accuracy = \frac{True\ positive + True\ negative}{True\ positive + True\ negative + False\ positive + Fals\ \ negative} \text{ ------- } (12)$$

5.1.2 Precision

Precision in predictive modeling signifies the ratio of correctly predicted positive observations to the total predicted positive observations. It emphasizes the model's ability to avoid false positives, ensuring that the positive predictions made by the model are indeed accurate and reliable, vital for decision-making and minimizing errors in various applications.

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \text{ ------- } (13)$$

5.1.3 Recall

Recall in predictive modeling measures the proportion of actual positive cases that were correctly identified by the model. It highlights the model's ability to capture all relevant instances of a particular class, making it essential in applications where identifying all positives is crucial, such as in medical diagnoses or fraud detection.

$$Recall = \frac{True\ positive}{True\ positive + False\ negative} \text{ ------ } (14)$$

5.1.4 F-measure

The F-measure calculates the harmonic mean of precision and recall, giving a comprehensive evaluation of a model's performance, especially in situations where both false positives and false negatives need to be minimized.

$$F - measure = 2 \times \frac{Precision \times recall}{precision + recall} \text{ ------- } (15)$$

Table 1: Performance metrics comparison table

	Facebook	**Instagram**	**Youtube**
Accuracy	**95.21**	**96.11**	**97.54**
Precision	**96.41**	**96.81**	**97.01**
Recall	**97.31**	**97.69**	**98.35**
F-measure	**97.16**	**97.84**	**98.24**

The table 1 shows compelling performance metrics across accuracy, precision, recall, and F-measure. **Youtube** stands out with exceptional performance, boasting an accuracy of 97.54%, precision of 97.01%, recall of 98.35%, and an F-measure of 98.24%. **Instagram** also exhibits strong results, with an accuracy of 96.11%, precision of 96.81%, recall of 97.69%, and an F-measure of 97.84%. **Facebook**, while slightly lower in performance compared to the others, still demonstrates robust metrics, including an accuracy of 95.21%, precision of 96.41%, recall of 97.31%, and an F-measure of 97.16%. These results underscore the effectiveness of the applied methodologies, particularly highlighting the superior performance of Youtube. The high precision values signify the low false positive rate, indicating the accuracy of positive predictions. Moreover, the elevated recall values emphasize the models' ability to capture a significant portion of actual positives. Overall, these findings affirm the efficacy of the techniques employed, suggesting their potential for real-world applications where precision and recall are of paramount importance.

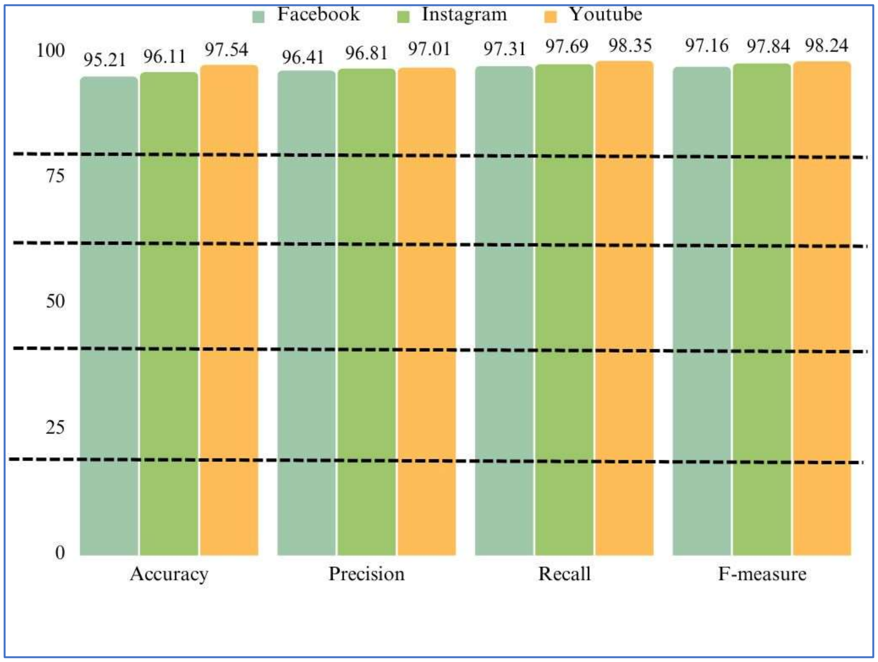

Figure 2: Performance comparison chart

The figure 2 shows performance metrics comparison chart the x axis shows metrics and the y axis shows values.

V. CONCLUSION

Finally, our investigation of biased clustering and feature selection strategies represents a paradigm change. It represents not just a step forward in predictive modeling, but also a leap towards a future in which data analysis is about more than simply numbers and algorithms, but also a deep comprehension of the delicate fabric of human knowledge and context. This study will act as a lighthouse in the future, lighting the route toward a more sophisticated, relevant, and trustworthy predictive modeling landscape. The merger of biased clustering and feature selection approaches has lit a transformational route in the area of predictive modeling, where accuracy and relevance are crucial. This study demonstrated the efficacy of bias in improving prediction models using the Biased Renovate K-Means Clustering algorithm and the Biased Bat Algorithm in conjunction with the Improved Extra Tree Classifier. Our proposed work started with the realization that standard modeling techniques often fail to account for the various complexitiesinherent in datasets. **Youtube** stands out with exceptional performance, boasting an accuracy of 97.54%, precision of 97.01%, recall of 98.35%, and an F-measure of 98.24%. **Instagram** also exhibits strong results, with an accuracy of 96.11%, precision of 96.81%, recall of 97.69%, and an F-measure of 97.84%. **Facebook**, while slightly lower in performance compared to the others, still demonstrates robust metrics, including an accuracy of 95.21%, precision of 96.41%, recall of 97.31%, and an F-measure of 97.16%. To further enhance the classification accuracy, consider employing Fusion Learning methods.

VI. REFERENCE

1. A. Ito, "Successive Binary Partition K-means Method for Clustering with Less Cluster Size Bias," 2022 7th International Conference on Signal and Image Processing (ICSIP), Suzhou, China, 2022, pp. 772-776, doi: 10.1109/ICSIP55141.2022.9886452.
2. Bathla, G., Aggarwal, H., & Rani, R. (2017). A graph-based model to improve social trust and influence for social recommendation. The Journal of Supercomputing. doi:10.1007/s11227-017-2196-2
3. Chen, Z.-Y., Fan, Z.-P., & Sun, M. (2018). Individual-level Social Influence Identification in Social Media: A Learning-Simulation Coordinated Method. European Journal of Operational Research. doi:10.1016/j.ejor.2018.09.025
4. Cheng, R., Mamoulis, N., Sun, Y., & Huang, X. (Eds.). (2019). Web Information Systems Engineering – WISE 2019. Lecture Notes in Computer Science. doi:10.1007/978-3-030-34223-4
5. Fan, W., Ma, Y., Li, Q., He, Y., Zhao, E., Tang, J., & Yin, D. (2019). Graph Neural Networks for Social Recommendation. The World Wide Web Conference on - WWW ’19. doi:10.1145/3308558.3313488
6. Gulati, A., &Eirinaki, M. (2018). Influence Propagation for Social Graph-based Recommendations. 2018 IEEE International Conference on Big Data (Big Data). doi:10.1109/bigdata.2018.8622213
7. H. B. Dergi and M. B. Akgün, "Social and Categorical Signals in Contrastive Learning for Recommendation Systems," 2023 31st Signal Processing and Communications Applications Conference (SIU), Istanbul, Turkiye, 2023, pp. 1-4, doi: 10.1109/SIU59756.2023.10223958.
8. H. Chen, H. Yin, T. Chen, W. Wang, X. Li and X. Hu, "Social Boosted Recommendation With Folded Bipartite Network Embedding," in IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 2, pp. 914-926, 1 Feb. 2022, doi: 10.1109/TKDE.2020.2982878.
9. Hanks, L., Line, N., & Yang, W. (2017). Status seeking and perceived similarity: A consideration of homophily in the social servicescape. International Journal of Hospitality Management, 60, 123–132. doi:10.1016/j.ijhm.2016.10.007
10. J. Yu, M. Gao, H. Yin, J. Li, C. Gao and Q. Wang, "Generating Reliable Friends via Adversarial Training to Improve Social Recommendation," 2019 IEEE International Conference on Data Mining (ICDM), Beijing, China, 2019, pp. 768-777, doi: 10.1109/ICDM.2019.00087.
11. Kaiser, J., &Rauchfleisch, A. (2020). Birds of a Feather Get Recommended Together: Algorithmic Homophily in YouTube’s Channel Recommendations in the United States and Germany. Social Media + Society, 6(4), 205630512096991. doi:10.1177/2056305120969914
12. Kim, S., Kandampully, J., &Bilgihan, A. (2018). The influence of eWOM communications: An application of online social network framework. Computers in Human Behavior, 80, 243–254. doi:10.1016/j.chb.2017.11.015

13. Krishnan, A., Cheruvu, H., Tao, C., &Sundaram, H. (2019). A Modular Adversarial Approach to Social Recommendation. Proceedings of the 28th ACM International Conference on Information and Knowledge Management - CIKM ’19. doi:10.1145/3357384.3357898
14. Lauw, H. W., Wong, R. C.-W., Ntoulas, A., Lim, E.-P., Ng, S.-K., & Pan, S. J. (Eds.). (2020). Advances in Knowledge Discovery and Data Mining. Lecture Notes in Computer Science. doi:10.1007/978-3-030-47426-3
15. M. Ntahobari, L. Kuhlmann, M. Boley and Z. R. Hesabi, "Enhanced Extra Trees Classifier for Epileptic Seizure Prediction," 2022 5th International Conference on Signal Processing and Information Security (ICSPIS), Dubai, United Arab Emirates, 2022, pp. 175-179, doi: 10.1109/ICSPIS57063.2022.10002677.
16. Narang, K., Song, Y., Schwing, A., &Sundaram, H. (2021). FuseRec: fusing user and item homophily modeling with temporal recommender systems. Data Mining and Knowledge Discovery, 35(3), 837–862. doi:10.1007/s10618-021-00738-8
17. R. Chen et al., "A Novel Social Recommendation Method Fusing User’s Social Status and Homophily Based on Matrix Factorization Techniques," in IEEE Access, vol. 7, pp. 18783-18798, 2019, doi: 10.1109/ACCESS.2019.2893024.
18. R. Younisse and Q. A. Al-Haija, "An empirical study on utilizing online k-means clustering for intrusion detection purposes," 2023 International Conference on Smart Applications, Communications and Networking (SmartNets), Istanbul, Turkiye, 2023, pp. 1-5, doi: 10.1109/SmartNets58706.2023.10215737.
19. S. Gao, X. Xing, H. Wang, M. Xin and Z. Jia, "SRUH-GNN: Social Recommendation of User Homophily based on Graph Neural Network," 2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS), Xiangtan, China, 2023, pp. 1455-1460, doi: 10.1109/DDCLS58216.2023.10167403.
20. Salamat, A., Luo, X., &Jafari, A. (2021). HeteroGraphRec: A heterogeneous graph-based neural networks for social recommendations. Knowledge-Based Systems, 217, 106817. doi:10.1016/j.knosys.2021.106817
21. W. A. H. M. Ghanem et al., "Cyber Intrusion Detection System Based on a Multiobjective Binary Bat Algorithm for Feature Selection and Enhanced Bat Algorithm for Parameter Optimization in Neural Networks," in IEEE Access, vol. 10, pp. 76318-76339, 2022, doi: 10.1109/ACCESS.2022.3192472.
22. Zhou, Z., Xu, K., & Zhao, J. (2018). Homophily of music listening in online social networks of China. Social Networks, 55, 160–169. doi:10.1016/j.socnet.2018.07.001