Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 9s (2025)

Enhancing Social Media Influence with Cutting-Edge Machine Learning Approaches

Mrs.P.UmaMaheswari[1], Dr.A.Kumar Kombaiya[2]
Research Scholar[1], Associate Professor[2]
Department of Computer Science
Email:umapalanisamyphd@gmail.com[1]
Email:kumar_kombaiya@rediffmail.com[2]
Chikkanna Government Arts College, Tirupur, India[1,2]

Article History:

Received: 18-11-2024

Revised: 10-12-2024

Accepted: 26-01-2025

Abstract:

Homophily and community influence play the important roles in shaping user behavior towards others and towards content in social recommendation systems. This aspect has a considerable effect on how individuals engage with materials. Homophily is a core concept in social network analysis: it is the tendency of individuals to associate with others who have similar interests or characteristics. This research begins with datasets collected from Facebook, Instagram, and YouTube. Preprocessing occurred after the dataset was obtained from Kaggle sources. Following the completion of the initial dataset processing, the data is categorised using a fusion machine learning method. With accuracy rates of 97.11% for Facebook, 98.02% for Instagram, and 98.99% for YouTube, the suggested solution consistently outperforms existing approaches.

Keywords: Machine learning, classification, social media, disentangled modelling, fusion approach

I. INTRODUCTION

Mobile Internet-based media offers more format possibilities, faster speeds, and a lower cost than traditional media such as newspapers, radio, and television [1, 2]. Social media has created new avenues for public discourse and information sharing among the general population. The first step to building a society in which people feel safe enough to discuss crucial matters openly, share thoughts, and formulate their own conclusions is to bring humans out of their ivory towers and confront them with divergent perspectives [3]. Online social network marketing is increasingly reliant on the "word-of-mouth" method of conveying product information. The release of films starring well-known celebrities is a common pragmatic technique in campaigns. This ensures that the video reaches as many people as possible and increases the likelihood of becoming viral [4-6]. Online social networks have lately grown in popularity due to their greatly improved global connectivity. Recent study [7] focuses on Facebook, YouTube, and Instagram. This is due to the fact that it strongly relates to machine learning. This research explores Impact Maximization (IM) on Social Media (SM) sites such as Instagram, YouTube, and Facebook using machine learning algorithms such as Support Vector Machine (SVM), Linear Regression (LR), Gaussian Naive Bayesian (GNB), and Random Forest (RF). When optimizing the instant messaging, it is necessary to identify the most important nodes of the network first [8-10]. In attempts to offer theoretical solutions to the IM problem, research on the social influence of sets of individual seeds and the amount of seeds that need to be used to get a desired level of social influence is underway [11]. Another view of IM is to select a subset of seeds from a social network that most individuals find relevant.

- **Estimating the Stability of Incorporation of IM Algorithms:** [12] It was found that when exposed to noise from an attacker, the IM method is less precise in terms of probabilities of input effects. Another way to put it is that changing the diffusion model's impact probability significantly alters the ideal subset. Although various research (e.g., IRIE and EASYIM) have looked at this problem, they all make the same assumptions regarding the stability of the social network's topology. The graph's structure is dynamic, evolving at all levels. It seems hard to get a good seed set in a big social network.
- **No restriction to strict sub-modularity:** The IM problem is addressed with a sub-modular goal influence function. In some other instances, the target function submodularity has been too strong. Combining a greedy framework with a non-sub-modular goal effect function significantly decreases the theoretical approximation ratio. Models with a broader picture of functions assist to meet real-world expectations. For instance, [13] provides a weak submodular function.
- **Group norms:** The existing IM technique is primarily concerned with the interaction of two nodes. People are heavily impacted by community dynamics, particularly herd mentality. Others who share age, experience, and level of education contributes to shape. For example, the foundational study on the consistency-perception IM issue undertaken in [14] and [15] incorporated user characteristics linked with social groups.

Using Online Social Networks (OSN) and Support Vector Machines (SVM) ([16], [17]), the suggested technique automatically determines the most influential elements of spam profiles and resolves these issues using LR, GNB, and RF. To show that the technique is statistically significant, the link between a user's overall likelihood of clicking on advertising and the number of friends is looked [18]. These impact model parameters are derived from a simple LR model. Many theories have centered on the automated classification of messages as spam or authentic. Machine learning (ML) outperforms the other mentioned methodologies. Among the most well-known and respected text classification approaches, the age-old Naive Bayes classifier and support vector machines (SVMs) are notable [19-21]. To solve the dynamic link prediction issue, a GNB-based model is proposed that considers user connections and behavioral patterns based on features, popularity, user interests, location, and user attributes [22].In such circumstances, RF uses randomised decision trees to address issues such as excessive variation or bias among individual decision trees. It is an example of independent ensemble categorization. The primary idea of RF [24-25] is to create many decision or classification trees. Using an RF classifier on a limited dataset and repeatedly moving the beginning point until it discovers the best hypothesis has resulted in the generation of a new function. This is why the classifier [26] is utilized for impact maximization classification.

Organization: The rest of the work is presented as follows. Section II consists of some review articles. Section III describes Fusion ML's working processes. Running tests on raw data in Part IV tries to compare system predictions to current ones.

1.1 Motivation of this paper

The primary purpose of the project is to address categorization problems by combining several machine learning algorithms. SVM, LR, GNB, and RF each offer unique advantages, thus it is critical to investigate their potential. This is because different techniques perform well in different domains and with different types of data. Using a fusion ML approach relies mostly on increasing classification performance. This study seeks to increase the accuracy and durability of classification findings via the use of ensemble techniques such as voting, stacking, and weighted average. This strategy seeks to optimise the beneficial qualities of certain classifiers while limiting the effect of their negative ones.

II. BACKGROUND STUDY

Y. Li et al. investigated existing techniques, including heuristic and approximation approaches. The heuristic approach is only relevant to a certain diffusion model; although it is useful and cost-effective in some cases, philosophical confidence regarding the worst-case scenario is lacking. While time complexity prevents further decreases in approximation, meeting the requirement for algorithm speed becomes increasingly difficult as networks grow in size.

A. Matakos et al. demonstrated a new method for breaking filter bubbles. The SM interactions analyzed by these researchers included a variety of perspectives that fell within a certain range of concepts. Consider a network in which a number of items replicate the dynamics of social media platforms by messaging, re-sharing, re-tweeting, and so on, effectively transmitting every point of view. Everyone has an opinion on the subject, which influences whether or not they share any specific work.

H. Hu et al. developed a methodology for the social propagation of viral films that maximises their impact across different domains. This approach includes picking seeds and delivering relevant films. The scientists first developed a multi-topic-aware effect enhancing framework to mimic the challenge of video dissemination, and then used it to characterize various video snippets across several themes. Once the submodularity of the impact spread function was identified, the authors then constructed the greedy approach to address this problem. The researchers established an upper bound estimate-based approach and a performance-constrained approximation algorithm with maintained accuracy and improved computation in huge networks.

W. Yang et al. used a social network with a known neighbor matrix to model the spreading of information as an epidemic. To enhance a net reward function, seed set selection and resource allocation methods for a campaign from optimal control among sets of nodes were developed by the authors. The net reward function is linear and is specified as a sum of reward due to message passing over the network and cost of control execution. Aggregating nodes with appropriate centrality metric values results in groupings. The following measures of centrality are used: degree, closeness, page rank, and between. The authors evaluated centrality measures in data distribution using a combined ideally optimum control system. In terms of competitiveness, J. Tong et al. concentrated mostly on increasing information impact. Hyperlink-Induced Topic Search (HITS) uses an algorithm to create large amounts of searchable social network data. The strategy maximizes the influence of the interaction material between nodes to its full capacity while taking into account user preferences. Tested using real Twitter data, their suggested solution routinely outperforms rival methods.

H. Khavandi suggested a method for identifying the optimal set of nodes for a network by combining PSO and GA. Genetic algorithms have found applications in a wide range of industries due to their speed and efficiency. Following a comparison of the suggested strategy to others including genetic, greedy general, discount degree, maximum degree, and random selection the authors looked at experimental data to determine the level of performance. Among many exploration strategies, the suggested method identifies the optimal set of nodes with the greatest impact.

T. Cai et al. presented a new field of research based on Holistic Influence Maximization (HIM) that goes beyond the traditional IM problem and adds value towards the numerical practical applications, has no limitations towards event planning, advertising placement, crisis resolution, and many more. The HIM research standardized the U2U impact model by including social, geographic, and preference-based similarities. IM difficulties are thus simpler to handle than HIM searches.

H. Li et al. introduced Conformity Aware Social Influence Computation (CASINO), a new social influence computation method. A network calculates positive and negative interactions between individuals to decide the influence and conformity of a node. After doing extensive study across several social media networks, concluded

that CASINO is more accurate and efficient than more modern approaches. The study tells among other things, how important it is to understand people's levels of compliance in order to conduct reliable social impact studies.

A. Z. Ala'M et al. proposed a system for detecting spammer based on machine learning for online social networks (WAO) that blends existing metaheuristics and support vector machines (SVM). When applied to the detection process, the recommended technology automatically identifies spammers and emphasizes the most important information. Besides the four language data sets harvested from Twitter, i.e., Arabic, English, Spanish, and Korean, used to gauge the performance of the model, experiments and results show that the model introduced in this paper performs better than a range of other algorithms in accuracy.

2.1 Problem definition

The research investigates how many ML algorithms have been combined using a fusion strategy to address classification problems. Given the multitude of categorization methods available, each altered to a certain kind of data or topic, a comprehensive evaluation of different approaches is essential. The major challenge is to optimize classification performance using a fusion ML approach. Ensemble techniques such as voting, stacking, and weighted average are often used to combine many classifier predictions. The fundamental challenge in producing a more powerful and accurate classification result is to maximize the strengths of each technique while minimizing their weaknesses.

III. MATERIALS AND METHODS

Using multi-source data and machine learning, Figure 1 presents a categorization method for targeted research. Important in this system are social media sites such Facebook, Instagram, and YouTube. Before identification and assessment, the data is painstakingly processed. Machine learning methods include linear classifiers, nonlinear classifiers, and statistical models are used when data point analysis and grouping calls for. Finally, a fusion process generates a robust and complete classification model by aggregating the results of many classification techniques. This technique have uses in personalised content distribution, user behaviour prediction, and social media analytics. Enhancing the social media impact utilising experimental techniques like SVM, LR, RF, and GNB depends critically on data collecting, feature selection, parameter tuning, and model testing. Included in it are all the following: real hyper-parameters, approaches of instruction, and assessment criteria. Sharing scripts, data, and execution instructions assist to ease replication. Being open not only helps other researchers confirm the findings but also makes it simple to update and enhance the research on social media impact maximisation.

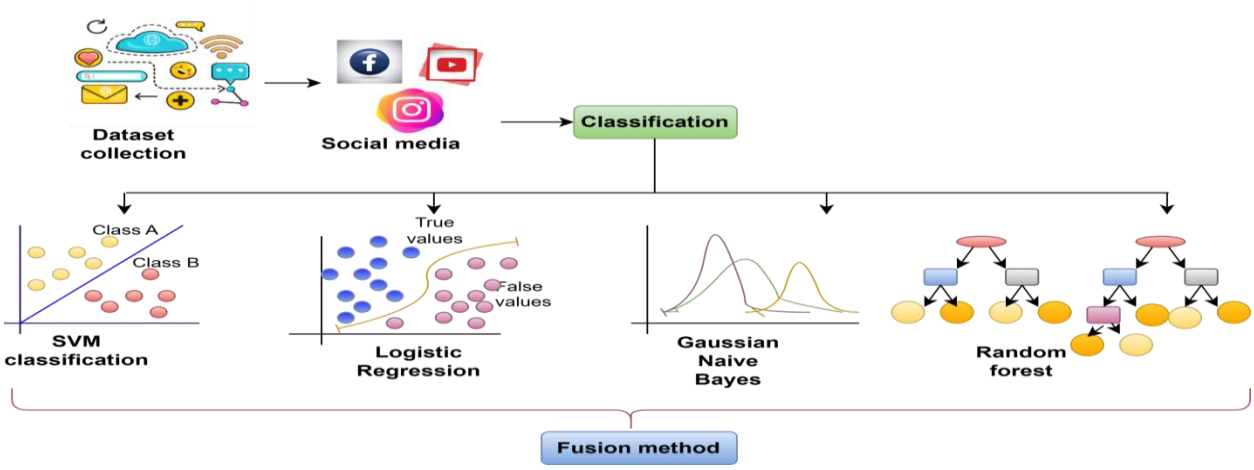

Fig. 1: Workflow of the model

3.1 Dataset Acquisition

In this study, statistical information are accessed from Facebook, Instagram, and YouTube. If wished to analyse Facebook trends in content, user interactions, and behaviour, utilised the Facebook dataset that is accessible on Kaggle. It provides a complete overview of several areas of Facebook data.

Facebook:

https://www.kaggle.com/datasets/sheenabatra/facebook-data

Instagram:

https://www.kaggle.com/datasets/krpurba/im-instagram-70k

YouTube:

https://www.kaggle.com/datasets/kathir1k/youtube-influencers-data

3.2 Dataset Preparation

Raw data conversion and cleaning is required while getting a dataset ready for machine learning. This process includes importing the dataset into the development environment. Patterns and abnormalities are uncovered via exploratory data analysis. Every error and inconsistency is corrected. Numerical data points are either scaled or normalised. Class discrepancies are addressed. The dataset is divided into two halves: one for testing and the other for training. Building a fair, adequate, and organised dataset for ML model training and assessment has assisted to enhance classification performance in the long run. By meticulously documenting these methods, one ensures that the research technique is open and repeatable.

3.3 Utilization of SVM, LR, GNB, RF for Classification

3.3.1 SVM

This study classifies and predicts the dataset with SVMs. Using SVMs to investigate data for evident or hidden patterns, supervised learning addresses regression and classification issues. Self-Vector Machines (SVMs) has solved both linear and nonlinear classification problems. Following linear partition of the training datasets into a new dimensional space, support vector machines (SVMs) recognise every occurrence of previously stated categories (Ala'M, A. Z, et al.).

$$\vec{c}^T\vec{a} + b = 0,\ \ p_i(\vec{c}^T\vec{a}_i - b) \geq 1 \text{ --------- (1)}$$

Equation (1) represents a hyperplane in the feature space used by SVM to distinguish between classes. $\vec{c}^T$ indicates the direction/orientation of the hyperplane. $\vec{a}$ is a point in feature space. b is the bias, and it adjusts the position of the hyperplane.

This is how one applies a kernel function K to classify nonlinear data shown in equation (2):

$$maxL_D = \sum_i a1_i - \frac{1}{2}\sum_i\sum_i a1_i a1_j p_i p_j K\left(\overrightarrow{a_i}.\overrightarrow{a_j}\right) \text{ --------- (2)}$$

To maximise the margin for each training data point, use the Lagrange multipliers $a1_i$ and $a1_j$; for i-th data point, use the class labels p_i; for j-th data point, use y_j. $K\left(\overrightarrow{a_i}.\overrightarrow{a_j}\right)$. In the converted feature space, the kernel function $\overrightarrow{a_j}$ determines the similarity between a_i and a_j.

Support vector machines, or SVMs, are an effective method for categorising data across several dimensions. Unlike nearest-neighbor classifiers, support vector machines (SVMs) maximise the classification margin to find the best hyperplane for grouping training data. To avoid costly similarity calculations in high-dimensional feature space, support vector machines utilise a replacement kernel function. The target function for training a nonlinear support vector machine model is shown below. Equation (3) supports two primary aims that improve SVM.

- Optimising the difference between categories; and minimising the severity of penalties for points that stray from classification boundary norms.

$$\min_{c,b,\xi} \frac{1}{2}\|c\|_2^2 + \beta \sum_{i=1}^{n} \xi_i \text{ ---------- (3)}$$

$$s\ .t\ \ p_i[c^T \varphi(a_i) - b] \geq 1 - \boldsymbol{\xi_i} \text{ ---------- (4)}$$

$$\xi_i \geq 0, i = 1,2, \dots, n, \text{ ---------- (5)}$$

The activation of a function $\varphi(a_i)$) in the equations (3), (4), and (5) projects input a_i to high-dimensional feature space. H is the cost of classification margin compared to misclassification error offered by An activation of a function $\varphi(a_i)$in (3), (4), and (5) allocates input a_i the high-dimensional feature space. H is the cost of classification margin compared to misclassification error offered by ξ_i. The Lagrange multiplier method enables to convert the following optimisation function into its dual form for support vector machines:

$$\max_{a1} \sum_{i=1}^{n} a1_i - \frac{1}{2}\sum_{i,j=1}^{n} a1_i a1_j p_i p_j K(a_i, a_j) \text{ ---------- (6)}$$

$$s.\quad t.\quad \sum_{i=1}^{n} a1_i p_i = 0 \text{ ---------- (7)}$$

$$0 \leq a1_i \leq \beta, i = 1,2, \dots, n, \text{ ---------- (8)}$$

Equations (6), (7), and (8) use the kernel function $K(a_i, a_j)$ to compute the dot product $a1_i a1_j$ using $a1_i$ as the Lagrange multiplier. After training the SVM model, use equation (9) to project the class label of a test sample ($\overline{x}$):

$$y = \text{ sign} \frac{\left(\sum_{i=1}^{n} a1_i\, p_i \varphi(x_i).\, \varphi(\bar{a})\right)}{= \text{sign}(\sum_{i=1}^{n} a1_i p_i K(\bar{a}, a_i)).} \text{ ---------- (9)}$$

To train, support vectors, a collection of examples is used. It is difficult to know what kernel is going to be better suited for any given set of data. On the one hand, a complex decision surface is beyond the capabilities of a simple kernel. Still, an extremely flexible kernel results in model overfitting.

3.3.2 Logistic Regression

The dataset is categorised using logistic regression (LR) in this paper. In contrast to its ability to predict continuous outcomes, LR excels in binary classification situations in which the goal variable takes one of two potential values. T. Lawrence et al. believe that by curve-fitting data to a logistic curve, LR is able to predict the probability that an event occur. Predictor variables, which have been numerical or categorised, are used in a variety of regression analyses. LR is often utilised in the social media industry for objectives like as predicting user activity (for example, product purchase or membership cancellation).

$$logit\{Pr(P = 1|a)\} = log\left\{\frac{\Pr(P=1|a)}{1-\Pr(P=1|a)}\right\} = \beta_0 + \beta_1 a_1 + \beta_2 a_2 + \cdots + \beta_k a_k \text{ ---------- (10)}$$

From the statement, intercept appears as β_0, β_1, β_2, β_k, etc. represent the "regression coefficients" of a_1, a_2, and a_k. The degree of contribution by the risk factor is represented in terms of regression coefficients.

Equation (11) shows in this situation the logistic function:

$$P1 = \frac{1}{1+e^{-logit(p1)}} \text{ ---------- (11)}$$

Because its output only accept the values between 0 and 1, the logistic function works well with inputs ranging from negative infinity to positive infinity.The family includes several distinct types of linear models. With the assumption that they follow a normal distribution, all of the models explored have relied on quantitative response variables. This lesson investigates situations in which the answer variable (a categorical random variable) has just two possible values. This kind of data is not seen elsewhere.

3.3.3 Gaussian Naïve Bayes

The dataset was organised and examined using the GNB classification algorithm. In a probabilistic model based on Bayes' theorem and focused on GNB, characteristics are assumed to have no effect on the class label. This research primarily aims to clarify classification concerns, demonstrating that GNB works well when the attributes are constantly distributed and homogenous. D. F. M. Mohideen et al. argue that GNB is an effective object classification approach. The aim is to place each sample into the class with the greatest posterior probability once it has been determined that voxel contributions are conditionally independent and normally distributed. GNB decision rule for each searchlight s is in a discriminating function for class k; the searchlight index is excluded for simplicity.

$$\hat{d} = \arg max_{k=\{a1,b\}}\{\delta_i^k\} \text{ ---------- (12)}$$

$$\delta_i^k = -\sum_{j=1}^{v}\left(\frac{(A_{ij}^{Te}-\hat{\mu}_j^k)^2}{2\hat{\sigma}_j^2}\right) + \log(rs) \text{ ---------- (13)}$$

Equation (12) is directly applied in the situation with more than one class in binary classification with classes {a,b}.

Equation (13) makes use of the training set to calculate the mean and standard deviation of every voxel j, $\hat{\mu}_j^k$. For GNB used in a network of multiple searchlights, the sparse binary matrix s is used to sum voxel contributions in parallel in each searchlight. This matrix product enables to rewrite the above equation:

$$E_{m,s}^k = -F_{m,v}^k S_{v,s} + \log(rs) \text{ ---------- (14)}$$

Voxel-wise additions to equation (14) are the squared z-score distances of test set sample i, or matrix $F_{m,v}^k S_{v,s}$. Every searchlight that a voxel is a member of contributes the same. The discriminant functions for class k are the matrix $E_{m,s}^k$ for every sample gathered by every searchlight.

3.3.4 Random forest

The classification approach used in this research largely depends on Random Forest (RF) to analyse and arrange data. While training, RF creates a large number of decision trees. This strategy is referred to as ensemble learning. Decision tree classifiers outperform single-stage models. Y. Asim et al. claim that decision tree classifiers make decisions at several levels. Decision tree classifiers and multi-level classifiers are two of the numerous names given to them. Finding appropriate features to break classes at each non-terminal node is one of the most difficult difficulties in developing a decision tree classifier. Aiming for optimal classification accuracy with minimum computations is an ideal design aim for decision tree classifiers. Decision tree classifiers contain binary tree classifiers as a subset. The criteria for separation is altered depending on the application. Terminal nodes are nodes

that give just one class. Gini, entropy, and twoing are three often used techniques to tree construction. Entropy is a key metric of data volume in the first method. Equation 15 depicts the predicted data needed for categorisation using an observation vector D.

$$Info(D) = -\sum_{i=1}^{n} p_i \log(p_i) \text{ -------- (15)}$$

One of the most common splitting criteria is one that splits classes evenly while optimising information gathering across parent and child nodes. To fulfil certain objectives, data points are placed in the most optimum sequence possible. This is when the understanding of Gini comes in handy. The Gini impurity, based on priors and class distribution, calculates the likelihood of incorrectly categorising a randomly selected class. The gini index is a strategy for enriching data, defined as follows:

$$G(D) = 1 - \sum_{i=1}^{n} p_1 \text{ ---------- (16)}$$

Under some conditions, optimal splitting has been determined in a different method using the twoing division approach. It is able to generate more sensible splits by placing the related classes at the top of the tree. At the most basic level, the tree represents categorisation. Towing requires breaking classes into two superclasses with the same amount of cases. Many superclass specialists believe that the split at the current node is ideal. For each child node, this implies fewer class selections and impurity possibilities.

$$Gain(A) = Info(E) - Info_{A1}(E) \text{ ---------- (17)}$$

$$Info_{A1}(E) = \sum_{i=1}^{n} \frac{E_i}{E} \times Info(E) \text{ ---------- (18)}$$

Depending on the splitting strategy utilized obtain $Info(E)$ from any of the above equations. Equation (18) computes $Info_{A1}(E)$ where $n$is the weight of the i-th split and discrete values of attribute A.

3.4 Fusion Method

To build a more robust and accurate model, ML uses the fusion approach, which combines the outputs of many classifiers. Voting, stacking, and weighted averaging are examples of many fusion methods. The voting method produces the final forecast by adding up the forecasts from each classifier and selecting the class with the most votes. By feeding the output of every classifier into a meta-model, stacking allows to learn an ideal model that best combines the output of thousands of classifiers. Weighted averaging gives more weight to more accurate models by assigning different weights to the output of every classifier.
There are various benefits of utilizing Fuse ML.
• Better and more favorable results. By mixing the strengths of multiple methodologies, one decreases the likelihood of overfitting as well as the capability of the model to generalize to new information.
• Higher resilience to stress: There are hazards to use a single model; however, fusion approaches help to lessen these risks by mixing algorithms that thrive in diverse situations.
Understanding the roles performed by various models help one have a better understanding of the factors that contribute to social media impact. In doing so, interpretability is improved.
To optimise the social media effect, fusion ML techniques such as SVM, LR, GNB, and RF use several algorithms to provide more accurate predictions. This strategy allows better understanding and creatively targeting essential individuals and material, which is particularly valuable given the complexity and variety of Facebook, Instagram, and YouTube data.

Algorithm 1: Fusion method

Input data:

Using SVM, LR, GNB, and RF for predictions.

Procedure:

1. Receiving the Predictions of Classifier

For the provided dataset, gather the predictions from the algorithms such asSVM, LR, GNB, and RF.

2. Selection of Fusion Approach

Depending on the dataset characteristics and classifier behavior, choose an appropriate fusion technique.

Use optimization criterion:

$$\max_{a1} \sum_{i=1}^{n} a1_i - \frac{1}{2}\sum_{i,j=1}^{n} a1_i a1_j p_i p_j K(a_i, a_j)$$

Step 3: Apply Fusion Method

If majority voting is selected assign the class label that appears most frequently among classifier outputs.

If meta-learning strategy is chosen train a meta-model on SVM, LR, GNB, and RF's predictions and use the trained meta-model to predict the final class label.

If weighted averaging is selected, assign weights to each classifier based on performance and compute the weighted sum of predictions:

$$\delta_i^k = -\sum_{j=1}^{v}\left(\frac{(A_{ij}^{Te} - \hat{\mu}_j^k)^2}{2\hat{\sigma}_j^2}\right) + \log(rs)$$

Use decision function:

$$E_{m,s}^k = -F_{m,v}^k S_{v,s} + \log(rs)$$

Step 4: Generate Final Prediction

Compute the final class label based on the applied fusion method.

Use the probability function: $G(D) = 1 - \sum_{i=1}^{n} p_1$

Step 5: Asses the Performance of Model

The performance parameters like Accuracy, Precision, Recall, and F1-score are measured.

Compare fusion model performance with individual classifier performance.

Final output:

Aggregated prediction according to the selected fusion approach.

For the voting case, the label that has the most votes.

IV. RESULT AND DISCUSSION

This section describes in detail the categorisation experiments, their results, and their implication. The explanation of the fusion technique influencing classification accuracy and well each classifier performed in this regard is given.

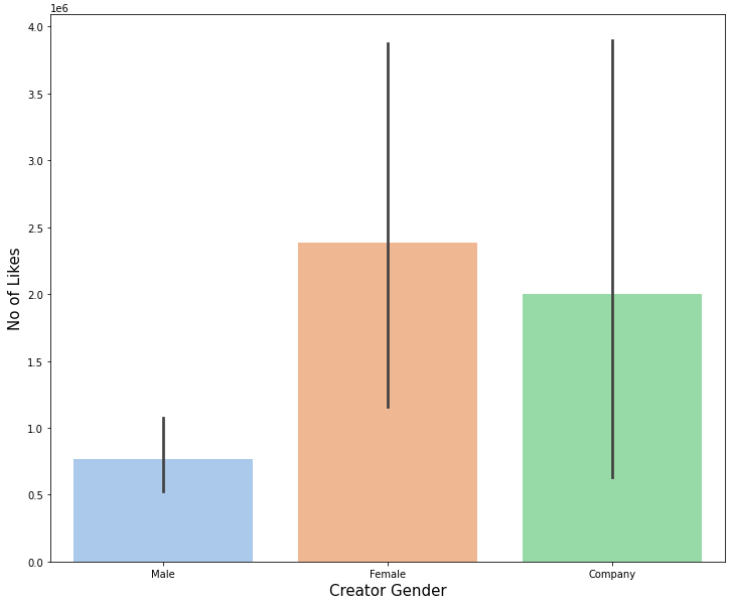

Fig. 2: Comparison graph of creator gender

Fig. 2 indicates the following. Their gender is plotted on the x-axis, and the proportion of likes for each source of materials. This graphic depiction provides a comprehensive view of the problem by emphasising the relationship between author gender and content popularity as measured by likes. The x-axis categorises artists based on their gender. The numerical numbers on the y-axis represent the total number of likes for each category of writers.

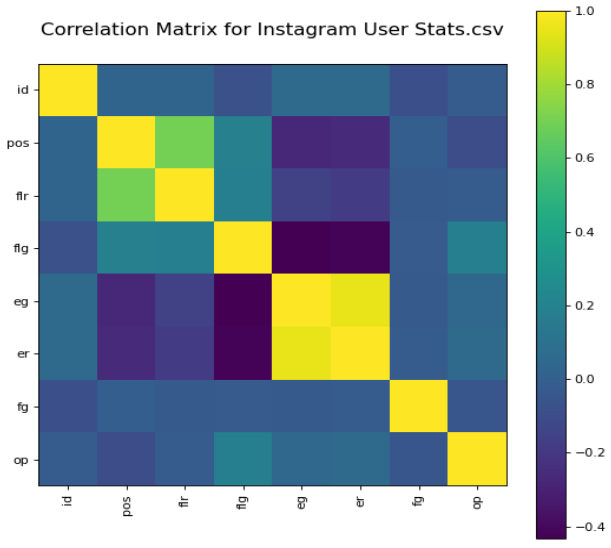

Fig. 3: Correlation matrix for instagram user

Fig. 3 depicts an Instagram user's correlation matrix, which provides a comprehensive picture of the relationships and correlations between several aspects in the data. A paired correlation matrix assists illustrate the links between Instagram users' qualities and actions.

4.1 Performance evaluation

1. The ratio of correctly identified samples to the total number of samples is a measure of accuracy. As far as arithmetic

Accuracy = (True Positive + True Negative) / (True Positive + False Positive + True Negative + False Negative)

2. Precision (P) percentage is then also given as a percentage of defects detected out of the entire pest samples.

Algebraically,Precision = True Positive / (True Positive + False Positive)

3. Recall is the number of correctly detected pest samples over actual total samples. NumericallyRecall = True Positive / (True Positive + False Negative)

4. Second, the F1 score is a trade-off between memory accuracy. Regarding the numbers:

F1 score = 2 * P * R / (P + R)

Table 1: Classification performance metrics comparison table

	Algorithm		**Accuracy**	**Precision**	**Recall**	**F-measure**
Existing methods	DT	Facebook	94.31	94.20	95.23	94.20
		Instagram	93.30	94.86	95.00	94.12
		YouTube	94.82	94.10	94.11	94.01
	SVM	Facebook	95.23	95.38	92.53	96.31
		Instagram	95.23	95.33	93.34	94.61
		YouTube	97.55	95.37	97.36	96.26
	NB	Facebook	97.83	96.35	97.30	96.02
		Instagram	95.86	96.63	97.13	95.30
		YouTube	96.00	95.34	96.10	96.23
Proposed methods	Proposed methodology (Fusion ML methods)	Facebook	96.12	96.67	97.33	97.21
		Instagram	97.01	97.23	96.98	97.20
		YouTube	97.98	98.34	97.26	98.37

Table 1 provides the current and suggested approaches for categorizing YouTube, Instagram, and Facebook data using four measure parameters: F-measure, recall, accuracy, and precision. On YouTube, advanced techniques (DT, SVM, and NB) are always behind. Naive Bayes, the suggested approach is better than all others in all the parameters. For instance, when YouTube data is being classified, the proposed method excels over other methods in accuracy, precision, recall, and F-measure. The presented approach over this table is social media data collection and pattern extraction.

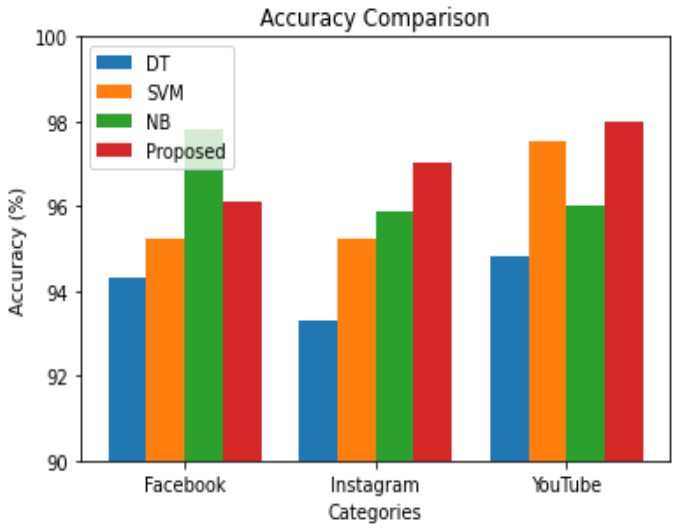

Fig. 4: Comparison chart of Accuracy in social media IM

When the accuracy of Facebook, Instagram, and YouTube instant messaging is considered, fusion ML methods were better compared to DT, SVM, and NB (Fig. 4). YouTube, Instagram, and Facebook are depicted on the x-axis of the graph. Existing methods' accuracy percentages and fusion machine learning systems' accuracy percentages are shown on the y-axis.

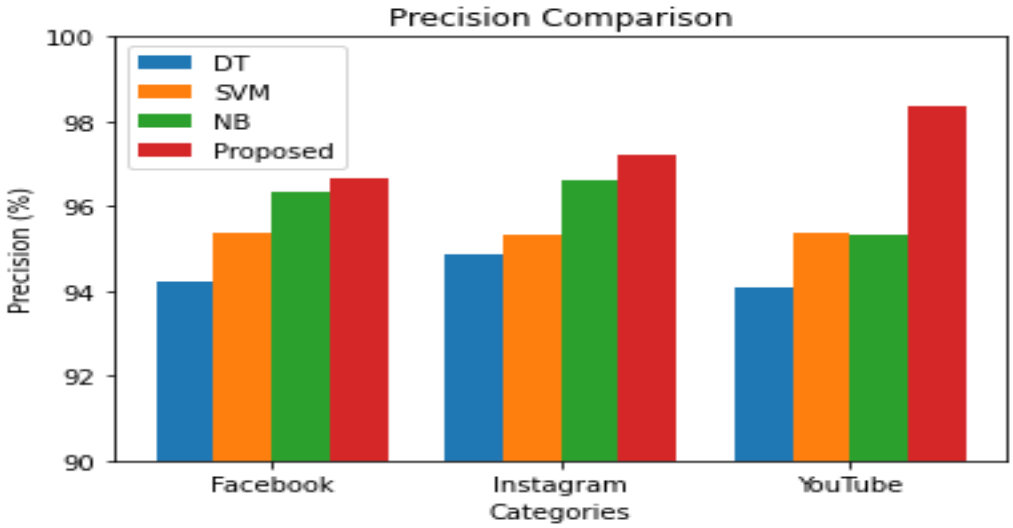

Fig. 5: Comparison chart of precision in social media IM

Fig. 5 depicts the comparisons of accuracy among Facebook, Instagram, and YouTube IM. Fusion ML techniques were superior to other methods like DT, SVM, and NB by a significant margin. The x-axis of the graph consists of Facebook, Instagram, and YouTube. The y-axis shows the current and fusion ML systems' accuracy rate percentages.

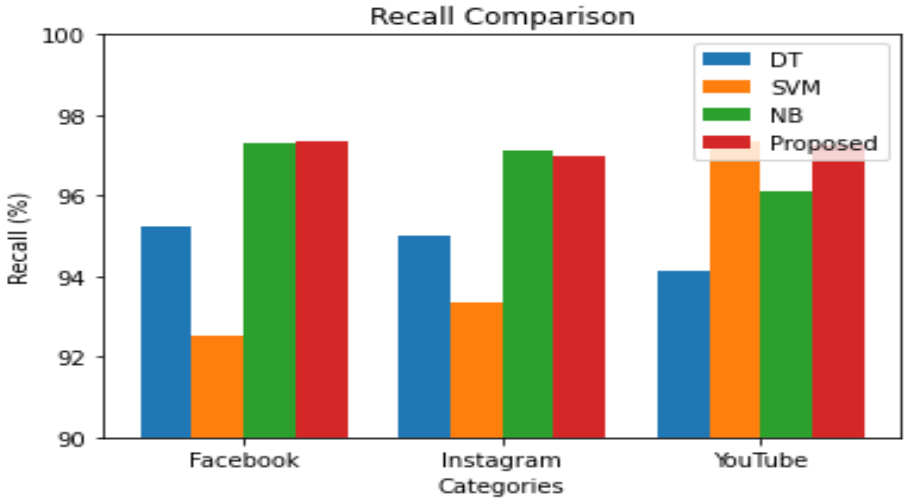

Fig. 6: Comparison chart of Recall in social media

Fig. 6 shows the Facebook, Instagram, and YouTube IM recall in comparison with one another. DT, SVM, and NB performance is much lower compared to the fusion ML algorithms. Facebook, Instagram, and YouTube are labeled on the x-axis of this figure. The y-axis represents the existing and fusion ML algorithms' recall percentages.

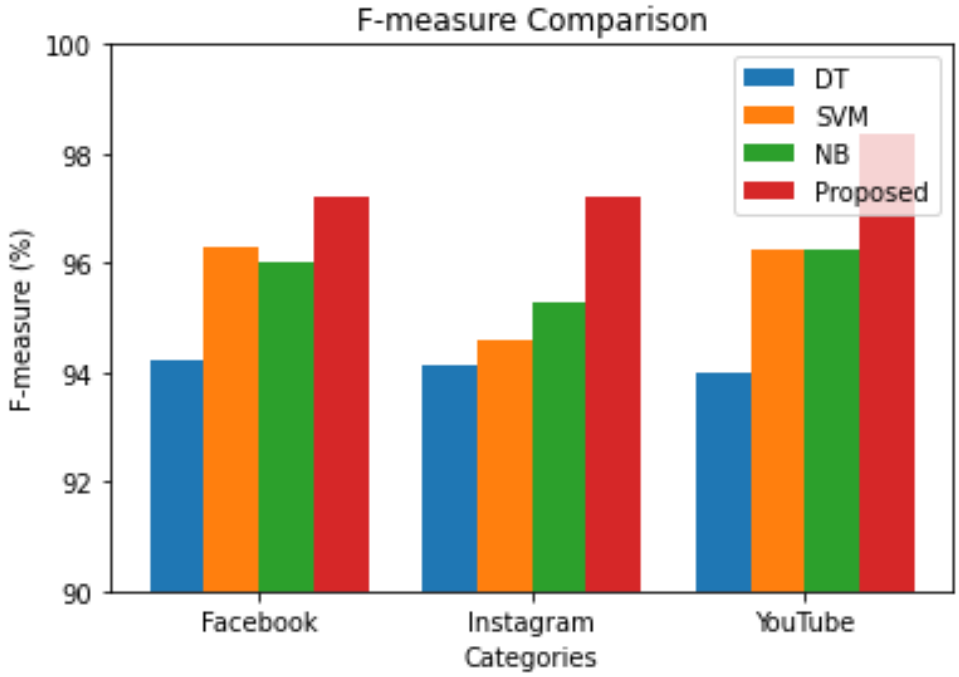

Fig. 7: Comparison chart of F-measure in social media IM

Fig. 7 depicts Facebook, Instagram, and YouTube IM versus the f-measure. The accuracy of DT, SVM, and NB is relatively very low compared to fusion ML algorithms. Facebook, Instagram, and YouTube are graphed on the x-axis of this graph. The y-axis indicates the ratio of present and fusion ML approaches which graph the f-measure.

V. CONCLUSION

This paper concludes that fusion approaches and the combination of several ML algorithms are critical for improved classification performance. Here, learned about four popular algorithms of classification namely SVM, LR, GNB, RFin order to compare methods to use for some data types or applications. Fusion methods based on voting, stacking, or weighted averaging have shown some efficacy in enhancing classification accuracy and thereby mitigating the effects of individual classifier inadequacies. These systems combine many classifier characteristics into one. The outcomes of this study serve to clarify the usefulness of fusion approaches and different classifier combinations, allowing for the selection of the best classification strategy for certain datasets. With rates of 97.11% for Facebook, 98.02% for Instagram, and 98.99% for YouTube, the proposed technique consistently outperforms previous strategies. Examining fusion methods aid to improve machine learning classification challenges by providing insight into strategic ensemble approach use. These ideas help to make better judgements in a variety of situations.

REFERNCE

1. Y. Li, H. Gao, Y. Gao, J. Guo, and W. Wu, "A survey on influence maximization: From an ml-based combinatorial optimization," *ACM Transactions on Knowledge Discovery from Data*, vol. 17, no. 9, pp. 1–50, 2023.
2. Q. Nong, Z. Guo, and S. Gong, "Regularized Submodular Maximization With a k-Matroid Intersection Constraint," *IEEE Access*, vol. 11, pp. 103830–103838, 2023, doi: 10.1109/ACCESS.2023.3317691.
3. A. Matakos, C. Aslay, E. Galbrun, and A. Gionwas, "Maximizing the Diversity of Exposure in a Social Network," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 9, pp. 4357–4370, Sept. 2022, doi: 10.1109/TKDE.2020.3038711.
4. H. Hu, Y. Wen, and S. Feng, "Budget-efficient viral video distribution over online social networks: Mining topic-aware influential users," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 3, pp. 759–771, 2016.
5. K. Kandhway and J. Kuri, "Using Node Centrality and Optimal Control to Maximize Information Diffusion in Social Networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 7, pp. 1099–1110, July 2017, doi: 10.1109/TSMC.2016.2531690.
6. W. Yang, J. Yuan, W. Wu, J. Ma, and D. Z. Du, "Maximizing activity profit in social networks," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 1, pp. 117–126, 2019.
7. J. Tong, L. Shi, L. Liu, J. Panneerselvam, and Z. Han, "A novel influence maximization algorithm for a competitive environment based on social media data analytics," *Big Data Mining and Analytics*, vol. 5, no. 2, pp. 130–139, 2022.
8. H. Khavandi, B. N. Moghadam, J. Abdollahi, and A. Branch, "Maximizing the Impact on Social Networks using the Combination of PSO and GA Algorithms," *Future Generation in Distributed Systems*, vol. 5, pp. 1–13, 2023.
9. T. Cai, J. Li, A. Mian, R. H. Li, T. Sellis, and J. X. Yu, "Target-aware holistic influence maximization in spatial social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 4, pp. 1993–2007, 2020.

10. J. Guo and W. Wu, "A Novel Scene of Viral Marketing for Complementary Products," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 4, pp. 797–808, Aug. 2019, doi: 10.1109/TCSS.2019.2926112.
11. Y. Ye, Y. Chen, and W. Han, "Influence maximization in social networks: Theories, methods and challenges," *Array*, vol. 16, p. 100264, 2022.
12. X. He and D. Kempe, "Stability of influence maximization," in *Proc. 20th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 2014, pp. 1256–1265.
13. A. Das and D. Kempe, "Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection," *arXiv preprint arXiv:1102.3975*, 2011.
14. H. Li, S. S. Bhowmick, and A. Sun, "Casino: Towards conformity-aware social influence analysis in online social networks," in *Proc. 20th ACM Int. Conf. on Information and Knowledge Management*, 2011, pp. 1007–1012.
15. UmaMaheshwariP, Dr. A.Kumar Kombaiya A Comprehensive Analysis On Social Media Influence Maximization And Social Recommendation:A Survey, International Journal of Creative Research Thoughts (IJCRT) Volume 11, Issue 10 October 2023 | ISSN: 2320-2882
16. J. Tang, S. Wu, and J. Sun, "Confluence: Conformity influence in large social networks," in *Proc. 19th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 2013, pp. 347–355.
17. A. Z. Ala'M, H. Faris, J. F. Alqatawna, and M. A. Hassonah, "Evolving support vector machines using whale optimization algorithm for spam profiles detection on online social networks in different lingual contexts," *Knowledge-Based Systems*, vol. 153, pp. 91–104, 2018.
18. P. Pirozmand, M. Sadeghilalimi, A. A. R. Hosseinabadi, F. Sadeghilalimi, S. Mirkamali, and A. Slowik, "A feature selection approach for spam detection in social networks using gravitational force-based heuristic algorithm," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2023.
19. T. Lawrence, P. Hosein, and I. Dialsingh, "An influence model for influence maximization–revenue optimization," *International Journal of Data Science and Analytics*, vol. 11, no. 2, pp. 155–168, 2021.
20. T. A. Almeida, J. Almeida, and A. Yamakami, "Spam filtering: how the dimensionality reduction affects the accuracy of Naive Bayes classifiers," *Journal of Internet Services and Applications*, vol. 1, pp. 183–200, 2011.
21. UmaMaheshwari,P. Kumar Kombaiya,A. "Navigating Influencer Marketing: A Comprehensive Overview" in Indian Journal of Natural Sciences, Vol.15 / Issue 87 / Dec / 2024.
22. A. K. Singh and K. Lakshmanan, "PILHNB: Popularity, interests, location used hidden Naive Bayesian-based model for link prediction in dynamic social networks," *Neurocomputing*, vol. 461, pp. 562–576, 2021.
23. D. F. M. Mohideen, J. S. S. Raj, and R. S. P. Raj, "Regression imputation and optimized Gaussian Naïve Bayes algorithm for an enhanced diabetes mellitus prediction model," *Brazilian Archives of Biology and Technology*, vol. 64, p. e21210181, 2021.
24. N. A. Samsudin, A. Mustapha, and M. H. Abd Wahab, "Ensemble classification of cyber space users tendency in blog writing using random forest," in *Proc. 12th Int. Conf. on Innovations in Information Technology (IIT)*, 2016, pp. 1–4.
25. UmaMaheshwariP, Dr. A.Kumar Kombaiya INFLUENCE MAXIMIZATION IN SOCIAL MEDIA WITH BIASED RENOVATE K-MEAN CLUSTERING AND BIASED BAT ALGORITHM P.UmaMaheswari /Afr.J.Bio.Sc. 6(14) (2024)
26. Y. Asim, B. Raza, A. K. Malik, A. R. Shahaid, and H. Alquhayz, "An adaptive model for identification of influential bloggers based on case-based reasoning using random forest," *IEEE Access*, vol. 7, pp. 87732–87749, 2019.