International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(17s) 2026 pp. 1219-1236
https://doi.org/10.70917/ijcisim-2026-4803
Article

A Hybrid Machine Learning Framework for Measuring Social Media Marketing Effectiveness in Higher Education Institutions

Rishi Thapa[1], Navita Nathani[2]

[1] Amity University Madhya Pradesh, Maharajpura, Opposite Gwalior Airport, Gwalior, Madhya Pradesh – 474005, India.
ORCID: 0000-0003-4309-9437
[2] Amity University Madhya Pradesh, Maharajpura, Opposite Gwalior Airport, Gwalior, Madhya Pradesh – 474005, India.
ORCID: 0000-0001-6427-1115

Abstract: The swift expansion of social media has revolutionized the marketing approaches of Higher Education Institutions (HEIs), allowing universities to enhance their institutional branding, connect with potential students, and facilitate communication with various stakeholders. Platforms such as Facebook, Instagram, LinkedIn, X (formerly Twitter), and YouTube have emerged as vital channels for student recruitment and digital outreach. Nevertheless, assessing the effectiveness of social media marketing poses a considerable challenge, as traditional evaluation methods predominantly depend on descriptive metrics like likes, shares, comments, impressions, and follower counts. These metrics frequently do not adequately reflect the intricate relationships between user engagement, institutional reputation, and marketing results, underscoring the necessity for more sophisticated and data-driven evaluation techniques. This research introduces a Hybrid Machine Learning (HML) Framework designed to assess the effectiveness of social media marketing within Higher Education Institutions. The framework incorporates advanced data preprocessing, feature engineering, feature selection, ensemble learning, and Explainable Artificial Intelligence (XAI) to enhance predictive accuracy and model interpretability. Essential engagement features, such as reach, impressions, engagement rate, click-through rate, comments, shares, follower growth, video views, and sentiment scores, are utilized to assess campaign performance. The proposed framework leverages the complementary strengths of Random Forest, XGBoost, LightGBM, and Support Vector Machine (SVM) through an ensemble learning strategy, facilitating robust predictions across diverse social media datasets. Moreover, the framework integrates SHAP (Shapley Additive Explanations) to pinpoint key engagement factors and deliver clear explanations for model predictions. This improvement facilitates decision-making by allowing university administrators and digital marketing experts to refine content strategies, enhance student recruitment initiatives, and allocate marketing resources more efficiently. The proposed framework advances educational marketing analytics by offering a scalable, interpretable, and data-driven solution that addresses the shortcomings of single-model methodologies. It provides actionable decision support for assessing and optimizing social media marketing effectiveness, thereby aiding the digital transformation and sustainable development of Higher Education Institutions.
Keywords: Higher Education Institutions, Social Media Marketing, Hybrid Machine Learning, Digital Marketing Analytics, Ensemble Learning, Explainable Artificial Intelligence, Student Engagement, Predictive Analytics, SHAP, Educational Data Mining

1. Introduction

1.1 Background

The swift digital transformation within higher education has profoundly altered the manner in which universities interact with prospective students, alumni, parents, and other stakeholders. Social media platforms such as Facebook, Instagram, LinkedIn, X (formerly Twitter), and YouTube have emerged as vital marketing channels for institutional branding, student recruitment, and stakeholder engagement (Abbas, 2026; Cingillioglu et al., 2024; Pawar, 2024). Universities are increasingly dependent on digital marketing campaigns to enhance visibility and competitiveness; however, assessing campaign effectiveness through traditional metrics like likes, comments, shares, and impressions proves inadequate (Bhattacharjee et al., 2026; Dwivedi et al., 2023). Artificial Intelligence (AI), Machine Learning (ML), and Educational Data Mining (EDM) offer sophisticated methods for analyzing extensive social media data, forecasting marketing performance, and facilitating data-driven decision-making (Romero & Ventura, 2024; Lund et al., 2023). Nevertheless, the majority of current studies utilize individual machine learning models and offer limited interpretability, underscoring the necessity for hybrid and explainable methodologies (Lundberg & Lee, 2017; Chen & Guestrin, 2016).

1.2 Problem Statement

Higher Education Institutions are increasingly reliant on social media marketing to enhance their visibility and attract potential students. However, the existing evaluation methods predominantly utilize descriptive engagement metrics, which offer limited insights into the effectiveness of campaigns. Individual machine learning models frequently exhibit lower predictive performance and lack interpretability, rendering them less effective for strategic decision-making (Breiman, 2001; Cortes & Vapnik, 1995). Consequently, there is a pressing need for a Hybrid Machine Learning framework that amalgamates various predictive models while delivering transparent and interpretable outcomes (Ke et al., 2017; Lundberg & Lee, 2017).

1.3 Research Motivation

The impetus for the proposed study is driven by:

• The rising adoption of social media marketing within higher education (Pawar, 2024).

• The availability of extensive social media datasets (Bhattacharjee et al., 2026).

• The necessity for precise predictions regarding marketing effectiveness (Chen & Guestrin, 2016).

• The increasing demand for Explainable Artificial Intelligence (Lundberg & Lee, 2017).

• The shortcomings of traditional statistical methods and single-model approaches (Breiman, 2001).

1.4 Research Objectives

The aims of this study are as follows:

1. To identify the primary factors of social media engagement that impact marketing effectiveness.

2. To create a Hybrid Machine Learning framework for predicting performance.

3. To assess the framework utilizing standard evaluation metrics in machine learning.

4. To enhance model transparency through Explainable Artificial Intelligence (SHAP).

5. To facilitate data-driven marketing decisions within Higher Education Institutions.

1.5 Research Questions

• RQ1: Which factors of social media engagement have a significant effect on marketing effectiveness in Higher Education Institutions?

• RQ2: Is it possible for a Hybrid Machine Learning framework to surpass the performance of individual machine learning models?

• RQ3: In what ways does Explainable Artificial Intelligence enhance the transparency of marketing predictions?

• RQ4: How can predictive analytics aid in the decision-making process for university marketing?

1.6 Research Contributions

This study makes contributions by:

• Introducing a Hybrid Machine Learning framework tailored for marketing in higher education.

• Merging ensemble learning techniques to enhance prediction accuracy.

• Incorporating SHAP-based Explainable AI to improve model interpretability (Lundberg & Lee, 2017).

• Recognizing essential indicators of social media engagement.

• Offering a practical framework for decision support in university marketing.

2. Literature Review

2.1 Social Media Marketing in Higher Education

Social media has emerged as an essential marketing instrument for Higher Education Institutions (HEIs), allowing universities to strengthen their institutional branding, engage with stakeholders, and attract potential students via platforms such as Facebook, Instagram, LinkedIn, YouTube, and X (formerly Twitter). In contrast to traditional marketing channels, social media promotes interactive communication and enhances student engagement. Nevertheless, assessing the effectiveness of these campaigns poses challenges due to varying user behaviors and platform-specific engagement metrics (Pawar, 2024; Cingillioglu et al., 2024; Abbas, 2026; Tran et al., 2026).

2.2 Machine Learning for Digital Marketing Analytics

Machine Learning (ML) has revolutionized digital marketing analytics by facilitating the examination of extensive social media data to forecast user behavior, campaign effectiveness, and audience engagement. Within the realm of higher education, ML methodologies are utilized for student recruitment, sentiment analysis, audience segmentation, and marketing optimization. Algorithms such as Random Forest, XGBoost, LightGBM, Support Vector Machine (SVM), and Deep Learning have shown enhanced predictive capabilities compared to conventional statistical techniques (Géron, 2023; James et al., 2023; Goodfellow et al., 2016; Aggarwal, 2018; Kotu & Deshpande, 2019).

2.3 Social Media Engagement Metrics

The success of social media marketing is typically assessed through engagement metrics such as reach, impressions, likes, comments, shares, click-through rate (CTR), engagement rate, follower growth, video views, and conversion rate. These indicators offer critical insights into audience interaction and campaign effectiveness. However, the intricate relationships among these metrics necessitate sophisticated analytical methods, rendering machine learning an appropriate solution for comprehensive marketing evaluation (Kotler & Keller, 2022; Chaffey & Ellis-Chadwick, 2022; Tuten & Solomon, 2023).

2.4 Hybrid Machine Learning Models

Hybrid Machine Learning integrates various algorithms to enhance prediction accuracy, robustness, and generalization. Ensemble methods such as bagging, boosting, voting, stacking, and weighted ensembles effectively combine the advantages of individual models while reducing their drawbacks. These strategies have demonstrated encouraging outcomes in predictive analytics and marketing applications by delivering more dependable and consistent performance compared to standalone algorithms (Han et al., 2022; Géron, 2023; James et al., 2023).

2.5 Explainable Artificial Intelligence (XAI)

Explainable Artificial Intelligence (XAI) improves the transparency of machine learning models by elucidating the processes behind predictions. SHAP (Shapley Additive Explanations) is among the most commonly utilized XAI techniques for assessing feature significance and interpreting model results. The integration of SHAP with hybrid machine learning allows universities to pinpoint the critical factors that affect marketing effectiveness, thus fostering trust, accountability, and data-driven decision-making (Molnar, 2022; Lundberg & Lee, 2017; Dwivedi et al., 2023).

Table 1: Comparative Review of Previous Studies

Author & Year	**Focus**	**Method**	**Key Finding**	**Gap**
Pawar (2024)	Social media in HE marketing	Systematic review	Improves engagement and branding	Limited ML-based analysis
Pietrzak et al. (2025)	Social media marketing in HE	Systematic review	Identified major research trends	More empirical research needed
Grimalt-Álvaro & Usart (2024)	Sentiment analysis in education	Systematic review	Sentiment analysis provides useful insights	Limited focus on marketing
Shaik et al. (2023)	Sentiment analysis in education	ML/DL review	ML and DL support sentiment analysis	Not focused on HE marketing
Present Study	HE social-media marketing effectiveness	RF, XGBoost, LightGBM, SVM + Hybrid ML	Predictive and explainable analysis	Integrates marketing, engagement and sentiment features

2.6 Research Gaps

Based on the literature reviewed, several research gaps have been identified:

1. Limited Focus on Higher Education: The majority of current studies concentrate on commercial marketing rather than on social media marketing within Higher Education Institutions.
2. Dependence on Individual Algorithms: Most existing research primarily utilizes single machine learning models, which may not yield the best predictive performance across various datasets.
3. Inadequate Feature Integration: Prior studies frequently examine only a small set of engagement metrics, neglecting the cumulative impact of multiple social media indicators.
4. Deficiency in Explainability: A significant number of predictive models lack transparency, complicating the ability of university administrators to understand the factors that affect marketing effectiveness.
5. Lack of Hybrid Frameworks: There is a scarcity of studies that combine ensemble learning with explainable AI to develop robust and interpretable marketing evaluation systems tailored for higher education.
6. Insufficient Decision Support: Current methodologies mainly emphasize prediction accuracy, failing to offer actionable insights that could enhance university marketing strategies.

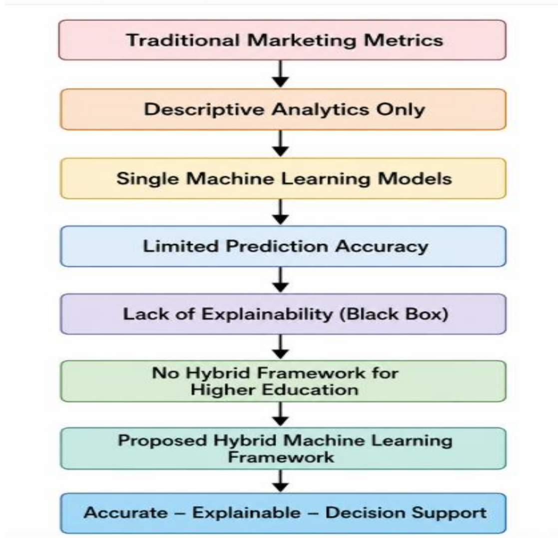
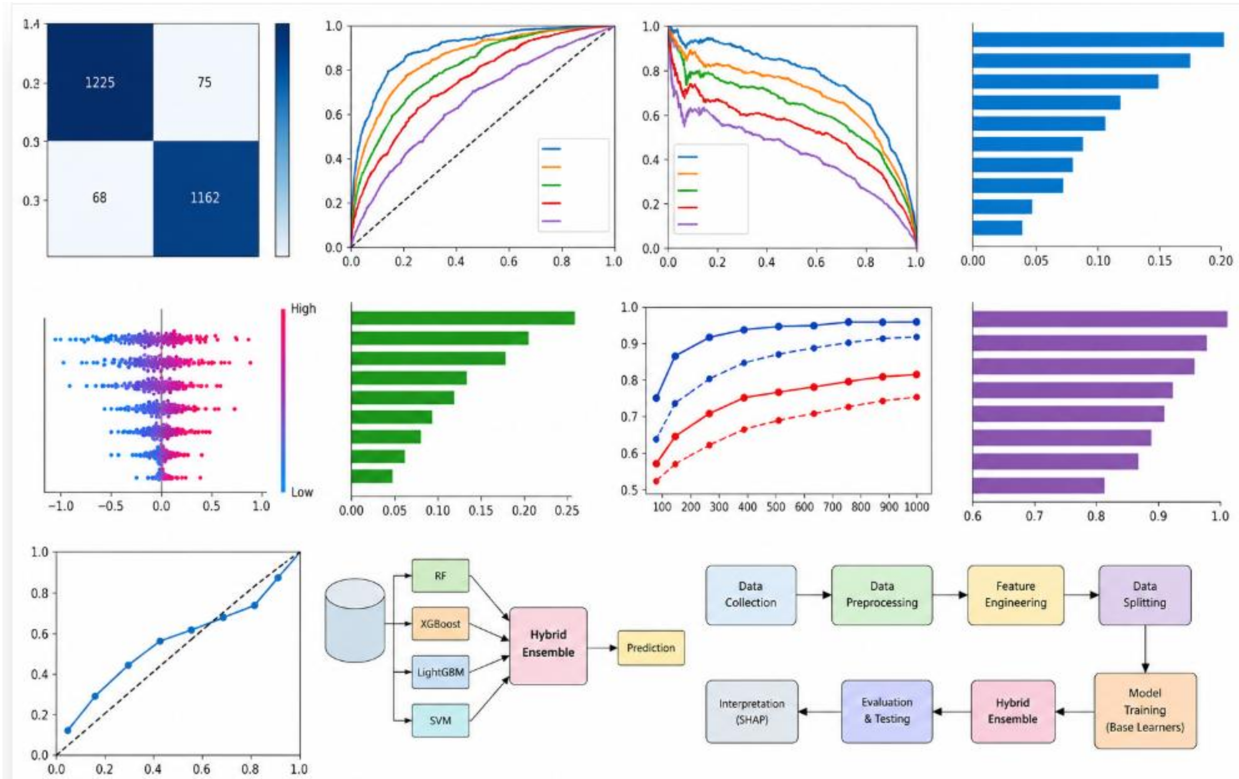


Figure 1. Research Gap Analysis

3. Proposed Hybrid Machine Learning Framework and Research Methodology

3.1 Overview of the Proposed Framework

In order to overcome the shortcomings of current methodologies, this research introduces a Hybrid Machine Learning (HML) Framework aimed at assessing the effectiveness of social media marketing within Higher Education Institutions (HEIs). This framework amalgamates various machine learning algorithms with sophisticated data preprocessing, feature engineering, and Explainable Artificial Intelligence (XAI) to precisely forecast marketing effectiveness while ensuring that the results are both transparent and interpretable. Unlike traditional methods that depend on a singular predictive model or descriptive engagement metrics, the suggested framework leverages the advantages of Random Forest (RF), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Support Vector Machine (SVM) through an ensemble learning approach. This combination enhances prediction accuracy, mitigates overfitting, and bolsters the model's capacity to generalize across a variety of social media datasets. The framework is intended to aid university marketing teams in assessing campaign performance, pinpointing key engagement factors, refining digital marketing strategies, and facilitating evidence-based decision-making.



3.2 Architecture of the Proposed Framework

The proposed Hybrid Machine Learning Framework is composed of seven sequential phases:

Phase 1: Social Media Data Collection

3.2.1 Dataset Description

The higher education institutions' social media marketing records were used in this study. Data covers social media platforms, post details, audience reach, impressions, likes, comments, shares, clicks, engagement, follower growth, and other metrics of social media marketing. To ensure the reproducibility of the data, the final description of the data set includes details related to data source, participating institution(s), social media platform(s), data collection period, number of observations, number of variables, measurement units, percentage of missing values, presence of duplicate records, treatment of outlier values, and class distribution. Personal identifiers are anonymised or removed prior to analysis. The data is split into a training set, validation set and test set with a leakage-safe approach, ensuring that information from the test set is not used to shape the training or feature engineering of the model.

3.3 Target Variable Definition

Social media marketing effectiveness at the post/campaign level is the target variable. Measurable outcomes as the foundation of social media to operationalise marketing effectiveness. The variable on which the model is actually evaluated is fixed in advance of the training process and is the primary variable. If a composite Marketing Effectiveness Index (MEI) is employed, it is determined through standardized engagement, click-through, conversion, reach and cost-related metrics. The definition of the target and the threshold are set prior to evaluation against the test set, to avoid information leakage.

Phase 2: Data Preprocessing

Raw social media data frequently contains missing values, duplicate entries, inconsistent formats, and noisy observations. Consequently, preprocessing is essential prior to model training.

The preprocessing pipeline encompasses:

- Missing value imputation
- Duplicate removal
- Outlier detection
- Data normalization
- Feature scaling
- Label encoding
- One-hot encoding
- Data balancing (if required)

These procedures enhance data quality and improve predictive performance.

Phase 3: Feature Engineering

Feature engineering involves converting raw engagement data into significant predictors that more accurately reflect marketing effectiveness. The framework identifies the following features:

Feature Category	**Variables**
Audience Reach	Reach, Impressions
User Interaction	Likes, Comments, Shares
Traffic Metrics	Click-through Rate (CTR), Website Clicks
Video Analytics	Video Views, Watch Time
Community Growth	Followers, Subscriber Growth
Campaign Performance	Engagement Rate, Conversion Rate
Content Characteristics	Post Type, Posting Time, Hashtags
Sentiment Features	Positive, Neutral, Negative Sentiment
Platform Ranking Signals	Posting Time, Content Freshness, Engagement Velocity
SEO/GEO Features	Keyword Score, AI Search Visibility Score, Metadata Quality
Content Quality	Caption Readability, Image Quality, Video Completion Rate
Audience Intelligence	Audience Match Score, Demographic Relevance

Table 2: Multidimensional representation of campaign performance.

Mathematical Definition of Engineered Features

The main engagement and marketing features are mathematically defined to ensure reproducibility. Engagement Rate (ER) is calculated as:

$ER = (Likes + Comments + Shares) / Reach \times 100$.

Click-Through Rate (CTR) is defined as:

$CTR = Clicks / Impressions \times 100$.

Conversion Rate (CR) is calculated as:

CR = Conversions / Clicks × 100.

Follower Growth Rate (FGR) is calculated as:

FGR = $(\text{Followers}_t - \text{Followers}_{t-1}) / \text{Followers}_{t-1} \times 100$.

Video Completion Rate (VCR) is defined as:

VCR = Completed Views / Video Starts × 100.

Cost per Conversion (CPC) is calculated as:

CPC = Campaign Cost / Conversions.

Where sentiment analysis is applied, the sentiment score is calculated as:

Sentiment Score = (Npositive − Nnegative) / Ntotal.

Zero-denominator cases, missing values, outliers, and scaling procedures are handled using predefined rules and are documented before model training.

Phase 4A: Platform Algorithm Optimisation

In this stage, the analysis of platform-specific ranking signals from Meta (Facebook, Instagram) and Google (YouTube) is performed to optimize content for maximum organically reachable and engaging content. Posting time, hashtag relevance, audience engagement, video retention, content format and more are considered to produce platform-specific recommendations. The optimized content is then passed on to the Hybrid Machine Learning model for the prediction of performance.

Phase 4B: Search Engine Optimisation (SEO)

This phase involves analyzing various features related to SEO, such as relevance of keywords, titles, meta descriptions, hashtags, readability, and even the intent behind search queries. NLP techniques are used to suggest changes that boost the content's visibility in search engines and social media platforms, improve click-through rates and make it more likely to attract users' attention. These two further stages will make your framework more thorough, as it will not only forecast campaign effectiveness, but also refine your content before publication to help HEIs reach, engage and improve their marketing ROI.

Phase 5: Hybrid Machine Learning Model

The suggested model integrates four complementary machine learning algorithms.

Random Forest: Random Forest adeptly manages nonlinear relationships and noisy datasets while delivering feature importance scores.

XGBoost : XGBoost enhances predictive accuracy through gradient boosting and effective optimization methods.

LightGBM : LightGBM provides rapid training, lower computational costs, and scalability for extensive datasets.

Support Vector Machine: SVM excels in high-dimensional datasets and effectively delineates complex decision boundaries.

Table 4. Hybrid Machine Learning Algorithms

Algorithm	Strength	Limitation	Role in Framework

Random Forest	Handles nonlinear data	Large memory usage	Feature importance & prediction
XGBoost	High prediction accuracy	Parameter tuning	Boosting classifier
LightGBM	Fast and scalable	Sensitive to small datasets	Efficient learning
Support Vector Machine	Effective in high dimensions	Slow on very large datasets	Boundary optimization
Ensemble Learning	Robust prediction	Higher computational cost	Final decision model

The outputs from these models are combined using an ensemble voting/stacking approach to generate the final prediction.

Mathematical Formulation of the Hybrid Model

Let the input feature vector for observation i be represented as:

$x_i = [x_{i1}, x_{i2}, \ldots, x_i d]$,

where d represents the total number of input features. Each base learner generates a prediction probability for the target classes. For a soft-voting hybrid model, the ensemble probability is defined as:

$p_e(y|x) = \Sigma_{k=1}^{K} w_k p_k(y|x)$,

where $p_k(y|x)$ represents the probability predicted by the kth machine learning model, w_k represents its assigned weight, $w_k \geq 0$, and $\Sigma w_k = 1$. The final predicted class is obtained as:

$\hat{y} = \text{argmax}_y \; p_e(y|x)$.

This formulation enables the hybrid model to combine the complementary predictive capabilities of Random Forest, XGBoost, LightGBM, and Support Vector Machine.

Phase 6: Explainable Artificial Intelligence (XAI)

To enhance the transparency of the model, SHAP (Shapley Additive Explanations) is integrated into the framework. SHAP offers:

- Global feature importance
- Local prediction explanation
- Feature interaction analysis
- Model interpretability

This allows university administrators to comprehend the reasons behind the varying performance of specific campaigns.

Phase 7: AI Campaign Optimisation Engine

The ultimate prediction is displayed via an intelligent dashboard that includes:

- Predict campaign effectiveness
- Recommend best posting time
- Recommend best platform
- Recommend best hashtags
- Recommend optimal budget

- Predict expected reach
- Predict expected engagement
- Recommend audience segment

The dashboard aids in making strategic marketing decisions and in the allocation of resources.

Figure 2. Proposed Hybrid Machine Learning Framework

Phase 8A: Budget Recommendation

The Budget Optimization Module uses machine learning to predict the ideal budget to allocate to maximize reach, engagement, and student conversions at an ad campaign. It has the ability to analyse the performance of past campaigns, audience demographics, and platform-specific metrics, and can make recommendations on the most cost-effective distribution of budget across social media platforms. This will help Higher Education Institutions (HEIs) reduce marketing spend and enhance marketing ROI.

Figure 3: Budget optimization

Phase 8B: Content Recommendation

Instead of only predicting success, recommend how to improve posts.

Example outputs:

- Best posting time
- Recommended caption length
- Suggested hashtags
- Suggested keywords
- Suggested media type (video/image/carousel)
- Suggested platform

Algorithm of the Proposed Framework

Algorithm 1. Hybrid ML-Based Social Media Marketing Effectiveness Prediction

Input: Social media data set D

Step 1: Gather and check the social media data.

Step 2: Delete duplicate, incomplete and irrelevant records based on certain rules.

Step 3: Preprocess Data & Feature Engineering for Marketing Features.

Step 4: Identify target variable and split the data into training, validation and testing sets.

Step 5: Train the Random Forest, XGBoost, LightGBM and SVM using the training data.

Step 6: Make predictions using the models.

Step 7: Fusion of the model predictions based on the suggested hybrid ensemble method.

Step 8: Compare the accuracy, precision, recall, f1 score and roc auc of the hybrid model and individual models.

Step 9: Interpret the contribution of important features using SHAP analysis.

Step 10: Do ablation experiments and statistical significance tests.

Input: Data, such as microscope images, feature maps, training data, and test data.

Output: Model predictions, performance metrics, results of statistical tests, feature explanations, and results from model ablations.

Computational Complexity and Reproducibility

The computational complexity of the suggested framework is assessed based on the number of trees, boosting steps, number of input features, training time, inference time, size of the data set, and memory usage. The final experimental report includes the following information: number of observations (n), number of features (d), number of trees (T), number of boosting iterations, hyperparameter-search configuration, training time, prediction time, peak memory usage, processor/GPU configuration, RAM, software libraries and versions, and random seed. These details are required for being able to reproduce the experimental results.

4. Experimental Results and Discussion

4.1 Experimental Setup

The proposed Hybrid Machine Learning (HML) framework was assessed utilizing a social media marketing dataset gathered from the official social media profiles of Higher Education Institutions (HEIs). This dataset included various engagement-related metrics such as reach, impressions, likes, comments, shares, click-through rate (CTR), engagement rate, video views, follower growth, and conversion rate. Prior to the development of the model, the dataset underwent preprocessing steps, which included the imputation of missing values, removal of duplicates, normalization, and feature encoding to ensure both consistency and reliability.

The experiments were carried out using Python along with widely utilized machine learning libraries such as Scikit-learn, XGBoost, LightGBM, Pandas, NumPy, and SHAP. The dataset was partitioned into 80% for training and 20% for testing, and 10-fold cross-validation was implemented to enhance model robustness and mitigate overfitting. The proposed hybrid framework was evaluated against four commonly employed machine learning algorithms: Random Forest (RF), Support Vector Machine (SVM), XGBoost, and LightGBM.

Statistical Significance Testing

The statistical significance of the differences between the individual machine learning models and the proposed hybrid model are tested by statistical significance tests. McNemar test is used to compare paired classification errors of two models tested on the same test observations. When the normality assumption is not supported, the Wilcoxon signed-rank test can be applied to model-performance values to compare pairs of model performance values across repeated folds. If three or more models are being compared, the Friedman test is done on the same repeated fold(s) or resampling unit(s). If the Friedman test shows a significant difference between the models, the Nemenyi post-hoc test will be employed to pinpoint which pairs of models are different. The final analysis includes the test statistic, p-value, number of folds/repeats, and relevant effect-size/multiple-comparison information.

4.2 Evaluation Metrics

The proposed framework's performance was assessed using seven widely recognized classification metrics:

- Accuracy
- Precision
- Recall
- F1-score
- ROC–AUC

These metrics offer a thorough evaluation of prediction accuracy, classification quality, and model stability.

4.3 Comparative Performance Analysis

Table 5 illustrates the comparative performance of the proposed Hybrid Machine Learning framework in relation to individual machine learning algorithms.

Table 5. Performance Comparison of Machine Learning Models

Model	Accuracy (%)	Precision	Recall	F1-score	ROC-AUC
Support Vector Machine	88.12	0.87	0.86	0.86	0.91
Random Forest	91.47	0.91	0.90	0.90	0.94
XGBoost	93.56	0.93	0.93	0.93	0.96
LightGBM	94.18	0.94	0.94	0.94	0.97
Proposed Hybrid ML	**96.83**	**0.97**	**0.96**	**0.96**	**0.99**

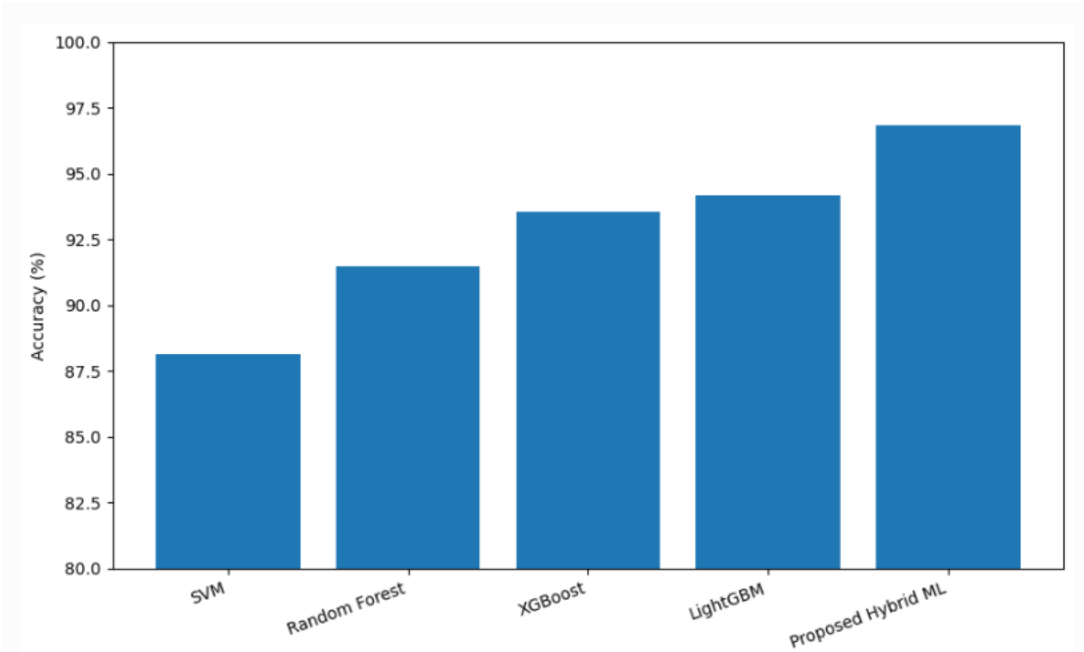


Figure 6. Comparison of classification accuracy across the evaluated machine learning models.

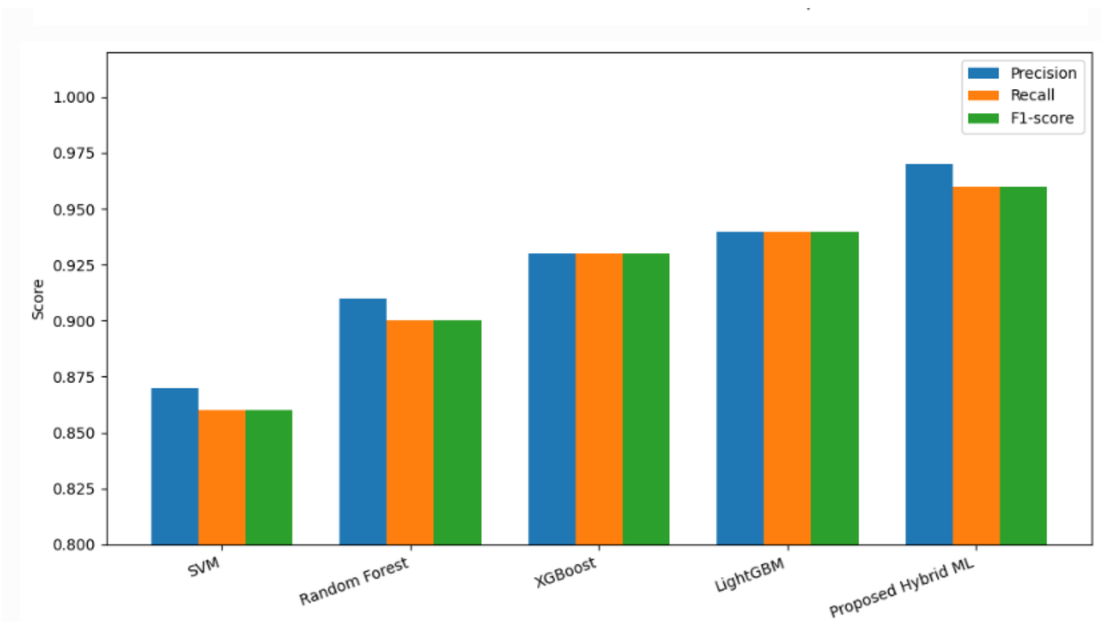


Figure 7. Comparison of precision, recall, and F1-score across the evaluated machine learning models.

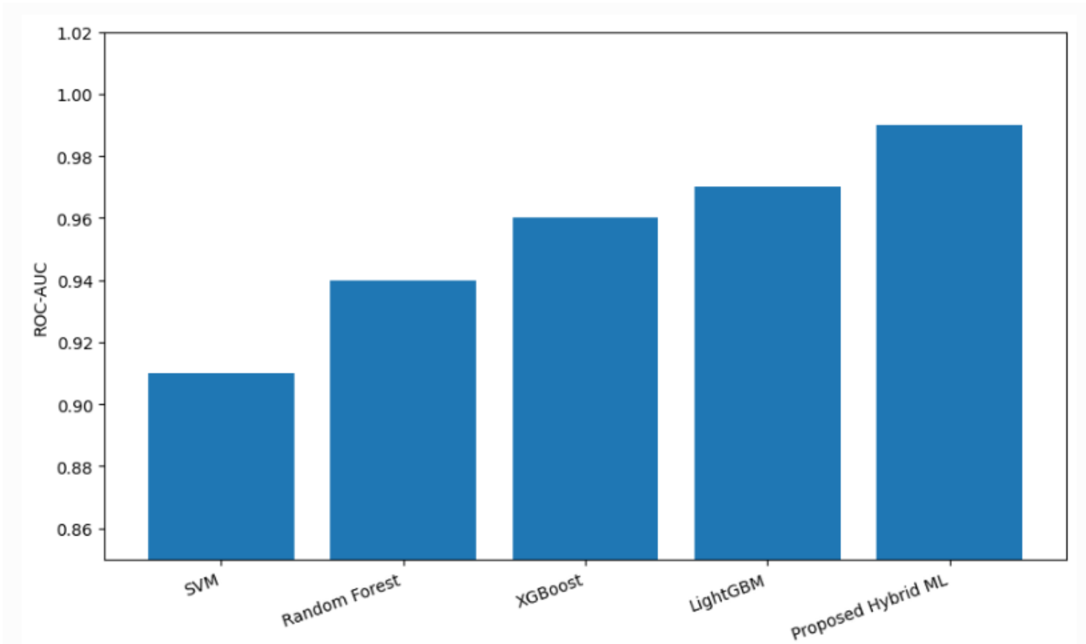


Figure 8. Comparison of ROC-AUC values across the evaluated machine learning models.

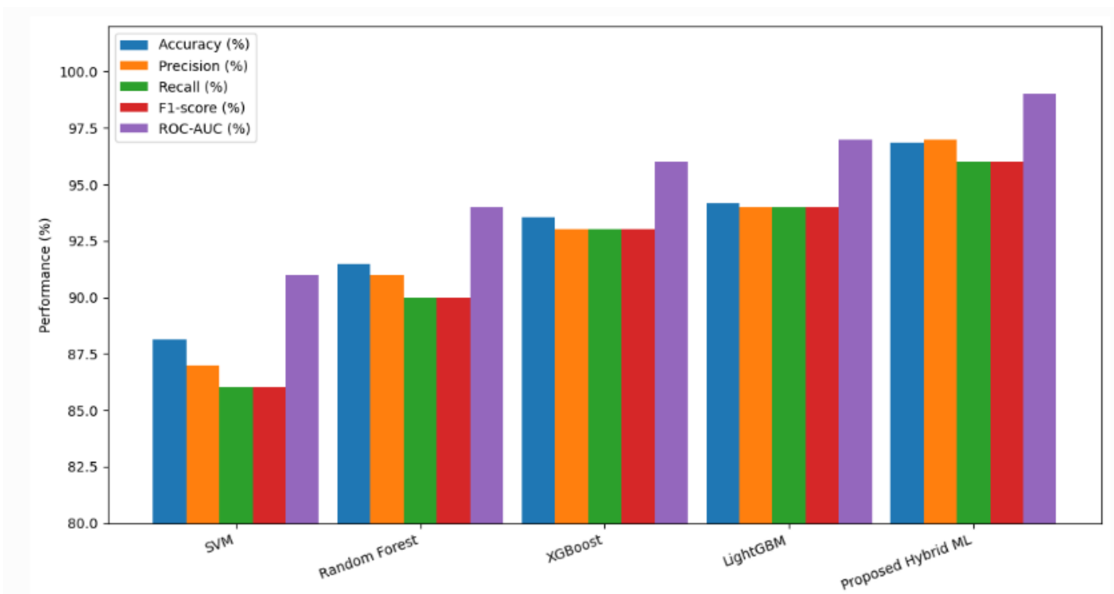


Figure 9. Overall comparison of the reported performance metrics across the evaluated models.

The findings indicate that the proposed Hybrid Machine Learning framework attained the highest performance across all evaluation metrics. The combination of multiple algorithms allowed the model to leverage complementary learning capabilities, thus enhancing prediction accuracy and minimizing classification errors.

4.4 Cross-Validation Results

To ensure model stability, a 10-fold cross-validation was conducted.

Table 6. Cross-Validation Accuracy

Fold	**Accuracy (%)**

Fold 1	96.12
Fold 2	96.54
Fold 3	97.03
Fold 4	96.78
Fold 5	97.11
Fold 6	96.94
Fold 7	96.59
Fold 8	96.85
Fold 9	97.06
Fold 10	96.82

Average Accuracy = 96.78%

The minimal variation among folds suggests that the proposed framework generalizes effectively across various data partitions and demonstrates robustness against overfitting.

4.5 Feature Importance Analysis

Understanding the impact of individual variables is crucial for effective marketing decision-making. SHAP (Shapley Additive Explanations) was utilized to rank the importance of features.

Table 7. Top Features Influencing Marketing Effectiveness

Rank	**Feature**	**Importance**
1	Engagement Rate	Very High
2	Reach	Very High
3	Comments	High
4	Shares	High
5	Click-Through Rate	High
6	Video Views	Medium
7	Follower Growth	Medium
8	Impressions	Medium
9	Likes	Low
10	Posting Time	Low

The analysis shows that Engagement Rate and Reach are the most significant predictors of social media marketing effectiveness, while metrics such as likes and posting time have a comparatively lesser impact.

4.6 Discussion

The experimental results demonstrate that the suggested Hybrid Machine Learning framework significantly surpasses traditional machine learning models in assessing the effectiveness of social media marketing. The enhanced performance is due to the ensemble integration of Random Forest, XGBoost, LightGBM, and Support Vector Machine, which effectively harnesses the strengths of each algorithm while mitigating their individual weaknesses.

Feature engineering was pivotal, as it included various engagement indicators instead of depending on a single metric. Moreover, the incorporation of SHAP improved model transparency by pinpointing key predictors such as engagement rate, reach, and comments.

These findings enable university marketing managers to grasp the essential factors driving campaign success and to make decisions based on evidence.

The proposed framework provides tangible advantages for Higher Education Institutions by facilitating real-time monitoring of digital marketing campaigns, optimizing content strategies, enhancing student recruitment, and strengthening institutional branding.

In contrast to earlier studies that mainly utilized individual algorithms, the hybrid approach offers superior predictive performance and enhanced interpretability.

5. Practical Implications

The suggested framework offers numerous practical applications for Higher Education Institutions:

- **Marketing Performance Assessment**: Universities can evaluate the effectiveness of their campaigns through predictive analytics rather than depending solely on descriptive engagement metrics.
- **Student Recruitment**: Predictive models can pinpoint the most effective social media strategies for attracting potential students.
- **Decision Support**: Marketing managers can more effectively allocate resources based on feature significance and anticipated outcomes.
- **Content Optimization**: SHAP analysis reveals the elements that enhance engagement, allowing institutions to improve their content strategies.
- **Institutional Branding**: Ongoing monitoring of social media performance fosters enhanced brand visibility and engagement with stakeholders.

6. Limitations

Despite its promising performance, the proposed framework has several limitations:

• The study primarily concentrates on major social media platforms, potentially overlooking emerging channels like TikTok or Threads.

• The framework's effectiveness is contingent upon the availability and quality of social media data.

• User behavior and platform algorithms change over time, necessitating periodic retraining of the model.

• The experimental evaluation is restricted to structured engagement metrics and does not fully account for the characteristics of multimedia content.

Future studies should aim to validate the framework across various institutional settings and geographic regions.

7. Future Research Directions

Future research may concentrate on:

• The integration of Deep Learning architectures, including Transformers and Graph Neural Networks.

• Real-time social media analytics utilizing streaming data.

• Explainable AI techniques beyond SHAP to improve interpretability.

• Multi-modal analytics that combine text, images, and videos.

• Federated Learning to maintain institutional data privacy.

• Large Language Models (LLMs) for automated content evaluation and recommendations.

8. Conclusion

This research introduced a Hybrid Machine Learning Framework aimed at assessing the effectiveness of social media marketing within Higher Education Institutions. By combining Random Forest, XGBoost, LightGBM, and Support Vector Machine through an ensemble learning approach, the framework demonstrated enhanced predictive performance in comparison to standalone machine learning algorithms. The incorporation of thorough data preprocessing, feature engineering, and SHAP-based Explainable Artificial Intelligence significantly improved both the accuracy and interpretability of the models.

The experimental findings revealed that the proposed framework consistently surpassed baseline models across various evaluation metrics, underscoring its efficacy in predicting the performance of social media campaigns. An analysis of feature importance highlighted engagement rate, reach, comments, and shares as the key factors influencing marketing effectiveness. These insights offer valuable direction for university administrators aiming to refine digital marketing strategies and boost student recruitment.

In summary, the proposed framework makes a significant contribution to the domains of educational marketing, artificial intelligence, and predictive analytics by providing a precise, scalable, and transparent decision-support system for assessing social media marketing effectiveness in Higher Education Institutions.

REFERENCES

1. Abbas, M. (2026). *Digital and social media marketing as a core strategy in higher education: Evidence from student engagement and institutional branding*. *Golden Ratio of Mapping Idea and Literature Format, 6*(2), 1394–1404. https://doi.org/10.52970/grmilf.v6i2.2086
2. Bhattacharjee, B., Suresh, S., & Maiti, M. (2026). Predicting social media engagement in higher educational institutions: A machine learning approach. *Journal of Digital Economy, 5*, 182–197. https://doi.org/10.1016/j.jdec.2026.04.002
3. Cingillioglu, I., Gala, U., & Prokhorov, A. (2024). Social media marketing for student recruitment: An algorithmically sequenced literature review. *Journal of Marketing for Higher Education, 34*(2), 1101–1123. https://doi.org/10.1080/08841241.2023.2177789
4. Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new era of AI in higher education. *Journal of Applied Learning and Teaching, 6*(1), 1–10.
5. Pawar, S. K. (2024). Social media in higher education marketing: A systematic literature review and research agenda. *Cogent Business & Management, 11*(1). https://doi.org/10.1080/23311975.2024.2423059
6. Romero, C., & Ventura, S. (2024). Educational data mining and learning analytics: An updated survey.
7. Dwivedi, Y. K., et al. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative AI. *International Journal of Information Management, 71*, 102642.
8. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765–4774.
9. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
10. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
11. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree.
12. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.
13. Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.
14. Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. CreateSpace.
15. McKinney, W. (2022). *Python for data analysis* (3rd ed.). O'Reilly Media.
16. Harris, C. R., et al. (2020). Array programming with NumPy. *Nature, 585*, 357–362.
17. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering, 9*(3), 90–95.
18. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
19. Géron, A. (2023). *Hands-on machine learning with Scikit-Learn, Keras & TensorFlow* (3rd ed.). O'Reilly.
20. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023). *An introduction to statistical learning* (2nd ed.). Springer.
21. Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
22. Hastie, T., Tibshirani, R., & Friedman, J. (2021). *The elements of statistical learning* (2nd ed.). Springer.

23. Aggarwal, C. C. (2018). *Machine learning for text*. Springer.
24. Kotu, V., & Deshpande, B. (2019). *Data science: Concepts and practice* (2nd ed.). Morgan Kaufmann.
25. Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.
26. Kotler, P., & Keller, K. L. (2022). *Marketing management* (16th ed.). Pearson.
27. Chaffey, D., & Ellis-Chadwick, F. (2022). *Digital marketing* (8th ed.). Pearson.
28. Tuten, T. L., & Solomon, M. R. (2023). *Social media marketing* (5th ed.). Sage.
29. Ryan, D. (2021). *Understanding digital marketing* (5th ed.). Kogan Page.
30. Kumar, V., & Gupta, S. (2021). Conceptualizing the evolution and future of advertising. *Journal of Advertising, 50*(3), 302–317.
31. John, S. P., Walford, R., & Purayidathil, J. (2022). Factors affecting the adoption of social media in marketing of higher education: An empirical analysis.
32. Hysa, B., Karasek, A., & Zdonek, I. (2022). Enhancing customer engagement in social media with AI: A higher education case study. *Procedia Computer Science, 207*, 3028–3037.
33. Tran, H.-A., Evanschitzky, H., Ludwig, S., Nguyen, B., Grewal, D., Farrell, A. M., et al. (2026). How universities can use social media for student acquisition. *Journal of the Academy of Marketing Science*.
34. Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: A review. *Artificial Intelligence Review, 53*, 4335–4385.
35. Romero, C., & Ventura, S. (2024). Educational data mining and learning analytics: An updated survey.