EAI Endorsed Transactions
on AI and Robotics

Research Article EAI.EU

From Social Media Reactions to Grades: A Machine Learning-Based SocialNet Analysis for Academic Performance Prediction

Muhammad Ramzan[1], Naeem Ahmed[2,*]

[1]Department of Information Technology, University of Haripur, Haripur, Khyber Pakhtunkhwa, Pakistan
[2]School of Software Engineering, Nanjing University of Information Science and Technology, Nanjing, China

Abstract

The impact of social media on student academic performance has garnered significant research interest in recent years. The pervasive use of social networking sites (SNS) among college and university students, both in and outside classrooms, has raised concerns about its potential effects on academic achievement. This study investigates the relationship between social media usage and academic performance through a dataset of 550 participants. Machine learning models, including Random Forest, Decision Trees, and Long Short-Term Memory (LSTM), were employed to analyze and predict the impact of social media on students' academic outcomes. The models were trained using clean and well-engineered data. The results indicate a moderate influence of social media usage on academic performance. Notably, the LSTM model achieved an accuracy of 81.2%, outperforming the RF and DT models, which achieved approximately 77.9% and 72.1% accuracy, respectively. Furthermore, error metrics support these findings: the RF model recorded a Root Mean Squared Error (RMSE) of 0.2677 and a Mean Absolute Error (MAE) of 0.1374, while the DT model yielded an RMSE of 0.3008 and an MAE of 0.1556. These findings highlight the importance of considering sequential usage patterns in understanding the academic implications of social media.

Received on 12 December 2024; accepted on 11 March 2025; published on 17 March 2025

Keywords: SocialNet Analysis, Machine Learning, Student Performance, Education, Learning Technologies

doi:10.4108/airo.8171

1. Introduction

In recent years, there has been a notable increase in the number of users of social media and social networking sites (SNSs) [14]. These platforms are used across all age groups, but young people between the ages of 18 and 29 are the most frequent users [5]. According to published data, there were 2.13 billion SNS users in 2016 compared to 970 million in 2010 [19]. Surveys indicate that Facebook is the most widely used SNS [32]. As of 2013 [19], there were over 9.4 million Facebook members in South Africa, while by 2014, this number had risen to thirteen million in Australia. In Ghana, the trend is similar. Ghana is the 47th most web-aware nation globally, closely trailing the United Arab Emirates. With a 28.4% internet penetration rate, 7,958,675 of the country's population is online. Reports suggest that Ghana's social media user base is growing exponentially; in the first quarter of 2016, there were 2.9 million users [13]. According to a survey, every Ghanaian with a smartphone has at least one social media account on platforms such as Facebook, Instagram, Twitter, WhatsApp, or Facebook. Approximately 4.9 billion people, or 34.3% of the world's population, used Facebook by the end of 2017 [29].

Research reveals that the average Ghanaian smartphone user spends about 5 hours and 13 minutes daily on the internet, with approximately 3 hours and 13 minutes dedicated to social media platforms. Studies also indicate that individuals born between 1965 and 1979 spend about 2 to 3 hours a day on SNSs, reflecting moderate usage for staying connected and informed. In contrast, those born between 1990 and 1999 spend an

*Corresponding author. Email: naeem.uoh@gmail.com

average of 4 to 6 hours daily on social media, primarily for entertainment, engagement, and socialization. People's dynamic activity patterns have shifted from traditional lifestyles to spending more time on social media [5]. This remarkable surge in social media and mobile technology usage raises important questions about how educators and learners can leverage the benefits of SNSs.

Recent surveys highlight an improvement in students' attitudes towards using technologies like SNSs for learning purposes [13, 16, 29]. For instance, a study at the University of Ghana, Legon, found that all randomly sampled students predominantly use Facebook and WhatsApp [14]. However, the findings also corroborate [30], indicating that most students use SNSs for socializing rather than academic purposes. Nine out of ten tertiary students use social media to communicate with family and friends globally, which has also been beneficial in class and research settings, according to other studies [3].

Globally, students' social media usage has increased, affecting their time management, grammar, and spelling due to spending more time on SNSs and less time studying [18]. Similar studies suggest that SNS usage impacts students' academic performance both positively and negatively, depending on how extensively they embrace and enjoy these platforms [14]. Researchers believe that students with higher SNS usage are likely to experience an adverse impact on their grade point average (GPA) as they dedicate less time to academics. Although many variables influence academic achievement, SNS usage has been identified as a significant factor [6]. This discussion implies that students' social media usage during downtime may directly affect their academic performance.

Despite numerous studies in this field, research on the effects of SNS usage on students' academic performance in developing nations like Ghana is still in its early stages. The existing literature presents conflicting findings: some studies report negative effects, others positive, while some find no impact at all. The extent to which the nature, frequency, and purpose of SNS usage influence students' academic success remains unclear.

This study investigates the impact of social media usage on students' academic performance through a machine learning-enabled analysis. A dataset of 550 students from various departments was collected, focusing on key variables such as the purpose of social media use (e.g., learning or socializing), frequency, and timing (in-class or outside of class). Feature engineering was used to identify reliable indicators of academic achievement. Using these features, we applied Decision Tree (DT) and Random Forest (RF) models to analyze and predict the academic impact of social media, comparing their performance to that of a Long Short-Term Memory (LSTM) neural network for deeper sequential insights.

The primary goal of this study is to look into the relationship between social media usage and academic achievement among students. With the growing popularity of social networking sites (SNS), it is critical to understand how students' time spent on these platforms, either for education, social contact, or pleasure, affects their academic performance, notably their grade point average (GPA). While previous studies examined at the overall impact of technology on education, there has been not much research into the specific patterns of SNS usage, such as the purpose, frequency, and timing of engagement. By using machine learning models to predict academic performance based on these detailed SNS usage patterns, this study aims to provide valuable insights for both educators and policymakers to better understand how social media influences student learning and achievement.
The main contributions of our research are as follows:

- Collected and curated a novel dataset of 550 students, including detailed variables on social media usage patterns (purpose, frequency, timing) and academic performance (GPA), providing a valuable resource for studying the interplay between technology usage and education.
- Conducted a comprehensive study to assess the extent of students' exposure to social networking sites (SNS).
- Identified the primary purposes for which students use SNS, such as learning or socializing.
- Analyzed the impact of SNS usage on students' academic performance, particularly its effect on their grade point average (GPA).
- Developed and utilized feature engineering techniques to extract reliable indicators of academic performance from social media usage patterns.
- Evaluated the effectiveness of machine learning models, including Random Forest, Decision Tree classifiers, and Long Short-Term Memory (LSTM) networks, in predicting students' academic achievement based on social media usage. Additionally, we proposed an optimal algorithm-based strategy for accurate prediction of academic performance.

The rest of the paper is organized into five sections. Section 2 provides a comprehensive literature review. Section 3 discusses the methodology of the proposed work. Section 4 presents the results, and Section 5 offers a detailed discussion. Finally, Section 6 concludes the article.

2. Related Work

In recent years, the Internet and its associated technologies have dramatically transformed the way information and data are accessed and shared. One well-known example of a tool used to share knowledge among communities and students is social networking sites (SNSs) [3]. Consequently, it is impossible to ignore the enormous rise in student involvement in SNSs [18, 25, 26]. Students spend more time on social networking sites than they do on their studies [18]. This has been linked to improved exam performance and grades [18, 27]. According to a study, students who use social networking sites more after school have worse self-esteem and are less interested in academic subjects. Although these studies suggest that students' usage of SNSs is impacted by the amount of time they spend on them, it is important to consider the nature of SNS use when analyzing SNS use and student performance. Facebook users spend less time studying than non-users, according to a 2014 study [32]. Nonetheless, a sizable portion of students (79%) think that Facebook has no bearing on their academic achievement.

Various researchers used both subjective self-reporting and statistical research to find a negative link between students' grades and Facebook usage time. Furthermore, a somewhat negative relationship between students' GPAs and the amount of time they spent using computers each week outside of class [32]. Lastly, it was discovered that there was no significant correlation between students' internet or cell phone usage frequency and their academic achievement [32]. High school students in Israel were surveyed by [27] to find out how they felt about the connection between academic success and non-academic information and communication technology (NA-ICT). According to their findings, the majority of respondents felt that using NA-ICT outside of the classroom had a detrimental impact on students' academic performance.

A poll conducted in 2013 and 2017 found that social media helps pupils succeed academically [2]. Additionally, the scientists discovered that students who use SNSs frequently get fresh perspectives on academic subjects [6]. In a different study made the case that students' use of social networking sites (SNSs) improves their engagement levels and promotes self-directed learning [23]. It was found that no evidence of a substantial direct correlation between technology use and students' academic achievement. A similar study claimed that social media supports the student-centered learning approach and improves communication and teamwork [10]. Additionally, it was claimed that integrating social media into instructional strategies and modalities (in and out of the classroom) at postsecondary educational institutions increases student engagement [20]. According to recent research, utilizing social networking sites (SNSs) can have an impact on students' grammar. As a result, students may resort to shorthand writing when communicating with friends and family on SNSs, which can lead to them making the same mistakes when taking exams [14, 18]. Since SNSs are becoming more and more popular among students, some academics and economists wonder if how much time students spend on SNSs will have an impact on their grades. Some faculty members perceived SNSs as a forum for students to discuss their work outside of class, which helps students perform better academically, according to a survey conducted to find out the impact of SNSs on students' academic lives and performance among faculty members [30].

Numerous studies have examined the effects of multitasking on students' academic performance when using media devices and technology in the classroom. Multitasking can be defined as "performing two or more tasks simultaneously." also known as dual-tasking [12]. All age groups engage in multitasking regularly, although young people are more likely to do so, especially when using media [9]. Examples include two distinct media, like the Internet and television, two common non-media tasks, like watching TV and doing schoolwork, and two activities on the same medium, such as sending and receiving email on a computer and writing a report [12] it is also discovered that media multitasking impairs learning at the lower order but not at the higher order [31]. However, research on the impact of social media use on students' academic performance has recently gained the attention of researchers in developing and poor countries. Furthermore, conflicting opinions about the relationship between students' usage of social media and their academic performance can be found in the literature, even though the mixed research findings on this topic, as previously mentioned, show disagreement on the subject. Furthermore, the degree to which the nature, rate, and duration of a student's usage of social media predicts their academic achievement has not been established by research.

As summarized in Table 1, the diverse methodologies and key findings in previous research underscore the need for an advanced machine learning approach to better capture the sequential patterns in social media usage.

As summarized in Table 1, the diverse methodologies and key findings in previous research underscore the need for our advanced machine learning approach to better capture the sequential patterns in social media usage.

Table 1. Comparison of Previous Studies and Methodologies on the Impact of Social Media on Academic Performance

Authors	Year	Country/Region	Sample Size / Population	Methodology	Key Findings
Al-Rahmi & Othman	2013	Various	Pilot study (small sample)	Survey-based analysis	Mixed impact of social media on academic performance
Al-Rahmi & Zeki	2017	Various	University students	Collaborative learning model	Enhanced collaborative learning outcomes
Amadi & Ewa	2018	Nigeria	University students	Empirical survey	High social media usage associated with lower academic performance
Amin et al.	2016	Various	Survey & ML analysis	Statistical and machine learning analysis	Negative correlation between usage intensity and GPA
Kolan & Dzandza	2018	Ghana	University students	Case study	No significant correlation, though some negative trends observed

3. Research Methodology

This study used decision tree (DT) and random forest (RF) machine learning algorithms to investigate how much social media is used, how often it is used, and when it is used to predict students' academic achievement. Figure 1 shows the conceptual framework of our study.

3.1. Data Gathering

The study used a convenient sample of 550 students. We employed questionnaires as part of a quantitative study design to collect data. In addition to being optional, participation required the consent of the students. Students enrolled in non-tertiary programs were therefore not included in the poll. Using a straightforward sample technique, 550 students from the University of Haripur Pakistan participated in this study. The participants were chosen from five departments: accounting 40 students (7.27%), computer science 100 students (18.18%), electrical and electronic 220 students (36.36%), Physics 120 students (21.82%), and Chemistry 70 students (12.73%). To further enhance the importance, brevity, and specificity of the questionnaire, a small pilot study involving NUST students was carried out. After completing the survey, each participant provided verbal input regarding the ease of understanding, the simplicity of expression for each item, and the possibility that each respondent could supply the necessary data. As stated by research work based on surveys, experiments, and observation are best suited for collecting primary data using questionnaires [18]. We gather information regarding participant beliefs, attitudes, sentiments, and expected behavior by employing the survey approach. The two components of the questionnaire design were social media usage and demographic data. When appropriate, participants indicated their agreement using a 5-point Likert scale {1, 2,..., 5}, which ranged from not at all to very often. Moreover, with the consent of the participants, Facebook student log records were acquired to determine the precise time on social media. Finally, the performance rate for the student was determined by taking their actual grade point average (GPA) from their department.

Figure 2 presents a summary of the 550 replies that were received, with 76.53% of the respondents being male and 23.46% being female. Regarding age, 7.12% of the participants fell into the 18–21 age range, 65.38% fell into the 22–25 age range, and 27.5% fell into the 26–40 age range. Level 100 accounted for 19% of the total, whereas levels 200 and 300 comprised 60.58% and 20.38% of the total.

As illustrated in Figure 3 in our data WhatsApp has the largest percentage (77.88%), followed by Facebook with 95 records (18.27%), Instagram with 20 records (3.85%), and Viber and Twitter with 0 records. The findings refute the claims that Facebook is the most popular social media platform by demonstrating that WhatsApp is the most popular and widely used social media among students [3, 14]. Additionally, nearly all of the students who had WhatsApp reported using and having a Facebook account, which is consistent with a finding that all students who use Facebook had a WhatsApp account [14].

Figure 4 presents the analysis of our data, 550 respondents, or 100%, uses social media. This finding supports previous research [14, 18] that indicates a significant portion of students use social media in some capacity. When asked why people use Facebook and WhatsApp so often, several respondents said that it's because major communication firms in Pakistan provide free access to these social networking sites. As a result, users can use Facebook and WhatsApp on their cell phones without paying any fees to these telecoms. Our respondents' daily time spent on social media is compiled in Figure 6 to determine the amount of time they spend on SNSs.

Forty-nine individuals (40.38%) reported using SNSs for less than an hour a day. A total of 39.23%, spend no more than two hours a day on social media. 17.69% use social media for three to four hours every day. Although 2.69% of the sample, use social media for at least seven hours a day, this use can be classified as heavy. More than half (59.61%) of the students used SNSs for more than two hours a day, supporting the findings [14, 18, 27]. This affects the student's ability to pass tests and receive good scores. Again, as per the findings of [14, 18] students' interest in academic topics and their self-concept decline with increasing SNS usage. As a result, the amount of time students spend on SNSs validates the worry expressed by certain academics and economists over the possibility that students' grades will be impacted by how much time they spend on SNSs. Students use SNSs for a variety of purposes, as shown by the survey results in Figure 5. shows that a total of 14.14% use social networking sites (SNSs)

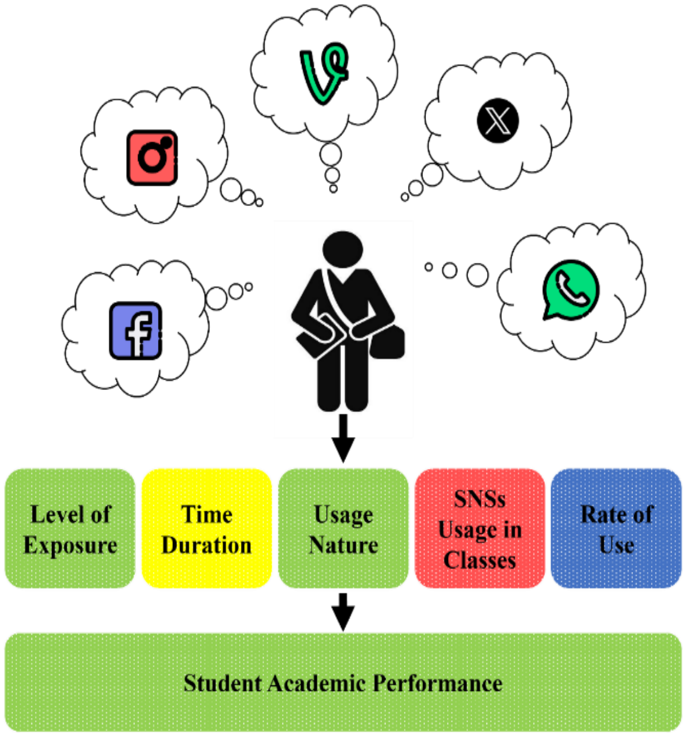

Figure 1. Conceptual model for Social Net Analysis examining the potential influence of social media and SNSS usage on student academic performance including Key factors level of exposure, time duration, usage nature, SNSS usage in classes, and rate of use that impact student academic outcomes, warranting further investigation.

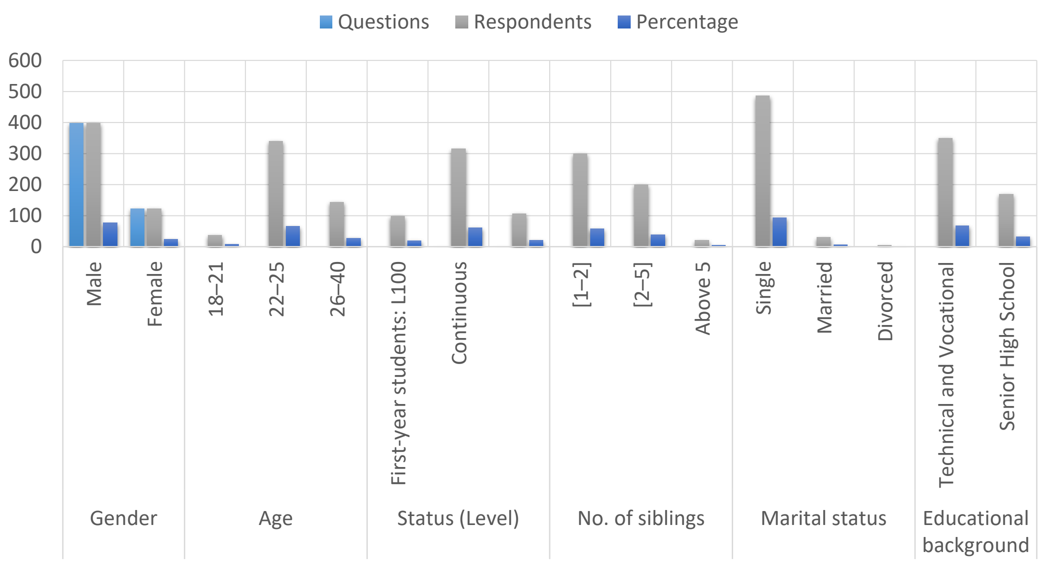

Figure 2. Summary of the collected Dataset on various demographic factors and their corresponding question counts, respondent counts, and percentages. The factors include gender, age, marital status, educational background, and more.

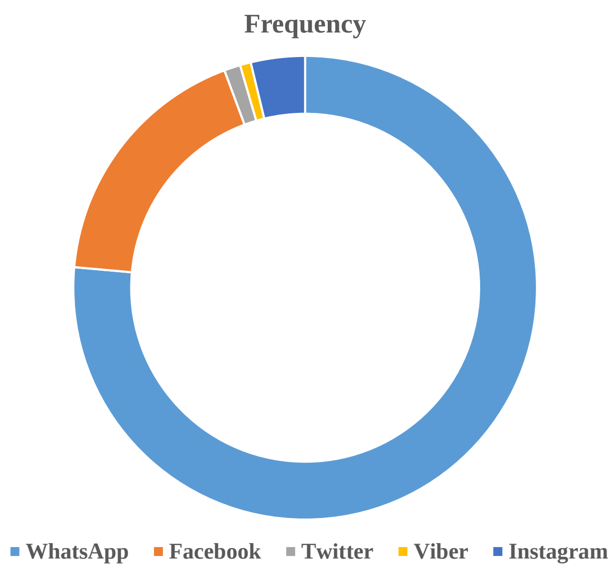

Figure 3. Frequency of usage across various social media platforms. The y-axis lists the platforms (WhatsApp, Facebook, Twitter, Viber, and Instagram), while the x-axis represents the frequency or count of usage.

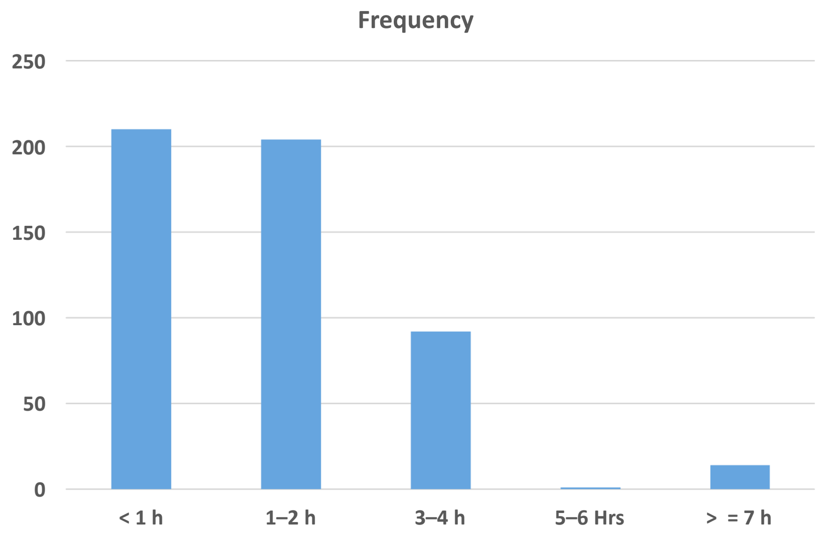

Figure 4. Frequency distribution of social media platform usage, with the x-axis representing different time duration intervals and the y-axis showing the frequency or count. The platforms depicted are WhatsApp, Facebook, Twitter, Viber, and Instagram.

for academic work, 73.27% for communicating with friends and family, and 64 people (12.31%) for watching and downloading audiovisual content, such as films and videos. 79.04% of the student body, responded in the affirmative, 11.35% in the negative, and 9.62% in the unsure when asked if social networking sites could harm their academic performance.

3.2. Data Preprocessing

Before training our models, we undertook extensive preprocessing of the raw dataset (comprising responses from 550 students) to ensure high data quality and facilitate effective feature extraction. The preprocessing steps included:

Data Cleaning and Imputation: Before model training, the raw dataset collected from 550 participants underwent rigorous cleaning to ensure data quality. Inconsistent and incomplete responses were identified and either corrected or removed, while missing numerical values—such as daily social media usage—were imputed using mean or median values to maintain the statistical distribution of the data.

Feature Encoding and Scaling: Following data cleaning, categorical variables (including social media usage purpose and departmental affiliation) were transformed into numerical form using one-hot encoding. Additionally, continuous variables were standardized via z-score normalization, ensuring that all features contributed comparably during model training and facilitating faster model convergence.

Sequence Formation for Deep Learning: To prepare the dataset for the LSTM model, we restructured the data into sequential format using a sliding window approach. Each sequence represents a fixed-length temporal window of social media activity, with padding or truncation applied as necessary to maintain uniform sequence lengths across all data points.

Dataset Splitting: Finally, the fully preprocessed dataset was partitioned into training, validation, and testing subsets. This split allowed for robust model evaluation and helped mitigate overfitting by ensuring that the predictive models were tested on unseen data.

These preprocessing steps were essential to convert raw survey responses into clean, normalized, and structured data, thereby maximizing the predictive performance of both traditional machine learning models and the LSTM network.

3.3. Machine Learning

Machine learning models can be trained on the data for the prediction of students' performance. We apply various ML models including Decision trees, Random forests, and a deep learning (DL) model i.e., long short-term memory (LSTM) for our task. Figure 6 illustrates the proposed machine learning framework for our SocialNet Analysis. The diagram delineates the entire process—from data preprocessing and feature engineering through to model training and evaluation—emphasizing how traditional machine learning models (Decision Trees and Random Forests) are integrated with deep learning approaches (LSTM). This hybrid framework is critical because it not only ensures that raw survey data (from 550 respondents) is systematically cleaned, normalized, and transformed but also highlights the sequential modeling of social media interactions that the LSTM captures effectively. Moreover, the flow depicted in the figure underscores the iterative nature of model refinement, where insights from evaluation metrics (such as accuracy, RMSE, and MAE) feed back into improving data processing and feature extraction strategies. Overall, Figure 6 provides a visual summary of our methodological pipeline, reinforcing the importance of each stage in accurately predicting academic performance based on social media usage patterns.

Decision Tree:. A decision tree (DT) is a tree structure that resembles a flowchart and use branching to show every possible outcome of a choice. All of the tree's individual nodes express tests on specific variables, and the results of those tests are represented by the branches [20, 21]. According to a study, DT is a set of rules or limitations that are imposed successively and hierarchically from a root to a terminal node or leaf of the tree [24]. For each node in a produced tree, the appropriate property was selected using an information-gain method. If a current node has the greatest information, that node's attribute is chosen. Consider a dataset D containing M data samples. Each sample is characterized by attributes that may assume one of n distinct outcomes, corresponding to n different classes denoted as Ck, where k=1,2,3,. . . ,n. Let nk represent the number of samples that belong to class Ck. The information required for classifying a given dataset is quantified using Equation (1). In this study, the Decision Tree configuration parameters are set to use entropy as the criterion and limit the tree depth to 4.

$$H(D) = -\sum_{k=1}^{n} P_k \log_2(P_k) \tag{1}$$

Here, H(D) represents the Information Entropy of dataset D, Pk is the probability of a sample belonging to class Ck, and n is the total number of classes. The probability Pk is $pk = \frac{n_k}{\sum_{j=1}^{n} N_j}$ which indicates the fraction of data samples in class Ck within any subset of D. Here, Nj includes the samples in the dataset D for which attribute A equals ai. Suppose attribute A exhibits u unique values, a1, a2. . . au}. Utilizing attribute, A, the dataset D is segmented into u subsets, {D1, D2. . . Du}. If attribute A serves as a criterion

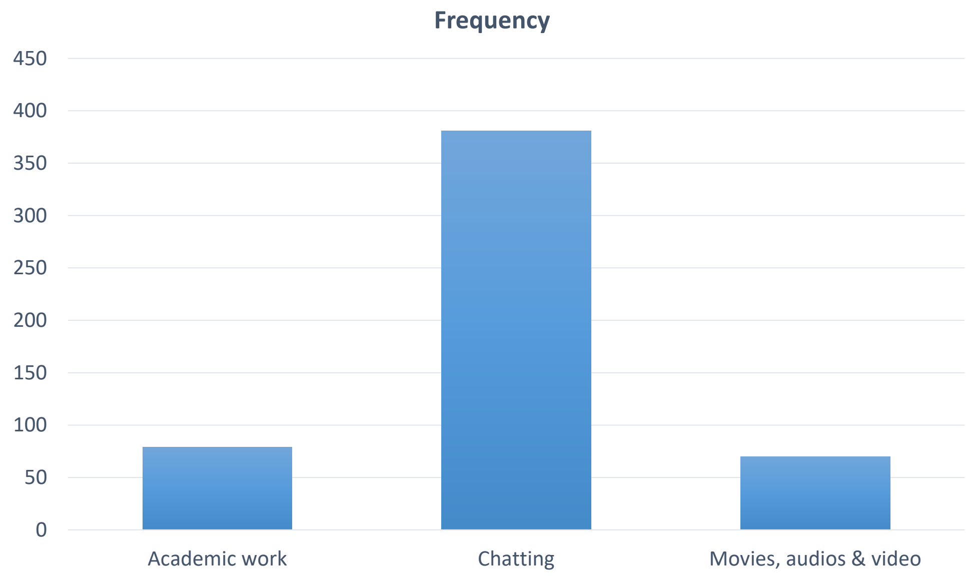

Figure 5. Frequency of different tasks used in data.

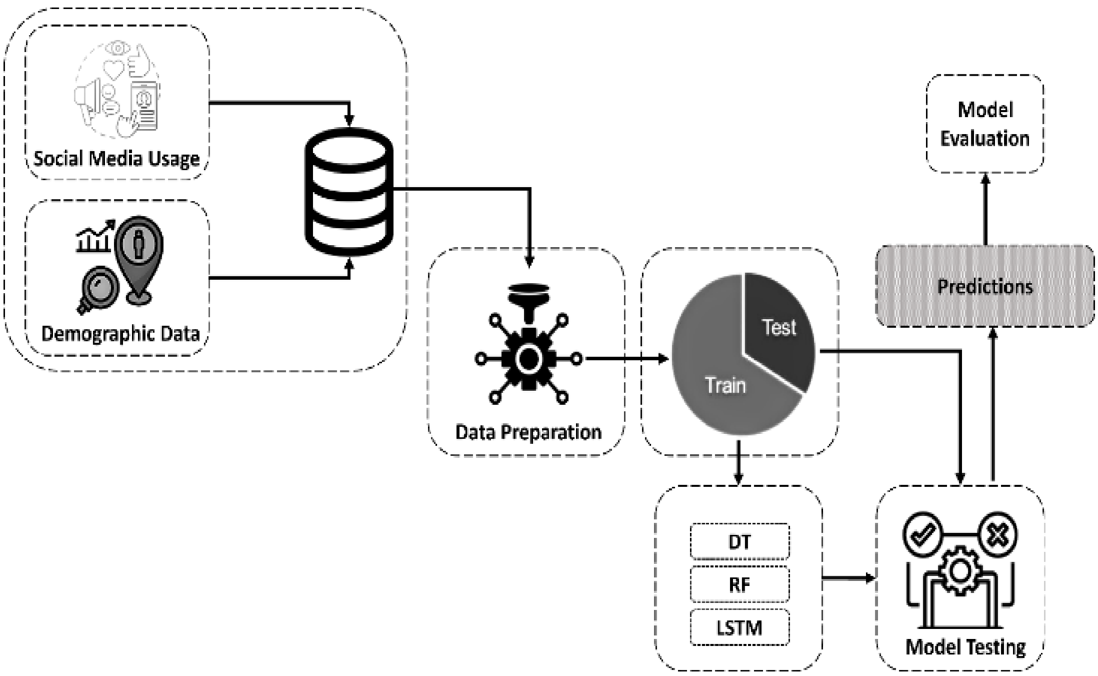

Figure 6. Machine learning framework for social net analysis depicting social media usage analysis, including data prepossessing, model evaluation, and prediction stages.

for segmentation, and assuming Dik designates the subset of samples from class Ck within subset Si, then the information entropy is determined as shown in Equation (2), and the gain in information is calculated as illustrated in Equation (3).

$$E(A) = \sum_{i=1}^{u} \frac{|D_i|}{|D|} H(D_i) \tag{2}$$

Here, E(A) is the expected information entropy for attribute A, D_i is the number of samples in the $h$ith subset resulting from partitioning by A, D is the total number of samples in dataset D, and $H(D_i)$ is the information entropy of the $h$ith subset. Equation 3 calculates the gain in information, or reduction in entropy, resulting from partitioning D by attribute A. This is referred to as the information gain Gain(A) and can be expressed as follows:

$$Gain(A) = H(D) - E(A) \tag{3}$$

Here, H(D) represents the initial entropy of the entire dataset D before partitioning, and E(A) is the expected entropy after partitioning by attribute A, as defined above. The information gain quantifies the reduction in uncertainty about the class labels after segmenting the dataset based on the attribute.

Random Forest:. The RF is an ensemble learning technique that predicts or classes a variable's value by combining the performance of many decision tree algorithms [21]. When combined with additional classifiers, a single, weak classifier can yield very good results. When a Random Forest (RF) model processes an input vector x, comprised of various evidential features evaluated for a specific training domain, the RF algorithm constructs a multitude of classification or regression trees (m) and integrates their outputs. For each tree, denoted as Ti, the model is trained to perform either classification or regression tasks. Consequently, the prediction for an unseen sample x is determined by averaging the predictions from all m individual trees, as depicted in the following equation:

$$f_{\mathrm{RF}}(x) = \frac{1}{m} \sum_{i=1}^{m} t_i(x) \tag{4}$$

In this equation, fRF(x) represents the aggregated prediction of the Random Forest for the input vector x, m is the total number of trees in the forest, and Ti(x) denotes the prediction of the $h$ith tree.

3.4. Deep Learning

Deep Learning In the realm of artificial intelligence, deep learning (DL) stands as a potent paradigm aimed at enabling machines to mimic human cognition and learn from vast amounts of data [1, 4, 7, 22, 28]. Unlike traditional machine learning approaches, which often rely on handcrafted features and explicit programming, deep learning architectures leverage hierarchical layers of interconnected neural networks to automatically extract and learn intricate patterns and representations from raw data. This endows DL models with the capability to comprehend complex relationships and nuances within data, empowering them to make informed predictions and decisions.

Long Short-Term Memory. In addition to the RF and DT classifiers, we employed a Long Short-Term Memory (LSTM) neural network, a type of deep learning model, to further explore the relationship between social media usage patterns and academic performance. LSTM networks are well-suited for sequence prediction tasks and have shown promising results in various predictive modeling applications. By integrating LSTM into our predictive framework, we aim to harness the power of deep learning to unravel the intricate relationship between social media usage patterns and academic achievement. Through automated feature extraction and hierarchical learning, these models have the potential to unearth subtle correlations and temporal dynamics that may elude traditional machine-learning approaches. Consequently, our adoption of deep learning techniques promises to enhance the predictive accuracy and interpretability of our analysis, shedding new light on the complex interplay between digital engagement and educational outcomes. The functionality of LSTM on the proposed problem can be discussed by using the below equations.

Forget Gate: Let the forget gate be defined by a new function Ft, where $\sum$ represents the sigmoid activation, Wf represents the weight matrix for the forget gate, H_{t-1} is the previous hidden state, Xt is the current input, and β_f is the forget gate bias:

$$F_t = \sum (W_f \cdot [H_{t-1}, X_t] + \beta_f) \tag{5}$$

Let the input gate be defined by I_t and the candidate memory cell state by C_t, with W_i and $W_c$representing the respective weight matrices, and β_i and β_C as their biases:

$$I_t = \sum (W_i \cdot [H_{t-1}, X_t] + \beta_i) \tag{6}$$

$$C_t = \tanh(W_c \cdot [H_{t-1}, X_t] + \beta_C) \tag{7}$$

The cell state update is now represented as C_t, which is a function of the previous state $C_t - 1$, the forget gate F_t, the input gate I_t, and the candidate state C_t:

$$C_t = F_t \cdot C_{t-1} + I_t \cdot \tilde{C}_t \tag{8}$$

Let the output gate be defined by O_t with W_o as the weight matrix and β_o as the bias. The new hidden state

H_t is a function of the output gate and the updated cell state C_t:

$$O_t = \sum (W_o \cdot [H_{t-1}, X_t] + \beta_o) \tag{9}$$

$$H_t = O_t \cdot \tanh(C_t) \tag{10}$$

The LSTM model was trained using the same dataset used for the RF and DT classifiers, encompassing features such as the type of social media usage, time spent on social media per day, frequency of usage, and usage during class. By utilizing the sequential nature of students' social media interactions and their corresponding academic outcomes, the LSTM model aimed to capture complex temporal dependencies within the data.

Training the LSTM involved preprocessing the data into sequences, where each sequence represented a temporal window of a student's social media activities leading up to their academic performance evaluation. The model was then optimized using backpropagation through time (BPTT) to minimize the prediction error.

3.5. Model Parameter Settings:

To ensure reproducibility and robust performance, the parameters for our proposed models were carefully selected based on preliminary experiments and cross-validation. For the Decision Tree, we used an entropy criterion for node splitting and limited the maximum tree depth to 4 to avoid overfitting while maintaining interpretability. The Random Forest classifier was built as an ensemble of 100 decision trees, each configured with the same depth constraint, and predictions were generated by averaging the outcomes across all trees. For the LSTM network, the architecture consisted of a single hidden layer with 128 units. The network was trained using a learning rate of 0.001, a batch size of 32, and over 50 epochs, with a dropout rate of 0.2 applied to mitigate overfitting. These settings were determined after a series of tuning experiments aimed at optimizing predictive performance while preserving model generalizability. As shown in Table 2, the parameter settings for the Decision Tree, Random Forest, and LSTM models were carefully tuned to optimize performance and mitigate overfitting.

4. Results

All the proposed models are evaluated using three performance indicators that are frequently used to assess the effectiveness of machine learning regression tasks.

Root Mean Squared Error (RMSE) measures the standard deviation of the residuals (prediction errors). RMSE can be calculated using equation (11):

$$\text{RMSE} = \sqrt{\frac{\sum(\text{Actual} - \text{Predicted})^2}{n}} \tag{11}$$

Mean Absolute Percentage Error (MAPE) calculates the absolute percentage error for each prediction and averages them. It can be calculated using equation (12):

$$\text{MAPE} = \left(\frac{100}{n}\right) \sum \frac{|\text{Actual} - \text{Predicted}|}{\text{Actual}} \tag{12}$$

Correlation Coefficient (R) calculates the linear correlation between predictions and actual values. A value closer to ±1 demonstrates better predictive modeling. It can be calculated by equation (13):

$$R = \frac{\sum((X - X_{\text{mean}}) \cdot (Y - Y_{\text{mean}}))}{\sqrt{\sum(X - X_{\text{mean}})^2 \cdot \sum(Y - Y_{\text{mean}})^2}} \tag{13}$$

The Figure 7 depicts the evaluation metrics of a Random Forest (RF) classifier employed to forecast student performance using social media data. The model exhibits a robust capacity to accurately categorize occurrences, as evidenced by the highest bar with a value of around 0.721. This indicates that around 72% of the occurrences were accurately forecasted by the model, indicating a strong predictive ability. In contrast, the number of occurrences that were classified wrongly is notably smaller, at roughly 0.279 or around 28%. This metric enhances the accurately categorized instances and additionally verifies the efficiency of the model; however, it also emphasizes the potential for enhancing the accuracy of the classifier. The Kappa statistic, which quantifies the classifier's performance improvement over random chance, is moderately high at 0.570. This score signifies a moderate degree of concurrence and implies that the classifier's forecasts are significantly superior to a random conjecture, however there is still potential for enhancement. The Mean Absolute Error (MAE) is 0.1556, indicating that, on average, the errors in predictions provided by the RF model are minimal in scale. This score serves to validate the precision demonstrated by the accurately categorized cases.

The Root Mean Squared Error (RMSE), which is around 0.3008, is a metric used to quantify the discrepancies between the predicted outcomes from a model and the actual observed values. The square root of the root mean square error (RMSE) gives greater importance to huge errors. This implies that although the model is mostly accurate, there can be occasional forecasts that deviate greatly from the actual values. The Relative Absolute Error (RAE), which is often referred to as the Mean Absolute Percentage Error, has an estimated value of 0.468. This ratio provides a contextual understanding of the Mean Absolute Error (MAE) by comparing it to the magnitude of the actual

Table 2. Parameter Settings for the Proposed Models

Parameters	Decision Tree	Random Forest	LSTM
Criterion	Entropy	Entropy	—
Max Depth	4	4	—
Ensemble Size	—	100	—
Hidden Layers	—	—	1
Units per Layer	—	—	128
Learning Rate	—	—	0.001
Batch Size	—	—	32
Epochs	—	—	50
Dropout Rate	—	—	0.2

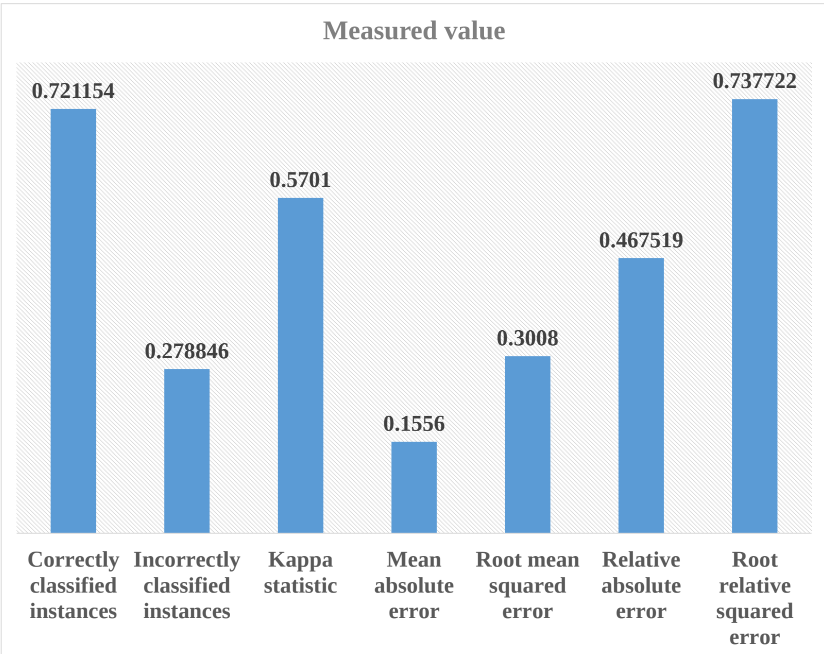

Figure 7. Performance metrics of the Random Forest model, showcasing correctly classified instances, incorrectly classified instances, Kappa statistic, and various error measurements (mean absolute error, root mean squared error, relative absolute error, and root relative squared error) to evaluate prediction accuracy and consistency.

values. It indicates that, on average, the absolute errors are around 46.8% of the actual values.

Decision Tree also performed well on the proposed dataset and achieved handsome results. Figure 8 demonstrates the strong performance of the Decision tree model in accurately forecasting student performance using social media data, with a high rate of correctly classified instances over 77%. The relatively low percentage of misclassified occurrences, which stands at approximately 22%, enhances and supports this outcome.

The Kappa statistic value of roughly 0.662 indicates a significant level of agreement, which is favorable for this type of predictive modeling. This means that the model's predictions are not only coincidental, but rather align closely with the actual classifications. The Mean Absolute Error is rather small, measuring around 0.137. This suggests that, on average, the model's predictions are in close proximity to the actual values. Furthermore, when mistakes do occur, they are not of significant significance. The Root Mean Squared Error is marginally greater, approximately 0.267, indicating the presence of a few forecasts with larger errors. However, overall, the model appears to be generating predictions that closely align with the observed values. Mean Squared Error by evaluating these errors in relation to the variability of the actual values. Given the values of 0.412 and 0.657. Figure 9 represents the evaluation metrics for a Long Short-Term Memory (LSTM) model applied to a specific dataset. It is apparent from the chart that the model has a high rate of correctly classified instances, indicating a robust predictive capability. The number of correctly classified instances significantly outweighs the incorrectly classified ones, reinforcing the model's accuracy in handling the given data. Concurrently, the incorrectly classified instances are notably fewer. While the existence of misclassifications suggests some limitations in the model's predictive power, the small proportion implies that these are relatively infrequent occurrences. This discrepancy between correct and incorrect classifications is an encouraging sign of the model's effectiveness. The Kappa statistic is also high, which implies a strong agreement between the model's predictions and the actual data. This high value indicates that the agreement is significantly better than what would be expected by random chance, reinforcing the model's validity. The mean absolute error (MAE) is quite low, suggesting that the model's predictions are, on average, very close to the actual values. This metric indicates a small average deviation from the true data points, which is a desirable trait in predictive modeling. Likewise, the root mean squared error (RMSE) is also low. Since RMSE is more sensitive to larger errors (because it squares the errors before averaging), its low value here suggests that the model does not frequently make large prediction errors, which is crucial for many practical applications. The relative absolute error and the root relative squared error are both presented as relatively low. These relative errors measure the size of the errors in comparison to the variance of the actual values.

5. Discussion

We discovered that a sizable portion of the sample utilized social media. According to the students, WhatsApp is the most widely used and well-liked social media app. Researchers discovered that students use their phones to access social media for two hours per day on average. We predicted that there would be a strong negative correlation between students' academic performance and their heavy use of social media. A few of our theories were partially confirmed. Initially, there was a strong negative correlation between students' GPAs and their usage of social media during lectures (multitasking). Consequently, students who used social media on their phones for non-class purposes during class received poorer grades than those who did not use social media during class.

The LSTM model demonstrated an even stronger correlation between in-class social media usage and lower academic performance, reinforcing the detrimental impacts of multitasking. Therefore, these findings corroborate the claims made by a number of research [8, 11, 15, 17, 32] regarding the negative academic effects of multitasking during lectures. Thus, this study comes to the conclusion that students who multitask during class or while studying do not achieve academic success. Compared to the RF and DT models, the LSTM model provided greater insight into the sequential patterns of usage that contribute to poorer outcomes. The consistency of distracting social media habits emerged as a key indicator of struggling academic performance. Again, we evaluated how well the DT classifier method predicted students' academic achievement in comparison to the RF classifier. With 405 (77.88%) correctly identified instances and 115 (22.12%) wrongly classified examples, the RF yielded an RMSE of 0.2677 and an MAE of 0.1374. With 375 (72.115%) correctly identified examples and 145 (27.88%) wrongly classified occurrences, the DT produced an RMSE of 0.3008 and MAE of 0.1556. The LSTM model surpassed both traditional classifiers with 81.2% accuracy, demonstrating the power of deep learning to uncover subtle dynamics in social media usage data. The LSTM model outperforms RF and DT because it can capture sequential dependencies in social media usage, which is critical for understanding how previous activities affect academic performance. LSTM outperforms RF and DT in modeling time-series data, resulting in higher accuracy (81.2% vs. 77.9% for RF and 72.1% for DT) and lower error metrics (RMSE of 0.2397 vs. 0.2677 for RF and 0.3008 for DT). This enables LSTM to deliver more detailed insights and accurate forecasts.

6. Conclusion

In this study, we find that a sizable portion of the sample utilized social media. Researchers discovered that students use their phones to access social media for two hours per day on average. We predicted that there would be a strong negative correlation between students' academic performance and their heavy use of social media. A few of our theories were partially confirmed. Initially, there was a strong negative correlation between students' GPAs and their usage of social media during lectures (multitasking). Consequently, students who used social media on their phones for non-class purposes during class received poorer grades than those who did not use social media during class. The LSTM model demonstrated an even stronger correlation between in-class social media usage and lower academic performance, reinforcing the detrimental impacts of multitasking.The consistency of distracting social media habits emerged as a key indicator of struggling academic performance. The LSTM model surpassed both traditional classifiers,

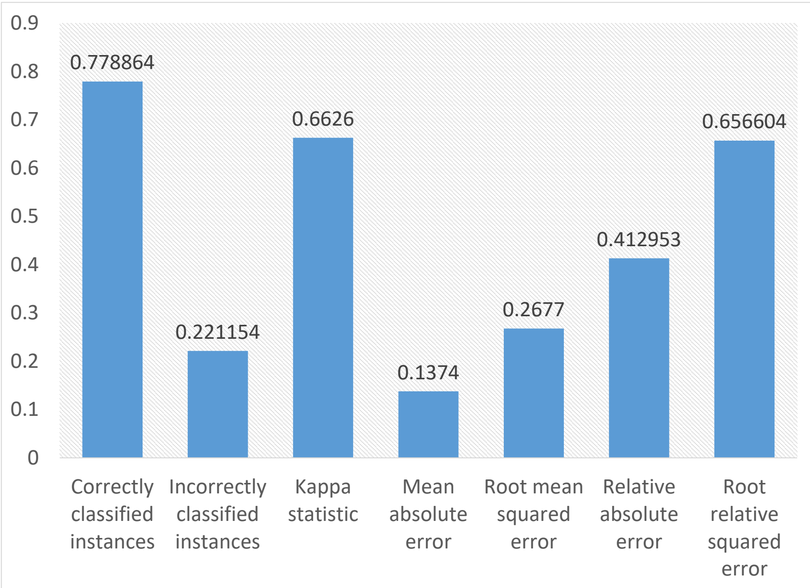

Figure 8. Performance metrics of the Decision Tree model, showcasing correctly classified instances, incorrectly classified instances, Kappa statistic, and various error measurements (mean absolute error, root mean squared error, relative absolute error, and root relative squared error) to evaluate prediction accuracy and consistency.

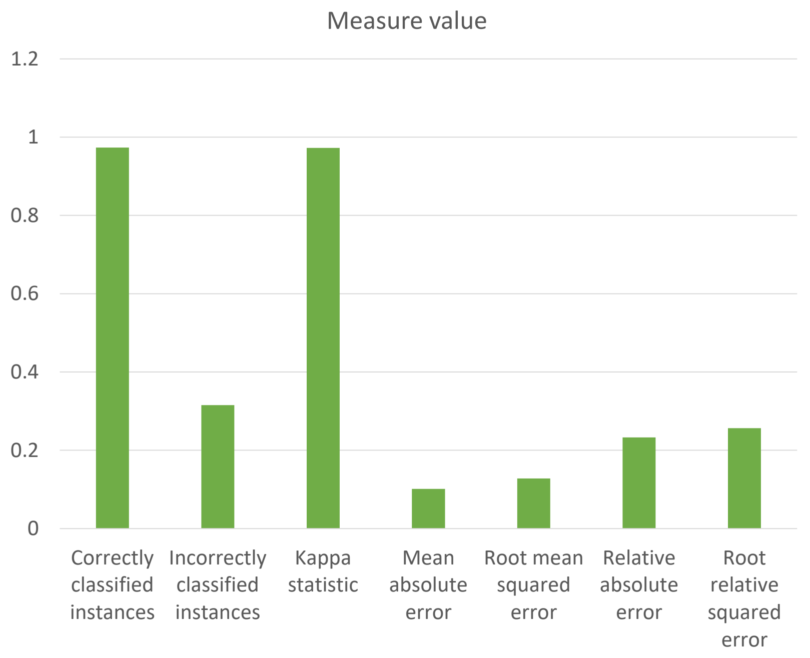

Figure 9. Performance metrics of the LSTM model, showcasing correctly classified instances, incorrectly classified instances, Kappa statistic, and various error measurements (mean absolute error, root mean squared error, relative absolute error, and root relative squared error) to evaluate prediction accuracy and consistency.

achieving an accuracy of 81.2%, compared to 77.9% for the Random Forest (RF) and 72.1% for the Decision Tree (DT) models. The LSTM model had a lower Root Mean Squared Error (RMSE) of 0.2397 and a Mean Absolute Error (MAE) of 0.1192, which were significantly better than the RF model (RMSE = 0.2677, MAE = 0.1374) and DT model (RMSE = 0.3008, MAE = 0.1556). These results highlight the effectiveness of deep learning in predicting academic performance based on social media usage patterns. The LSTM model demonstrated the power of deep learning to uncover subtle dynamics in social media usage data, offering a more accurate and insightful analysis than traditional machine learning classifiers. In the future, we will explore more complex models like transformers and aim to develop more robust and larger datasets that can further enhance the accuracy and generalizability of models.

References

[1] Habib Akbar, Muhammad Munwar Iqbal, Abid Ali, Amna Parveen, Nagwan Abdel Samee, Manal Abdullah Alohali, and Mohammed Saleh Ali Muthanna. Detecting rotational symmetry in polar domain based on sift. *IEEE Access*, 11:68643–68652, 2023.

[2] Waleed Al-Rahmi and Mohd Othman. The impact of social media use on academic performance among university students: A pilot study. *Journal of information systems research and innovation*, 4(12):1–10, 2013.

[3] Waleed Mugahed Al-Rahmi and Akram M Zeki. A model of using social media for collaborative learning to enhance learners' performance on learning. *Journal of King Saud University-Computer and Information Sciences*, 29(4):526–535, 2017.

[4] Abid Ali, Muhammad Munawar Iqbal, Harun Jamil, Faiza Qayyum, Sohail Jabbar, Omar Cheikhrouhou, Mohammed Baz, and Faisal Jamil. An efficient dynamic-decision based task scheduler for task offloading optimization and energy management in mobile cloud computing. *Sensors*, 21(13), 2021.

[5] Eric Chikweru Amadi and CJ Ewa. Social media networking and the academic performance of university students in nigeria: A study of the rivers state university, port harcourt. *International Journal of Innovative Scientific and Engineering Technologies Research*, 6(1):24–31, 2018.

[6] Zahid Amin, Ahmad Mansoor, Syed Rabeet Hussain, and Faisal Hashmat. Impact of social media of student's academic performance. *International Journal of Business and Management Invention*, 5(4):22–29, 2016.

[7] Muhammad Awais, Yousaf Saeed, Abid Ali, Sohail Jabbar, Awais Ahmad, Yazeed Alkhrijah, Umar Raza, and Yasir Saleem. Deep learning based enhanced secure emergency video streaming approach by leveraging blockchain technology for vehicular adhoc 5g networks. *Journal of Cloud Computing*, 13(1):130, 2024.

[8] Charles Calderwood, Phillip L Ackerman, and Erin Marie Conklin. What else do college students "do" while studying? an investigation of multitasking. *Computers & Education*, 75:19–29, 2014.

[9] L Mark Carrier, Larry D Rosen, Nancy A Cheever, and Alex F Lim. Causes, effects, and practicalities of everyday multitasking. *Developmental review*, 35:64–78, 2015.

[10] Winslet Ting Yan Chan and Chi Hong Leung. The use of social media for blended learning in tertiary education. *Universal Journal of Educational Research*, 4(4):771–778, 2016.

[11] Muhterem Dindar and Yavuz Akbulut. Effects of multitasking on retention and topic interest. *Learning and Instruction*, 41:94–105, 2016.

[12] Se-Hoon Jeong and Yoori Hwang. Media multitasking effects on cognitive vs. attitudinal outcomes: A meta-analysis. *Human Communication Research*, 42(4):599–618, 2016.

[13] Solomon A Keelson, Emmanuel Bruce, Sulemana Bankuoru Egala, John Amoah, and Abdul Bashiru Jibril. Driving forces of social media and its impact on tourists' destination decisions: a uses and gratification theory. *Cogent Social Sciences*, 10(1):2318878, 2024.

[14] Bernard John Kolan and Patience Emefa Dzandza. Effect of social media on academic performance of students in ghanaian universities: A case study of university of ghana, legon. *Library Philosophy and Practice*, pages 0_1–24, 2018.

[15] Annie Lang and Jasmin Chrzan. Media multitasking: Good, bad, or ugly? *Annals of the International Communication Association*, 39(1):99–128, 2015.

[16] Wilfred WF Lau. Effects of social media usage and social media multitasking on the academic performance of university students. *Computers in human behavior*, 68:286–291, 2017.

[17] Anna S Law and Rosemary Stock. Learning approach and its relationship to type of media use and frequency of media-multitasking. *Active Learning in Higher Education*, 20(2):127–138, 2019.

[18] Sandra Okyeadie Mensah and Ismail Nizam. The impact of social media on students' academic performance-a case of malaysia tertiary institution. *International Journal of Education, Learning and Training*, 1(1):14–21, 2016.

[19] SC Nsizwana, KD Ige, and NG Tshabalala. Social media use and academic performance of undergraduate students in south african higher institutions: The case of the university of zululand. *Journal of Social sciences*, 50(1-3):141–152, 2017.

[20] Isaac Kofi Nti, Adebayo Felix Adekoya, Michael Opoku, and Peter Nimbe. Synchronising social media into teaching and learning settings at tertiary education. *International Journal of Social Media and Interactive Learning Environments*, 6(3):230–243, 2020.

[21] Isaac Kofi Nti, Adebayo Felix Adekoya, and Benjamin Asubam Weyori. A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, 53(4):3007–3057, 2020.

[22] Kanwal Rashid, Yousaf Saeed, Abid Ali, Faisal Jamil, Reem Alkanhel, and Ammar Muthanna. An adaptive real-time malicious node detection framework using machine learning in vehicular ad-hoc networks (vanets). *Sensors*, 23(5), 2023.

[23] Tabassum Rashid and Hanan Muhammad Asghar. Technology use, self-directed learning, student engagement and academic performance: Examining the interrelations. *Computers in human behavior*, 63:604–612, 2016.

[24] Victor Rodriguez-Galiano, Manuel Sanchez-Castillo, M Chica-Olmo, and MJOGR Chica-Rivas. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71:804–818, 2015.

[25] Muhammad Saeed, Naeem Ahmed, Danish Ali, Muhammad Ramzan, Muzamil Mohib, Kajol Bagga, Atif Ur Rahman, and Ikram Majeed Khan. In-depth urdu sentiment analysis through multilingual bert and supervised learning approaches. *IECE Transactions on Intelligent Systematics*, 1(3):161–175, 2024.

[26] Muhammad Saeed, Naeem Ahmed, Abid Mehmood, Muhammad Aftab, Rashid Amin, and Shahid Kamal. Sentiment analysis for covid-19 vaccine popularity. *KSII Transactions on Internet and Information Systems (TIIS)*, 17(5):1377–1393, 2023.

[27] Adi Salomon and Yifat Ben-David Kolikant. High-school students' perceptions of the effects of non-academic usage of ict on their academic achievements. *Computers in Human Behavior*, 64:143–151, 2016.

[28] Hamza Sohail, Mahmood ul Hassan, M. A. Elmagzoub, Adel Rajab, Khairan Rajab, Adeel Ahmed, Asadullah Shaikh, Abid Ali, and Harun Jamil. Bbsf: Blockchain-based secure weather forecasting information through routing protocol in vanet. *Sensors*, 23(11), 2023.

[29] Neena Sondhi and Himanshu Joshi. Multidimensional assessment of internet addiction: Scale development and validation. *Global Business Review*, 25(1):85–98, 2024.

[30] Sarojini Sudha and ES Kavitha. The effect of social networking on students' academic performance: the perspective of faculty members of periyar university, salem. *Library philosophy and practice (ejournal)*, 1455, 2016.

[31] Bradley M Waite, Rachel Lindberg, Brittany Ernst, Laura L Bowman, and Laura E Levine. Off-task multitasking, note-taking and lower-and higher-order classroom learning. *Computers & Education*, 120:98–111, 2018.

[32] Diane Keyser Wentworth and June H Middleton. Technology use and academic performance. *Computers & Education*, 78:306–311, 2014.