International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

ANALYSIS OF SOCIAL MEDIA USAGE AMONG STUDENTS USING MACHINE LEARNING

[1]P.Maheswari
Department of statistics, Sri Durga Malleswara Siddhartha Mahila Kalasala, Vijayawada.

[2] K.Ramya Sree
Department of statistics, Sri Durga Malleswara Siddhartha Mahila Kalasala, Vijayawada.

ABSTRACT

Social media platforms have become an integral part of students' daily lives, influencing communication, academic performance, behavior, and mental wellbeing. This study presents a machine learning–based analytical framework to examine social media usage among students. The proposed system analyzes behavioral data such as time spent, platform preference, activity frequency, and academic performance indicators. Classification algorithms are used to categorize usage patterns and predict their impact on academic outcomes. Performance is evaluated using accuracy, precision, recall, and F1-score. Results show machine learning can effectively analyze student behavior and identify usage patterns associated with academic outcomes.

Keywords

Social Media Analysis, Student Behavior, Machine Learning, Classification Models, Predictive Analytics, Educational Data Mining

1. INTRODUCTION

Social media has transformed communication and information sharing in modern society. Students extensively use platforms for networking, learning, entertainment, and collaboration. While social media provides educational benefits such as access to knowledge and peer learning, excessive or unstructured usage may negatively affect academic performance and productivity.

Analyzing social media usage manually is difficult due to the large volume of behavioral data generated daily. Machine learning techniques enable automated analysis by identifying patterns and relationships within datasets. By applying predictive models, it is possible to classify students based on usage behavior and evaluate its impact on academic outcomes. This research aims to develop a machine learning system that analyzes social media usage among students and predicts its influence using statistical performance metrics.

2. EXISTING SYSTEM

Traditional approaches to studying social media usage among students include surveys, interviews, and observational studies. These methods have several limitations:

- Time-consuming data collection
- Small sample sizes
- Subjective responses
- Limited scalability

- Difficulty identifying hidden patterns

Because of these drawbacks, conventional methods are insufficient for large-scale behavioral analysis.

3. PROPOSED SYSTEM

The proposed system utilizes machine learning classification algorithms to analyze student social media usage and predict behavioral or academic outcomes.

System Objectives

- Analyze student usage patterns automatically
- Identify correlations between usage and academic performance
- Classify students into usage categories (low, moderate, excessive)
- Provide predictive insights

Advantages

- Automated analysis
- High scalability
- Reduced human bias
- Improved prediction accuracy

4. METHODOLOGY

A. Data Collection

Data is collected from student surveys, academic records, and digital usage logs. Features include:

- Daily social media usage time
- Platform preference
- Frequency of login
- Purpose of usage (study/entertainment)
- GPA or academic score
- Sleep hours

B. Data Preprocessing

- Handling missing values
- Removing inconsistencies
- Encoding categorical variables
- Feature normalization

C. Feature Selection

Relevant attributes are selected using correlation analysis and statistical importance ranking.

D. Model Training

Dataset is split into training (80%) and testing (20%) sets. Models are trained on labeled data.

E. Prediction

The trained classifier predicts whether social media usage has positive, neutral, or negative academic impact.

F. Evaluation

Performance is evaluated using classification metrics.

5. ALGORITHMS USED

1. Logistic Regression

Used for binary or multi-class classification of usage impact.

2. Decision Tree

Builds a tree-structured model for predicting student behavior.

3. Random Forest

Combines multiple decision trees to improve accuracy and reduce overfitting.

4. Support Vector Machine (SVM)

Separates classes using optimal hyperplanes.

5. Naive Bayes

Applies probabilistic classification based on Bayes theorem.

6. Performance Metrics

Accuracy = (TP + TN) / (TP + TN + FP + FN

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

F1 Score = 2 × Precision × Recall / (Precision + Recall)

Sample:

TP=88 TN=92 FP=12 FN=8

Accuracy=0.90

Precision=0.88

Recall=0.917

F1=0.898

7. RESULTS AND DISCUSSION

Experimental analysis shows that ensemble models such as Random Forest provide the highest classification accuracy. SVM performs well for balanced datasets, while Naïve Bayes is efficient for smaller datasets. Decision Trees offer interpretability but may overfit when data is noisy.

Model performance improves when:

- Dataset size increases
- Irrelevant features are removed
- Data is properly normalized

The results demonstrate that machine learning can effectively identify behavioral patterns in student social media usage.

8. CONCLUSION

This study presented a machine learning–based framework for analyzing social media usage among students. The proposed system successfully classifies usage behavior and predicts its academic impact using multiple algorithms. Performance evaluation metrics confirm the effectiveness of the model. Compared to traditional survey-based approaches, the proposed system provides automated, scalable, and accurate analysis. Future enhancements may include real-time monitoring systems, mobile applications, and deep learning models for behavioral prediction.

Future Scope

- Integration with real-time usage tracking tools
- Psychological impact prediction
- Personalized academic recommendations
- Early detection of excessive usage behavior

References

[1] T. Mitchell, Machine Learning, McGraw-Hill, 1997.

[2] C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[3] I. Witten et al., Data Mining, Morgan Kaufmann, 2011.

[4] L. Breiman, Random Forests, 2001.

[5] N. Cristianini, Support Vector Machines, Cambridge, 2000.

[6] A. Géron, Hands-On Machine Learning, O'Reilly, 2019.