Survey on Machine Learning Approaches to Study the Influence of Social Media on Academic Performance of Students

Kanupriya Guha[1], Sanjana M C[2], Maria Suman A[3]

[1]*Department. of Computer Science, Mount Carmel College, Autonomous, Bengaluru, India*
[2]*Department. of Computer Science, Mount Carmel College, Autonomous, Bengaluru, India*
[3]*Department. of Computer Science, Mount Carmel College, Autonomous, Bengaluru, India*

ABSTRACT

The rise in usage of social platforms among the students, both outside and within lecture halls, has sparked debate about its impact on academic achievement. These platforms provide an equal opportunity and threat to the student community. There is an upsurge in the frequency of usage of social media which has a direct impact on the learning and performance of the students. As the majority of students are drawn towards social media and prioritize it over academic activities which results in a decline in their performance. In addition, the prolonged usage of social platforms has a negative impact on daily activities Thus study aims to understand the usage pattern of social media and its influence on the student community through various machine learning and statistical models.

Keywords: Machine Learning, Supervised Learning, Unsupervised Learning

1. INTRODUCTION

Social media is the websites and applications which facilitate the thought of making and sharing thoughts, information, interests, and other varieties of expression through the building of virtual networks and communities. It's an inextricable aspect of everyday life. It allows users to communicate and connect with others without having to satisfy head-to-head. People use social media to remain involved with family and friends, and various communities.

Social media has a clear and direct purpose for businesses that sell a product or service and are searching for ways to advertise their brands. But the most obvious use of social media is done by teenagers or young adults who are either in schools or colleges or have just joined the workforce. It is prominently used by students as they can communicate with their friends.

The number of social media users in India as of 2020 is 518 million and it is expected to rise to 1.5 billion by 2040**.** According to polls, 90% of 13 to 17-year-olds have used social media. Seventy-five percent say they have at least one active social media account, with 51 percent saying they check it at least once a day. According to a survey by Nasir, Khatoon, & Bharadwaj, in the year 2018, the majority of social media users in India are college students, with 33% of girls and 27% of boys. This report also predicted that India will have 370.77 million internet users in 2022. The use of social media has impacted students more as it has both negative and positive effects on them.

But the most impact of social media is on the student's academic performance, as it is the dimension of a student's achievement across various academic subjects and is assessed based on the scores or grades achieved in various subjects. Over the years various studies have found that students who spend more time on social media are likely to exhibit poor academic performance.

Using the prediction technique with the help of the parameters like academic performance, frequency of use of social media, duration of usage of social media, number of social media enrolled, the purpose of using social media, concentration and writing ability, the level of addiction can be detected and, in the person, can be accessed by the level of addition from which the relationship between social media and academic performance can be derived.

2. LITERATURE REVIEW

Basil C.E. Oguguo1 et. al. [1] Here, the influence of using social platforms on senior high school students' academic achievement was probed in this study. By a multi-stage sampling technique, the study gathered a total of 150 students from five schools which included 70 males and 80 females. Data collection was done using the Social Media Questionnaire (SMQ) and the Students' Accounting Achievement Proforma (SAAP). As per the findings, students commonly use social platforms to meet new acquaintances, find out more about their projects, find other educational materials and keep up with the newest trends and news. The Cronbach Alpha technique was used to calculate the internal consistency dependability of the instruments based on the gathered data. The study questions were answered using mean and standard deviations, and the hypotheses were tested using a t-test at the 0.05 level of significance. Based on the technique following was found, on average 2 to 4 hours were spent on social media by students every day and there was no notable impact of students' frequency of social media use on their mean academic results in accounting, but students' gender had a remarkable influence on their mean academic achievement in accounting.

Mrs. Balakamakshi T Y et.al. [2] This research focuses on the impact of E-learning on students' academic learning outcomes at the college level. The study's goal was to determine the consequences of E-learning on student interest and learning. From various colleges of Chennai, 250 female students' responses were collected using convenience sampling technique, percentage analysis, and the Chi-square test were used to analyze the same. According to the report E-learning, gives students more time flexibility. Based on findings, it was also found out that e-learning allows students to better manage their time and motivates them to learn independently.

Isaac kofi Nti et.al. [3] This study looked at the relationship between students' academic performance and their use rate, nature of use, usage in class, and level of exposure to a social networking site. A convenient sampling technique was used to sample 550 people, with an effort made to address the literature's varied perspectives on social media use and academic performance. They used a predictive framework based on Decision Tree (DT) and Random Forest (RF) machine learning algorithms to forecast the student's GPA and found that students' academic performance is influenced in part by their use of social networking sites and their use in class.

Kiran Shahzadi et. al. [4] investigated the ways in which social media sites influence students' academic performance and assess the impact of social media on the educational system. They have used student-por.csv and student-mat.csv datasets in this paper, and they worked on a classifier to predict the impact of social media on students' academic performance. They used K nearest-neighbor (KNN), support vector machine (SVM), and Linear regression technique in Python for this purpose since they forecast somewhat better and correctly about the student's academic performance. The findings show that there is a strong link between social media and academic performance among students. This article demonstrates that the majority of students have cell phones and internet access, therefore they use 40 minutes for 4 hours daily and do not pay attention to their studies, all of these online behaviors have an impact on a student's academic achievement.

Benedicta Krampah Akoto Bitherman et.al. [5] determined the influence of social media usage on undergraduate students at Ghana Communication Technology University (GCTU) in Accra's academic performance. For this investigation, the survey method was used. The goals of this study were to determine the purpose of visits to various social media platforms, analyze the average time spent on these platforms, determine the consequences of these platforms on students' academic performance, and

investigate relevant issues in platform utilization. For the study, 800 pupils were chosen as a sample. For sample selection, a random sampling procedure was used. The data were analyzed using the Statistical Package for the Social Sciences (SPSS version 20). Based on the data, it was found that, despite public perceptions about teenage usage of social media, the majority of students were interested in using social media for academic objectives.

Xin Su et. Al. [6] The purpose of this study was to investigate the links between student social media use, engagement, and academic achievement, with the premise that student involvement would mediate the relationship between social media use and academic performance. Using survey data from 307 Chinese college students, it was discovered that student engagement played a mediating role in the link between social media use and academic performance. Moreover, the findings show that the use of social media by college students has little impact on their academic performance. The findings add to our knowledge of the relation between social media use, student engagement, and academic success in Chinese higher education.

Jeanna Mastrodicasa, [7] stated that both in negative and positive ways the social media has impacted college students in an exceedingly different kind of ways. Knowing the quality of the time spent on social platforms and also the activities college students engage in is important for higher education administration. Researchers have begun to study the consequences on students' well-being empirically, and preliminary findings suggest that more research is needed.
There is no one conclusion to draw from the research about social media sites and their impact on college students. Some researchers have found it to have a positive influence on activities, while others have found the opposite of it, this has made a bit of chaos for student affairs organizations to understand from which to determine the results.

Agwi Uche Celestine et al. [8] This research studies the impact of social media sites on student academic performance at Samuel Adegboyega University. The study was led by four research questions and three hypotheses. The study adopted a descriptive survey design. The population used as a sample were students from Samuel Adegboyega University in Edo state. A validated questionnaire with 0.96 reliability was used for data collection. Frequency counts, percentage, and mean were the quantitative statistics used. To determine the hypothesis's significance, inferential statistics were used. The F-statistic was used to test whether the hypotheses were significant at 0.05 using the Analysis of Variance (ANOVA) approach. The findings showed that there is a relation between time spent on social media sites and academic work. It was also discovered that the nature of the student's social media activity had no considerable impact on the student's academic achievement. Furthermore, the study reveals that a student's gender has no bearing on how they use and interact with social media.

Wilfred W.F. Lau, [9] The motive of this study was to see if and how the two behaviors, social media use and social media multitasking affect academic achievement in university students Multitasking on social media has also become more common. But little is known about how these to behaviors influence university students' academic performance. It was found that usage of social platforms for academic activities didn't predict much of academic performance as measured by cumulative grade point average in a sample of 348 undergraduate students at a comprehensive university in Hong Kong, but usage of the social platforms for non-academic purposes and also social media multitasking was affecting student's academic performance negatively was predicted.

Xing Xu et al. [10] states that access to the Internet is becoming increasingly convenient for students, which has a civilizing effect on their studies and living. This research aims to reveal the link between Internet usage and academic success, as well as to use machine learning to predict student academic achievement based on consumption statistics. From the genuine Internet usage data of 4000 students, a set of features, including online duration, Internet traffic volume, and connection frequency, were extracted, computed, and normalized. These features were utilized to predict academic achievement using 3 common machine learning algorithms: decision tree, neural network, and support vector machine. The findings show that behavioral discipline is critical to academic performance. Academic performance is positively correlated with Internet connection frequency factors, whereas academic

Address for correspondence:

Mail Id: knupriya.guha@gmail.com , mcsanjana2001@gmail.com , mariasuma@mccblr.edu.in

performance is inversely correlated with Internet traffic volume features. In terms of online time features, the amount of time spent on the internet causes unanticipated differences in performance amongst datasets. Furthermore, as the number of features increases, the algorithms' prediction accuracy improves. The findings suggest that Internet usage data can be used to distinguish and predict academic success in students.

3. GENERAL METHOD
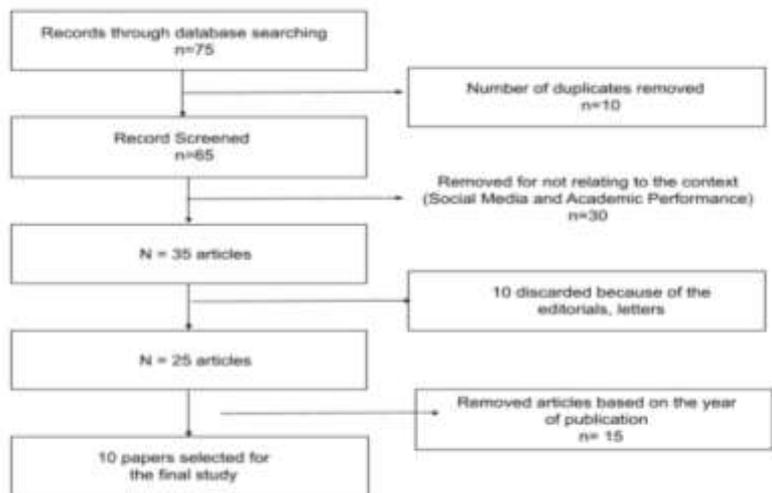
3.1 DATABASE SOURCE

The articles were collected from various platforms like Google Scholar, Research gate, and Scrip.

3.2 METHODOLOGY

"*Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA model) is used in filtering the articles. It is a minimum set of items for systematic reviews and meta-analyses that is based on evidence. The flow diagram demonstrates how information flows through the various stages of a Systematic Review. It shows the number of records found, how many were included and how many were excluded, as well as the reasons for exclusions.*"

Initially, articles were gathered from various database sources. Few of the articles were discarded as out of context, to avoid redundancy, and to consider the papers after 2010. We identified 75 articles and removed 10 articles to avoid duplication. 65 articles were screened out of which 30 were discarded as they did not have the complete text and were not matching with the theme/focus of our study, that is, social media and academic performance. And also 10 editorials and letters were omitted. In the remaining 25 papers, based on the year of publications the articles were eliminated, resulting in a final selection of 10 papers for the study.

4. APPROACHES IN MACHINE LEARNING

"*Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed*".

It is a type of artificial intelligence that allows software applications to become more accurate at predicting outcomes. For any child to learn to recognize objects/persons, no procedure is explained to that child to introduce the properties of the objects/persons in question and then decide what they are. This child is only shown more than one instance of these objects/persons. Then the human brain automatically starts to identify these features over time (consciously) and learns to recognize objects/persons. The machine learning model does the same.

There are four different types of machine learning: -

Supervised Learning - Supervised learning problems can be grouped into regression (output=numeric) and classification (output=object) problems depending on the output. It is a type in which a model is trained on a labelled data set. That is, it has both input and output parameters.
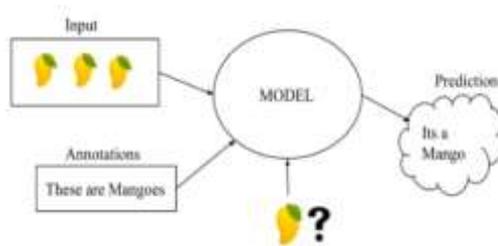

Figure1. Diagrammatic representation of supervised learning

We train an Artificial Intelligence to expect output by providing it an input when we train it in supervised learning. If the output produced by Artificial Intelligence differs from the output we provide, it will recalculate. This technique is continued with the data set until the Artificial Intelligence error rate is as low as possible.

Cancer detection is an example of supervised learning. Using past data, it learns to forecast cancer. Inputs (results from cancer testing) and outputs are included in this training data (cancer or not cancer).

Some of the essential supervised learning algorithms are:
- K-Nearest Neighbors
- Naive Bayes
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVM)
- Decision Trees-Random Forests
- Neural Networks

Unsupervised learning problems can be grouped as clustering and association problems. Unlike supervised learning in unsupervised learning, education data are unlabeled. The system tries to learn without any instructors. It is a type of machine learning that transforms raw data into organized/meaningful data.
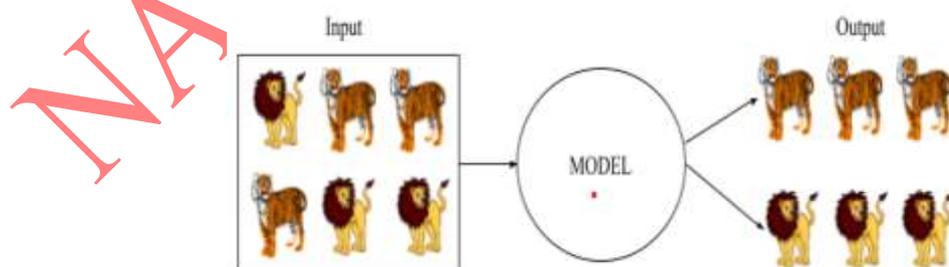

Figure2. Diagrammatic representation of unsupervised learning

The computer is taught using data that is fully unlabeled. Its primary applications are descriptive modelling and pattern detection. We have the input data and no corresponding output variables in this learning style.

Unsupervised learning aims to model the structure or distribution that underpins data about which we have no prior knowledge. Unsupervised learning is distinguished from supervised learning by the absence of proper outputs and instructors. When we don't know what to look for in the data, these algorithms are generally useful. The model learns by observing the data and identifying patterns. When we provide the model a data set, it creates clusters to detect patterns and correlations between variables. It cannot, however, add labels to the cluster.

Semi-supervised learning is a machine learning algorithm that falls between supervised and unsupervised learning. It uses a combination of labelled and unlabeled datasets throughout the training phase and constitutes a middle ground between supervised and unsupervised learning methods.
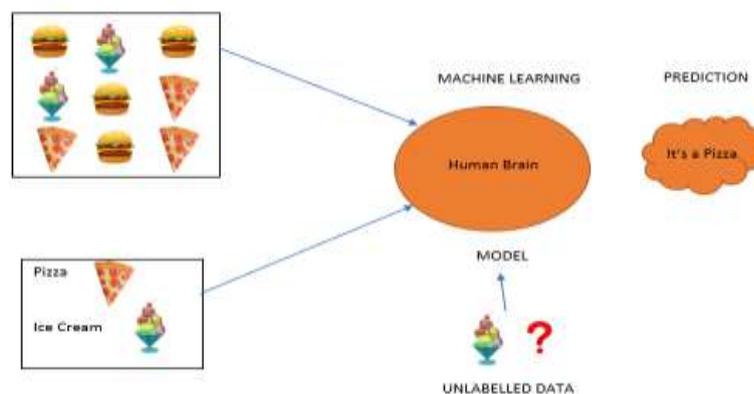

Figure3. Diagrammatic representation of semi-supervised learning

The notion of semi-supervised learning is introduced to solve the shortcomings of supervised and unsupervised learning methods. Semi-supervised learning's fundamental goal is to make good use of all accessible data rather than just labelled data, like supervised learning does. Similar data is first clustered using an unsupervised learning technique, which then aids in the labelling of unlabeled data into labelled data. It's because acquiring tagged data is more expensive than acquiring unlabeled data.

Reinforcement learning is a feedback-based process in which an AI agent (a software component) explores its surroundings automatically by striking and trailing, taking action, learning from its experiences, and improving its performance. The purpose of a reinforcement learning agent is to maximize the rewards for each good behavior and to minimize the punishments for each negative activity

There is no labelled data in reinforcement learning, unlike supervised learning, and agents learn solely from their experiences.

The reinforcement learning process is comparable to that of a human being; for example, a youngster learns different things through his daily encounters. Playing a game in which, the environment is the game, the motions of an agent at each step define states, and the agent's goal is to acquire a high score is an example of reinforcement learning. Agent receives feedback in the form of incentives and punishments.

4.1 MACHINE LEARNING ALGORITHMS

4.1.1 RANDOM FOREST

Random Forest is a machine learning algorithm that uses the supervised learning method. It can be used for both classification and regression issues. *"It is based on ensemble learning, which is the process of integrating numerous classifiers to solve a complex problem and improve the model's*

performance". The larger the number of trees in the forest, more precise it is, and the problem of overfitting is avoided.
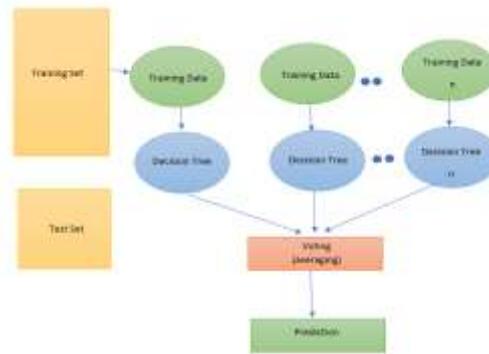

Figure5. Diagrammatic representation of Random Forest

4.1.2   SVM

SVMs are the most widely used classification algorithm in machine learning. Their mathematical underpinning is crucial in laying the groundwork for the geometrical division between the two classes. We'll look at how support vector machines function by looking at how they're implemented in Python, and then we'll look at some of the most relevant applications.
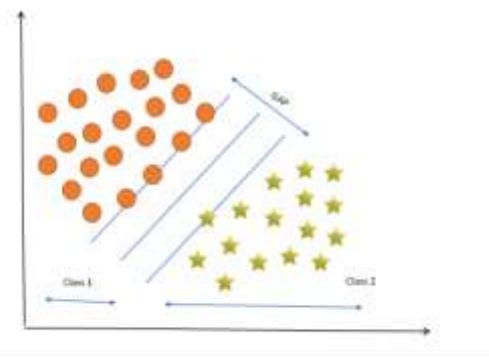

Figure 6. Diagrammatic representation of SVM

4.1.3   DECISION TREE

"*A decision tree is a flowchart-like structure in which each internal node represents a feature test (e.g., whether a coin flip will land heads or tails), each leaf node represents a class label (decision made after computing all features), and branches represent feature combinations that lead to those class labels.*" The categorization rules are represented by the pathways from root to leaf. The diagram below depicts the fundamental flow of a decision tree with labels (Rain (Yes), No Rain (No)).
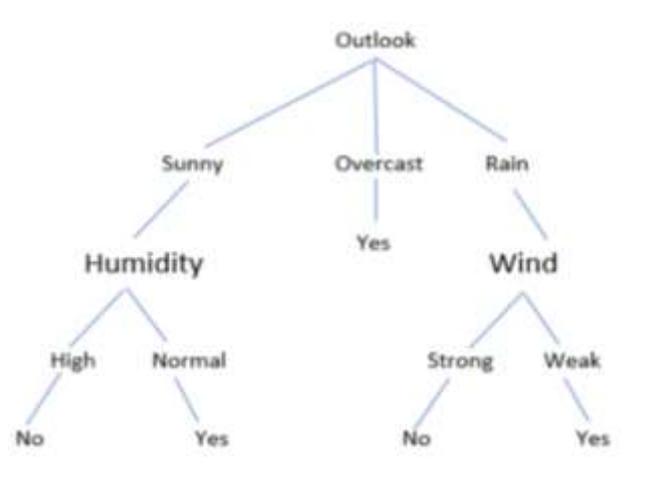
Kanupriya Guha[1], Sanjana M C[2], Maria Suman A[3] "Survey Literature on Machine Learning Approaches to Study the Influence of Social Media on Academic Performance"

Figure6. Diagrammatic representation of Decision Tree

REFERENCES

[1] Oguguo, Basil Chinecherem, et al. "Influence of Social Media on Students' Academic Achievement." *International Journal of Evaluation and Research in Education (IJERE)*, vol. 9, no. 4, 1 Dec. 2020, p. 1000, files.eric.ed.gov/fulltext/EJ1274777.pdf, 10.11591/ijere.v9i4.20638.

[2] Balakamakshi T Y, DR R Savithri, "Effect of e– learning on student's academic performance at college level" -- Palarch's Journal Of Archaeology Of Egypt/Egyptology 18(1), ISSN 1567-214x

[3] Isaac Kofi Nti, et al. "Prediction of Social Media Effects on Students' Academic Performance Using Machine Learning Algorithms (MLAs)." *Springerprofessional.de*, 21 Aug. 2021, www.springerprofessional.de/en/prediction-of-social-media-effects-on-students-academic-erforma/19589174.

[4] Shahzadi, Kiran, et al. "Machine Learning Methods for Analysing the Impact of Social Media on Students Academic Performance." *Journal of Applied Artificial Intelligence*, vol. 1, no. 1, 30 Dec. 2020, pp. 13–22, 10.48185/jaai.v1i1.96. Accessed 17 Feb. 2021.

[5] Krampah, Benedicta, et al. *Impact of Social Media Usage on Academic Performance of Undergraduate Students in Ghana Communication Technology University*. 4 Dec. 2021.

[6] Su, Xin, and Jiatao Huang. "Social Media Use and College Students' Academic Performance: Student Engagement as a Mediator." *Social Behavior and Personality: An International Journal*, vol. 49, no. 10, 6 Oct. 2021, pp. 1–8, 10.2224/sbp.10797. Accessed 21 Oct. 2021.

[7] Mastrodicasa, Jeanna, and Paul Metellus. "(PDF) the Impact of Social Media on College Students." *ResearchGate*, Feb. 2013, www.researchgate.net/publication/270550330_The_Impact_of_Social_Media_on_College_Students. Accessed 19 Mar. 2022.

[8] Uche, Agwi, and Ogwueleka Nonyelum. "IMPACT of SOCIAL MEDIA on STUDENTS' ACADEMIC PERFORMANCE." *International Journal of Scientific & Engineering Research*, vol. 9, no. 3, 2018, www.ijser.org/researchpaper/IMPACT-OF-SOCIAL-MEDIA-ON-STUDENTS-ACADEMIC-PERFORMANCE.pdf.

[9] Lau, Wilfred W.F. "Effects of Social Media Usage and Social Media Multitasking on the Academic Performance of University Students." *Computers in Human Behavior*, vol. 68, Mar. 2017, pp. 286–291, www.sciencedirect.com/science/article/pii/S0747563216307841, 10.1016/j.chb.2016.11.043.

[10] Xu, Xing, et al. "Prediction of Academic Performance Associated with Internet Usage Behaviors Using Machine Learning Algorithms." *Computers in Human Behavior*, vol. 98, Sept. 2019, pp. 166–173, 10.1016/j.chb.2019.04.015.

AUTHOR'S BIOGRAPHY

Kanupriya Guha[1], Sanjana M C[2], Maria Suman A[3] **"Survey Literature on Machine Learning Approaches to Study the Influence of Social Media on Academic Performance"**

Kanupriya Guha, Department of Computer Science, Mount Carmel College Autonomous, Mail ID: knupriya.guha@gmail.com

Sanjana M C, Department of Computer Science, Mount Carmel College Autonomous, Mail ID: mcsanjana2001@gmail.com,

Maria Suman A, Department of Computer Science, Mount Carmel College Autonomous, Mail ID: mariasuman@mccblr.edu.in