ijamrsd.com
JAMRSD
ISSN: 2583-6404
Journal of Advanced Multidisciplinary Research Studies and Development
Peer Reviewed Referred Journal | Bi-monthly | Open Access

# MACHINE LEARNING-BASED BEHAVIORAL CLASSIFICATION AND CORRELATION ANALYSIS ON THE IMPACT OF SOCIAL MEDIA AND OTT PLATFORMS ACROSS GENERATIONS

**Shubham Vijay Pol[1], Dr. Meenal Ambavane[2]**
[1] *Student, Kirti M. Doongursee College, Dadar, Mumbai.*
*Email: shubhampol3006@gmail.com*
[2] *Assistant Professor, Department of BAMMC, Kirti. M. Doongursee College, University of Mumbai.*
*Email: meenal.ambavane@despune.org*

**Abstract**
This research explores how social media and OTT (Over-the-Top) platforms influence behavioral and cognitive patterns among individuals from different generations using machine learning techniques. The study applies Decision Tree and Random Forest algorithms to survey data comprising 500 responses, focusing on habits such as screen time, sleep patterns, outdoor activities, and attention span. The dataset underwent cleaning, label encoding, and class balancing before modeling. Results showed that the Decision Tree achieved higher accuracy (0.889) and precision (0.955) compared to the Random Forest (accuracy 0.852). Correlation analysis revealed a positive relationship between excessive screen time and dissatisfaction, and a negative correlation between outdoor activity and late-night OTT usage. The impact of extended digital usage on one's lifestyle and concentration spectrum ought to be flagged, underscoring the increasing digital dependence of the youth.
**Keywords:** Machine Learning, Social Media Activity, Decision Tree, Random Forest, Digital Lifestyle, Attention, Correlation, and Instant Gratification.

► *Corresponding Author: Shubham Vijay Pol*

## Introduction

The influence of social media and OTT platforms on communication, entertainment, and other aspects of daily living has been profound and pervasive. While the affordances of connectivity and accessibility to remote places are benefits, abuse of such platforms has been linked to the disruption of the users behavioral and cognitive functions. Users are ditching face-to-face communication for digital interaction, and the events of binge-watching, as well as prolonged screen time, are increasingly becoming the norm. Such practices are likely to result in sleep deprivation, increased irritability, and shortened attention.

This study employs machine learning algorithms to analyse behavioural attributes collected through a structured survey. It seeks to identify generational trends and patterns linked to digital consumption and their impact on users' psychological and lifestyle balance. The goal is to provide data-driven insights into how technology influences focus, satisfaction, and engagement across different age groups.

**Literature Review**

The burgeoning of the digital world has greatly affected how people sleep, focus, and interact with one another, sleep health journal (2024) says that social media has sleep health consequences, as using social media and other digital platforms for extended periods of time before bed is associated with taking longer to fall asleep and having poorer quality sleep. BMC public health (2024) has noted that there is a deterioration of focus among those who use multiple digital services, as there are disruptive notifications, task switching, and rapid consumption of media. Other studies such as JAACAP (2023) have suggested digital screen behaviors and social media use are motivated by the dopamine responses the brain releases as a result of the instant gratification that social media platforms provide.

Interpersonal relationships are to have also changed in these modern times. Emerald (2020) has reported a decline in socialization through in-person meetings and the use of digitally mediated social interactions, which may have potential consequences for social/emotional development. It is these potential social and psychological changes that are the driving force for predictive modeling in trying to understand and categorize users.

The application of machine learning techniques in understanding modeling of human actions is of considerable depth. Quinlan (1986) developed one technique, Decision Trees (DT), which can factor recursively and predict outcomes as a function of variable values. Choice of outcome and factor values, he partitioned recursively until he arrived at a prediction. DT have a strong success rate in solving classification problems relative to interpretation of outcomes versus their behavioral attributes, in sleep, social media, and attention level usage. However, one DT is a single tree. Overfitting becomes a major problem. Noise in data is a common problem. To solve these problems, Breiman (2001) developed a more aggregate approach called Random Forests (RF). Multiple DTs are more computationally expensive. Prediction accuracy and robustness are overshadowed. Problem of solving with several DTs of high variance is produced. Behavioral data of high dimension is a common problem in social media usage RF are also effective solving outcomes of cognitive data.

The blending of behavioral studies and data science to the social media age is a frontline application of machine learning. DTs and RFs can in some instances accurately classify social media usage patterns to predict sleep problems and attention deficit disorders. As modeled data becomes more complex, these two fields allowed psychological data research to advance into computation. This leta machine learning and behavioral studies advance computational psychology. Being able to gain actionable insights to counter the adverse impacts of digital technology on human well-being.

**Objectives**

- Design and implement different kinds of dashboards and heatmap visualizations to show data relationships and predictive patterns.
- Understand correlations of different age demographics and their impacts of social media and OTT platforms including lifestyle and mental pattern changes.
- Create user personas based on behavioral and digital activity traits using treebased discriminative methods.
- Measure and compare model performance using various K-metrics (i.e. Accuracy, Precision, Recall, and F1score).

## Methodology

### Research Design:

There were 500 respondents of different ages and backgrounds who were surveyed and their data were collected online using a quantitative and survey-based design approach**.**

### Data Preprocessing:

During data cleaning, the Timestamp column was deleted, and the instances of missing values were replaced with mode values. Format the text entries for the survey question responses, then standardize them appropriately. Label Encoders were utilized to convert the data into numeric categorical responses. SMOTE was used to alleviate class imbalance and balance the dataset.

### Algorithmic Framework:

Two supervised models—**Decision Tree** and **Random Forest**—were trained to classify respondents based on their behavioral tendencies. The dataset was split in an **80–20 ratio** for training and testing. Model reliability was verified through **10-fold cross-validation**.

### Evaluation Metrics:

Accuracy, Precision, Recall, F1-score, and confusion matrices were computed to measure performance. Correlation analysis and feature importance visualization were employed to interpret model behavior.

## Results

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Decision Tree | 0.889 | 0.955 | 0.808 | 0.875 |
| Random Forest | 0.852 | 0.875 | 0.808 | 0.840 |

The Decision Tree model achieved the highest accuracy and precision, while both models showed similar recall, indicating consistency in detecting behavioral classes.

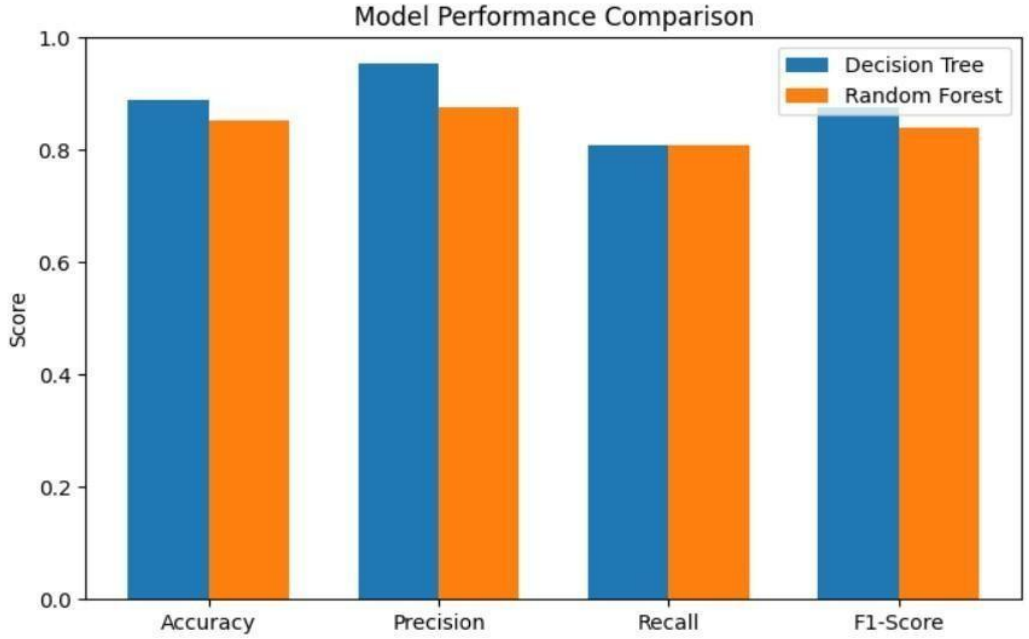


Figure 1: Model Performance Comparison Graph

This bar chart compares the performance of Decision Tree and Random Forest models across four evaluation metrics — Accuracy, Precision, Recall, and F1-Score. The Decision Tree achieved the highest accuracy (0.889) and precision (0.955), indicating stronger classification capability, while the Random Forest showed consistent recall, reflecting stable and balanced prediction behavior.

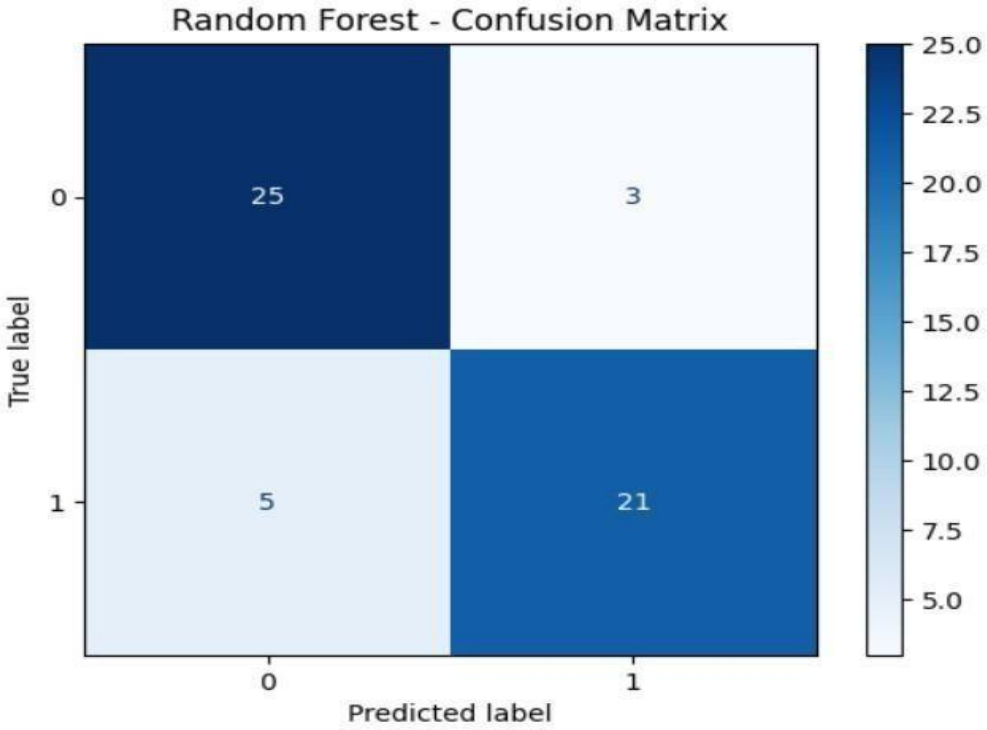


Figure 2: Confusion Matrix (Random Forest)

The confusion matrix illustrates the prediction outcomes of the Random Forest model. It shows that the model correctly classified most instances, with 25 true negatives and 21 true positives, while only a few cases (3 false positives and 5 false negatives) were misclassified. This confirms that the model performed reliably, effectively distinguishing behavioral categories with minimal errors.

**Feature Importance:**

The Random Forest model identified the following as top predictors of behavioral classification:

- Phone use before sleeping
- OTT/social media usage at night
- Binge-watching frequency
- Outdoor activity frequency

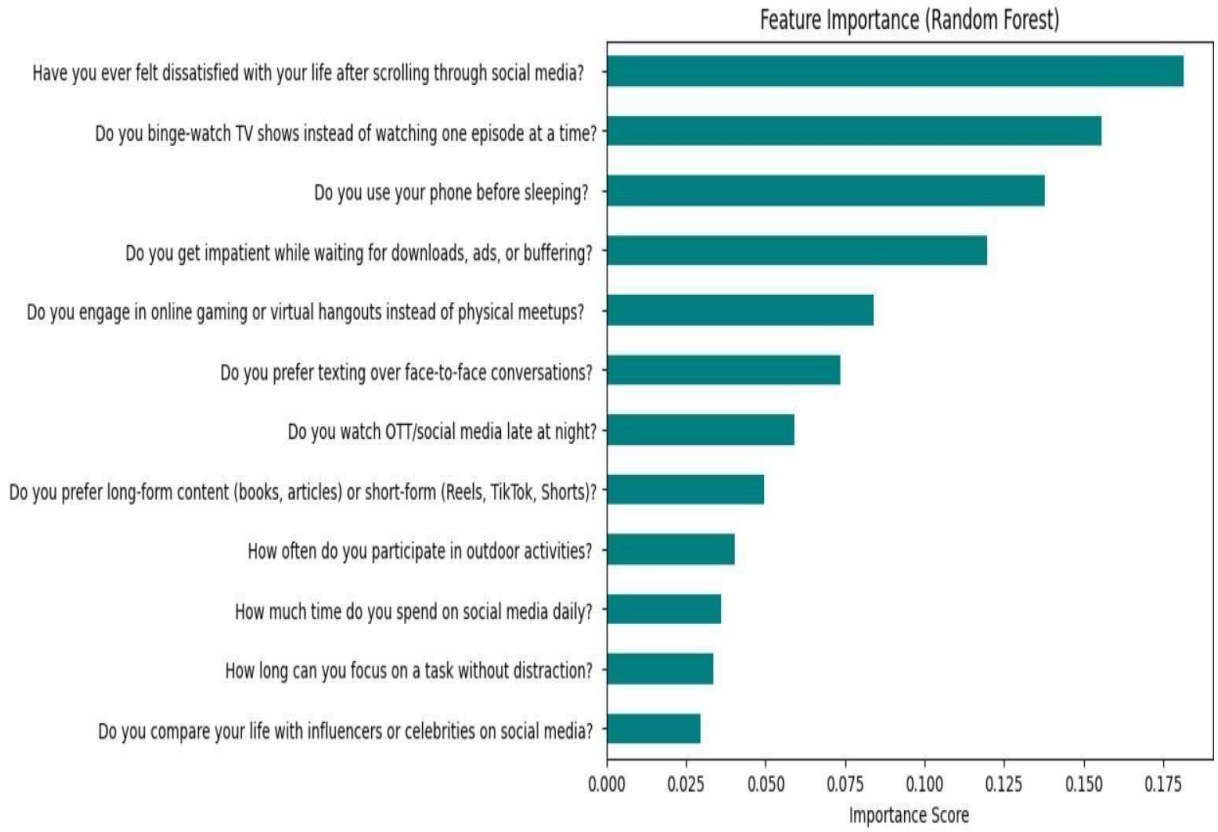


Figure 3: Feature Importance (Random Forest)

The feature importance graph from the Random Forest model shows that dissatisfaction after scrolling, binge-watching, and phone use before sleeping are the most influential factors affecting user behavior. Moderate importance is seen for online gaming, texting preference, and late-night social media use, while factors like screen time and outdoor activity have lesser impact. This indicates that excessive digital engagement strongly influences behavioral outcomes.

**Correlation and Visualization**

Correlation analysis revealed significant relationships:

- Strong positive correlation between screen time and dissatisfaction after scrolling
- Moderate negative correlation between **outdoor activity** and **late-night OTT usage**
- The **21–30 age group** showed the strongest association with impatience and higher short-term reward preference

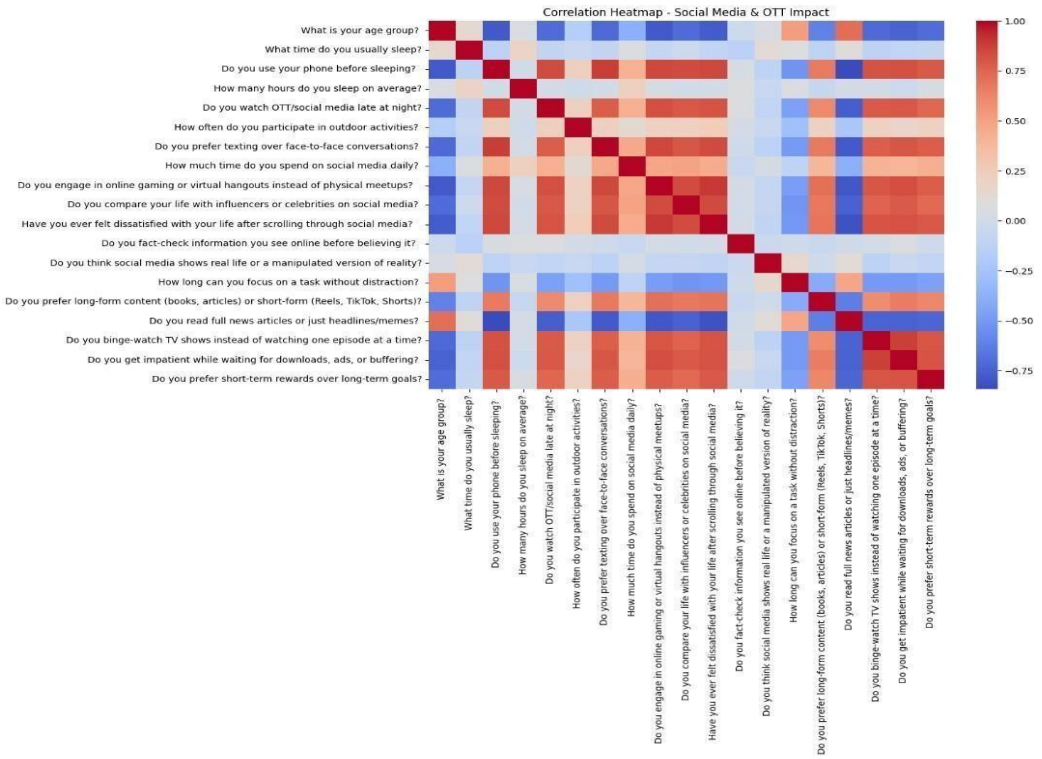


Figure 4: Correlation Heatmap

This figure presents the pairwise correlations among screen-related symptoms and behavioral variables. The color gradient reflects the strength and direction of relationships, with red indicating strong positive correlations and blue indicating negative associations. This visualization supports the identification of symptom clusters and potential feature redundancy, informing both exploratory analysis and model refinement.

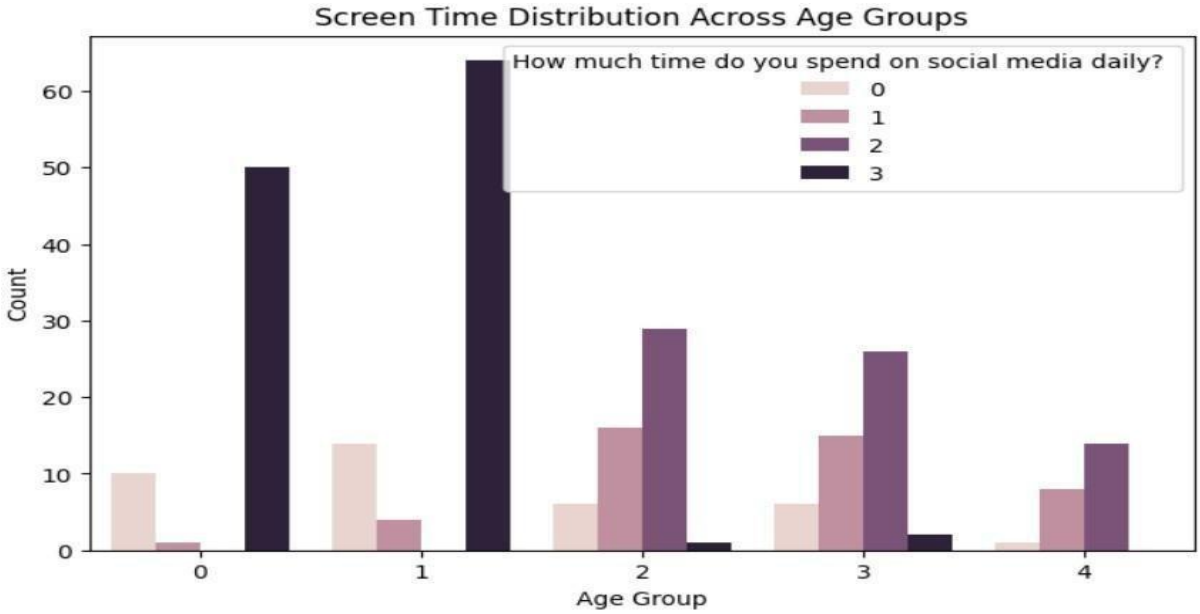


Figure 5: Screen Time Distribution Across Age Groups

This bar chart displays daily social media usage across different age groups. Usage is categorized into three levels: less than 1 hour, 1–3 hours, and more than 3 hours. Younger age groups (10s and 20s) show higher counts in the ">3 hours" category, while older groups (50s and 60s) tend to spend less time online. The visualization highlights a clear decline in screen time with increasing age, suggesting age-related differences in digital engagement.

**Result and Conclusion**

The study demonstrates that machine learning can effectively classify behavioral tendencies influenced by digital consumption. The Decision Tree outperformed Random Forest slightly, reflecting its adaptability to categorical survey data. Younger respondents exhibited higher screen time, late-night activity, and a stronger preference for instant gratification.

These findings align with previous studies highlighting the decline in focus and patience due to constant media stimulation. Correlation visualization further supports that excessive digital usage contributes to decreased outdoor participation and disturbed sleep.

The research successfully achieved its objectives, demonstrating that Decision Tree and Random Forest algorithms can classify behavioral traits and reveal correlations between media habits and age. With Decision Tree achieving **88.9% accuracy**, it proved more efficient for this dataset. The analysis confirms that social media and OTT exposure influence behavioral stability, attention span, and lifestyle quality, especially among younger users.

Balancing digital and real-world activities is crucial to maintaining cognitive health and sustainable media consumption habits.

**Future Scope**

Future work can expand the dataset geographically and employ ensemble deep learning methods to predict long-term behavioral changes. Including sentiment and emotion analysis could provide further insight into the psychological impact of prolonged digital exposure.

**References**

1. Sleep Health Journal (2024). *Impact of Social Media on Sleep Quality.*
2. Emerald (2020). *Social Media and Changing Socialization Patterns.*
3. BMC Public Health (2024). *Declining Attention Span in Digital Users.*
4. JAACAP (2023). *Dopamine and Instant Gratification in the Digital Age.*
5. Breiman, L. (2001). *Random Forests.* Machine Learning, 45(1), 5–32.
6. Quinlan, J. R. (1986). *Induction of Decision Trees.* Machine Learning, 1(1), 81–106.