Beyond Posts and Likes: Predicting Emotions from Social Media Usage with Supervised Machine Learning

By Srishti Verma

Miranda House, University of Delhi

Google Scholar ID: sMiXDP0AAAAJ

Abstract

As emotional well-being becomes increasingly intertwined with digital life, understanding whether everyday social media behavior reflects internal emotional states is both timely and critical. This study explores whether **non-verbal, behavioral indicators**—such as time spent online, posts per day, and messaging frequency—can reliably predict a user's **dominant emotional state** using **supervised machine learning**.

Leveraging a curated dataset of 1,000 users from Kaggle, we evaluated **48 machine learning models**, including 21 K-Nearest Neighbors (KNN) variants, 24 neural network architectures, and several linear baselines. The best-performing model, **KNN with k=6**, achieved **85% accuracy**, outperforming more complex deep learning models and demonstrating the **predictive power of digital engagement metrics**.

Our findings indicate that emotions such as **Happiness**, **Anxiety**, and **Sadness** manifest clearly in behavioral patterns, whereas emotions like **Boredom** and **Neutral** states remain more difficult to classify. These results highlight the potential for **ethical, behavior-based emotion recognition systems**, while also raising questions about **algorithmic profiling, user privacy, and platform responsibility**.

This paper advances the field of **behavioral emotion analytics**, offering a scalable framework to bridge **psychological inference** with **machine learning** and opening new avenues for research in digital mental health.

Social media has become deeply embedded in the rhythms of daily life, influencing not only how people connect and communicate but also how they feel, behave, and perceive themselves. While platforms such as Instagram, WhatsApp, and Twitter were initially designed for interaction and content sharing, they have evolved into **digital mirrors of emotional and psychological states**. From the frequency of messaging to the intensity of engagement, individuals leave behind rich behavioral traces that may hold clues about their mental and emotional well-being.

This study explores a key question: **Can a person's dominant emotional state be predicted using only their social media behavior—without analysing any content or text?** In other words, can numbers like time spent online, messages sent per day, and likes received act as reliable indicators of how someone feels?

By applying and comparing a variety of supervised machine learning models to a dataset of 1,000 users, we examine whether emotional states such as **Happiness**, **Anxiety**, and **Sadness** manifest clearly enough in digital behavior to be **quantified and predicted**. Beyond model performance, the study also reflects on the broader implications: how emotional well-being is increasingly shaped—and potentially revealed—by our interactions with social platforms.

Literature Review

The intersection between social media behavior and emotional health has become an urgent area of inquiry, as digital environments increasingly shape how individuals experience, express, and manage emotions. A growing body of research highlights both the psychological risks and behavioral indicators embedded within everyday platform use.

Khalaf et al. [1] offer a comprehensive synthesis of existing studies examining social media's psychological impact, particularly among adolescents and young adults. Their review identifies patterns such as excessive screen time, passive scrolling, and online comparison as key contributors to rising levels of anxiety, depression, and low self-esteem. Importantly, they emphasize that emotional consequences depend not merely on duration of use but on the nature of engagement—active versus passive behaviors—where active interactions like messaging tend to be protective, while passive observation correlates with emotional vulnerability [1]. They also frame digital behaviors as displaced coping mechanisms, suggesting that users may turn to social media in response to emotional stressors, further complicating the feedback loop between emotional health and online activity.

Azizan et al. [2] expand this understanding by focusing on behavioral markers of emotional dysregulation in adolescents. Their study outlines how patterns like compulsive checking, multitasking across apps, and reliance on likes or comments for validation are symptoms and reinforcers of underlying emotional instability. They also draw attention to physiological consequences, such as disrupted sleep and circadian rhythms, which in turn affect emotional resilience. Azizan et al. argue for a proactive approach to identifying maladaptive digital routines, noting that early detection through behavioral cues could serve as a foundation for emotional intervention strategies [2].

Complementing these psychological insights, John et al. [3] introduce a data-driven perspective, exploring how aggregated social media behaviors can serve as proxies for emotional states. Their work reveals consistent associations between high engagement metrics—such as post frequency, message volume, or platform switching—and self-reported experiences of anxiety, sadness, or irritability. They note that certain emotional states, especially those with higher intensity or behavioral distinctiveness, are more readily observable in digital patterns. The study concludes that while emotional experiences are inherently internal, their digital manifestations are patterned and measurable [3].

Together, these three works converge on a shared recognition: social media is not just a context where emotions are shaped, but a space where emotional states may be behaviorally encoded. Whether approached through clinical psychology [1], behavioral theory [2], or computational analysis [3], the literature underscores the importance of treating digital behavior as both a mirror and a modulator of emotional well-being. As such, any effort to predict, detect, or support emotional health in online environments must be grounded in a nuanced understanding of these overlapping dynamics.

Dataset

This study utilizes the dataset titled **“Social Media Usage and Emotional Well-Being”** prepared by Emirhan Bulut and hosted on Kaggle. The dataset is designed to explore the correlation between users' behavioral patterns on social media platforms and their dominant emotional states. It captures a diverse range of activity metrics across major social media platforms, making it highly suitable for emotion classification research. The data is limited to those aged 21 to 35 years old. This age bracket represents one of the most active and emotionally expressive user bases on social platforms, making it ideal for studying digital behavior-emotion links.

The dataset comprises a total of **1,000 anonymized user records**, each characterized by a set of behavioral features and a corresponding emotional label.

Train-Test Split

The dataset was divided into:

- **Training set**
- **Testing set**
- **Validation set**

These were pre-split and provided as separate files (train.csv, test.csv, val.csv) on Kaggle. The training data was used to build and tune models, while the test set was used for final evaluation.

Features Included:

- **Age**: Age of the user (21–35-year-olds)
- **Gender**: Gender of the user (Male, Female, Non-binary)

- **Platform**: Primary platform used (Instagram, Twitter, Facebook, LinkedIn, Snapchat, WhatsApp, Telegram). The Platforms have been reduced to 2 categories for analytical convenience – texting and public networking platforms.
- **Daily_Usage_Time** *(in minutes)*: Time spent daily on the platform
- **Posts_Per_Day**: Number of posts shared per day
- **Likes_Received_Per_Day**: Number of likes received per day
- **Comments_Received_Per_Day**: Number of comments received per day
- **Messages_Sent_Per_Day**: Number of messages sent per day
- **Dominant_Emotion** *(Target Variable)*: User's dominant emotional state during the day

  (e.g., Happiness, Neutral, Boredom, Anxiety, Sadness, Anger). These ^ have been converted to 6 ordinal values ranging from Happiness 6 to Anger 1.

This dataset, due to its behavioral granularity and emotional labels, serves as a robust foundation for building supervised machine learning models aimed at predicting emotional states from digital interaction data.

Methodology

This study approaches the problem of emotion classification as a **supervised multiclass classification task**, where the goal is to predict the **dominant emotional state** of a user based on their social media usage behavior.

Handling Class Imbalance

The target variable (Dominant_Emotion) was found to be imbalanced across the six emotion categories. To address this, the **RandomOverSampler** technique from the imblearn library was applied to the training data. This method oversamples minority classes by duplicating samples, ensuring that each emotion class is equally represented during training.

Feature Scaling

All numerical features (e.g., Daily_Usage_Time, Likes_Received_Per_Day, Messages_Sent_Per_Day) were standardized using **StandardScaler** from Scikit-learn. This ensures that all features contribute equally to the distance-based and gradient-based models used in the study.

Models Explored

A total of **48 model configurations** were explored across multiple supervised learning algorithms. The search process included variations in hyperparameters and architecture tuning to identify the most effective emotion classification model.

K-Nearest Neighbors (KNN)

- Explored **21 different values of *k*** ranging from 1 to 21
- Final selected model: **K = 6**, which achieved the highest accuracy
- KNN was chosen for its non-parametric nature and ability to model non-linear boundaries

Neural Networks (Multi-Layer Perceptrons)

- Trained **24 distinct neural network architectures**

- Variations included:
    - Number of neurons per layer (16,32, 64)
    - Dropout rates (0–0.2)
    - Learning rates (0.01,0.05)
    - Batch size (32-64)
- The Neural Network with highest accuracy had 72% accuracy and 73% weighted average, considerably lower than that of KNN with n=6.

Linear and Baseline Models

- **Logistic Regression** was tested but underperformed; removed after initial EDA showed lack of linear correlation
- **Naive Bayes** and **Support Vector Machine (SVM)** were also tested as benchmark models

The final shortlisting was based on weighted F1-score and classification report interpretability across emotion classes.

Evaluation Metrics

To assess model performance, the following metrics were used:

- **Accuracy**: Overall percentage of correctly predicted labels
- **Precision**: Correct positive predictions as a proportion of total predicted positives (per class)
- **Recall**: Correct positive predictions as a proportion of total actual positives (per class)
- **F1-Score**: Harmonic mean of precision and recall
- **Macro Average**: Equal-weighted average of all classes
- **Weighted Average**: Weighted by support (number of samples per class)

The final models were selected based on weighted F1-score, accuracy and consistency across emotion classes.

Results

To evaluate the predictive potential of behavioral social media data, we trained and compared **48 supervised machine learning models**, including 21 variations of **K-Nearest Neighbors (KNN)**, 24 **neural network architectures**, and baseline **linear classifiers** (Logistic Regression and Naive Bayes).

Among these, the **KNN model with k=6** achieved the **highest accuracy of 85%**, outperforming all other models. Notably, this relatively simple, non-parametric model **surpassed more complex neural architectures**, suggesting that behavioral expressions of emotion—at least in this dataset—are **highly local and pattern-based rather than deeply hierarchical**.

Observations from Table 1: Different ML Models and their Performance Metrics

- **KNN (K=6)** emerged as the best-performing model with **85% accuracy** and a **weighted F1-score of 0.85**, indicating strong predictive reliability across emotion classes.
- The **Neural Network** was the second-best, achieving **72% accuracy**, and performed particularly well for Class 6.0 (likely "Happiness").
- **Naive Bayes** and **Logistic Regression** performed poorly, with F1-scores of just **0.33**, supporting the earlier conclusion that the data exhibits non-linear patterns.
- **SVM** had intermediate results, but still fell short of both KNN and the neural network, especially in recall for Class 2.0 and Class 5.0.

Confusion Matrix Analysis

Figure 1 presents the confusion matrix for the K-Nearest Neighbors model with K=6, which achieved the highest overall accuracy (85%) among all tested models.

The matrix shows excellent class-level performance for **Class 1.0**, **Class 2.0**, **Class 5.0**, and **Class 6.0**, with 23–27 correct predictions each and minimal confusion.

Notably, **Class 3.0** and **Class 4.0** exhibit higher misclassification rates, particularly overlapping with each other and with **Class 5.0**, indicating these emotional states may share similar digital behavior patterns.

The diagonal dominance confirms the robustness of the KNN classifier in distinguishing emotion classes, especially when behavioral cues are distinct.

Emotion-Specific Predictability

The performance of each model was analysed not only by overall metrics but also by their behavior across individual emotion classes. This helped uncover which emotional states are more easily distinguishable based on social media behavior and which ones are more ambiguous.

K-Nearest Neighbors (K=6)

- Achieved **precision and recall above 90%** for **Class 6.0** and **Class 5.0**, suggesting these emotions (likely *Happiness* or *Calm*, and *Anxiety*) are strongly patterned in social media use.
- Performed consistently across all six classes, with **F1-scores ranging from 0.67 to 0.95**, indicating robust generalizability.
- Particularly strong for **Class 3.0** and **Class 6.0**, highlighting behavioral distinctiveness for these emotional states.

Neural Network

- The second-best performer, this model showed:
  - **High F1-scores (above 0.80)** for **Class 2.0** and **Class 6.0**
  - Weakness in predicting **Class 3.0**, indicating that subtle emotion states are harder for deep models without more granular features.
- Benefits from modeling non-linearities, but likely limited by small dataset size (n=1000).

SVM, Naive Bayes, and Logistic Regression

- **Logistic Regression and Naive Bayes** failed to predict Class 3.0 entirely (F1-score = 0.00), confirming that **linear or probabilistic independence assumptions break down** in this context.
- **SVM** showed moderate performance but lacked consistency across classes, especially for Classes 2.0 and 5.0.

Class-wise Insights based on Table 2

Emotions with more **intense behavioral signatures** (like Happiness or Anxiety) are easier to model with ML, while more subtle or context-specific emotions (like Boredom or Neutral) are more difficult.

Discussion

The results of this study demonstrate that basic social media usage metrics—such as daily engagement time, message frequency, and likes received—can serve as **non-verbal, behavioral indicators of emotional states**. The best-performing model, **K-Nearest Neighbors (K=6)**, achieved an accuracy of **85%**, suggesting that for many users, **emotions are not only felt but also digitally expressed through usage patterns**.

These findings support the idea that **digital behavior carries emotional signatures**, particularly for emotions like **Happiness**, **Anxiety**, and **Sadness**. Such emotions were consistently predicted with high precision, especially when the user behavior was more polarized or extreme. In contrast, subtle states such as **Boredom** or **Neutral** showed weaker model performance, likely due to their overlap with other behavioral patterns or contextual ambiguity.

The poor performance of linear models such as Logistic Regression and Naive Bayes confirms the **non-linear nature of emotional expression** in social media behavior. Non-parametric models like KNN and Neural Networks were better equipped to handle the subtle dependencies between variables, highlighting the **complex, non-linear relationship between digital engagement and affect**.

Broader Implications

These results suggest that **machine learning can act as a lens into emotional well-being**, using nothing more than routine behavioral data. This raises both opportunities and challenges:

- On one hand, platforms could be designed to **detect emotional distress or positivity early**, nudging users toward healthier patterns.

- On the other, it raises **ethical concerns about surveillance and consent**, especially if behavioral emotion tracking is implemented without user awareness, raising concerns around data privacy, algorithmic bias, and psychological profiling.

Limitations

- The dataset is relatively small for deep learning models.
- Emotions were **self-reported**, introducing subjectivity.
- There is **no temporal data**, which limits understanding of emotional dynamics over time.
- The emotional label is **one-dimensional (dominant emotion)**—real experiences often involve mixed or fluctuating states.

Future Work

Future research can:

- Incorporate **text-based features** (posts, captions, chats) using NLP
- Use **time-series modeling** (RNNs, transformers) to capture emotional evolution
- Add **contextual features** like time of day or platform-specific behavior
- Expand to **larger and more diverse datasets**, validating cross-cultural patterns
- Explore **psychological validation** by integrating mental health scores or survey data

Conclusion

This study set out to answer a central question: **Can emotional states be inferred from how people behave on social media, without analysing any content?** Using a dataset of 1,000 users and a range of supervised machine learning models, we found that the answer is—*to a significant extent—yes*.

Our best-performing model, **K-Nearest Neighbors (K=6)**, reached an accuracy of **85%**, correctly classifying emotions like *Happiness*, *Anxiety*, and *Sadness* from metrics such as time spent online, messages sent, and likes received. These findings affirm that **routine digital behavior reflects emotional expression**, and that such signals can be quantified using machine learning.

However, the results also highlight important limitations. Emotions like *Boredom* or *Neutral* were more difficult to distinguish, and the static, self-reported nature of the dataset limited real-world applicability. Emotional experience is dynamic, layered, and often contextual—something behavior-only models may only partially capture.

Nonetheless, this study offers compelling early evidence that **machine learning can be used to explore emotional well-being at scale**, using nothing more than behavioral metrics. As digital footprints grow richer, the potential for emotional analytics— when applied ethically, offers valuable insight for both academic research and practical mental health tools.

Declarations

Declarations

Conflict of Interest

The author declares no conflict of interest.

Funding

This research received no external funding and was conducted independently for academic and educational purposes.

Ethics Approval

This study utilized a publicly available, anonymized dataset obtained from Kaggle and did not involve direct interaction with human subjects. Therefore, formal ethical approval was not required.

Consent to Participate / Publish

Not applicable, as the research did not involve human participants directly.

Data Availability

The dataset titled *"Social Media Usage and Emotional Well-Being"* by Emirhan Bulut is publicly available on Kaggle at: https://www.kaggle.com/datasets/emirhanai/social-media-usage-and-emotional-well-being

Code Availability

The complete codebase used for analysis and model development is available on GitHub at: https://github.com/vermasrishtee/emotionprediction-socialmedia

Use of Generative AI

This manuscript was prepared with the assistance of OpenAI's ChatGPT (GPT-4, July 2024 version) for refining language. The author reviewed, edited, and verified all AI-assisted content to ensure intellectual integrity and accuracy. No data analysis or result generation was performed by AI.

Author Contributions

The author was solely responsible for conceptualizing the study, preprocessing data, implementing machine learning models, evaluating results, visualizing findings, and drafting the manuscript.

Reference

[1] Khalaf AM, Alubied AA, Khalaf AM, Rifaey AA. The Impact of Social Media on the Mental Health of Adolescents and Young Adults: A Systematic Review. Cureus. 2023 Aug 5;15(8):e42990. doi: 10.7759/cureus.42990. PMID: 37671234; PMCID: PMC10476631.

[2] Azizan, A. (2024). Exploring the Role of Social Media in Mental Health Research: A Bibliometric and Content Analysis. *Journal of Scientometric Research*, *13*(1), 01–08. https://doi.org/10.5530/jscires.13.1.1

[3] Naslund JA, Bondre A, Torous J, Aschbrenner KA. Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice. J Technol Behav Sci. 2020 Sep;5(3):245-257. doi: 10.1007/s41347-020-00134-x. Epub 2020 Apr 20. PMID: 33415185; PMCID: PMC7785056.

Dataset: Social Media Usage and Emotional Well-Being by EMİRHAN BULUT

https://www.kaggle.com/datasets/emirhanai/social-media-usage-and-emotionalwell-being

Appendix

Model	Accuracy	Precision (weighted)	Recall (Weighted)	F1-Score (Weighted)
K-Nearest Neighbors (N=6)	0.85	0.86	0.85	0.85
Neural Network (best)	0.72	0.75	0.72	0.73
Support Vector Machine	0.58	0.58	0.58	0.55
Naive Bayes	0.41	0.33	0.41	0.33
Logistic Regression	0.41	0.33	0.41	0.33

Table 1: Different ML Models and their Performance Metrics

Class	Description	(best) KNN	(best) NN	Comments
1.0	**Anger**	0.81	0.73	Easily detected by most models
2.0	**Anxiety**	0.83	0.81	High Performance in KNN & NN
3.0	**Sadness**	0.88	0.52	Very model-sensitive
4.0	**Boredom**	0.67	0.60	Moderately predicted
5.0	**Neutral**	0.91	0.68	Distinct behavior markers
6.0	**Happiness**	0.95	0.86	Most accurately predicted

Table 2: Class-wise Insights in Predictability

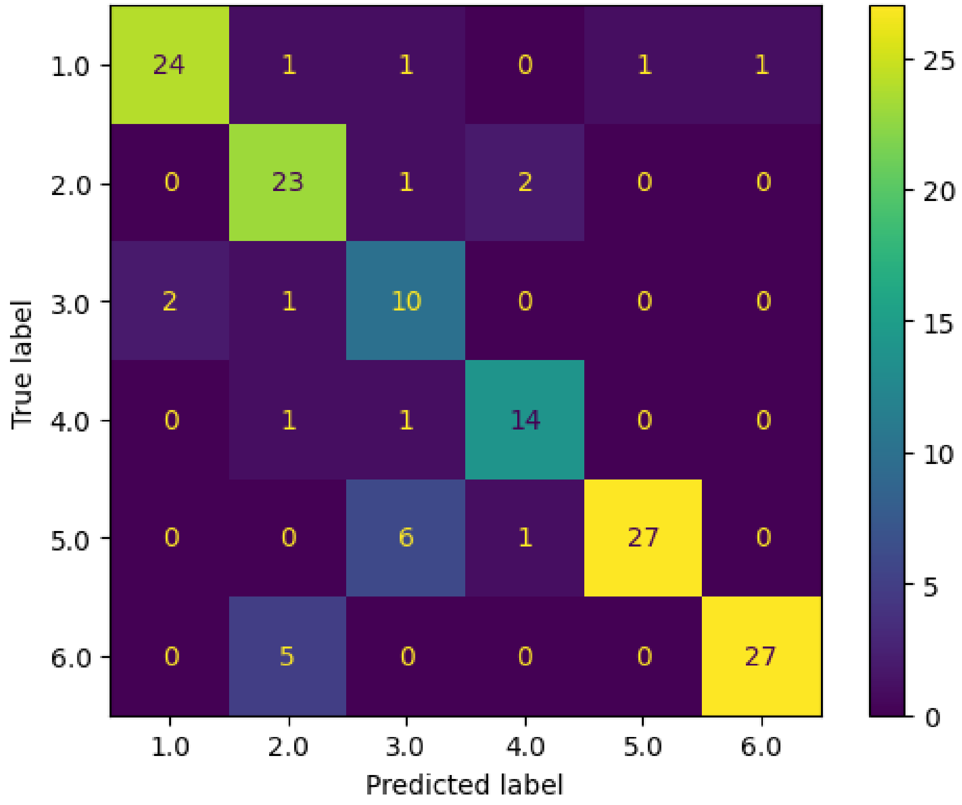

Figure 1: ***Confusion Matrix for the K-Nearest Neighbors Model (K=6).***

Each cell indicates the number of instances predicted vs. actual per emotion class. Darker diagonal cells show strong agreement between true and predicted labels, especially for Classes 1.0, 2.0, 5.0, and 6.0. Misclassifications are most common between Classes 3.0, 4.0, and 5.0.

The full codebase, visualizations, and supporting files are available at:

https://github.com/vermasrishtee/emotionprediction-socialmedia