Journal of Neurodevelopmental Cognition

Journal of Neurodevelopmental Cognition 5(1) (2024) 48-58
ISSN: 2645-565X
http://www.jncog.sbu.ac.ir

10.48308/jncog.2024.105194

Psychological and Demographic Determinants of Social Media Influence: Developing Predictive Models to Identify Influencers

S. M Mahdi Firouzabadi[1], G. Reza Jafari[2*], Reza KhosrowAbadi[1]

1 Institute for Cognitive and Brain Science, Shahid Beheshti University, Tehran, Iran.
2 Department of Physics, Shahid Beheshti University, Tehran, Iran.

Abstract

This study explores psychological and demographic characteristics distinguishing social media influencers from non-influencers and investigates the predictive potential of psychological features for influence. Using a diverse dataset containing age, gender, NEO personality scores, and a revised active/passive engagement scale of 1,214 Iranian participants, we aim to uncover significant feature differences and construct a predictive model for influence classification. Our statistical analyses reveal significant differences between influencers and non-influencers in key variables, including age and active/passive engagement and Neuroticism. However, machine learning models indicate that while distinct psychological characteristics are associated with influence, their predictive power shows promise but may be limited without additional behavioral or content-based metrics. This study contributes to the understanding of psychological factors in social influence and the feasibility of machine learning models for influencer identification.

Keywords: Social Network, Social influence, Machine Learning, Social Cognition

* Corresponding author

Email addresses: g_jafari@sbu.ac.ir

Received: October 2024 *Accept: December 2024*

1. Introduction

In recent years, social media platforms have become central to personal and professional communication, influencing numerous aspects of social behavior. Social media influencers, who attract significant engagement, have become prominent figures shaping opinions and trends. Previous research has linked social media engagement styles with personality and psychological characteristics, suggesting that certain personality traits and demographic factors may distinguish influencers from general users. (Gu, 2024 ; SharifiFard, 2024) However, limited studies have quantitatively explored these differences or assessed whether these traits can predict social influence.

This study investigates whether influencers demonstrate unique psychological or demographic properties compared to non-influencers. Our primary hypothesis is that influencers and non-influencers will exhibit distinct patterns in demographic and psychological variables, including age, gender, NEO personality traits, and engagement styles. A secondary question explores the feasibility of using psychological features to identify influencers via machine learning models. While our machine learning models indicate that the current dataset does not provide a strong predictive basis for influencer identification, our statistical analyses reveal significant differences in certain psychological and demographic features, such as age and active/passive engagement. This study will employ statistical analyses to find significant feature differences and build predictive models for influencer identification.

2. Literature Review

Prior research highlights the role of personality traits in social media behavior. Studies have shown that individuals with high extraversion and openness scores are more active on social platforms, while agreeableness and conscientiousness relate to content-sharing behavior. Furthermore, differences in engagement patterns, including frequency and duration of usage, may indicate higher social media influence.

Influencer marketing has emerged as a pivotal strategy in contemporary digital marketing, leveraging individuals with substantial social media followings to shape consumer behavior and brand perception. The effectiveness of this approach is underscored by studies demonstrating that influencers can significantly impact audience engagement and purchasing decisions through authentic content and personal connections. To optimize influencer marketing campaigns, it is essential to develop models that elucidate the mechanisms by which influencers affect audience behavior. Such models facilitate the identification of key factors driving consumer engagement, enabling marketers to tailor strategies that enhance brand credibility and foster deeper consumer relationships. (Shaimm et. al. 2024, Gu et. al. 2024)

Sharifi Fard et al. explored the relationship between the Big Five personality traits and happiness in Iranian society, highlighting the mediating role of problematic Instagram use, which aligns with our focus on NEO Big Five traits and social media behavior. Their results confirm that neuroticism had a strong positive role in Instagram problematic use among Iranian youth. (Sharifi Fard et al.,2024) NEO personality traits, particularly extraversion and openness, significantly influence active engagement on social media, while agreeableness and conscientiousness drive

content-sharing behaviors (Lin et. al.,2024). Pirzade highlighted the role of neuroticism, avoidant identity style, and the need to belong as psychological factors driving celebrity worship among adolescents, providing insights into the psychological underpinnings of social media influence. (Pirzade et. al, 2024) In the study by Lampropoulos Openness and extraversion emerged as the two most significant positive predictors of social media use. (Lampropoulos et. al. 2024)

Recent studies, such as one employing machine learning to explore social isolation and loneliness across schizophrenia, bipolar disorder, and general communities (Abplanalp, et. al., 2024), highlight the potential of predictive models in understanding complex social behaviors. Despite growing interest in using machine learning to explore user behavior, few studies quantitatively distinguish influencers based on psychological profiles, and fewer still have used machine learning to predict influence from such variables.

3. Data and Methods

Dataset Description

The dataset used in this study comprises responses from 1,214 social media users (N=1214, Age=19-68 years, Mean Age=31, STD=10, Male=371, Female=843), selected as part of a cross-sectional study conducted in Tehran from September to November 2023. The sample was gathered through convenience sampling and includes a range of demographic, behavioral, and psychological characteristics, as well as information about participants' social media usage. To be eligible, participants had to meet the following criteria: they had to be at least 18 years of age, fluent in Farsi, and able to read and write competently, and they had to provide informed consent. Data were collected through online questionnaires, ensuring anonymity; participants' names were not recorded, and participation was entirely voluntary. This study analyzed questionnaires completed by participants aged 19 to 60 years. Before submitting, participants were encouraged to review their responses for accuracy.

Key features of the dataset include age, gender, and scores from the NEO Five-Factor Inventory (NEO-FFI), a 60-item personality questionnaire assessing five main personality traits: Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. The NEO-FFI, an efficient and widely accepted tool for personality assessment, has shown high validity and reliability (Costa & McCrae, 1992). Its Farsi adaptation, validated in an Iranian sample, demonstrated satisfactory internal consistency, with intercorrelations among the scales ranging from 0.56 to 0.87 (Garousi Farshi, 2001). This makes the NEO-FFI a cost-effective option for personality assessment in both time and resources, as it provides reliable results with minimal criticism in comparison to other personality inventories (Garousi Farshi, 2001).

Another key measure in the dataset is the revised scale of social media engagement, a questionnaire adapted and validated for Iranian society (Firouzabdi et. al., 2024). Based on the Social Media Engagement Questionnaire developed by Pagani and Mirabello (Pagani et. al.,2011), this revised scale measures both passive and active engagement. Passive engagement includes behaviors like browsing without interaction, while active engagement includes interactions such as commenting or posting. The scale uses a 5-point Likert scale, where responses range from 1 ("Strongly disagree") to 5 ("Strongly agree").

The dataset also includes valuable information about participants' online presence in social networks including Follower counts, engagement metrics, and platform usage across multiple social media platforms were recorded and combined into an Influence Index. This index was used to categorize participants into influencers and non-influencers, providing a structured basis for comparing psychological and demographic characteristics between the two groups.

Defining the Influence Index and Identifying Influencers

Angelini et. al study of adolescents' perceptions of friendship quality demonstrates that **active social media use** correlates positively with enhanced friendships, aligning with the idea that higher **usage frequency** increases visibility and engagement, key drivers of influence. Additionally, the study notes that accessibility and diverse interactions on social media foster better peer connections, supporting the importance of **platform diversity** in broadening reach and amplifying influence across different social groups. Together, these findings provide empirical backing for the components of the Influence Index, emphasizing their critical role in measuring social influence. (Angelini et. al. , 2024)

To quantify social influence, we developed an **Influence Index** that integrates multiple dimensions of social media engagement, including follower counts, following behavior, usage frequency, and platform diversity. These dimensions were selected based on empirical evidence and theoretical findings that highlight their roles in shaping social influence online. Our Influence Index approach follows the practice of weighting components to reflect their relative importance in assessing influence, supported by existing literature on social media metrics and influence modeling.

To achieve comparability across features, each platform-specific follower count (Most used social platforms in Iran including Instagram, Telegram, Twitter, LinkedIn, Rubika, and Bale as local platforms) was percentile-normalized, allowing us to aggregate follower metrics across diverse social platforms into a **Total Followers** score. This process is commonly used to control for outliers, ensuring the index is not unduly impacted by exceptionally high or low follower counts. Research has demonstrated that follower count is a central indicator of social influence as it represents the audience size accessible to an individual, often correlated with perceived credibility and reach. Similarly, the **Total Following** was calculated using percentile-normalized data across platforms, representing a user's networking behavior and social connectivity. Following others can indicate active engagement with the community and network openness, factors known to influence social reach and interaction. We also included **Average Daily Use** (total time spent on social media across platforms) and **Average Daily Checks** (number of times platforms are accessed daily) as key indicators of engagement level. These behavioral components capture the intensity of platform engagement, which is closely associated with influence and the potential to create consistent, visible interactions. Finally, **Platform Count**, representing the diversity of platforms a user engages with, was included to account for cross-platform influence, where engagement on multiple platforms increases visibility and reach.

Weighting Components of the Influence Index

Based on these theoretical foundations, we applied a weighting scheme to the components of the Influence Index to reflect their relevance in measuring influence. **Total Followers** received

the highest weight (40%), as audience size is a primary determinant of social influence, strongly linked with perceived authority and reach. **Total Following** was assigned a weight of 20%, capturing the significance of network connectivity without overemphasizing it relative to audience size.

Average Daily Use and **Average Daily Checks** were each weighted at 20% and 10%, respectively, in line with findings that active and frequent engagement enhances user visibility and potential influence. Finally, **Platform Count** was given a 10% weight, balancing the benefit of cross-platform reach while recognizing that it is secondary to core engagement metrics. This weighted model allowed for a robust yet interpretable index to differentiate users.

To construct the Influence Index, we applied a weighted formula to each component, reflecting its relative importance in measuring influence:

$$Influence\ Index = (w_1 \times Total\ Followers) + (w_2 \times Total\ Following) +$$

$$(\boldsymbol{w_3} \times \boldsymbol{Average\ Daily\ Use}) + (\boldsymbol{w_4} \times \boldsymbol{Average\ Daily\ Checks}) + (\boldsymbol{w_5} \times \boldsymbol{Platform\ Count})$$

Total Followers received the highest weight (w_3= 0.4) due to its strong association with audience size and perceived authority. Total Following w_3= 0.2) reflects connectivity while avoiding excessive emphasis relative to audience size. Average Daily Use (w_3= 0.2) and Average Daily Checks $w_4 = 0.2$) capture active engagement and visibility, while Platform Count given the lower weight ($w_5 = 0.1$) balances the benefit of cross-platform reach while recognizing that it is secondary to core engagement metrics. This weighted model allowed for a robust yet interpretable index to differentiate users.

Users in the top 20% of the Influence Index were categorized as "influencers," while the remaining 80% were labeled "non-influencers." This classification enabled structured group comparisons and informed predictive modeling.

Statistical Comparison of Influencer and Non-Influencer

To assess differences in psychological and demographic features between influencers and non-influencers, we conducted a Mann-Whitney U test for each feature. This non-parametric test is particularly suitable for our data, as it does not assume a normal distribution and is robust to outliers, making it an effective choice for comparing traits such as personality scores and engagement metrics that may vary significantly in distribution. Given the multiple comparisons across several features, we applied the Bonferroni correction to control for Type I errors, ensuring that only truly significant differences were highlighted.

The analysis revealed that **Age** (adjusted p < 0.00001), **Neuroticism** (adjusted p = 0.026), **Active Engagement (Active_Revised)** (adjusted p = 0.0000087), and **Passive Engagement (Passive_Revised)** (adjusted p < 0.00000003) had statistically significant differences between influencers and non-influencers. This underscores their importance as meaningful predictors of influencer status. The significant difference in Age aligns with studies indicating that social

influence varies across age groups, with younger users often demonstrating greater activity on social media, which can enhance follower counts and engagement metrics. Active engagement scores are particularly relevant, as influencers tend to exhibit higher active behaviors (e.g., posting, commenting), which directly contribute to audience interaction and visibility. On the other hand, passive engagement scores, which measure how often users engage in passive activities such as reading posts or browsing social media, also show significant differences. This suggests that influencers may engage more frequently in both active and passive social media activities, possibly reflecting their broader involvement in the social media ecosystem

Feature	p-value, Adjusted p-value
Neuroticism	2.905678e-03 2.615110e-02 *
Extraversion	4.638250e-02 4.174425e-01
Openness	3.560584e-02 3.204525e-01
Agreeableness	1.140834e-01 1.000000e+00
Conscientiousness	4.658845e-01 1.000000e+00
Age	8.303608e-10 7.473247e-09 *
Active_Revised	9.628311e-07 8.665480e-06 *
Passive_Revised	3.268664e-19 2.941798e-18 *
Sex	5.072700e-01 1.000000e+00

***Table 1**: Summary of p-values and adjusted p-values for features distinguishing influencers and non-influencers.*

The significant difference in Neuroticism is noteworthy, as it may indicate that influencers tend to exhibit higher levels of emotional sensitivity, which could influence how they engage with their audience or respond to social feedback.

Two additional features, **Openness** (adjusted p = 0.32) and **Extraversion** (adjusted p = 0.42), exhibited p-values close to the 0.05 threshold in the unadjusted analysis, although they did not reach significance after the Bonferroni correction. These borderline results suggest a potential trend where influencers may score higher in these personality dimensions, which is in line with existing theories suggesting that openness to experience and extrovert behaviors may enhance social interactions (Lin et. al.,2024) and facilitate engagement with diverse followers. However, given the adjusted p-values slightly exceeding the threshold, we interpret these as suggestive trends rather than definitive findings. Further investigation with a larger dataset or additional

engagement metrics could help clarify the roles of these personality traits in distinguishing influencers from non-influencers.

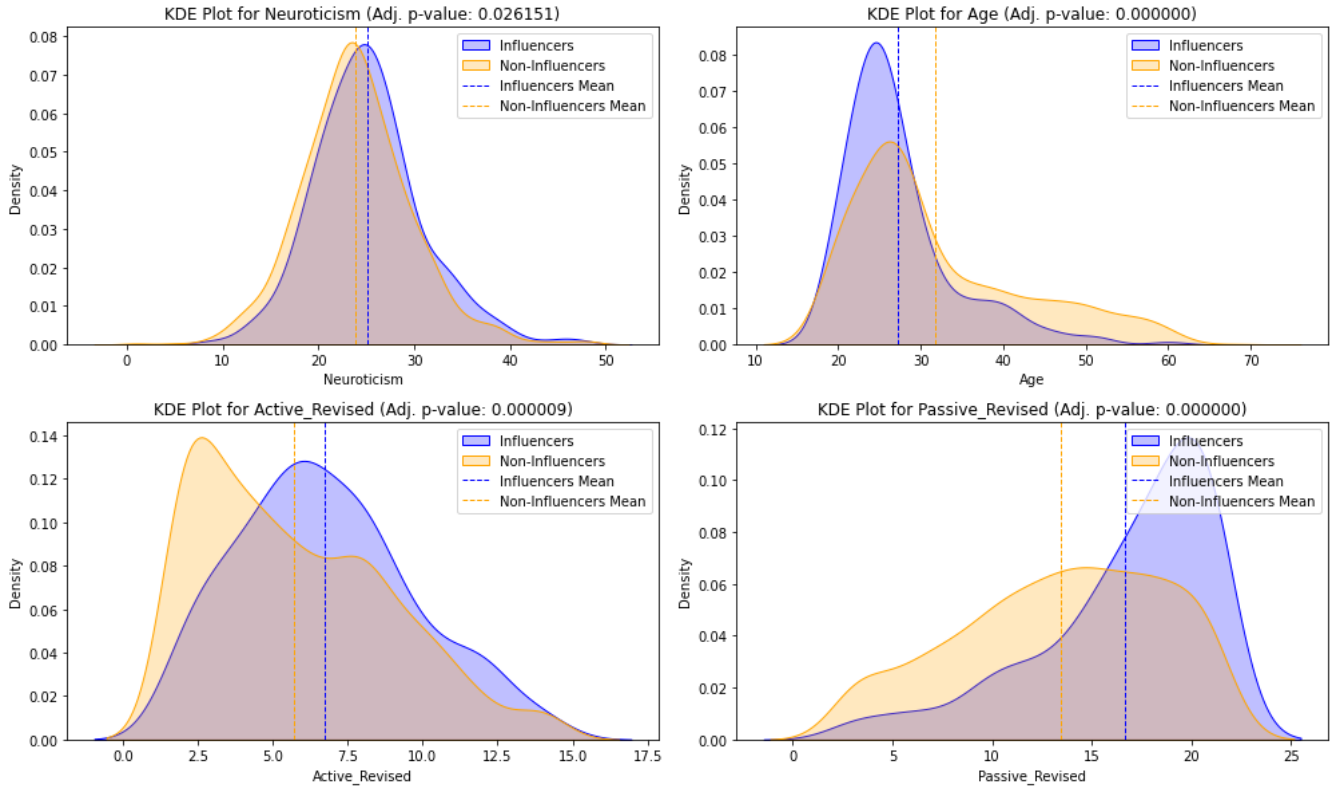

***Figure 1**: "Kernel Density Estimates (KDE) comparing distributions of key psychological and demographic features between influencers and non-influencers, with mean indicators and adjusted p-values highlighting statistically significant differences.*

Machine Learning Models Predicting Social Influence

To predict the influence index and classify influencers versus non-influencers, several machine learning models were used, each chosen for their unique suitability to specific aspects of this task. Input features for these models included the significant demographic and psychological variables identified in the statistical analysis. Each model was trained using K-fold cross-validation, with Mean Absolute Error (MAE) as the primary evaluation metric.

The **Multi-Layer Perceptron (MLP)** model was used to predict the continuous Influence Index. MLPs are a type of neural network particularly suited to capturing complex, non-linear relationships between input features and targets. Our MLP consisted of two hidden layers with 32 and 16 units, respectively, using ReLU activation for non-linearity and L2 regularization to prevent overfitting. The final layer matched the number of target variables, allowing the model to predict both influencer status and the Influence Index. The MLP's capacity for non-linear

modeling and feature interaction makes it a suitable choice for our dataset, which includes both psychological and behavioral variables that likely have complex interdependencies.

For further robustness, we used **Random Forest** and **Gradient Boosting Machines (GBM)** to predict the continuous Influence Index. Random Forest, an ensemble of decision trees, is well-known for handling high-dimensional data with mixed data types and for its resilience to overfitting due to averaging across multiple trees. Random Forests also provide feature importance scores, aiding interpretability and feature selection. On the other hand, GBM models build sequentially, learning from previous errors to reduce prediction bias. Separate GBM models were trained for each target variable, allowing the model to specialize in predicting each outcome. Both methods have proven effective in influence prediction studies due to their flexibility and high performance in non-linear data scenarios.

For the binary classification of influencers (predicting "Is_Influencer"), we employed **Support Vector Machines (SVM)** and **Logistic Regression with L1 and L2 regularization**. SVM, with a linear kernel, was chosen for its ability to find a hyperplane that best separates influencers from non-influencers in the feature space, handling high-dimensional data and ensuring generalizability. Logistic Regression models with L1 and L2 regularization were used to reduce model complexity and prevent overfitting. L1 regularization performs feature selection by driving coefficients of less relevant features toward zero, while L2 regularization penalizes large coefficients, yielding a simpler model. Logistic Regression is particularly interpretable, providing insight into feature impact on influencer likelihood, and complements the non-linear nature of the other models. Together, these classification models provide a balanced approach, capturing both linear and non-linear decision boundaries in the influencer classification task. Results indicate that the current dataset does not offer a strong predictive basis for distinguishing influencers based solely on psychological and demographic input features.

Results of Model Performance for Influence Prediction

The machine learning models used for predicting the Influence Index and classifying influencers versus non-influencers yielded varied results, reflecting moderate success in some cases and limited predictive power overall. The **Mean Absolute Error (MAE)** was the primary evaluation metric for regression models, measuring the average absolute difference between predicted and actual values. Lower MAE indicates better model performance. For the Influence Index prediction task, the Random Forest model achieved the best test MAE (0.1703), followed by Gradient Boosting (0.1719) and MLP (0.1808). Despite Random Forest's comparatively lower MAE, it still only marginally outperformed the baseline MAE of 0.1846, highlighting the need for additional data or feature refinement to improve predictive accuracy.

For the binary classification of influencer status, updated results indicate that **Gradient Boosting** slightly outperformed other models, achieving the highest **F1 Score** (0.3483), **Balanced Accuracy** (0.5894), and **AUC** (0.6881). However, **Logistic Regression with L1 and L2 regularization** showed comparable performance, with AUCs of 0.7017 and 0.7023, respectively, indicating moderate discrimination between influencers and non-influencers. Precision for Logistic Regression (L1: 0.6830, L2: 0.6401) was relatively high, but recall remained very low (L1: 0.1128,

L2: 0.1167), highlighting the difficulty of correctly identifying influencers. Meanwhile, **Random Forest** and **SVM** underperformed, with **SVM** showing no ability to discriminate between classes (F1 Score: 0.0000, Balanced Accuracy: 0.5000). Despite these results, the overall performance across models remains limited, emphasizing the need for additional data and feature engineering to enhance predictive power.

Influence Index Prediction	Binary Classification
Average Test MAE MLP: 0.1809 Random Forest: 0.1703 Gradient Boosting: 0.1719 **Baseline MAE: 0.1846**	**Model**: **Precision, Recall, F1, Balanced_accuracy, Auc** SVM : 0.0000, 0.0000, 0.0000, 0.5000, 0.6223 Logistic Regression (L1): 0.6830, 0.1128, 0.1799, 0.5412, 0.7017 Logistic Regression (L2): 0.6401, 0.1167, 0.1857, 0.5417, 0.7023 Random Forest: 0.4740, 0.1403, 0.2161, 0.5436, 0.6873 Gradient Boosting: 0.5185, 0.2628, 0.3483, 0.5894, 0.6881 (best)

Table 2: *Summary for the performance of 5 models for 2 tasks of Inflencer index prediction and binary classification of influencers and non-influencers*

The models indicate that the current dataset may not provide sufficient predictive power for distinguishing influencers based solely on psychological and demographic data. The limited performance of the models suggests that incorporating additional behavioral, social engagement, or network features, as well as further feature engineering, may be necessary to develop more accurate and robust models for identifying influencers.

4. Discussion and Conclusion

This study aimed to explore the psychological and demographic determinants of social media influence by developing predictive models to classify influencers and non-influencers based on an Influence Index. The findings highlight several key insights into the characteristics that distinguish influencers from non-influencers in the Iranian context.

Our results revealed that influencers exhibit significantly higher scores in both active and passive engagement. This aligns with previous studies that highlight the role of engagement in determining social influence (Lin et. al., 2024). A possible explanation for this result is the functionality of social media algorithms, which tend to prioritize users with higher overall activity, including both active interactions such as posting and commenting, as well as passive behaviors like viewing and browsing. This dual emphasis on engagement suggests that successful

influencers strategically leverage both active and passive elements to maximize visibility and reach.

Despite these insights, our machine learning models, including Logistic Regression, Random Forest, and Gradient Boosting, demonstrated limited predictive power in classifying influencers versus non-influencers using the current dataset. While some models, such as Gradient Boosting, achieved relatively higher F1 scores and AUC values, the overall performance suggests that psychological and demographic features alone are insufficient for accurate prediction. These findings emphasize the need for additional data and advanced feature engineering to capture the multifaceted nature of social influence more effectively.

Importantly, this study lays the groundwork for future research into developing more sophisticated metrics of social influence. The Influence Index introduced here provides a structured approach to quantify influence but can be further refined by incorporating additional dimensions, such as user-generated content quality, sentiment analysis, or network centrality metrics. Moreover, engineering new features—such as cognitive traits, behavioral patterns, and personality sub-dimensions—could significantly enhance the predictive capabilities of machine learning models.

To extend this work, future research should explore the inclusion of psychological and cognitive features beyond the NEO personality traits, such as decision-making styles, impulsivity, and attention span. Additionally, examining longitudinal data could provide insights into how social influence evolves over time, further enriching our understanding of influencers' traits and behaviors.

In conclusion, while the current study highlights the critical role of engagement in distinguishing influencers, it also underscores the complexity of modeling social influence. By building upon these findings and addressing the limitations, future studies can contribute to a more comprehensive understanding of what defines influence in the digital age.

Acknowledgment

Authors would like the Dr. Saeed Sadeghi for his insightful discussions and opinions in this research.

References

Abplanalp, S. J., Green, M. F., Wynn, J. K., et al. (2024). Using machine learning to understand social isolation and loneliness in schizophrenia, bipolar disorder, and the community. *Schizophrenia Research*, *10*, 88. https://doi.org/10.1038/s41537-024-00511-y

Angelini, F., Gini, G., Marino, C., & Van Den Eijnden, R. (2024). Social media features, perceived group norms, and adolescents' active social media use matter for perceived friendship quality. *Frontiers in Psychology*, *15*, Article 1222907. https://doi.org/10.3389/fpsyg.2024.1222907

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual.* Odessa, FL: Psychological Assessment Resources.

Ding, Y., et al. (2022). Dynamic tracking of state anxiety via multi-modal data and machine learning. *Frontiers in Psychiatry.* https://doi.org/10.3389/fpsyt.2022.757961

Firouzabadi, J., Jafari, K., & Khowrowabadi, M. (2024). Rethinking the metric for social media engagement: How cultural and temporal shifts shape online behaviors in Iran. *Scientific Reports* (under review).

Mehriar, M., & Ghazi Tabatabai, M. (2001). The use of NEO personality test and analysis of features and its factor structure among Iranian university students. *Journal of Humanities Research (Alzahra University), 11*, 173–198.

Gu, C., & Duan, Q. (2024). Exploring the dynamics of consumer engagement in social media influencer marketing: From the self-determination theory perspective. *Humanities and Social Sciences Communications, 11*, Article 587. https://doi.org/10.1057/s41599-024-03127-w

Jang, S., et al. (2022). Predicting personality and psychological distress using natural language processing: A study protocol. *Frontiers in Psychology.* https://doi.org/10.3389/fpsyg.2022.865541

Lampropoulos, G., Anastasiadis, T., Siakas, K., & Siakas, E. (2022). The impact of personality traits on social media use and engagement: An overview. *International Journal on Social and Education Sciences, 4*(1), 34–51.

Lin, H., Wang, C., & Sun, Y. (2024). How big five personality traits influence information sharing on social media: A meta-analysis. *PLOS ONE, 19*(6), Article e0303770.

Navarro, R., et al. (2022). Using machine-learning strategies to solve psychometric problems. *Scientific Reports.* https://doi.org/10.1038/s41598-022-23678-9

Pagani, M., & Mirabello, A. (2011). *Social Media Engagement Questionnaire.*

Pirzade, M., Peyvastegar, M., & Griffiths, M. D. (2024). Celebrity worship among adolescents is driven by neuroticism, avoidant identity style, and need to belong. *The Journal of Genetic Psychology, 185*(1), 1–14.

Sharifi Fard, S. A., Griffiths, M. D., Nabizadeh, S., Taheri, M., & Refaei, M. (2024). The relationship between five personality traits and happiness: The mediating role of problematic Instagram use. *Addiction & Health.*

Shamim, K., & Azam, M. (2024). The power of social media influencers: Unveiling the impact on consumers' impulse buying behaviour. *Humanities and Social Sciences Communications, 11*, Article 1461. https://doi.org/10.1057/s41599-024-03796-7

Walsh, C., et al. (2021). Applying machine learning approaches to suicide prediction using healthcare data: A conceptual framework. *Frontiers in Psychiatry.* https://doi.org/10.3389/fpsyt.2021.70791