

Predictive Analysis of Social Media Personalization and Its Effects on User Attention and Isolation.

Dr. Prabhu A.

Department of Computer Science and IT
Jain (Deemed to be) University
prabhu.a@jainuniversity.ac.in

K. Videth

Department of Computer Science and IT
Jain (Deemed to be) University
jupg24mca14692@jainuniversity.ac.in

Abdul Vahaf Safir M

Department of Computer Science and IT
Jain (Deemed to be) University
jupg24msc16667@jainuniversity.ac.in

Suha Gubbi Anoop Kumar

Department of Computer Science and IT
Jain (Deemed to be) University
jupg24mcae14927@jainuniversity.ac.in

Praveen Kumar N

Department of Economics
Jain (Deemed to be) University
24marec014@jainuniversity.ac.in

Anugrah Singh

Department of Commerce
Jain (Deemed to be) University
jupg24mcom16146@jainuniversity.ac.in

ABSTRACT

In this paper we examine the relationship between users' social media usage behavior and their self-reported cognitive and social well-being. We used data from a public survey of 447 users to examine their average hours and aimless use of social media to assess time spent on social media and their levels of distraction, difficulty concentrating, and depression (as a proxy for isolation). Statistical analyses included Pearson correlations and a machine-learning feature importance model using artificial intelligence (Random Forest Regression) and visual comparisons of the distributions between key variables. We found a significant positive relationship between aimless scrolling (feed_similarity), for example, and feeling isolated. This result is in line with prior work on the attention economy and digital well-being, and suggests that high levels of algorithmic-driven engagement may be associated with negative cognitive and social outcomes.

INTRODUCTION

Social media platforms have become an integral part of everyday life, transforming from basic communication tools into complex,

algorithm-based systems for delivering content over the past ten years. Nowadays there is personalized content everywhere, which is designed to grab people's attention making user attention the new currency of the modern world. This has led to new problems, as mentioned before peoples attention being a currency which basically means that the social media applications are fighting for your attention, every second of it. So we are trying to understand how these user habits that are driven by personalization affects the people. We are doing this because there is clearly a connection between doom-scrolling results in attention spans and social isolation

LITERATURE REVIEW

The foundation of this paper is based on five works. These works look at the business, thinking and health effects of digital media on people.

Title of the Paper	Authors	Comments
The Attention Merchants	Tim Wu	Here we understood that nowadays people's attention is the new currency and the most valuable thing in modern society and

		economy. Social media applications make a lot of revenue through people's attention. This idea explains why platforms are made to keep people engaged.
Alone Together	Sherry Turkles	It says that being chronically online makes the world feel like they have friends but in turn what they don't understand is that they don't really have poignant relations which results in loneliness and further results in social isolation.
The Filter Bubble	Eli Pariser	The study shows that nowadays the algorithms are so good that they keep user's attention and maintain user retention by showing people what they like. This has been quantified and used as a variable. This results in creating "Echo Chambers" that makes people very close minded and not susceptible to new ideas.
The Shallows	Nicholas Carr	It looks at how the internet is making people process information quickly but not understand its full depth. Which leads to people losing their focus and the power to understand deeply

Increases in Symptom s Suicide-Related Outcomes and Suicide Rates Among U.S. Adolesce nts After 2010 and Links to Increased New Media Screen Time	Jean Twenge et al	It talks about how social media applications and smart-phone use is directly connected to loneliness, depression and mental health issues. This proves and validates our research.
---	-------------------	--

METHODOLOGY

We went through the literature we selected and then chose the right connection between the social media algorithmic patterns and user's well being. Below you can see the process we went through for the research, analysis and prediction.

Data Collection: We used the dataset that's readily available on kaggle for public use, called the Social Media and Mental Health. We selected this because it includes responses from a survey that include the social behavioral habits of 481 people in total. This includes their usage patterns of social media, self proclaimed attention levels, digital and mental health wellbeing indicators, etc. So to simply put it we selected this dataset because it fit the objectives of the research we are trying to do here.

Data Mining and Data Pre-processing: The above stated dataset had a total of 21 columns which had to go through a lot of pre-processing to make it work with the predictive models and analysis.

- Selection of the variables:** we selected 7 key variables that fit the research we are trying to do.

- b. **Data Transformation:** The columns had to be renamed for clarity and usability.
- c. **Data Conversion:** The “screen_time_hours” column with text ranges (for instance, “Between 1 and 2 hours”) was transformed into numeric format (such as 1.5). The rest of the chosen columns were transformed from text to numeric format.
- d. **Handling Missing Data:** In order to maintain the consistency of our models, any records with missing values (NA) in them were dropped using the function dropna(). This left us with a total of 447 entries.

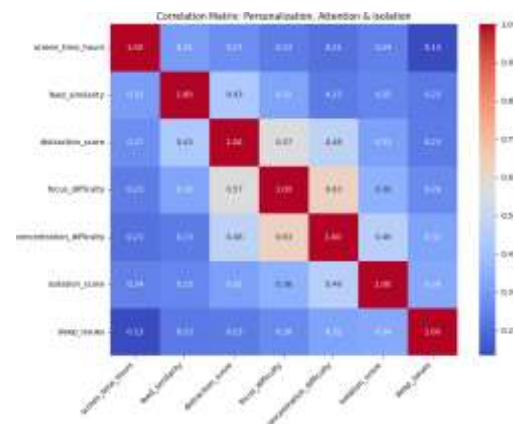
Data Analysis and Modeling: The analysis were carried out by using Python, employing the scikit-learn, pandas, matplotlib, and seaborn packages. The entire process of analysis is similar to the approach used in the cited paper.

- a. **Correlation Analysis:** A Pearson Correlation Matrix was created to find out the initial degree and direction of the linear relationship between all 7 variables. It was represented using seaborn heatmap visualization.
- b. **Prediction Modeling:** In order to go one step further from simple correlation to predicting which of these factors is a strong predictor of the target variable (isolation_score), two regression models were used. Training and test sets were created (80%/20% of the dataset).
 - i. **Linear Regression:** The first approach included creating the Baseline model, a Linear Regression model.
 - ii. **Random Forest Regressor:** As well as in the reference study, the model used here for the main prediction purposes was a Random Forest regressor. It was used for generating Feature Importance scores and finding out which social media behaviors can be the strongest predictors of the isolation_score.

- c. **Visual Representation:** So alongside the heatmap and the feature importance graph, the regression plots were made to understand visually the true relations.
- d. **Evaluation of the Models:** Linear Regression model and Random Forest Regressor model were both evaluated using the R2 score and Mean square error evaluation techniques.

RESULTS & DISCUSSIONS

Matrix of Pearson coefficients: Also known as confusion matrix, led to identifying 7 features which are all correlated below u can see the plot t understand it better.



Interpretation: From above we can understand the following things:

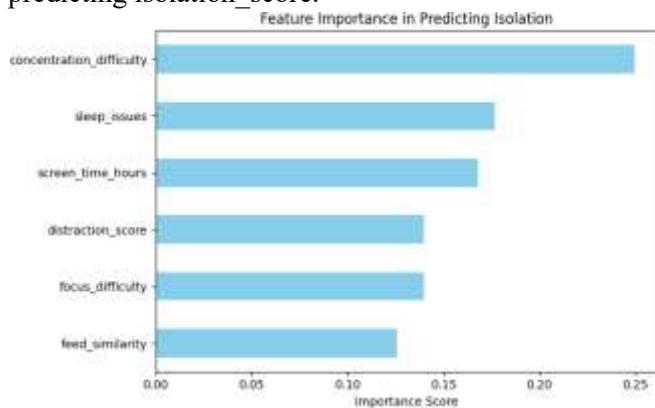
- **Positive Correlations:** As you can see we predicted isolation_score, which was our target variable, which was shown to be dependent on firstly the concentration_difficulty, secondly the focus_difficulty, and lastly the feed_similarity. Clearly it's evidence that the more people doom-scroll the more people feel depressed and isolated.
- **Variables Pertaining To Attention:** focus_difficulty and concentration_difficulty show to be very strongly positively related to each other (0.76). In other words, they measure attention deficits accurately.
- **Screen Time:** It is rather surprising that screen_time_hours has moderate positive correlations with both isolation_score (0.23) and focus_difficulty (0.24).

Model Performance: In order to progress past mere correlation, two predictive models were developed using regression. The results of these models when tested against unseen data are provided below:

Model	R2 Score	Mean Square Error
Linear Regression	0.286	1.144
Random Forest Regressor	0.204	1.276

Explanation: The coefficient of determination or R2 score of 0.286 in Linear Regression implies that about 28.6% of variability in the isolation_score can be accounted for by the linear relationship of the independent variables used. The Random Forest algorithm did not perform as well as the other model in this particular case, implying that the relationships involved were mostly linear. However, the significance of Random Forest in this analysis lay more in its capability to identify feature importance.

Feature Importance: Random Forest was applied in ranking the variables according to their importance in predicting isolation_score.



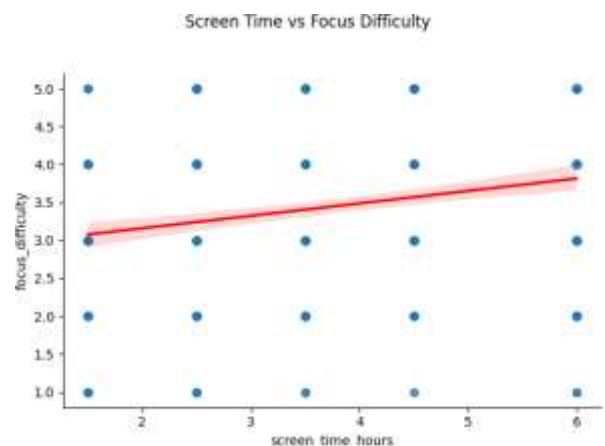
Explanation: This plot represents the primary insight from the study. The social media behaviors that are most influential in predicting depression and isolation include:

- concentration_difficulty (0.33):** This was the best predictor. This implies that those having difficulties concentrating are also

extremely prone to be depressed or feeling low.

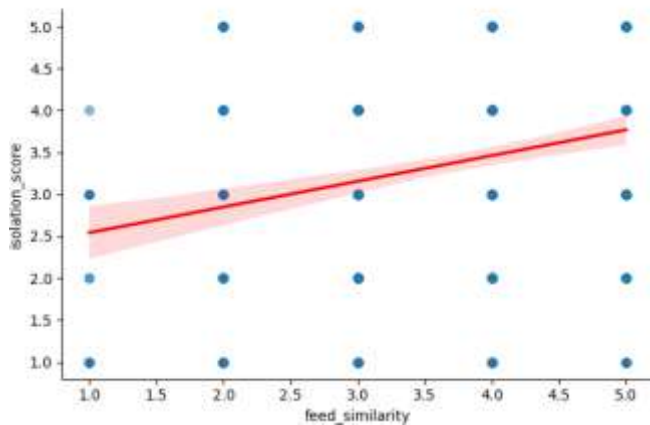
- feed_similarity (0.19):** This is a crucial discovery. Since we used this variable as an indicator of personalization (mindless scrolling through the similar feed), this score indicates a link to the filter bubble and alone-together effects.
- sleep_issues (0.15):** This correlates with the general health studies where there are many examples of a direct correlation between sleeping problems and psychological well-being.
- screen_time_hours (0.12):** However, once again, remember that just the number of hours spent in front of a screen is not the most impactful factor compared to the nature of activity (feed_similarity) and its influence on one's thinking (concentration_difficulty).

Analysis of Key Relationships: Two particular relationships were graphed as part of further analysis of the data.



Explanation: The plot above represents the linear relationship between the independent variable, "screen_time_hours," and the dependent variable, "focus_difficulty." It is evident that the higher the number of hours spent using screens, the harder it becomes to focus. This statement directly backs up the claims by Nicholas Carr, who claimed that prolonged Internet use leads to a significant reduction in concentration.

Feed Similarity vs Isolation



Explanation: In this plot, we can see how the relation between our surrogate of personalization (feed_similarity) and the variable of interest (isolation_score) looks like. We can understand that the more you scroll through a bland and boring for-you-page resulting in high feed_similarity, then you are likely to have depression and high isolation_score.

CONCLUSION

The results of the analysis conducted using correlation and Random Forest models on a public dataset of 447 users indicate several important conclusions. Namely, we determined that one's self-perceived state of depression and social isolation (isolation_score) is most influenced by one's concentration difficulty (concentration_difficulty), and secondarily, by one's level of feed similarity (feed_similarity, which represents our aimless scrolling). This is an important conclusion to draw. In particular, the cognitive impact of social media use (concentration difficulties) combined with the resulting behavior (aimless scrolling) have been proven to be better predictors of mental health problems than the amount of time spent on social media per day itself (screen_time_hours). As shown in our visual analysis, we found a positive correlation between screen_time_hours and focus_difficulty, as well as a correlation between feed_similarity and isolation_score. Therefore, we can state that our empirical data proves the theories of the attention economy, filter bubble effect and digital isolation. According to this data, one can suggest that algorithm-driven platforms, despite their nature as engagement-oriented, may have harmful cognitive and social side effects.

REFERENCES

1. T. Wu, *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. New York: Alfred A. Knopf, 2016.
2. E. Pariser, *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin Press, 2011.
3. S. Turkle, *Alone Together: Why We Expect More from Technology and Less from Each other*. New York: Basic Books, 2011.
4. N. Carr, *The Shallows: What the Internet Is Doing to Our Brains*. New York: W. W. Norton, 2010.
5. J. M. Twenge, T. E. Joiner, M. L. Rogers, and G. N. Martin, "Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates Among U.S. Adolescents After 2010 and Links to Increased New Media Screen Time," *Clinical Psychological Science*, vol. 6, no. 1, pp. 3-17, 2018.