Max Darling

MGT 665

Predicting Cognitive Level

Abstract

In today's digital age, increasing reliance on technology has raised concerns about its impact on cognitive function. Excessive screen time and constant digital distractions have contributed to shortened attention spans and impaired decision-making abilities. Continuous digital interruptions may reduce an individual's capacity to think deeply, often leading to more impulsive decisions (Frontiers in Cognition, 2023). The phenomenon known as "popcorn brain" further illustrates how sustained exposure to fast-paced digital media can hinder the brain's ability to focus and process information deeply (Real Simple, 2023).

To address this issue, I propose a machine learning model designed to predict an individual's cognitive performance based on key lifestyle factors, including sleep duration, stress levels, diet type, screen time, exercise frequency, and caffeine intake. By analyzing these variables, the model aims to detect patterns that correlate with cognitive performance levels, categorized as low, medium, or high. The dataset has 80,000 samples with diverse demographic attributes. The cognitive score that the machine will be learning from is calculated by a weighted formula. This approach offers a tool for individuals seeking to monitor and enhance their cognitive health.

Model development involves standard preprocessing steps, including data cleaning, standardization, and encoding, followed by training and testing using multiple algorithms such as multinomial logistic regression, random forest, decision tree, K-nearest neighbors, and XGBoost.

Introduction/Related Work

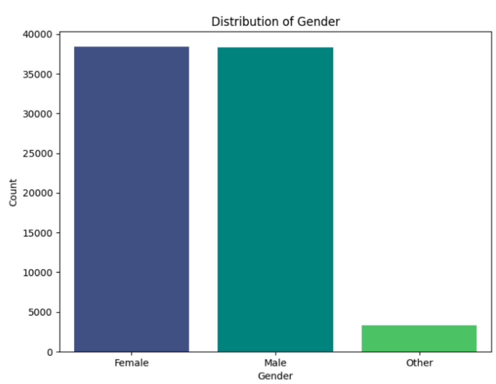
A person's cognitive level has many contributing factors to it. The cognitive process dimension is divided into six levels from low to high: remembering, understanding, applying, analyzing, evaluating, and creating (Anderson et al., 2001). The variables provided in this dataset contribute to these dimensions of cognitive level. Understanding these dimensions is not just an academic exercise, it has serious real-world implications, especially when considering broader societal trends in cognitive performance. Literacy rates in the United States present a troubling picture. Recent assessments show that more than 60% of students fail to read at grade level, with 54% of American adults reading below a sixth-grade level (US Literacy Statistics, 2024). These alarming statistics highlight a broader societal issue: diminished cognitive functioning may be more widespread than previously assumed. While literacy is only one measure of cognitive ability, it often reflects underlying deficits in memory, comprehension, and higher-order thinking skills, all of which are part of the cognitive process framework. Understanding the lifestyle and behavioral factors that correlate with cognitive levels is essential not just for individual awareness, but for shaping education policy, public health strategies, and workforce development. This report focuses on predicting cognitive level based on modifiable daily habits: sleep duration, stress levels, diet type, screen time, exercise frequency, and caffeine intake. By using the weighted formula to predict and showcase someone's cognitive level, we can better target interventions aimed at improving mental performance.

Methodology

The data being used consists of three categorical variables; gender, exercise frequency, and diet type. The remaining data is a float or an integer; sleep duration, stress levels, screen time, age, reaction time, and caffeine intake. After dropping unused columns such as AI predicted score and user ID, the three categorical variables will be encoded and the rest standardized. With no original missing values, when splitting the cognitive score given in the dataset into low, medium, and high, 165 NaN data points popped up. After removing these NaN data points, the following models will be trained and tested; multinomial logistic regression, random forest, decision tree, K-nearest neighbors, and XGBoost. These models will be evaluated by using accuracy, recall, precision, f1-score, and a confusion matrix.

Results

First, we will take a look at the exploratory data analysis of the dataset, showing correlations, and statistics of the variables.



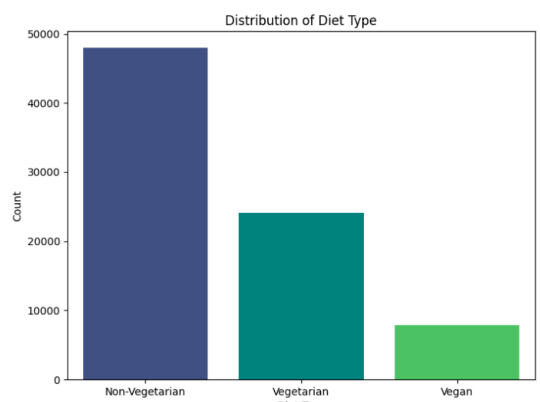


Figure 1

Distribution Chart of Gender

Figure 2

Distribution Chart of Diet Type

Figure 1 shows that we have a relatively even number of females and males, with a few ‘others’ for the gender variable. In **Figure 2**, there are almost 50,000 samples of Non-Vegetarian, around 25,000 Vegetarian samples, and a little less that 10,000 Vegan samples.

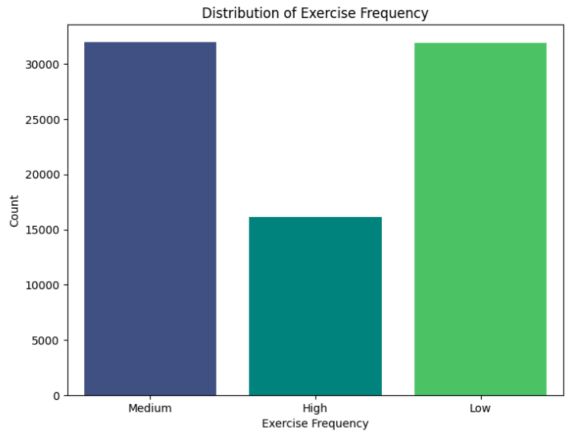


Figure 3

Distribution Chart of Exercise Frequency

Figure 3 displays over 30,000 samples having low and medium exercise frequency, with the remaining being high exercise frequency.

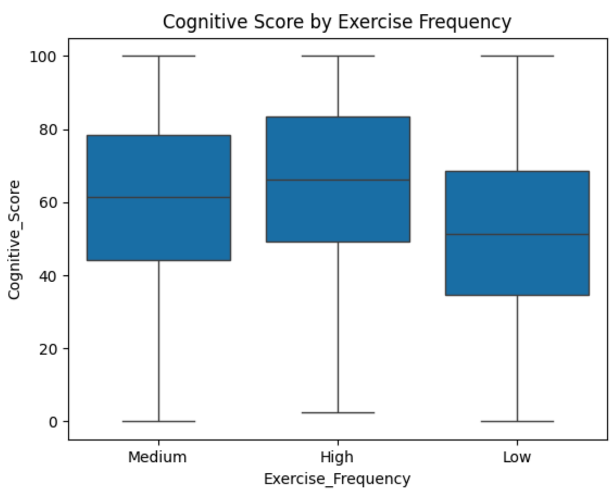


Figure 4

Boxplot of Cognitive Score by Exercise Frequency

Within **Figure 4**, is it clear that exercise frequency is a contributing factor to different cognitive levels. Looking at the high exercise frequency, the entire plot is higher than medium and low.

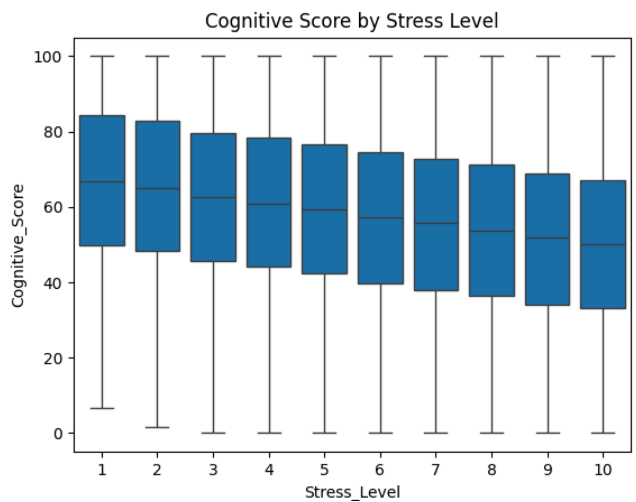


Figure 5

Boxplot of Cognitive Score by Stress Level

In **Figure 5**, we can see the correlation between these variables, as stress level goes up, the cognitive score boxplot goes down. Interesting to note that this decline seems to be linear.

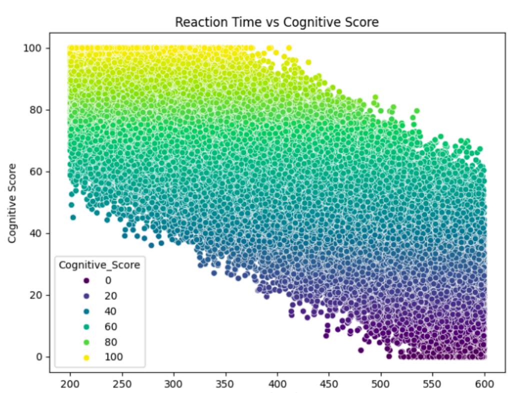


Figure 6

Scatterplot of Reaction Time vs Cognitive Score

Figure 6 shows a strong negative correlation between reaction time and cognitive score. As reaction time increases the cognitive score decreases.

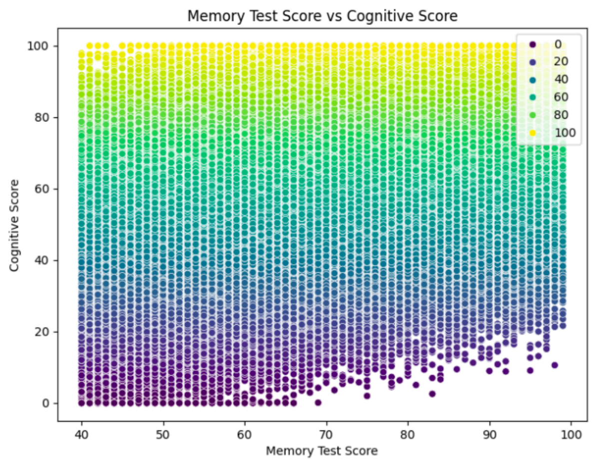


Figure 7

Scatterplot of Memory Test Score vs Cognitive Score

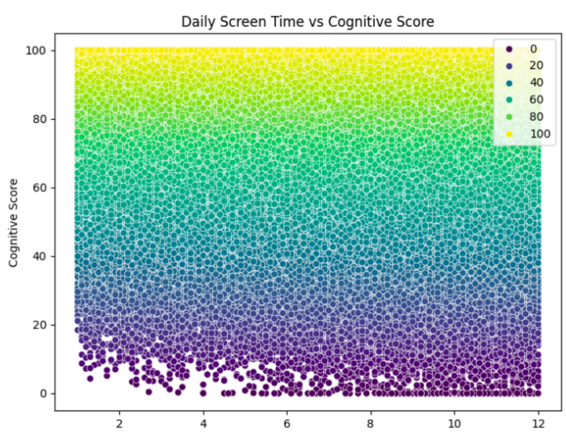


Figure 8

Scatterplot of Daily Screen Time vs Cognitive Score

Figure 7 and **Figure 8** shows the relationship of daily screen time and memory score to cognitive score. For **Figure 7**, as the memory test score reaches above 60, very low cognitive scores start to dissipate, forming a small correlation. Even when there is a very low memory score, we can see a few open spots in the high cognitive scores. In **Figure 8**, it is clear that as the screen time is less, there are less lower cognitive scores. Even though the plot does not show density it would be smart to assume that at these lower screen times, there is also an increase in higher cognitive scores.

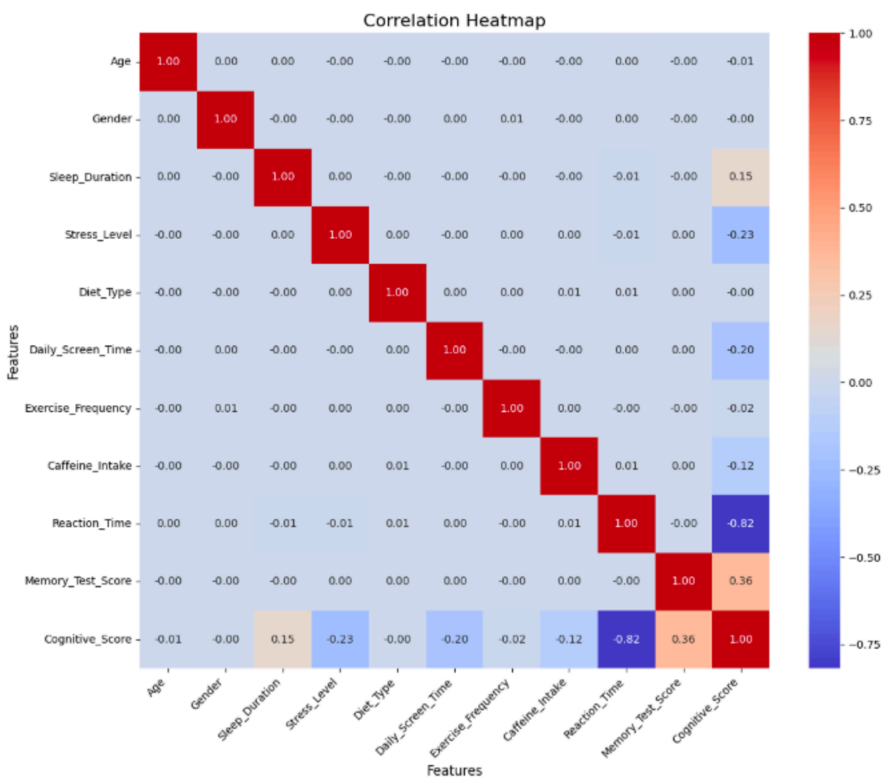


Figure 9

Correlation Heatmap

Figure 9 brings all of the previous analyses together. With no surprise, none of the variables are correlated besides with cognitive score. This heatmap proves the correlation assumptions from the previous figures. The top correlated variables with cognitive score are reaction time at -0.82, memory test score at 0.36, stress level at -0.23, daily screen time at 0.20, sleep duration at 0.15, and caffeine intake at -0.12. One thing to note, after seeing a correlation in **Figure 4,** the heatmap displays a -0.02 for exercise frequency.

Table 1

Evaluation Metrics of all Models

Evaluation Metrics

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.86	0.86	0.86	0.86
Decision Tree	0.89	0.89	0.89	0.89
Random Forest	0.94	0.94	0.94	0.94
XGBoost	0.97	0.97	0.97	0.97
K-Nearest Neighbors	0.88	0.88	0.88	0.88

Table 1 shows all evaluation metrics for all models used.

Discussion

The winning score was a 97% accuracy for the XGBoost model. With this model being the most complex, it comes at no surprise as the winner. Random Forest also had a great score of 94% across the board, scoring higher than the Decision Tree method, as Random Forest is a more intricate model using multiple decision trees to form its answer. Logistic Regression and K-Nearest Neighbors had the lowest scores with 86% and 88%, although not bad results, these scores were not as accurate as the other models. Some implications possible with this modeling is how the cognitive score is calculated originally by the data providers. They have a weighted formula that increases or decreases the computed cognitive score depending on what the variable is. With no scientific background provided to how this formula is weighted, there could be concerns with how accurate the given cognitive score is.

Conclusion

Having high confidence with an accuracy of above 85% in all models, predicting someone's cognitive score just got much easier. Cognitive health is a worsening situation and being able to expose it is step one. Assuming that the main implication is not a problem, this program can be used by any individual to assess their cognitive level and make changes to their lifestyle to improve themselves. It can also be used by doctors to develop specific plans to patients that score differently.

References

Frontiers in Cognition. (2023). *The impact of digital distractions on cognitive processing*. Frontiers in Cognition. https://www.frontiersin.org/journals/cognition/articles/10.3389/fcogn.2023.1203077/full

Real Simple. (2023, June 15). *Popcorn brain: How constant media exposure may be shortening your attention span—and how to refocus*. https://www.realsimple.com/popcorn-brain-shortening-attention-span-refocus-8600288

SamHarrison. (2022). *Human cognitive performance analysis* [Data set]. Kaggle. https://www.kaggle.com/datasets/samxsam/human-cognitive-performance-analysis

US Literacy Statistics. (2024). Mastermindbehavior.com. https://www.mastermindbehavior.com/post/us-literacy-statistics-842fb

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., and Wittrock, M. C. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives.* New York, NY: Longman.