Eye Tracking Data Analysis

August 19, 2023

Registration number: 2204567
Link to GitHub: https://github.com/laxmiprasannaadhurapalli/Empathy-Assessment

Executive summary (max. 200 words)	140
Main findings and Discussion (max. 600 words)	562
Conclusions (max. 300 words)	144
Total word count	846

Contents

Abstract

Hiring job candidates with a strong sense of empathy holds great significance for businesses, and this research delves into the potential of utilizing eye-tracking data for assessing empathy levels. Understanding the interplay between eye movements, cognitive functions, and overall performance is pivotal. The dataset consists of 502 CSV files from 60 individuals, encompassing a wide range of eye-tracking measurements. The process involved rigorous data preparation techniques, including feature selection, handling missing values, and data scaling. The refined dataset was employed to train a machine learning model, specifically the Random Forest algorithm. The model's effectiveness was gauged using metrics such as Mean Squared Error (MSE), R-squared, and Mean Absolute Error (MAE). The Random Forest model's capability to extract intricate patterns and insights from the data enhances the potential for more accurate predictions. This study underscores the utility of eye-tracking data in evaluating empathy during the recruitment of new employees.

1 Main Findings

1.1 Data Preprocessing:

The initial dataset consisted of 502 CSV files containing eye-tracking metrics collected from 60 participants. All these dataset files were combined and subjected to preprocessing to enhance its suitability for modeling. To streamline the data, irrelevant columns were removed, and missing values were managed using forward and backward filling. This approach was adopted based on feedback, which highlighted that eliminating null values in the time series was not appropriate. During the preprocessing stage, certain columns required modifications, such as replacing commas with periods to ensure consistency. Additionally, label encoding was applied to convert categorical features like eye movement type, validity left-right, and pupil diameter left-right from object format to float format. A notable observation was that the majority of participants exhibited "Fixation" as their predominant eye movement type, although other types like "Saccade" and "Unclassified" were also present. Interestingly, Participant 11 displayed the highest average score, while Participant 7 had the lowest average score in the dataset.

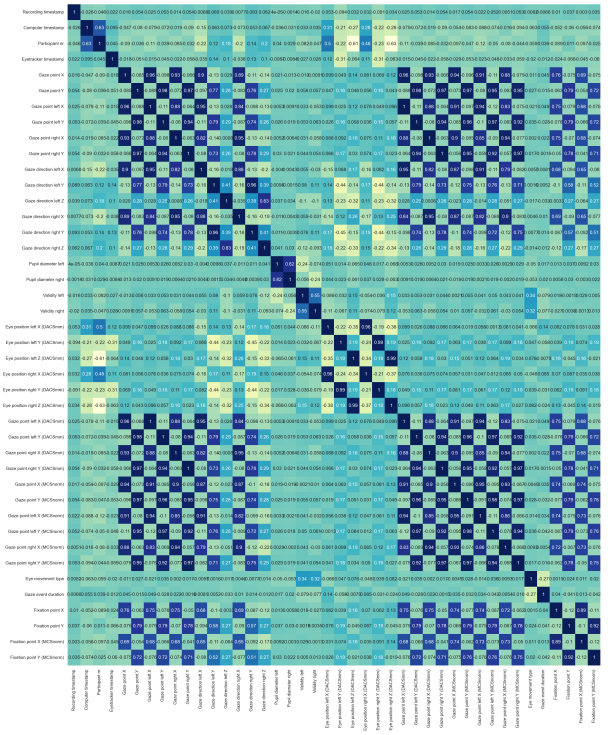

Figure 1: Correlation Heatmap

1.2 Model Performance

Utilizing the Random Forest model: The preprocessed data underwent training with the Random Forest model to forecast participants' scores, leveraging its capacity to grasp intricate patterns.

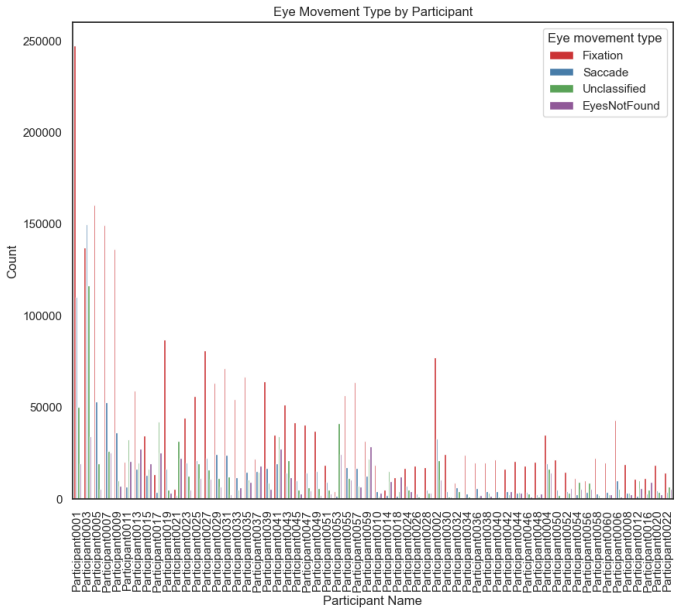

Figure 2: Participents Eye Type

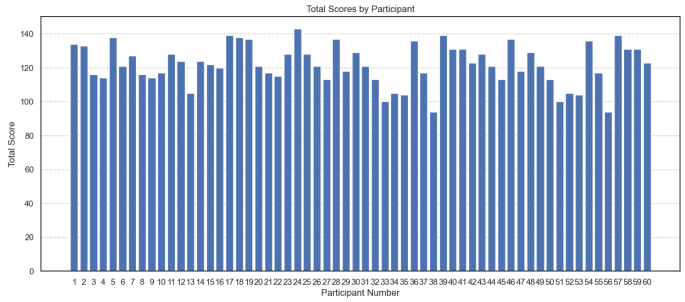

Figure 3: Total Scores by Participant

Outcome: The Random Forest model yielded an average Mean Squared Error (MSE) of 2.87, an R-squared value of 0.9828, and a Mean Absolute Error (MAE) of 1.03.

The dataset was divided into training and testing sets, with 20 percent of the data allocated for testing. Given the nature of the problem as a time series, a random state was not employed during data splitting to avoid randomness. The performance evaluation of the models took place on the testing set to gauge their effectiveness on previously unseen data.

2 Discussion

I used a program called the Random Forest model to help us understand how participants' scores can be predicted. This model is good at finding hidden patterns in the data. After training the model with our prepared data, we found that, on average, its predictions were off by about 2.87 points from the actual scores. I also discovered that the model's predictions were quite accurate, showing a similarity of about 98.28% to the real scores. On average, the model's guesses were off by around 1.3 points from the actual scores. To make sure our results were trustworthy, I divided our dataset into two parts: one for training the model and the other for testing how well it works. We put aside 20% of the data for testing. When we evaluated the model using this test set, it performed well, even on new data it hadn't seen before. This research suggests that we can use information from people's eye movements to make good guesses about their empathy levels. This could be a helpful tool when businesses are choosing new employees, as it gives us a new way to understand how well candidates might connect with others.

3 Conclusions

This study has demonstrated the potential of using eye-tracking data to assess empathy levels in individuals. By employing the Random Forest model, we were able to predict participants' scores with a good degree of accuracy. The results indicate that eye movements can provide valuable insights into a person's empathetic tendencies. These findings have significant implications, particularly for businesses aiming to recruit candidates with strong empathetic qualities. The ability to evaluate empathy through eye-tracking data adds a novel dimension to the selection process, enhancing the understanding of how well potential employees might interact and connect with others. While further research and refinement are necessary, this study opens the door to a new avenue for evaluating and enhancing empathetic qualities in individuals, ultimately contributing to more effective candidate assessment and potentially fostering more empathetic workplaces.

[**01**] [**02**] ref