Using Eye-Tracking Data to Determine Empathy Scores in Recruitment: A Predictive Modeling Approach

April 24, 2023

Registration number: 2202258
Link to GitHub: https://github.com/famesippa/CE888_datascience.git

Executive summary (max. 200 words)	112
Main findings and Discussion (max. 600 words)	595
Conclusions (max. 300 words)	165
Total word count	872

Contents

Abstract

The aim of this study was to examine the signals recorded by the eye-tracker device, then process and analyse the data to find out whether and how they can be used to determine an empathy score for HR. The experimental data was collected from the other research, which provided eye-tracker data for multiple participants through questionnaires to determine empathy scores before and after the experiment. The study will pay attention to seeking corelated data with empathy scores in order to generate the prediction score from a rational prediction model. Which can explain why low and high scores for job interviewees are used by HR to make hiring decisions in the recruitment process.

1 Main Findings

1.1 Methodology

We used the dataset and data from "EyeT4Empathy" [1] and the analysis method from "Making sense of data" [2] to approach the main goal of this study. Following the four main steps of problem definition, data preparation, implementation, and deployment, we got started. Briefly, we began by learning about the entire dataset, including the experimental procedure, the variables they produced. The two primary features that are generated, according to the experiment, are fixations and saccades, two of the two main parts of gaze dynamics. Fixations are the times when the eyes are fixed on a specific area of the image [3]. The size of the pupil is another characteristic of human eyes. Its alterations are related to our general cognitive states as well as how quickly information is processed [4].

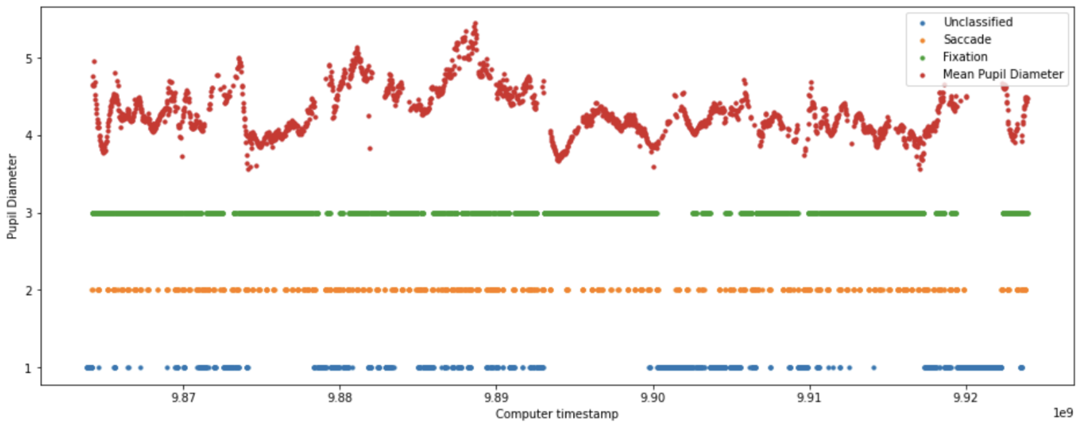

Figure 1: Sample of comparing gaze types and pupil diameter from one recording on one participant to .

After exploring the data is data preparation. The dataset was processed by cleaning the noise data, removing unnecessary data, and transforming the gaze types data to new features by faction fixation and saccade so that we can generate the final dataset, figure 2, which includes relevant data to use for prediction, by dividing it into train and test datasets at 80% and 20%, respectively.

Implementing the analysis is the next step. To find potential features for a prediction, we again performed feature selection by searching relationships using an SNS heatmap plot figure 3. Furthermore, to evaluate, it is necessary to compare accuracy before and after selection. This experiment will select feature which correlation value greater than 0.5. A statistical measurement of the linear relationship between two variables is provided by the heatmap.

Making predictions is a part of this task as well. It is critical to select an appropriate regression model by comparing the error, MEA (mean absolute error), MSE (mean square error), and R-square for each model. Additionally, the KFold cross-validation method was used in each prediction process because it may help in reducing overfitting. We use 10-fold cross-validation to improve performance, which will typically give a more accurate estimate of model performance as it uses more data for training and testing.

1.2 Result

The result is displayed in table 1. It displays the average prediction error. For DecisionTreeRegressor, MSE:55.25, MAE:2.4, R2:0.5 on non-selected features; MSE:49.22, MAE:1.97, R2:0.64 on

	Participant name	Total Score extended before	Total Score extended after	Recording name	Pupil diameter mean	Pupil diameter max	Pupil diameter min	Pupil diameter std	Pupil diameter mean total	Pupil diameter std total	Fixation fraction	Saccade fraction	Score diff	Group
0	8	98	116	1	4.285498	5.445	3.570	0.310255	4.133947	0.118013	0.624585	0.135797	18	0
1	8	98	116	2	4.159457	5.030	3.280	0.209858	4.133947	0.118013	0.772194	0.113765	18	0
2	8	98	116	3	4.080679	5.130	3.115	0.313496	4.133947	0.118013	0.697556	0.138379	18	0
3	8	98	116	4	4.010152	4.970	3.265	0.197150	4.133947	0.118013	0.569275	0.117118	18	0
4	10	100	117	1	3.253155	4.640	2.620	0.237729	3.282211	0.062083	0.746608	0.161174	17	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
295	59	136	131	8	2.956741	3.825	2.435	0.159706	2.954143	0.082884	0.374499	0.160144	-5	1
296	60	80	123	1	3.210304	3.660	2.665	0.123036	3.225335	0.067988	0.624483	0.128757	43	0
297	60	80	123	2	3.211766	3.640	2.850	0.108543	3.225335	0.067988	0.732936	0.138031	43	0
298	60	80	123	3	3.158862	3.765	2.505	0.133811	3.225335	0.067988	0.705426	0.120709	43	0
299	60	80	123	4	3.320406	3.690	2.740	0.128477	3.225335	0.067988	0.698769	0.130506	43	0

300 rows × 14 columns

Figure 2: This is final dataset ready to train model .

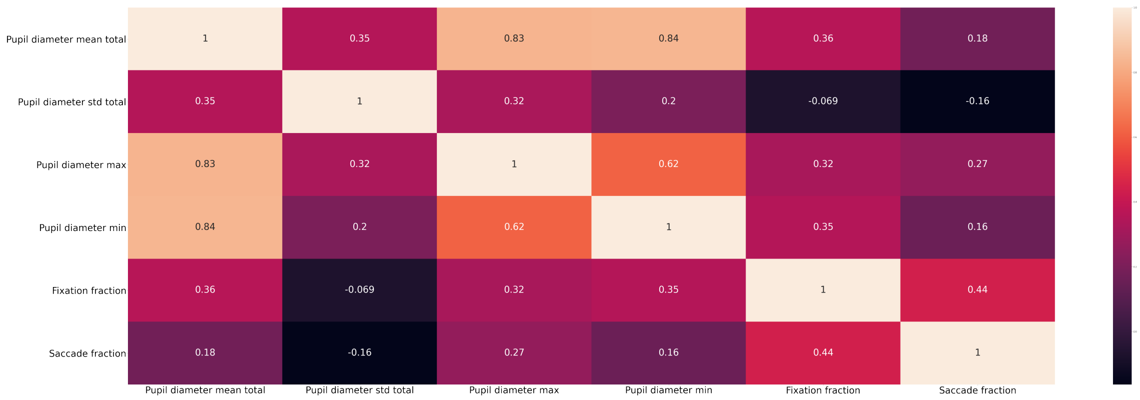

Figure 3: SNS Heatmap shows correlation coefficient.

selected features; and for the feature CatBoostRegressor model, MSE:15.65, MAE:2.47, R2:0.78 on non-selected features; and MSE:24.27, MAE:3.47, R2:0.82 on selected features. The box-plot on figure 4 shows the different between both potential models is variance which means CatBoost-Regressor can provide more accurate each time.

Another finding is the poor correlation between empathy score and certain features, such as pupil diameter, fixation fraction, and saccade fraction, show in table 2. which lacks eye-tracker information and a reflex score.

2 Discussion

Based on the results shown in Table 1, it can be observed that the lower MSE and MAE values and higher R2 value for the DecisionTreeRegressor model compared to the non-selected features clearly show that performance improved when using the selected features. The model's overall performance is still less precise than the CatBoostRegressor model, though. CatBoostRegressor modifies ordered boosting to prevent over-fitting because it is based on the theory of decision trees and gradient boosting [5][5]. The box-plot in Figure 4, it also supports accuracy between the two models.

Limiting participants had an impact on the prediction, and the raw data is full of unused and noisy components. So, only about 50 people were used to correct the eye-tracking data after the data had been cleaned. More precise participant information is required to increase accuracy, and each model's parameters must be tuned to produce the best outcome.

3 Conclusions

The findings of this study indicate that eye-tracking data can be used to determine an empathy score for human resources, which can be used to make more informed hiring decisions during the recruitment process. Multiple participants' empathy scores were determined using questionnaires,

Regression model	Average Prediction Error					
	Non-selected feature			Selected feature		
	MSE	MAE	R2	MSE	MAE	R2
DecisionTreeRegressor	55.25	2.40	0.50	49.22	1.97	0.64
LinearRegression	152.43	9.95	0.29	151.50	9.97	0.05
GradientBoostingRegressor	13.47	2.54	0.70	30.04	3.98	0.76
ElasticNet	151.88	10.05	0.22	151.88	10.05	0.03
SVR	149.99	9.95	0.27	149.94	9.91	0.06
BayesianRidge	152.49	10.08	0.30	152.72	10.09	0.05
CatBoostRegressor	15.65	2.47	0.78	24.27	3.47	0.82
XGBRegressor	20.50	3.10	0.66	43.36	4.80	0.69
LGBMRegressor	24.09	3.31	0.72	55.35	5.65	0.65

Table 1: Average error value for each models

Empathy score	Correlation coefficient							
	Pupil diameter mean	Pupil diameter max	Pupil diameter min	Pupil diameter std	Pupil diameter mean total	Pupil diameter std total	Fixation fraction	Saccade fraction
Before experiment	-0.32	-0.15	-0.41	0.19	-0.33	-0.21	-0.33	0.1
After experiment	0.02	0.073	-0.033	0.089	0.021	0.04	0.0095	-0.042

Table 2: Correlation coefficient each features

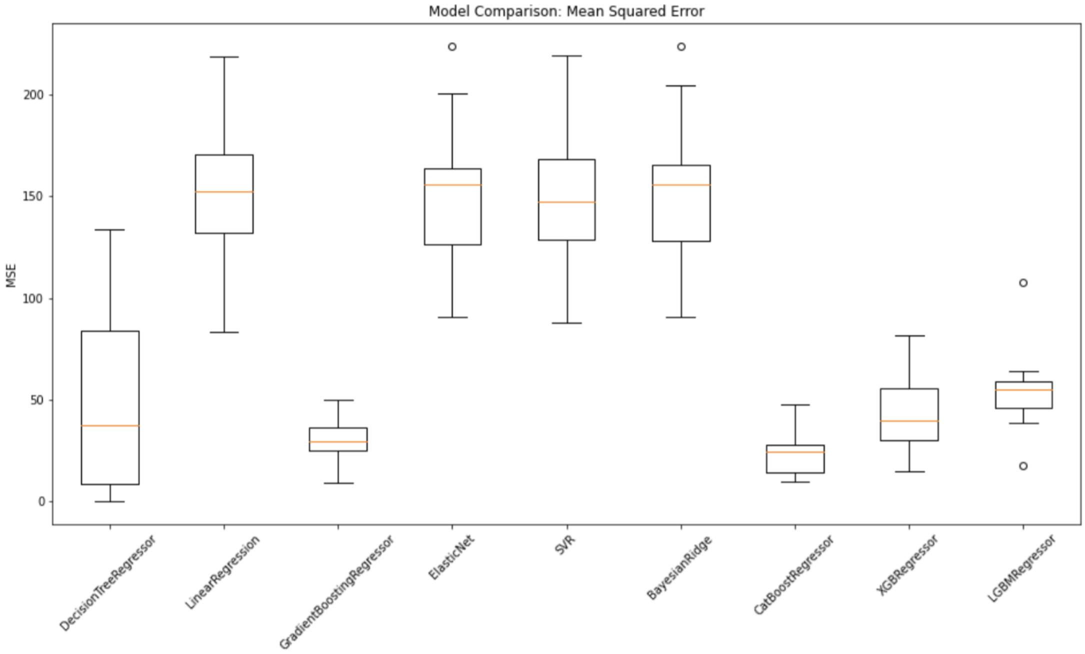

Figure 4: Example Box-plot MSE for each regression models.

which were then processed and analysed to identify correlations with eye-tracking data. The focus of the study was to develop a predictive model capable of generating a sensible prediction score based on the correlated data.

The evaluation of the predictive models revealed that the CatBoostRegressor model was more accurate than the DecisionTreeRegressor model. In addition, the study revealed the need for additional research to optimise the models and increase the number of participants in order to improve the accuracy of the predictions.

The study has provided valuable insights into the potential use of eye-tracking data in the recruitment process to determine empathy scores. These results have substantial ramifications for HR hiring decisions. However, empathy scores are not the only determinant of interviewer performance.

References

[1] P. Lencastre, S. Bhurtel, A. Yazidi, G. Mello, S. Denysov, and P. Lind, "Eyet4empathy: Dataset of foraging for visual information, gaze typing and empathy assessment," *Scientific Data*, vol. 9, p. 12, 2022.

[2] G. J. Myatt, *Making sense of data: a practical guide to exploratory data analysis and data mining*. John Wiley & Sons, 2007.

[3] C. Antoniades, U. Ettinger, B. Gaymard, I. Gilchrist, A. Kristjansson, C. Kennard, S. Noachtar, J. Sweeney, and N. Smyrnis, "An internationally standardised antisaccade protocol," *Vision research*, vol. 84, pp. 1–5, 2013.

[4] S. C. Otero, B. S. Weekes, and S. B. Hutton, "Pupil size changes during recognition memory," *Psychophysiology*, vol. 48, no. 10, pp. 1346–1353, 2011.

[5] S. Thiesen, "Catboost regression in 6 minutes. a brief hands-on introduction to...," 2019. Accessed: 2023-04-24.