Behavioral Patterns in Smartphone Usage Predict Big Five Personality Traits

Team Members

Ayush Das(IMT2019014)
Balkaran Singh(IMT2019016)
Daksh Agarwal (IMT2019505)
Rakshit Gupta (IMT2019516)

Under the guidance of Prof. Uttam Kumar and Dr. Tanujit Chakraborty

Abstract:

The paper focuses on understanding, quantification, and evaluating individual differences in behavior, feelings, and thoughts. They analyzed behavioral data obtained from 743 participants in 30 consecutive days of smartphone sensing (25,347,089 logging events). They computed variables (15,692) about individual behavior from five semantic categories (communication & social behavior, music listening behavior, app usage behavior, mobility, and general day- & nighttime activity). Using a machine learning approach (random forest, elastic net), they showed how these variables can be used to predict self-assessments of the big five personality traits at the factor and facet level. The results reveal distinct behavioral patterns that proved to be differentially predictive of big five personality traits. Overall, this paper shows how a combination of rich behavioral data obtained with smartphone sensing and the use of machine learning techniques can help to advance personality research and can inform both practitioners and researchers about the different behavioral patterns of personality.

Dataset:

There were 15694 variables in the dataset which roughly corresponded to the behavioral categories of communication, app usage, music consumption, general day- and nighttime activity (day- and nighttime dependency was treated as a distinct category in the analyses), and mobility along with 35 (5 factors and 30 facets) personality criteria. Gender, age, and education were solely used for descriptive statistics and were not included as predictors in any of the models

Preprocessing:

We applied a series of pre-processing steps before the machine learning analysis. First, we removed ids with 0 unique apps.(i.e ids that have 0 "daily_mean_num_unique_apps"). Furthermore, we excluded variables with less than 2 % unique values - as they would not add much information to the modeling process. In this step, we reduced the initial number of 15,692 variables to the final dataset of 1852 variables. Finally, we eliminated extreme outliers by excluding data points that are unreasonably far (greater than 100 times the median absolute deviation) from the sample median. This was done to minimize the impact of possible logging errors on the modeling process.
Finally, the dataset was merged with 35 (5 factors and 30 facets) personality criteria along with age, gender, and education.
The final dataset consists of 1852 predictor variables along with 35 (5 factors and 30 facets) personality criteria.

Models used:

We used Elastic net regularized linear regression and Non-linear tree-based random forest models. We trained machine learning models for the prediction of all personality factors. For model-benchmarking, we compared the predictive performance of elastic net regularized linear regression models with those of nonlinear tree-based random forest models. We performed additional pre-processing and hyperparameter tuning, using 5-fold cross-validation.

Evaluations:

We evaluated the predictive performance of the models based on the Pearson correlation (r) between the predicted values and the person-parameter trait estimates from the self-reported values of the respective personality trait variables. Additionally, we considered the root mean squared error (RMSE) and the coefficient of determination (R2) as measures of predictive performance.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

r = correlation coefficient

x_i = values of the x-variable in a sample

$\bar{x}$ = mean of the values of the x-variable

y_i = values of the y-variable in a sample

$\bar{y}$ = mean of the values of the y-variable

Descriptive Statistics:

Traits	Max	Min	SD	Median	Mean
O	2.1176	-1.9994	0.705676	-0.05054	-0.10515
C	2.2897	-2.2831	0.74396	-0.08094	-0.1399
E	2.131	-2.4442	0.741908	-0.00989	-0.0291
A	2.6439	-2.1099	0.752646	-0.05501	-0.1329
ES	2.5201	-2.2386	0.717534	-0.01712	-0.00855
ES1.Carefreeness	4.302	-4.302	1.302694	0.153585	0.268
ES2.Equanimity	5.021	-2.609	1.073448	0.590037	0.492
ES3.Positive.mood	5.595	-5.783	1.422575	0.946872	0.834
ES4.Self.consciousness	3.896	-3.572	1.172371	0.645814	0.776
ES5.Self.control	5.147	-3.554	1.00407	0.657779	0.739
ES6.Emotional.robustness	5.526	-3.778	1.183907	0.669119	0.786
E1.Friendliness	5.413	-1.912	1.27687	1.437697	1.37
E2.Sociableness	5.644	-4.495	1.752019	1.284502	1.283
E3.Assertiveness	5.613	-3.345	1.365761	0.442942	0.473
E4.Dynamism	5.94	-4.106	1.590665	1.216849	1.154
E5.Adventurousness.	5.27	-4.397	1.480331	0.434923	0.554
E6.Cheerfulness	6.088	-3.233	1.654809	1.969987	1.85
O1.Openness.to.imagination	5.329	-2.295	1.378862	1.259788	1.305
O2.Openness.to.aesthetics	4.61	-4.551	1.285815	0.375484	0.288
O3.Openness.to.feelings	6.041	-5.648	2.06622	2.010458	2.001
O4.Openness.to.actions	5.419	-2.745	1.397742	1.358319	1.443
O5.Openness.to.ideas	5.515	-1.631	1.425247	1.648909	1.584
O6.Openness.to.the.value.and.norm.system	4.863	-3.543	1.027483	0.917306	0.94
C1.Competence	5.665	-3.322	1.209091	0.847303	0.722
C2.Love.of.order	5.669	-4.34	1.554426	1.11267	1.16
C3.Sense.of.duty	5.504	-1.593	1.403827	1.94524	1.782
C4.Ambition	5.857	-2.391	1.673567	1.833244	1.902
C5.Discipline	5.747	-3.608	1.468755	1.478433	1.496
C6.Caution	5.747	-2.509	1.332855	1.524417	1.452
A1.Willingness.to.trust	5.416	-4.144	1.411357	0.366035	0.194
A2.Genuineness	4.246	-1.559	0.945509	0.999872	0.862
A3.Helpfulness	6.045	-2.474	1.373177	1.646816	1.667
A4.Obligingness	5.546	-1.864	1.29575	1.153354	1.046
A5.Modesty	5.106	-2.682	1.131235	0.801046	0.734
A6.Good.naturedness	6.402	-2.993	1.771593	2.134708	2.094

These are descriptives of demographic and personality trait variables for the 629 participants

Results:

The results show that levels of big five personality traits were successfully predicted from records of smartphone usage for the majority of factors and facets. The results also show that the non-linear random forest models on average outperformed the elastic net models in both prediction performance and the number of successfully predicted criteria

Openness

	Random Forest			Elastic Net	
	R score	R2 score	RMSE score	RMSE	R2 score
O	**0.225698**	**0.023989**	**0.682878**	**0.756787**	**0.027452**
O1 Openness to imagination	0.105548	-0.00977	1.386158	1.476162	-0.005648
O2 Openness to aesthetics	**0.192659**	**0.043862**	**1.246087**	**1.379363**	**0.054189**
O3 Openness to feelings	**0.155472**	**0.030383**	**2.117509**	**2.068301**	**0.017369**
O4 Openness to actions	**0.191268**	**0.036129**	**1.390994**	**1.663455**	**0.017826**
O5 Openness to ideas	**0.112265**	**0.045988**	**1.339332**	**1.464892**	-0.005473
O6 Openness to the value and norm system	0.159148	-0.009784	1.057639	1.003974	0

Conscientiousness

	Random Forest			Elastic Net	
	R score	R2 score	RMSE score	RMSE	R2 score
C	**0.227415**	**0.042322**	**0.763932**	**0.861514**	**0.011598**
C1 Competence	0.117269	0.007468	1.286418	1.409460	-0.017536
C2 Love of order	**0.231498**	**0.065125**	**1.578964**	**1.576171**	**0.021498**
C3 Sense of duty	**0.204856**	**0.030567**	**1.370139**	**1.453722**	0
C4 Ambition	**0.168715**	**0.043796**	**1.743762**	**1.716321**	**0.021249**
C5 Discipline	**0.181489**	**0.045685**	**1.485048**	**1.813091**	**0.017435**
C6 Caution	0.207596	-0.02134	1.347841	1.400674	-0.017469

Extraversion

	Random Forest			Elastic Net	
	R score	R2 score	RMSE score	RMSE	R2 score
E	**0.285598**	**0.093849**	**0.678736**	**0.801199**	**0.041489**
E1 Friendliness	0.254896	-0.02218	1.344724	1.416509	-0.017458
E2 Sociableness	**0.151589**	**0.162531**	**1.793164**	**1.831160**	**0.054598**
E3 Assertiveness	**0.287458**	**0.045303**	**1.347941**	**1.391931**	**0.045269**
E4 Dynamism	**0.262256**	**0.052893**	**1.615476**	**1.708976**	**0.037415**
E5 Adventurousness.	**0.257748**	**0.059575**	**1.539965**	**1.704921**	**0.038469**
E6 Cheerfulness	0.112659	0.032745	1.648299	1.781033	**-0.005335**

Agreeableness

	Random Forest			Elastic Net	
	R score	R2 score	RMSE score	RMSE	R2 score
A	**0.075863**	**0.005964**	**0.75004**	**0.774775**	**-0.025589**
A1 Willingness to trust	0.035521	-0.03672	1.413499	1.440454	-0.021459
A2 Genuineness	0.014812	-0.0327	0.988917	0.903919	-0.021114
A3 Helpfulness	0.027836	-0.01755	1.358526	1.492525	-0.016572
A4 Obligingness	0.072489	-0.01659	1.332774	1.349770	-0.023368
A5 Modesty	0.037173	-0.01173	1.234979	1.250199	-0.017787
A6 Good naturedness	0.033698	-0.03141	1.677354	1.874083	-0.022941

Emotional Stability

	Random Forest			Elastic Net	
	R score	R2 score	RMSE score	RMSE	R2 score
ES	**0.094125**	**0.017682**	**0.77792**	**0.745206**	**0.005441**
ES1 Carefreeness	**0.197365**	**0.06664**	**1.231299**	**1.345344**	**0.021459**
ES2 Equanimity	0.085529	-0.02205	1.112855	1.133538	-0.027789
ES3 Positive mood	**0.097436**	**0.005697**	**1.459127**	**1.436669**	**-0.005663**
ES4 Self consciousness	**0.324598**	**0.041027**	**1.089926**	**1.180353**	**0.021148**
ES5 Self control	0.127369	-0.03641	0.95536	1.150276	-0.035769
ES6 Emotional robustness	0.061576	-0.05674	1.282224	1.237314	-0.044189

Comparison between random forest and elastic net linear regression models.

Comparison of R2 scores in Random Forest and ElasticNet
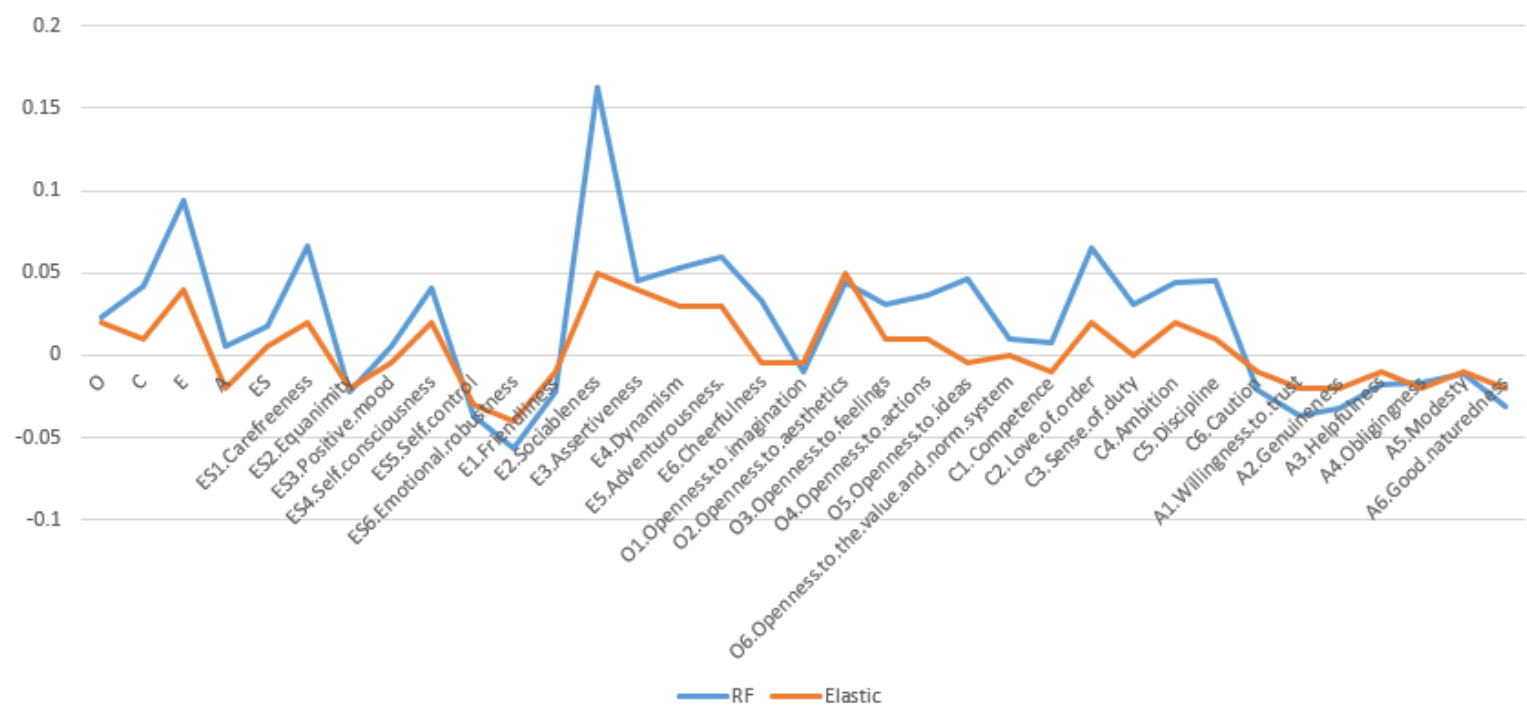
RMSE comparision b/w RandomForest and ElasticNet
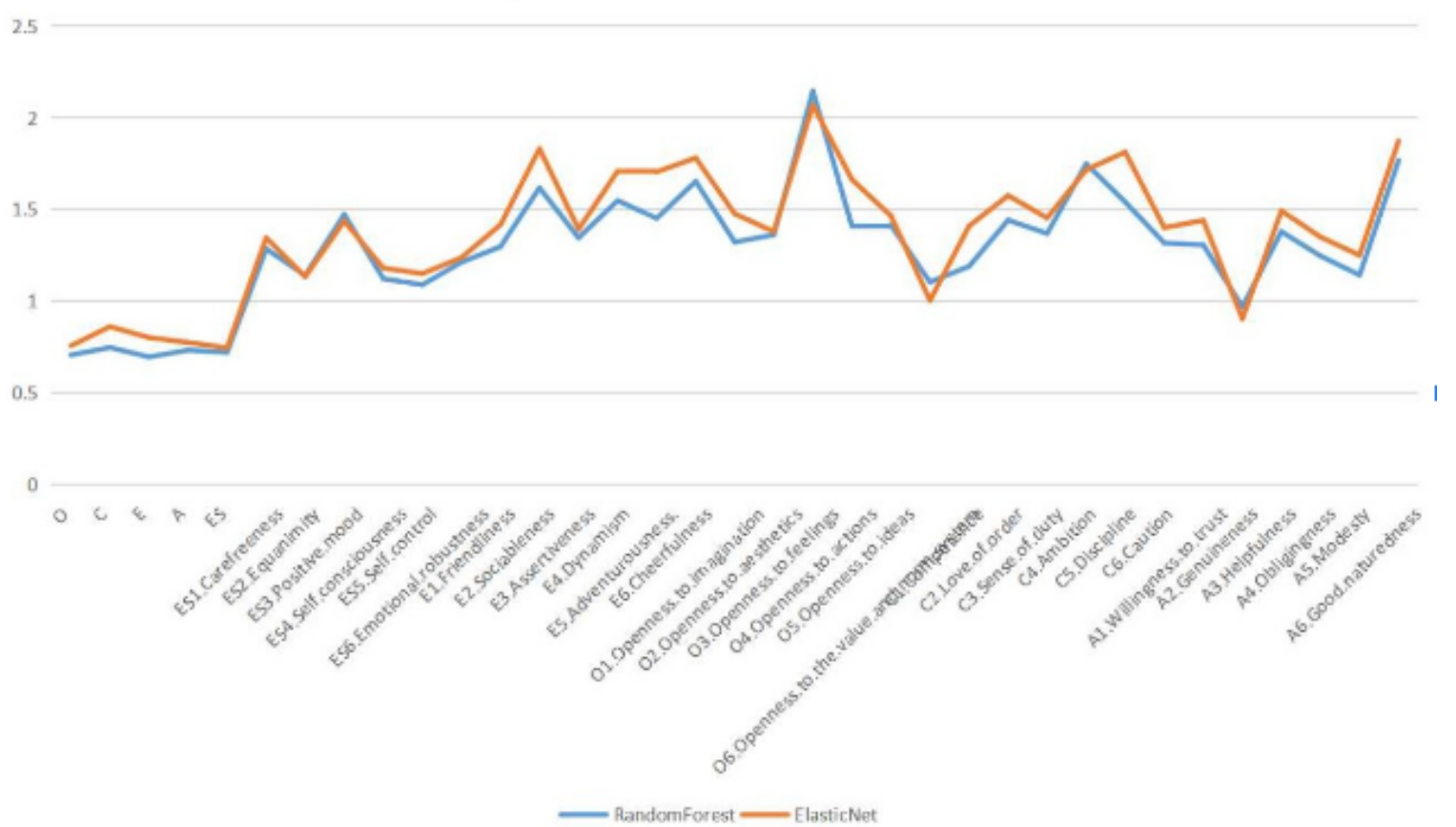
Conclusion:

Personality trait levels can be predicted from behavior, collected from smartphones. The paper demonstrated that distinct patterns of semantically different behaviors are predictive for individual levels of big five personality traits. The findings suggest that everyday behavior on smartphones can allow for inferences on individual personality trait levels. It also suggests that research in behavioral sciences and personality psychology, in particular, could benefit from the inclusion of diverse behavioral indicators. The results also show that the non-linear random forest models on average outperformed the elastic net models in both prediction performance and the number of successfully predicted criteria

Link to the project repository can be found <u>here</u>

References

<u>Stachl, Clemens, et al. "Predicting Personality from Patterns of Behavior Collected with Smartphones." Proceedings of the National Academy of Sciences, vol. 117, no. 30, 28 July 2020, pp. 17680–17687</u>

<u>"Repository: Predicting Personality from Patterns of Behavior Collected with Smartphones." Osf.io, 12 Oct. 2018, osf.io/kqjhr/, 10.17605/OSF.IO/KQJHR. Accessed 17 May 2022.</u>

<u>Brownlee, Jason. "How to Develop Elastic Net Regression Models in Python." *Machine Learning Mastery,* 6 Oct. 2020</u>