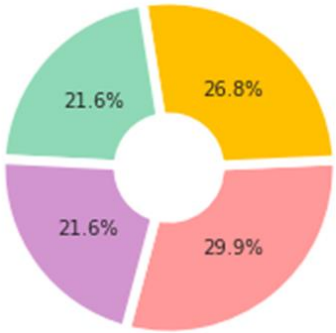
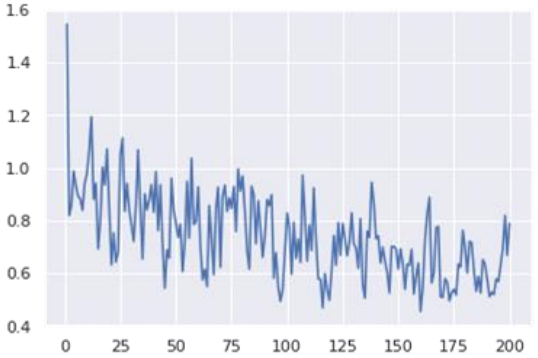
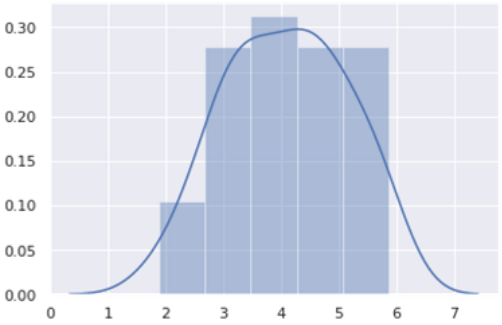


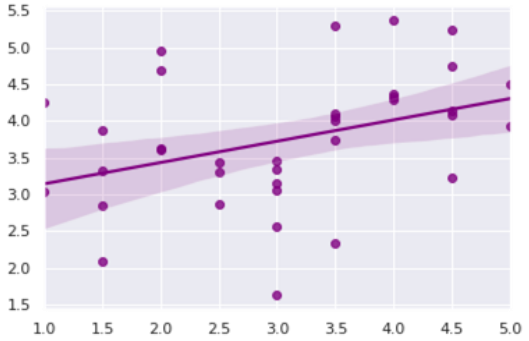
PREDICTIVE MODELING

-IMMERSION-

Andreea Al-Afuni, JiHoon Kim, Bati Yilmaz, and Angela Sofia Royo Romero







Freie Universität Berlin

CONTENTS

RESEARCH QUESTION

HYPOTHESIS

WHAT FACTORS
IMMERSION CAN BE PREDICTED
FACILITATE
BASED ON THE AROUSAL AND VALENCE RATING DATA
IMMERSION ?

WORKFLOW

ENVIRONMENT SETTING

WORKFLOW

PART I: Group and Data Selection

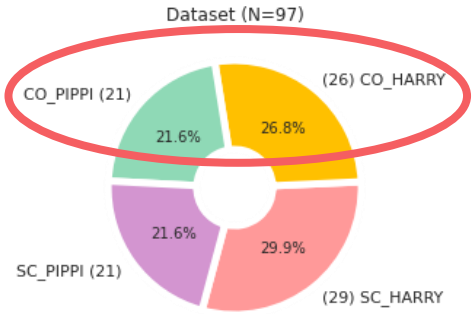
PART II: Data Preprocessing

PART III: Regression Model

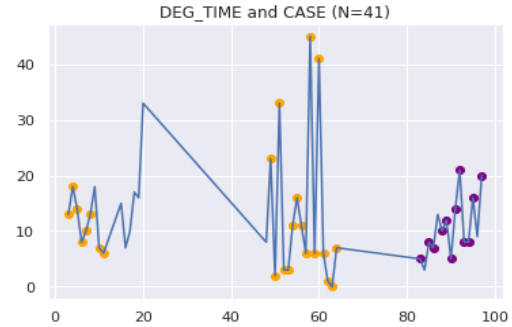
PART IV: Statistical Analysis with BIG FIVE

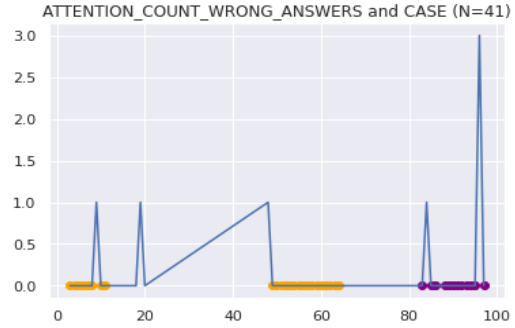
INDEPENDENT VARIABLE

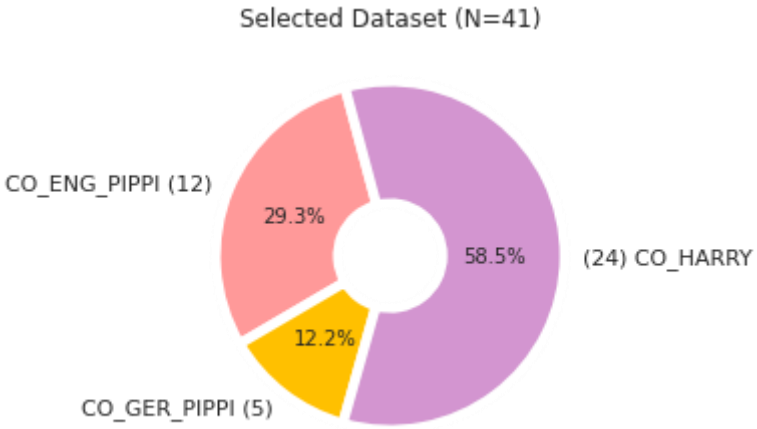
GROUP SELECTION & DATA SELECTION



bad data case list: [9, 19, 48, 84, 96, 87]







INDEPENDENT VARIABLE

AROUSAL RATING: 1 - 5 & VALENCE RATING: 1 - 7

HARRY 125 RATING DATA: (24,250)

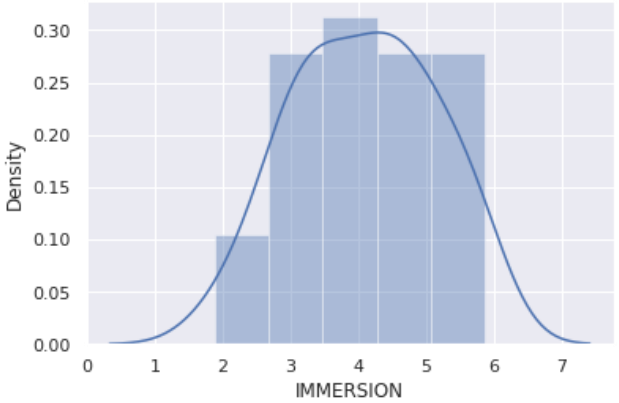
PIPPY 133 RATING DATA: (12,266)

DEPENDENT VARIABLE

READER RESPONSE

1. Focusing of attention
2. Text absorption
3. Imaginability
4. Spatial involvement
5. Reception termination
6. Suspense
7. Emotional involvement
8. General reading enjoyment
9. Identification
10. Parasocial interaction
11. Cognitive involvement
12. Thematic interest
13. Analytical reception
14. Ease of cognitive access

IMMERSION DISTRIBUTION



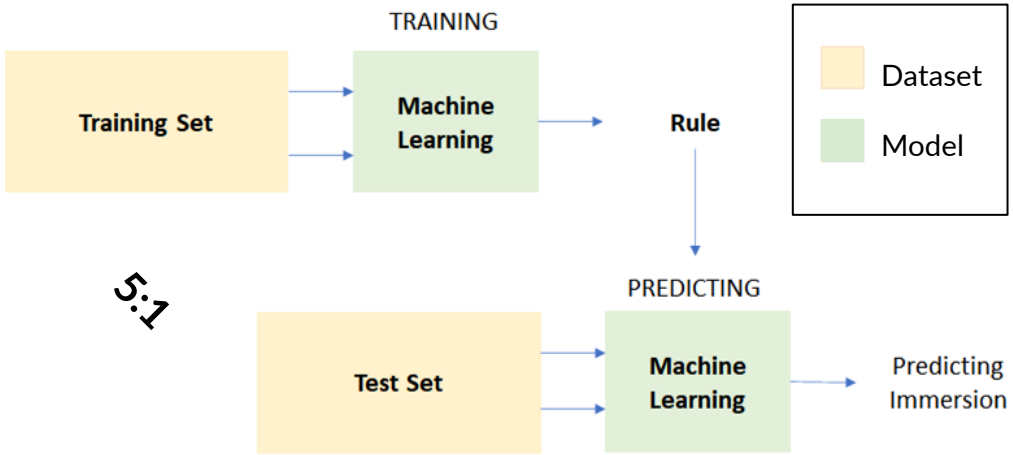
	Mean	Std	Min	50 %	Max
IMMERSION	4.0903	1.0554	1.8760	4.1875	5.8750

DEPENDENT VARIABLE

```
HARRY 125 RATING DATA: (24,1)
PIPPY 133 RATING DATA: (12,1)
```

PREDICTIVE MODELING

MODEL PIPELINE



REGRESSION MODELS

- → MLR — Multiple Linear Regression
- → k-NNR — k-Nearest Neighbors Regression
- → SVR — Support Vector Regression
- → XGB-R — XGBoost Regression
- → NN-R — Neural Network Regression

MODEL SCORE, MEAN ABSOLUTE ERROR (MAE)

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n}$$

CORRELATION, ORDINARY LEAST SQUARES (OLS)

R^2 coefficient of determination

P>[t] *p*-value, 0.05

BFI_OPENESS

OLS Regression Results

Dep. Variable:	IMMERSION	R-squared:	0.124
Model:	OLS	Adj. R-squared:	0.102
Method:	Least Squares	F-statistic:	5.797
Date:	Tue, 29 Sep 2020	Prob (F-statistic):	0.0206
Time:	16:42:13	Log-Likelihood:	-51.944
No. Observations:	43	AIC:	107.9
Df Residuals:	41	BIC:	111.4
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.5075	0.598	4.196	0.000	1.301	3.714
BFI_OPENNESS	0.3803	0.158	2.408	0.021	0.061	0.699

MODEL EVALUATION

MODEL EXECUTION EXAMPLE

XGB Regression
(TEXT: HARRY, CONDITION: COHERENT)

train score
expected_value 2.8712053 real_value [2.875]
expected_value 4.246798 real_value [4.25]
expected_value 5.7406764 real_value [5.75]
expected_value 5.6238847 real_value [5.625]
expected_value 4.750317 real_value [4.75]
expected_value 5.124679 real_value [5.125]
expected_value 2.369713 real_value [2.375]
expected_value 4.87198 real_value [4.875]
expected_value 3.7467732 real_value [3.75]
expected_value 4.1253195 real_value [4.125]
expected_value 3.2460938 real_value [3.25]
expected_value 3.2475824 real_value [3.25]
expected_value 4.6228137 real_value [4.625]
expected_value 5.4976606 real_value [5.5]
expected_value 3.2489967 real_value [3.25]
expected_value 5.746959 real_value [5.75]
expected_value 3.6226304 real_value [3.625]
expected_value 1.9332503 real_value [1.875]
expected_value 4.372049 real_value [4.375]
expected_value 3.7485611 real_value [3.75]
MAE: 1.2524397712945938

test score
expected_value 5.026205 real_value [5.875]
expected_value 4.187478 real_value [4.625]
expected_value 4.153034 real_value [5.25]
expected_value 4.596566 real_value [4.25]
MAE 0.6996761560440063

MODEL SCORE, MAE

TEXT	HARRY		PIPPI	
MODEL	Training set	Test set	Training set	Test set
MLR	2.6201e-15	0.7482	1.0214e-15	0.8470
k-NNR	0.0	0.9935	0.0	0.4256
SVR	1.0932	0.8244	0.8195	0.6250
XGB-R	1.2524	0.6997	0.8555	0.6250
NN-R	0.2608	1.0629	0.2419	0.6727

Score: MAE, Rounded to the fourth decimal place

MODEL OVERFITTING -> MODEL TUNING

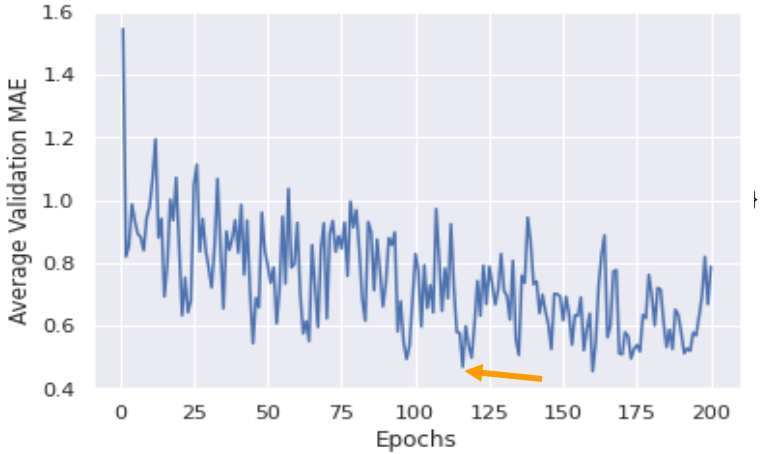
MODEL TUNING

HYPERPARAMETER TUNING

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 64)                16064
_________________________________________________________________
dense_1 (Dense)              (None, 64)                4160
_________________________________________________________________
dense_2 (Dense)              (N
_________________________________________________________________
dense_3 (Dense)              (N
=================================================================
Total params: 24,449
Trainable params: 24,449
Non-trainable params: 0
_________________________________________________________________
```

```
Model: "sequential_7"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_28 (Dense)             (None, 128)               32128
_________________________________________________________________
dense_29 (Dense)             (None, 64)                8256
_________________________________________________________________
dense_30 (Dense)             (None, 128)               8320
_________________________________________________________________
dense_31 (Dense)             (None, 1)                 129
=================================================================
Total params: 48,833
Trainable params: 48,833
Non-trainable params: 0
_________________________________________________________________
```

HYPERPARAMETER TUNING AND CROSS-VALIDATION

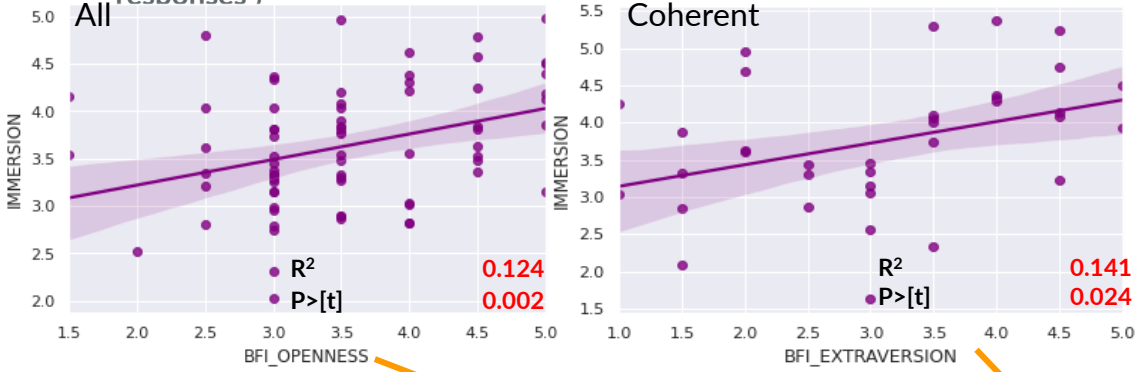


MODEL TUNING SCORE: LOSS, MAE

TEXT	HARRY		HARRY (TUNED)	
	LOSS	MAE	LOSS	MAE
TRAIN	0.0796	0.2608	0.7142	0.7523
TEST	1.4938	1.0629	0.3190	0.4619

CORRELATION BETWEEN BIG FIVE AND IMMERSION (Reader responses)

Correlation between BIG FIVE AND IMMERSION (14 Reader responses)



BIG FIVE AND READER RESPONSES (Coherent)

	FOCUS OF ATTENTION	TEXT ABSORTION	IMAGINABILITY	SPATIAL INVOLVEMENT	GENERAL READING ENJOYMENT	IDENTIFICATION	EASE OF COGNITIVE ACCESS
BFI EXTRAVERSION	0.039	0.059	0.006	0.094	0.083	0.210*	0.004
BFI AGREEABLENESS	0.009	0.007	0.137*	0.013	0.008	0.022	0.000
BFI CONSCIENTIOUSNESS	0.142*	0.017	0.002	0.036	0.003	0.015	0.001
BFI NEUROTICISM	0.000	0.030	0.010	0.007	0.053	0.001	0.003
BFI OPENNESS	0.003	0.030	0.012	0.034	0.026	0.138*	0.080

Score: R^2, * $P \leq 0.05$

BIG FIVE AND IMMERSION (2 Reader responses)

Dep. Variable:	IMMERSION	R-squared:	0.179
	coef	std err	t	P>|t|
Intercept	2.8917	0.469	6.171	0.000
BFI_EXTRAVERSION	0.3887	0.143	2.725	**0.010**

BFI_EXTRAVERSION
BFI_OPENNESS

DISCUSSION

- Predicting the immersion by looking at the valence/arousal rating data
- Addressing possible explanations
- Implications of the findings
- Application Future study (future direction)
  - Integrating EEG signals
  - Feature reduction
  - Applied BIG FIVE (iv: arousal/valence, dv:)
  - Small dataset problem

LIMITATIONS

- Terminological ambiguity (immersion, engagement, etc.)
- The total sample amount is limited (*N* = 97, 36, Pippi 12, Harry 24).
- In data set as well, more data set points are needed to make comprehensive models
- N-P Problem (combine them) 250 (N > P)
- Hyperparameter tuning optimization was not conducted in a general fashion. (cross-validation 1-by-1, classification we only used one)
- Elaborate statistical analysis and more variation in … (Lasso… x)

"All models are wrong, but some models are useful"

— George Box (Box and Draper 1987, pp. 424)*

THANK YOU!

*George Box is a retired statistics professor at the University of Wisconsin.