

AI Matchmaking based on Personality Insights

Virtual Brownbag, 19.05.2021



Who will talk to you?

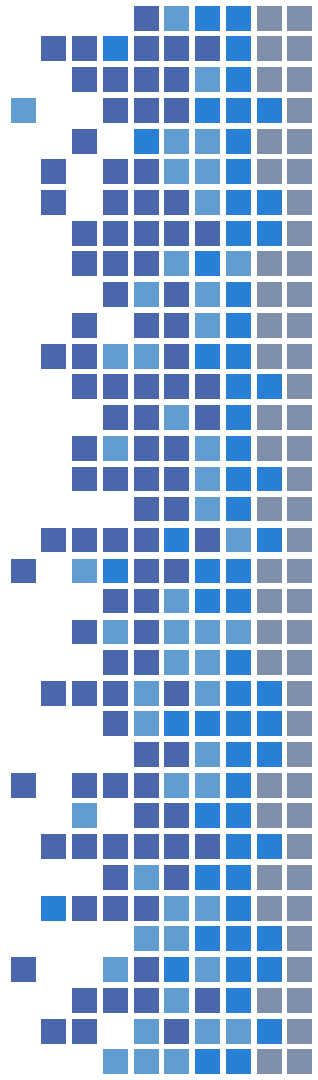


Tobias Olbrück



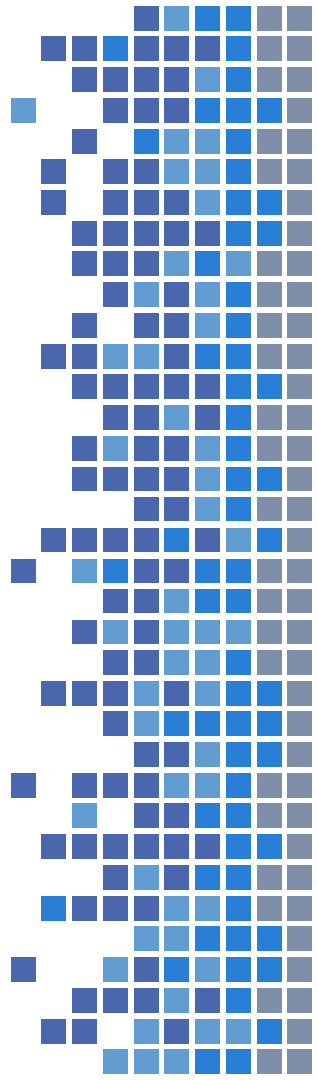
Malte Hain

Students of Information Systems (M.Sc.) at the University of Cologne



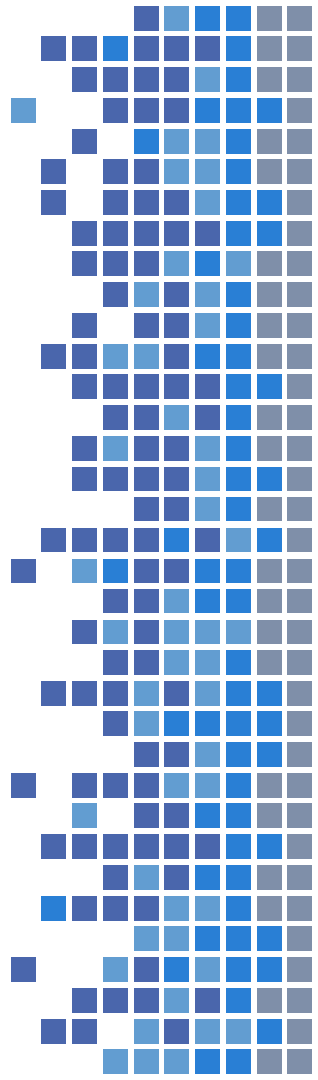
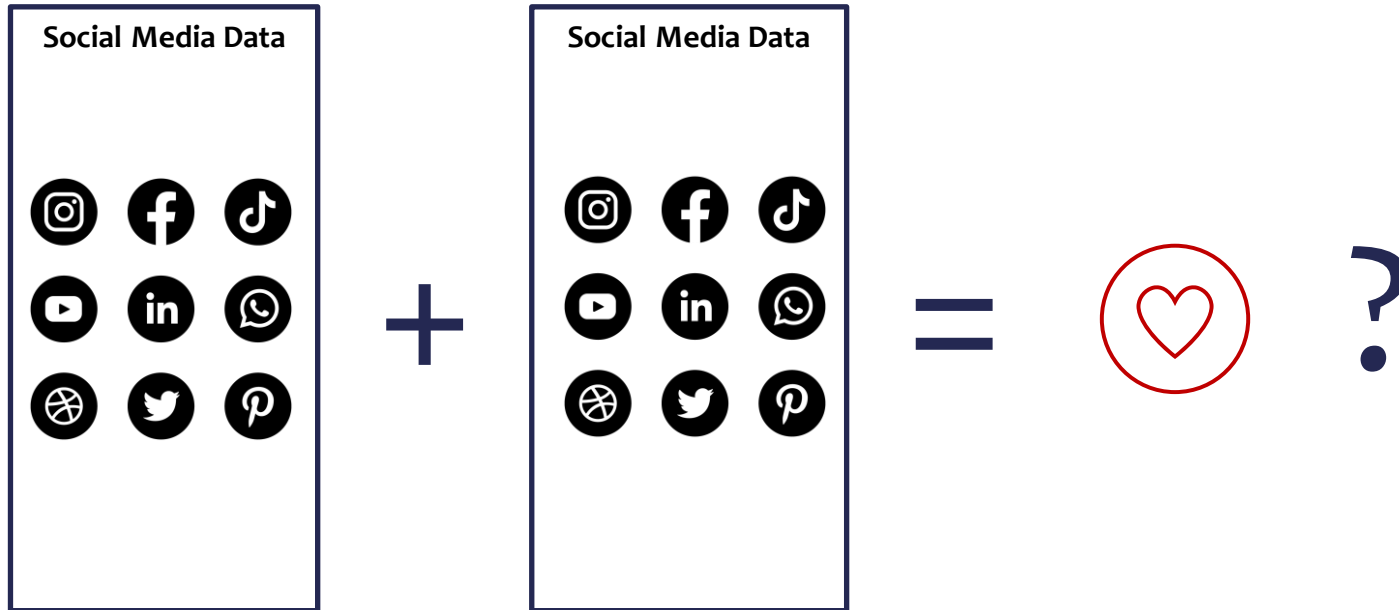
Outline

- Project Overview
- Methodology
- Results
- Live Demo
- Limitations & Possible Extensions

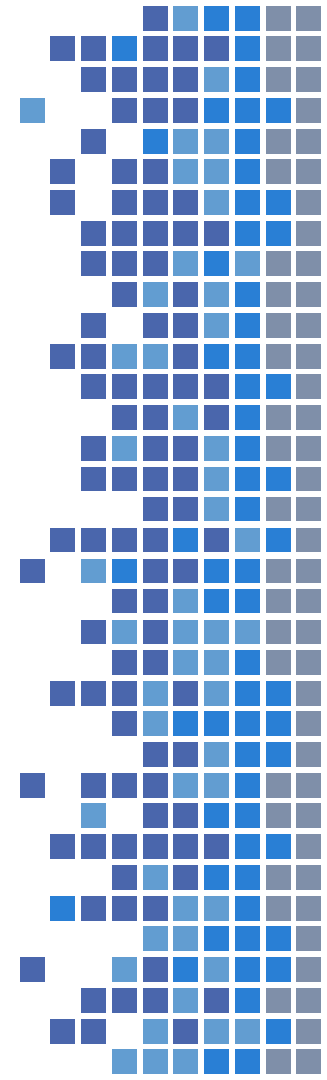


The Project in a Nutshell

Is it possible to train an AI couple matchmaking algorithm based on social media data?

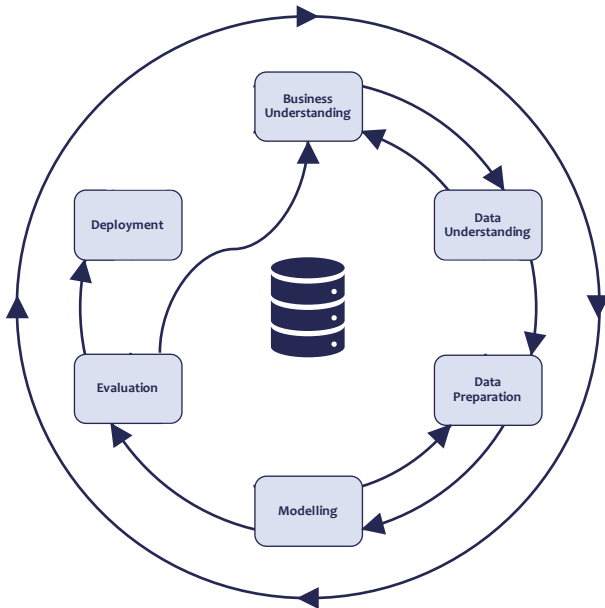


The Project in a Nutshell



Methodology: CRISP-DM

Cross Industry Standard Process for Data Mining (short CRISP-DM)



Business Understanding

- Understand underlying theories (e.g., Big 5 Personalities)

Data Understanding

- Search for couples in Twitter bios
- Fetch twitter profiles of known couples

Data Preparation

- Clean wrong couples
- Visualize personality traits of known couples

Modeling

- Build machine learning model to predict the fit between two people

Evaluation

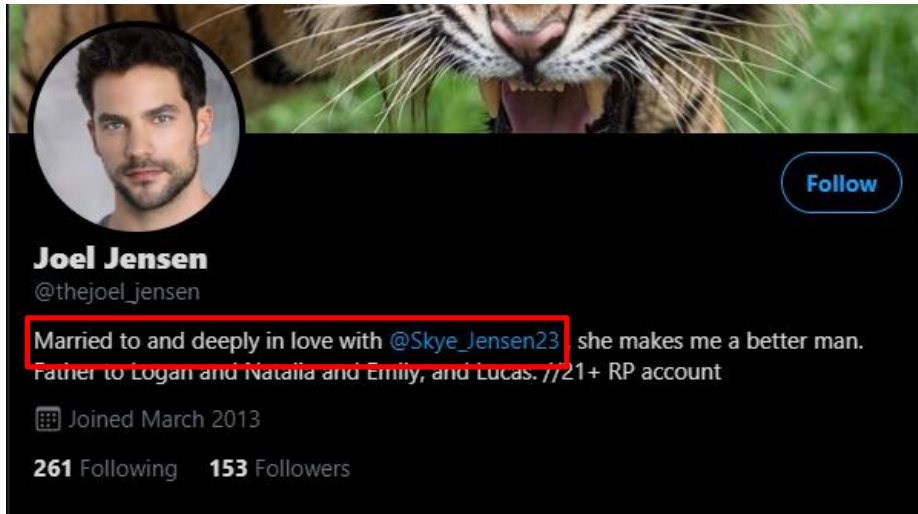
- Evaluate performance of model

Deployment

- Deploy machine learning model and website on docker image

Data Understanding – Data Fetching

1 Automated search for couples on Twitter based on their bio

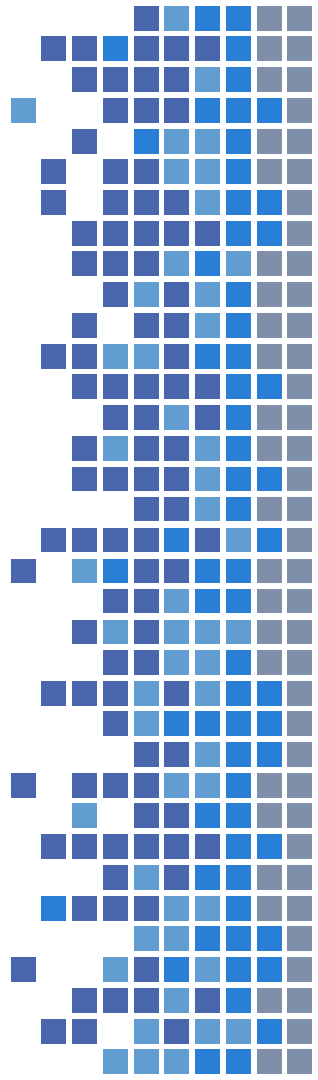


Data Understanding – Data Fetching

1 Automated search for couples on Twitter based on their bio

2 Manual & automated evaluation / cleaning of found couples

3 Fetch of complete Twitter timeline for all couples

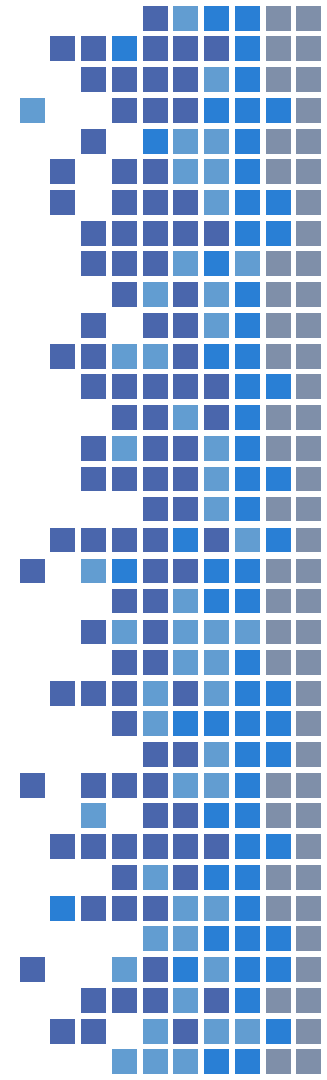


Data Preparation: Cleaning of wrong couples

Name	Profile Biography	Partner
PurpelTV	Catch me streaming on FB or Twitch! Married to @mrspurpelTV	mrspurpelTV
papaj_isasaint	married to @_iamtasha_ Father to Skylar and Noah #Cowboys #Spurs #WhiteSox	_iamtasha_
legomyborrego	Married to my high school sweetheart. Do Good, Be Kind, Live Happy. Owner of @jjbcinema	jjbcinema
real_akitaneru	demigirl lesbian // call me neru // married to @SilverRingz	SilverRingz
Jes_Ter	Just Jess! Married to the amazing @reb_mcq ,ù\$Ô[]è #LoveWins	reb_mcq
Ra3xWraith	Married to my best friend. Cat-Dad. Member of @ManCity OSC @BlueMoonDallas1.	ManCity

Explanation

- 4.500 found twitter users in total
- ~ 15 % sorted out manually
- 3.800 users / 1.900 couples as final data basis



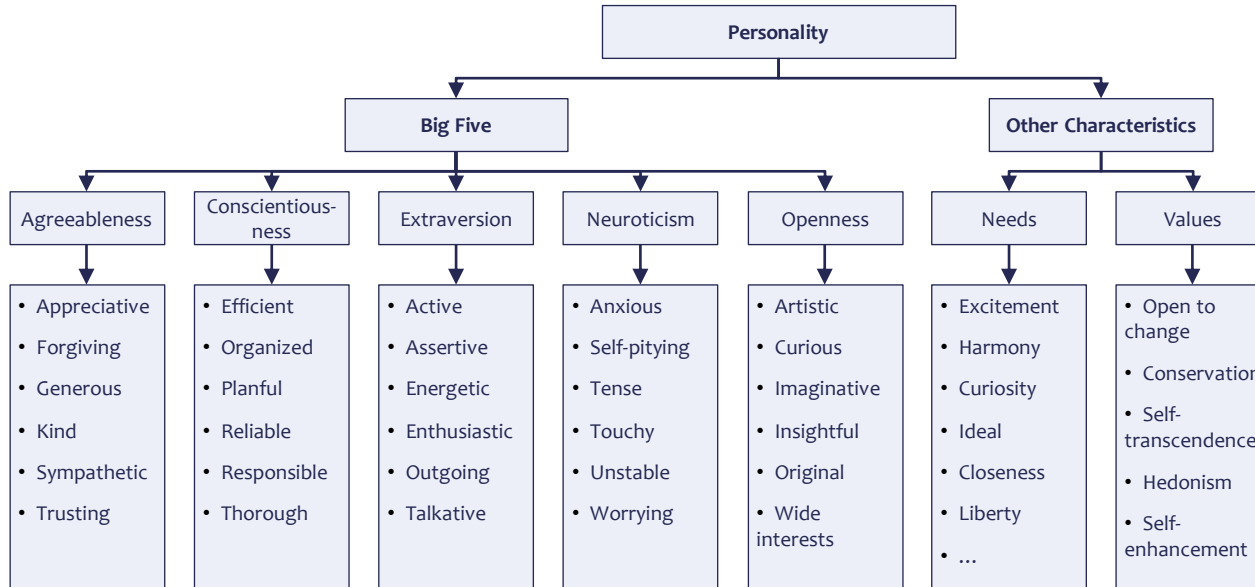
Data understanding: Feature Creation



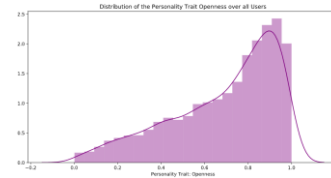
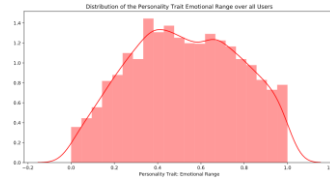
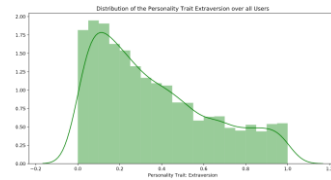
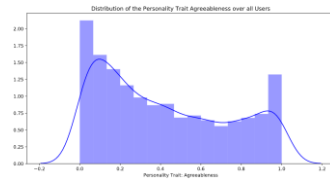
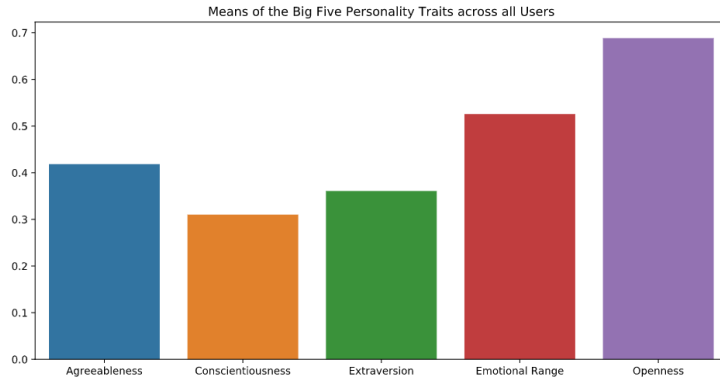
Personality is “a characteristic way of thinking, feeling, and behaving. Personality embraces moods, attitudes, and opinions“.



IBM
Watson



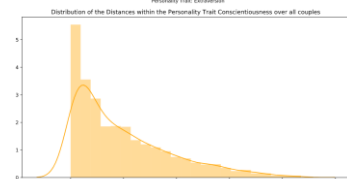
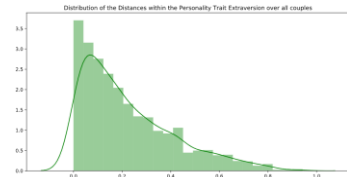
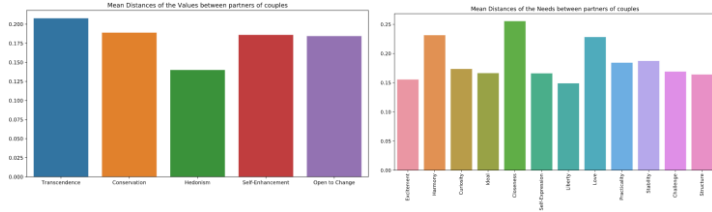
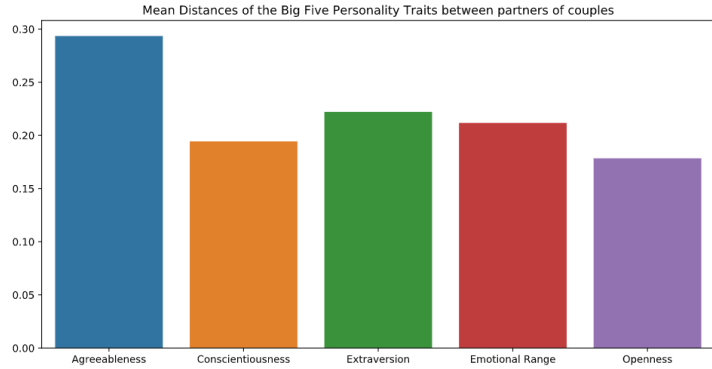
Data understanding: First Insights



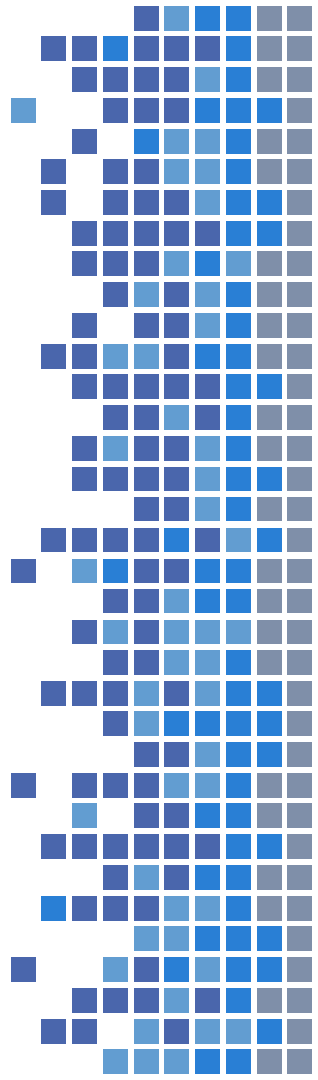
Key Messages

- Various distributions over all different personality features
- People are very open
- People have low values for extraversion
- Caution regarding bias: Twitter users do not represent the overall English-speaking population

Data understanding: First Insights



Key message: Similarities attract each other!
Biggest difference in agreeableness within couples.



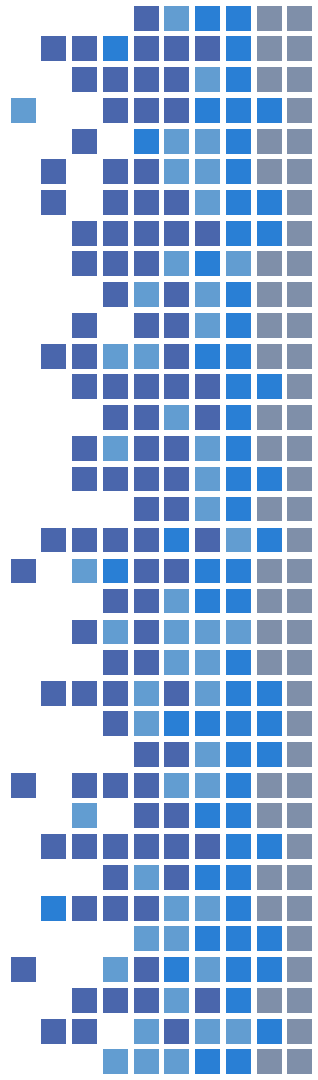
Modeling

Regression & Nearest Neighbors

- Idea: Predict “perfect” partner and find x closest real persons to it in our dataset
- Approach:
 - 1. Multi-target regression predicts Big Five Personality Characteristics, Needs & Values
 - 2. Distance metric specifies the x nearest neighbors to the regression output
- Tested multi-target regression algorithms:
Linear Regression, Decision Tree, Random Forest Regressor, Neural Network
- R^2 score ~ 0.3
MAE score ~ 0.18

Binary Classification

- Idea: Decision, if two person fit together as a couple
- Approach:
 - 1. Binary Classification based on Big 5 Personality Characteristics, Needs & Values
 - 2. Choose predicted couple with highest fitting probability
- Tested binary classification algorithms:
Support Vector Classifier, Logistic Regression, KNN, ...
- Classification accuracy ~ 0.75 %

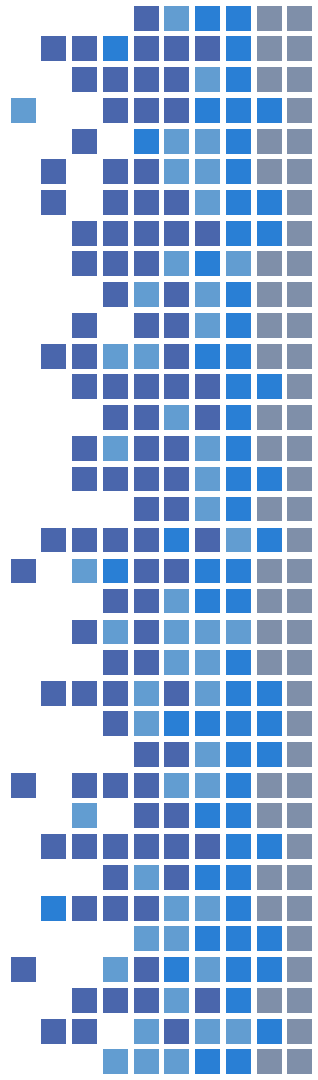


Evaluation

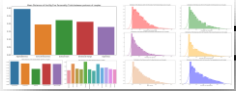
- Linear Regression Model supports general approach
 - Easy to understand
 - Significant and strong correlations
 - Comparatively good performance regarding domain difficulty
- Binary Classification with overall stronger performance



Binary Classification
as Model of Choice



Final Application



1.900 Couples



Home Find Matching Partner Find Matching Celebrity Check Your Partner

Your Twitter Profile

In order for us to analyze your Twitter profile for your personality, values and needs, it is important that your Twitter profile is public. You can find out how to change your profile from private to public: [help](#).

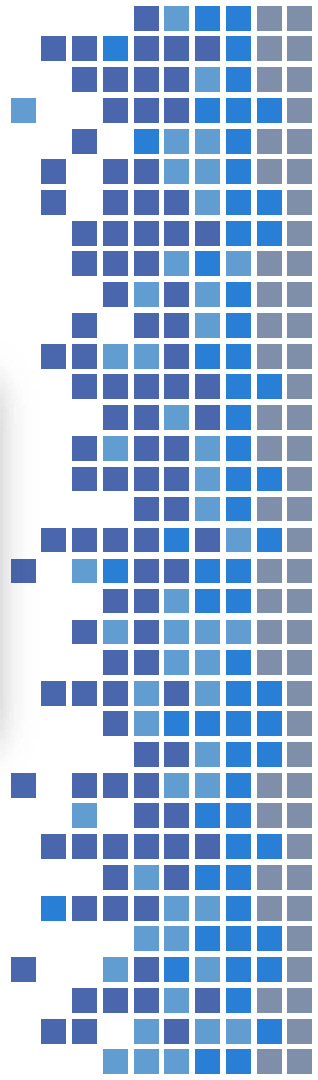
billgates

Your Big 5 Personality

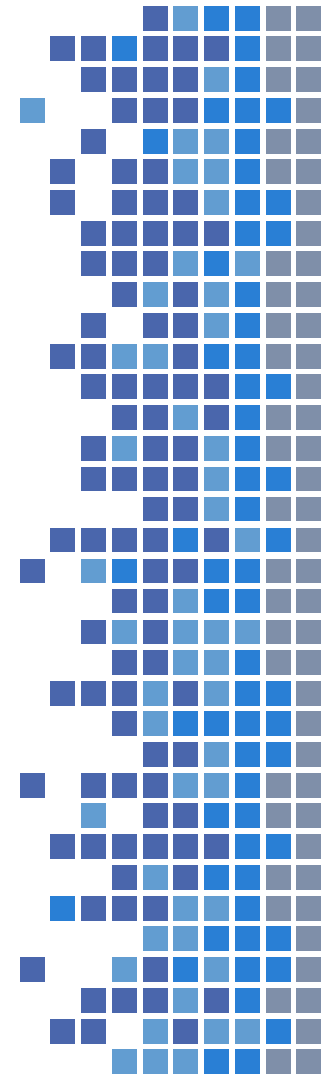
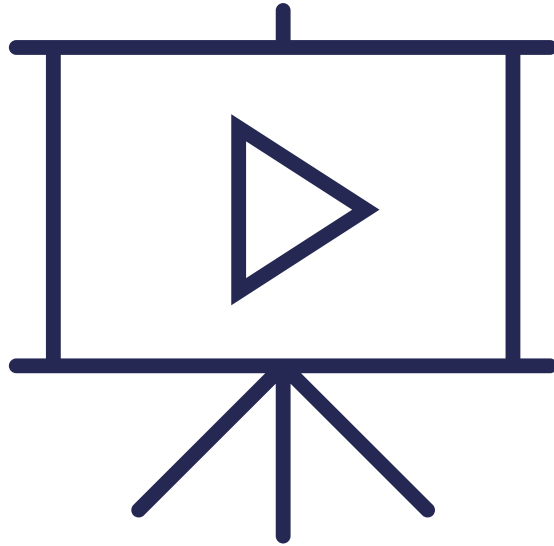
Agreeableness: 80.0%
Conscientiousness: 1.00%
Extraversion: 10.0%
Emotional range: 10.0%
Openness: 10.0%

Top 5 Matching Partner for billgates

#	Twitter User	Probability	Action
1	@khillodas	75.45%	Show Twitter Account
2	@hingley	74.9%	Show Twitter Account
3	@affecain	73.90%	Show Twitter Account
4	@ngswade	71.17%	Show Twitter Account
5	@amaalSivan	71.04%	Show Twitter Account

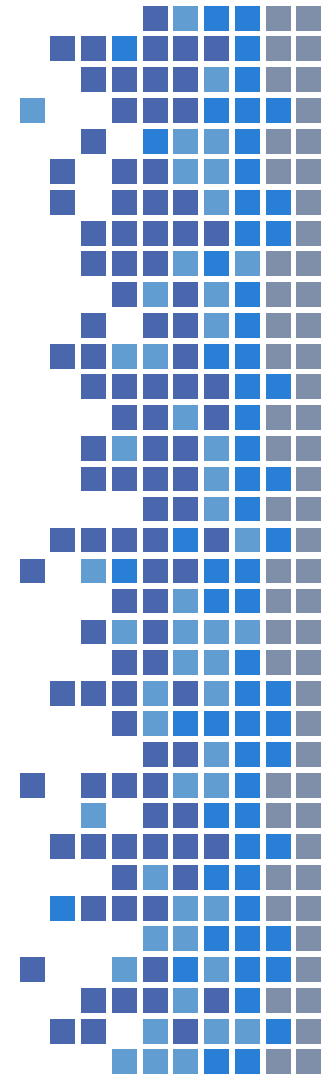


Deployment



Limitations & Possible Extensions

- Selected twitter users are biased:
 - The observed group is **not a random selection** of users and is restricted to twitter users. The users need to fulfill multiple requirements (bio with partner name, Tweets in English, only last 500 tweets are used)
 - **Shy or closed people** maybe do not share their relationship status and are less likely to tweet the required wordcount.
- Use **more couples as training** input and evaluate other data sources beside public Twitter timelines e.g., private conversation chats from messengers.
- Examine specifically **same-sex and opposite-sex couples** and its implications on the training process and the trained model.
- The model could be improved by **using a clustering approach** on the user group first and develop different models for the different cluster. This could help our modeling because we noticed distinct groups of users with different age, political views, sexual orientation, and other factors.





Thank you for your attention.

Do you have any questions?

