PERSONALITY TRAITS ANALYSIS USING

SOCIAL MEDIA

BY ANIKET GUPTA,
CHINMAY SHARMA,
AMEY KARAN,
KHOOSHI ASMI,
SUJAL DEODA

"PROBLEM STATEMENT"

Analyze candidates personality/behaviour using social media analysis, psychometric test and challenging puzzles.(Use ML)

ETHICS AND FAIRNESS

DOING IT THE RIGHT WAY

- There is a growing awareness of the sensitive nature of the data we expose to social media and the inferences that can be drawn from them.
- As a result, many APIs that can extract user data from these websites have become inaccessible to non-developers and require fulfilment of ethical guidelines for their usage.
- We have considered these APIs to be black boxes which would be available during operation.

THE KEY ELEMENTS OF APP

01 ANALYSIS USING SOCIAL MEDIA

02 PSYCHOMETRIC TEST

03 PUZZLE

WHY THE BIG 5 PERSONALITY TRAITS?

We did try to work out our analysis on the MBTI and DISC personality tests however the best results and analysis was only possible when considering the BIG 5 personality Test.

MBTI

MBTI gives a personality type rather than a personality trait.However while studying the personality of an individual to it a particular job we need an analysis of the traits of the individual.

DISC

Compared to Big 5, DISC has comparatively less scientific proof to support it, focus on behavior thus less applicable in emotional situations

WHAT'S NEW?

- We need to find the personality of a given potential employee on the OCEAN metric.
- The goal of analysis at each step is that we try to r**educe the ambiguity in predicted traits** .
- Each of the evaluations in our model helps cover the blindspots that may be present in the others, hence increasing the accuracy of predicted psychological profile.
- e.g. The puzzle section covers for profilicity of social media and peer influence in the questionnaire.

SOCIAL MEDIA ANALYSIS

ANALYSING

01

- After considering many references:
  [1, 2 , 3, 4 , 5, 6 , 7 , 8 , 9] we decided on utilising the model proposed in Christian, Suharthono et al.

Recent trends in deep learning based personality detection

Yash Mehta[1] · Navonil Majumder[2] · Alexander Gelbukh[2] · Erik Cambria[3]

Published online: 10 October 2019

Abstract

Recently, the automatic prediction of personality traits has receiv
ically, personality trait prediction from multimodal data has emer
field of affective computing. In this paper, we review significan
which have been employed for personality detection, with an e
based methods. This review paper provides an overview of the

Computers and Electrical Engineering 100 (2022) 107852

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Personality prediction model for social media using machine learning Technique☆

Murari Devakannan Kamalesh[a,*], Bharathi B[b]

[a] Research scholar, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India
[b] Sathyabama Institute of Science and Technology, Chennai, India

ABSTRACT

Predicting human behavior and personality from the social media applications like Facebook, Twitter and Instagram is achieving tremendous attention among researchers. Statistical information about the human thoughts expressed via status on social media is essential assets for research in predicting various human behaviour and personality. The current work mainly focuses on guessing user personality based on big five personality traits. An intelligent Sentence analysis model is built to extract personality features. In this article, a new Binary-Partitioning Transformer (BPT) with Term Frequency & Inverse Gravity Moment (TF-IGM) is proposed that identifies relationships among feature sets and traits from datasets. The proposed work outperforms the all feature extraction average baseline set on multiple social datasets. A maximum F1-score of 0.762 and accuracy of 78.34% on the Facebook dataset; 0.783 and 79.67%; on the Twitter dataset, 0.821; 86.84% on Instagram dataset is achieved.

Research | Open access | Published: 17 May 2021

Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging

Hans Christian, Derwin Suhartono, Andry Chowanda & Kamal Z. Zamli

Journal of Big Data **8**, Article number: 68 (2021) | Cite this article

28k Accesses | **29** Citations | Metrics

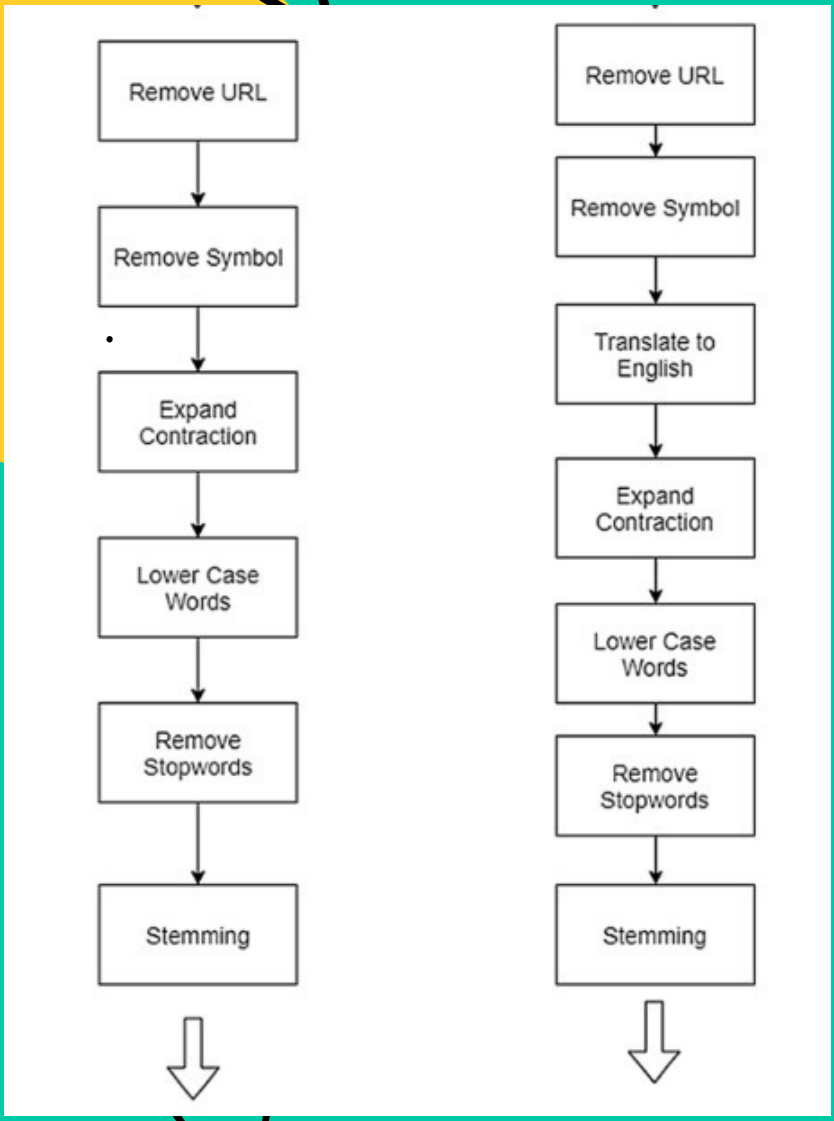
Abstract

The ever-increasing social media users has dramatically contributed to significant growth as far as the volume of online information is concerned. Often, the contents that these users put in social media can give valuable insights on their personalities (e.g., in terms of predicting job satisfaction, specific preferences, as well as the success of professional and romantic relationship) and getting it without the hassle of taking formal personality test. Termed personality prediction, the process involves extracting the digital content into features and mapping it according to a personality model. Owing to its simplicity and proven capability, a well-known

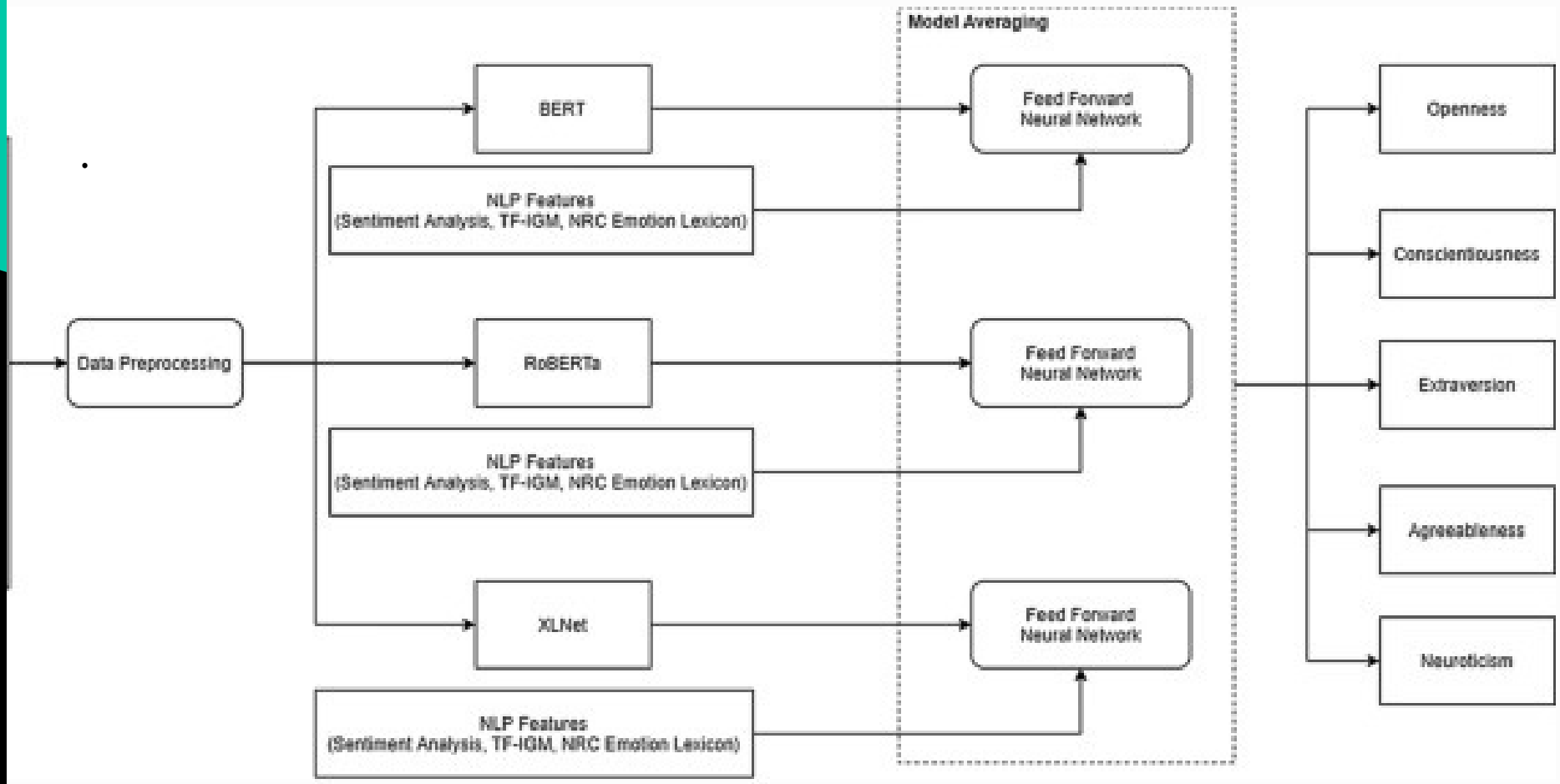
Table 1
Survey on prediction of personality traits.

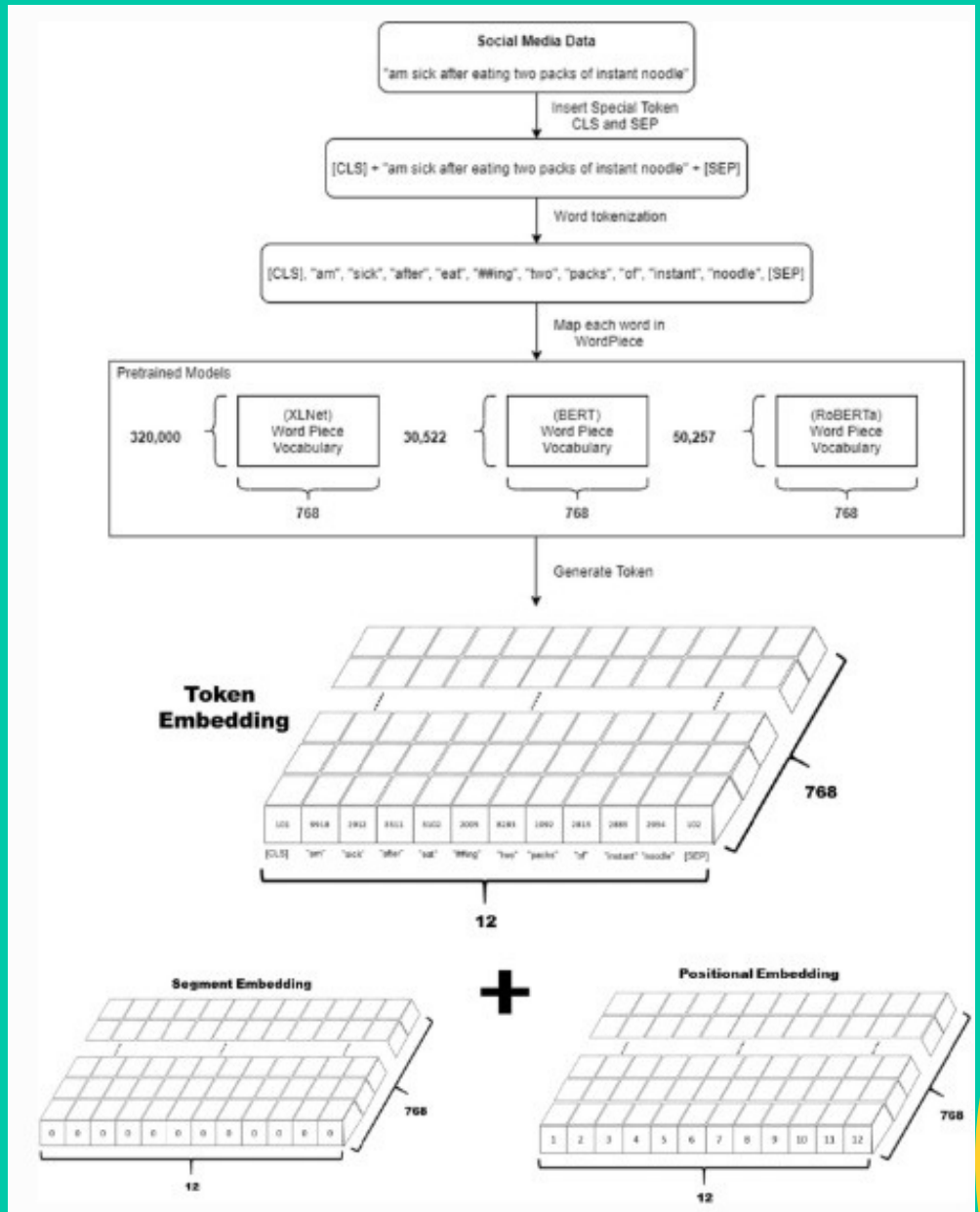
Author	Goals	Technique	Advantages	Disadvantages
Mehta, Y.; Majumder, N. (2020) [13]	Personality prediction	Deep Learning	High performance and faster execution	Needs Complicated structure
Fiok, K. et al. (2020) [14]	Sentence classification	MLM	Used in advanced NLP	Lower accuracy
Sukhbaatar et al. (2019) [15]	Text classification	Adaptive Span transformer	Reduce High complexity	fully connected nature; complicated in handling large data set
Liu and Lapata et al. (2019) [16]	Text classification	Hierarchical Transformer	Reduced memory usage	Cost; Not implemented for long text
Bharadwaj et al. (2018) [17]	online text	SVM, TF-IDF, LIWC etc	the best accuracy for all dimensions in MBTI	for the word analysis, a Less weightage factor value is given
Chaudhary et al. (2018) [18]	online text using MBTI model	Naïve Bayes, SVM, LR and Random Forest	Reduced memory usage	Lower accuracy
Kaur and Gosain (2018) [19]	Imbalance data set for sampling technique.	Decision tree	better performance using oversampling	More re-sampling methods are needed.
Gjurković and Šnajder (2018) [20]	Personality prediction of Reddit dataset.	SVM and MLP	Best performance in linguistic features extraction	age and gender are not considered
Buraya et al. (2017) [21]	personality prediction using a large Multisource dataset, NUS-MSS	Supervised	the best performance in Feature vectored	poor in performance for handling large data set
Ong et al. (2017) [22]	Big5 Model for Bahasa tweets	XGBoost, SVM	Faster Execution	Limited dataset of only.
Ngatirin et al. (2016) [23]	ML classifiers personality prediction based on a Twitter dataset	Random Forest, Random Tree, SVM	Better accuracy	Complicated data set

"RELATIVE COMPARISONS OF DIFFERENT MODELS FOR PERSONALITY PREDICTION"



- We have assumed that the API for the relevant social media platforms (Instagram, Twitter, LinkedIn's) are blackboxes and available for use during operation.
- We have implemented the required token preprocessing to use this data.
- This is fed into a multimodal deep learning architecture with pretrained language models BERT, RoBERTa and XLNET.
- Apart from this, additional NLP methods such as sentiment analysis, TF-IGM are utilized for feature extraction and concatenated to create our final feature vector


Model Averaging
Data Preprocessing
BERT
NLP Features
(Sentiment Analysis, TF-IGM, NRC Emotion Lexicon)
RoBERTa
NLP Features
(Sentiment Analysis, TF-IGM, NRC Emotion Lexicon)
XLNet
NLP Features
(Sentiment Analysis, TF-IGM, NRC Emotion Lexicon)
Feed Forward
Neural Network
Feed Forward
Neural Network
Feed Forward
Neural Network
Openness
Conscientiousness
Extraversion
Agreeableness
Neuroticism



Pre-trained models feature extraction

PSYCHOMETRIC TEST ANALYSIS

GOING BEYOND

- The test consists of 10 questions.
- Each question is recommended as a means to reducing the ambiguity of the most ambiguous trait after each question
- The hyper parameter of 0.05 for reducing ambiguity of a trait was obtained after trying combinations of choices on the free online tests
- The tuple ambiguity will be that obtained after the socil media analysis.

"You need the ability to look at all the different aspects of a problem."

02

03

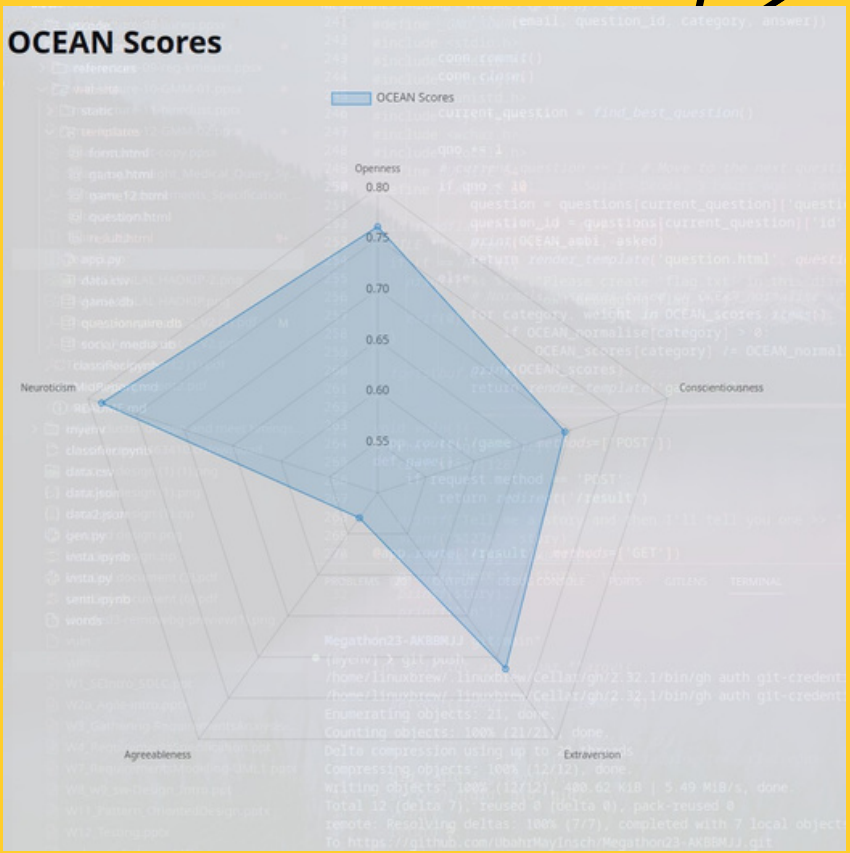
PUZZLE ANALYSIS

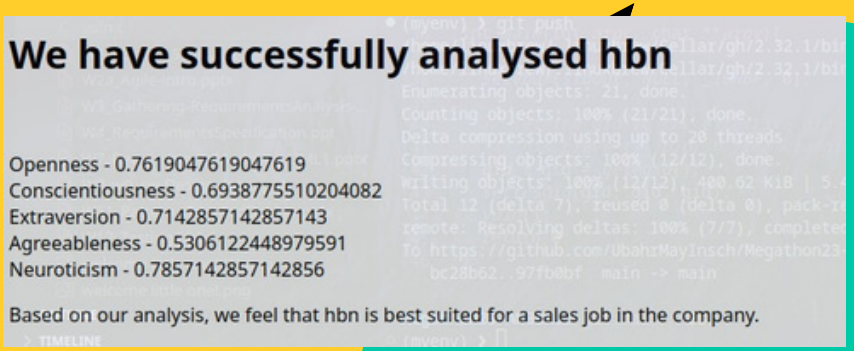
SUPER MARIO

ANALYSING CHOICES

- There is incentive for potential employees to give dishonest answers on their personality test and also to project a certain type of profile on their social media.
- A game or puzzle environment for psychometric evaluation sidesteps these issues by not revealing possible metrics for evaluation to the participants.

- Based on these guidelines a super mario clone will be used which also collects the data of the user playing
- The relevant data we are collecting includes adventure coins, Time Since the last death, Time taken to complete a level, Leaderboard position.
- In different levels we also control for showing the participant leaderboard data.
- As recommended in [1], we can use these to predict the psychological profile, which further helps reduce the ambiguity in our model.





"OUR SAMPLE OUTPUT"

- This sentiment was echoed by many of our references [.1 , 2, 3].
- Reference [3] also recommended guidelines for selecting a game which could be used for psychometric evaluation.

.

Table 2: Representing game type for each Big5 personality

Personality	Description	Game
Openness	Openness to experience refers to one's willingness to try new things as well as engage in imaginative and intellectual activities. It includes the ability to "think outside of the box."	Games with questions of visual elements/pictures for detecting if user is visual and active Practical tasks with strict information games for sensing and intuition Case studies with multiple step games if user is active, sensing, visual and sequential Responding positively for all Gamification elements is correlated to all working styles show extroversion Responding negatively is correlated to all working styles show neuroticism These Games should include badges, leader boards, points, levels, feedback, time track, stories progress bar and competition
Neuroticism	The overall emotional stability of an individual through how they perceive the world. It takes into account how likely a person is to interpret events as threatening or difficult.	
Conscientiousness	Describes a person's ability to regulate their impulse control in order to engage in goal- directed behaviors. It measures elements such as control, inhibition, and persistence of behavior.	
Agreeableness	How people tend to treat relationships with others. Unlike extroversion which consists of the pursuit of relationships, agreeableness focuses on people's orientation and interactions with others.	
Extroversion	Reflects the tendency and intensity to which someone seeks interaction with their environment, particularly socially. It encompassed the comfort and assertiveness levels of people in social situations.	

BIBLIOGRAPHY

- Apart from the papers linked in the slides,
- we have only utilized ChatGPT as an AI tool.
- We have also utilised the GitHub directory of Super Mario clone (“Mario in Python”) created in Python by Mar. This can eventually be adapted to our implementation with the necessary modifications.

THANKS A LOT