Universiteit Utrecht

Master Thesis

in the field of

Applied Data Science

Analyzing Personality Trait Intercorrelations: A Comparison between Model-Generated and Questionnaire-Derived Correlations

supervised by: Anastasia Giachanou

submitted on: 30th of June 2023

submitted by: Merel Hoekstra
6155391
Utrecht
m.m.hoekstra@students.uu.nl

Abstract

The primary emphasis of this paper lies in examining the intercorrelations between personality traits as identified through computational models, and investigating how these interconnections compare to those derived from questionnaires. Various computational models, including RoBERTa, DistilBERT, BERTweet, ALBERT, and XLNet, are implemented and evaluated using the PANDORA dataset. The intercorrelations found by these models are compared to the intercorrelations reported in the meta-analysis by van der Linden et al. (2010). This research reveals discrepancies between dataset intercorrelations and theory, and the models used to predict personality traits were unsuccessful in explaining score variation. The research emphasizes the need for further exploration, including improving model performance, refining preprocessing techniques, and utilizing annotated datasets for personality prediction.

Keywords: Automatic personality recognition, Big Five model, intercorrelations, personality questionnaires, computational models, PANDORA dataset.

Contents

Abstract I
1 Introduction 1
2 Related literature 4
2.1 Automatic Personality Recognition 4
2.2 Personality Models 5
2.2.1 MBTI 5
2.2.2 Big Five Model 5
2.3 Meta-Analysis and Earlier Psychological Research 7
3 Methodology 9
3.1 Intercorrelations Big Five Model 9
3.2 Models 10
3.2.1 Model selection 10
3.3 BERT models 11
3.3.1 DistilBERT 11
3.3.2 RoBERTa 11
3.3.3 BERTweet 12
3.3.4 AlBERT 12
3.4 XLNet 12
3.5 Experimental Settings 13
3.5.1 DistilBERT 13
3.6 Error analysis 13
4 Data 15
4.1 PANDORA Dataset 15
4.1.1 Data exploration 15
4.2 Data cleaning 16
4.2.1 Sampling 16
4.3 Text preprocessing 18
4.3.1 Stopwords 19
4.3.2 Ethical considerations 19
5 Results and Analysis 21
5.1 Intercorrelations Dataset 21
5.1.1 Comparing the Sample with Theory 22

5.2 Performance of Models 23
5.2.1 Metrics 23
5.2.2 Performance 24
5.3 Intercorrelations of the models' predictions 25
6 Discussion 31
6.1 Results 31
6.1.1 Intercorrelations sample vs intercorrelations theory 31
6.1.2 Model accuracy 32
6.1.3 Research Questions 32
6.2 Limitations 33
7 Conclusion 34
7.1 Future Work 35
Bibliography 36

1 Introduction

The understanding, quantification and evaluation of individual differences in behavior, feelings and thoughts have always been central topics in psychological science (Stachl, 2019). The interest in explaining those differences in human behaviors and feelings has led to the emergence of personality traits, which are psychological constructs that aim to explain the wide variety of human behaviors in terms of a few stable and measurable individual characteristics (Vinciarelli & Mohammadi, 2014).

In order to provide personality measures for personality traits, two models have been most commonly applied: the Myers-Briggs Type Indicator (MBTI) and the Big Five model (BF) (Radisavljević et al., 2022). The MBTI is a categorical personality assessment that focuses on four binary dimensions (Sharon et al., 2023), whereas the Big Five model is a trait-based model that measures personality along five continuous dimensions: openness, conscientiousness, extraversion, agreeableness, and neuroticism (De Raad, 2000).

Extensive research has been conducted on both the MBTI and the Big Five model, with a focus on comparing and contrasting the two frameworks while identifying their respective strengths and limitations (Elliott, n.d.). The MBTI is appreciated for its simplicity and ease of understanding, but it faces criticism for a lack of empirical support according to research (Elliott, n.d.). Conversely, the Big Five model has a strong empirical basis, yet it is also criticized for oversimplifying the complexity of personality traits (Fang et al., 2022)(Entringer et al., 2021). Of the two, the Big Five model is the most widely accepted and researched taxonomy of personality traits in psychology (Phan & Rauthmann, 2021).

Pyschological research on individual differences in behavior and personality traits is based on data obtained through self-report questionnaires as the primary source of data (Stachl, 2019). Traditionally, individuals have been asked to complete lengthy surveys as a means of assessing their personality traits. However, due to the increasing use and success of deep learning methods and the increase of available text data, it is now possible to compute personalities from digital texts (Zhao et al., 2022). This is called Automatic Personality Recognition (APR).

APR deals with the identification of a target individual's personality type through computational methods (Mushtaq & Kumar, 2022). Knowledge of an individuals' personality type has a broad spectrum of potential applications. Amongst others, it can be used for recommender systems, recruitment systems, online marketing, social network friend selection, and human resource management systems (Ramezani et al., 2022).

Due to the extensive range of applications, the interest in Automatic Personality trait Recognition has spiked in fields like psychology, neuropsychology, computer science, and other related domains (Zhao et al., 2022).

APR offers many advantages over self-report questionnaires. One notable advantage is its reduced time consumption compared to the lengthy process of filling out questionnaires (Fang et al., 2022). Specifically, APR excels in quickly analyzing large amounts of data and providing real-time or near-real-time analysis of personality traits.

Another advantage of APR is that it bypasses potential limitations inherent in self-report questionnaires, such as individuals being unable to fully explain or accurately perceive their own personality traits (Pedregon, 2012).

Despite the advantages of APR, it is accompanied by several challenges that require attention and resolution (Fang et al., 2022). Fang et al.(2022) discuss multiple challenges in the field of APR that demand the attention of the natural language processing research community. One particular challenge highlighted is the independent prediction of Big Five personality traits when using models for computation (Fang et al., 2022). This approach fails to consider the intercorrelations between these traits, which have been explored in psychological studies (Van der Linden et al., 2010). However, many APR studies neglect to examine these trait intercorrelations, unlike traditional questionnaire-based studies (Fang et al., 2022).

Evidence of intercorrelations between the Big Five personality traits was found in multiple psychological researches (Van der Linden et al., 2010). In 2010, Van der Linden et al. conducted a meta-analysis to examine the intercorrelations among the Big Five personality factors (Van der Linden et al., 2010). Their findings revealed several significant intercorrelations among the Big Five traits, which were derived from self-report questionnaires (Van der Linden et al., 2010).

Among the correlations discovered between the Big Five model traits, one example is those of the trait "Openness to Experience". This trait demonstrates positive associations with all other Big Five traits, with the exception of neuroticism, where it exhibits a negative correlation (Van der Linden et al., 2010).

By disregarding these intercorrelations, a challenge arises. Even when models have a good performance, there remains uncertainty regarding whether the models have made the correct predictions for the right reasons (Fang et al., 2022). This oversight raises questions about the validity of these automatic methods for recognizing personality traits accurately. Hence, the objective of this research is to examine the correlations between traits computed by a model and compare them to the intercorrelations observed in earlier survey-based studies.

This research paper will closely examine the aforementioned challenge by answering the following two research questions: *"What are the intercorrelations observed among personality traits when they are detected using computational models?"* and *"How do these correlations compare to the correlations found in self-report questionnaires?"*

Chapter 2 will review related literature and explore the intercorrelations between traits identified through questionnaires using a meta-analysis. Chapter 3 will outline the methods employed, while Chapter 4 will delve into the data utilized for automatic personality detection. Chapter 5 will present the intercorrelations discovered by the computational model. The paper will discuss and conclude the findings in Chapters 6 and 7. This research reveals discrepancies between dataset intercorrelations and theory, and the models used to predict personality traits were unsuccessful in explaining score variation.

2 Related literature

Research in the field of human behavior, specifically focusing on human emotion recognition and other related affective phenomena, is rapidly gaining momentum (Halim et al., 2019). This chapter will provide an overview of the related literature and delve into the details of Automatic Personality Recognition (APR), the focus of this research. Additionally, the Big Five model, which will be utilized in this study, will be further explained, along with an overview of the intercorrelations discovered among the Big Five traits.

2.1 Automatic Personality Recognition

As introduced before, APR deals with the identification of a target individual's personality type through computational methods, utilizing various sources (Mushtaq & Kumar, 2022). In other words, personality recognition is the process of extracting information from online content, such as text, and categorizing it based on a personality model (Christian et al., 2021). The extensive adoption of social media platforms has empowered individuals to openly share their perspectives and thoughts on a wide range of topics, including personal well-being, psychology, financial matters, social interactions, the environment, and even politics. In some cases, these digital written expressions can be used to characterize the individual's behavior and personality (Christian et al., 2021). Consequently, the surge in social media usage has intensified the investigation of these platforms as a means to gain deeper insights into individuals, allowing for a better understanding and assessment of their personalities. As social media continues to witness exponential growth, the interest in exploring these avenues for enhanced comprehension and personality evaluation has grown accordingly. One significant factor contributing to this interest is the cost-effectiveness of collecting data through these platforms.

Several previous studies have used social media to predict users' personality automatically (Adi et al., 2018). For example, in 2017, Tandra et al. attempted to build a system that can predict a person's personality based on users' Facebook user information, including 10,000 Facebook statuses (Tandera et al., 2017). Additionally, In

2015, Omheni et al. used an online environment to investigate practically the utility of annotation in reflecting an accurate user personality profile (Omheni, 2015). Another research, conducted by Ong et al. in 2017, focused on building a personality prediction system based on a Twitter user's information for Bahasa Indonesia, the native language of Indonesia (Ong et al., 2017).

Within these studies, various sources have been used for prediction (Mushtaq & Kumar, 2022). Skowron et al. in 2016 employed many different sources in their research, such as textual data, visual content, and users' meta features extracted from Twitter and Instagram to forecast personality traits (Skowron et al., 2016). They found that the joint analysis of users' simultaneous activities in social networking sites seems to lead to a consistent decrease in the prediction errors for each personality trait (Skowron et al., 2016). Notably, the scope of this research primarily centers around the automatic prediction of personality utilizing textual sources.

2.2 Personality Models

2.2.1 MBTI

The MBTI test was originally developed to measure people's personalities type (Myers et al., 1998).

The MBTI operates on the principle that variations in behavior between individuals can be described in terms of preferences between opposing characteristics (Behaz & Djoudi, 2012). These preferences form four fundamental dimensions of psychological life:

1. Introversion (I) vs. Extraversion (E)
2. Sensation (S) vs. Intuition (N)
3. Thinking (T) vs. Feeling (F)
4. Judging (J) vs. Perception (P)

2.2.2 Big Five Model

The Big Five Model, also known as the Five-Factor Model, is the most widely accepted personality theory held by psychologists today (Lim, n.d.). The theory states that personality can be boiled down to five core factors, known by the acronym CANOE or OCEAN (Lim, n.d.).

- Conscientiousness – impulsive, disorganized vs. disciplined, careful
- Agreeableness – suspicious, uncooperative vs. trusting, helpful
- Neuroticism – calm, confident vs. anxious, pessimistic
- Openness to Experience – prefers routine, practical vs. imaginative, spontaneous
- Extraversion – reserved, thoughtful vs. sociable, fun-loving

(Lim, n.d.)

Traditionally, a person is assigned a continuous score for each of these traits based on a Self-Report Questionnaire. Individuals are given a questionnaire that consists of a series of statements or questions related to different aspects of personality. (Satow, 2021)They are asked to rate the extent to which they agree or disagree with each statement based on their own self-perception. The questionnaire typically uses a Likert scale or a similar rating system, ranging from strongly agree to strongly disagree (Calabrese et al., 2012). The responses are then scored, and the scores on different items related to each trait are combined to provide an overall score for that trait.

Examples of questionnaires are Big Five Inventory (BFI) (John et al., 2010) and NEO Personality Inventory (NEO-PI) (Costa & McCrae, 2008)

According to Fang et al.(2022), the Big Five personality traits are a more preferable personality framework to use compared to the MBTI (Fang et al., 2022) This is because of a few reasons. Firstly, Big-5 provides a more realistic and accurate classification of personality traits by scoring individuals on a continuous spectrum, unlike MBTI's dichotomous approach (Fang et al., 2022). This continuous representation better captures inter-individual differences and preserves more information. Secondly, Big-5 is supported by a stronger empirical foundation than MBTI (Fang et al., 2022). Personality psychologists have expressed limited enthusiasm towards the MBTI McCrae and Costa Jr, 1989. Thorough and intricate analyses conducted by Stacker and Ross (Stricker & Ross, 1964) resulted in a critical evaluation of both the typology and the scales within the MBTI. These theorists argue that the Jungian concepts, which are intended to underlie the MBTI, have been subject to distortion (McCrae & Costa Jr, 1989). Thirdly, Big-5 has been extensively studied in relation to various social science constructs, such as emotions, styles, and mental illnesses, making it more relevant for research purposes compared to MBTI (Fang et al., 2022). Finally, Big-5 is rooted in natural language (lexical hypothesis), indicating that cues related to Big-5 traits are more prevalent in text data compared to MBTI cues (Fang et al., 2022).

Given the drawbacks of the MBTI, the Big Five model will be utilized to evaluate personality within computational models. As mentioned before, unlike the MBTI, the Big Five model employs continuous scores, which is advantageous when examining correlations and analyzing personality traits.

2.3 Meta-Analysis and Earlier Psychological Research

Traditionally, in studies from psychology, individuals' personality traits were measured through self-report questionnaires. These questionnaires often consisted of surveys that participants had to complete in order to determine their personality characteristics (Barbaranelli et al., 2003). What was often included in these types of research studies was an investigation of intercorrelations between Big Five personality traits found in subjects (Fang et al., 2022).

Van der linden, et al. (2010) present a meta-analysis (K= 212, total N= 144,117) on the intercorrelations among the Big Five personality factors (Van der Linden et al., 2010). They found the following correlations.

They found that 'Openness to Experience' is positively correlated with 'Conscientiousness', 'Extraversion', and 'Agreeableness', and negatively correlated with 'Neuroticism' (Van der Linden et al., 2010). Meaning that if someone is imaginative and spontaneous (Open to Experience), he/she is often also more sociable, trusting and disciplined, and less anxious and pessimistic. Furthermore, their research revealed that 'Conscientiousness' is positively correlated with 'Extraversion' and 'Agreeableness', and negatively correlated with 'Neuroticism' (Van der Linden et al., 2010). Hence, individuals who demonstrate characteristics of being disciplined and cautious (Conscientiousness) often display helpfulness and sociability as well. Similarly, 'Extraversion' was found to be positively correlated with 'Agreeableness' and negatively correlated with 'Neuroticism' (Van der Linden et al., 2010). This indicates that individuals who are outgoing and sociable (Extraversion) are likely to possess trustworthiness but exhibit lower levels of anxiety. Moreover, their study highlighted a negative correlation between 'Agreeableness' and 'Neuroticism' (Van der Linden et al., 2010), suggesting that individuals who are highly trusting and helpful (Agreeableness) tend to experience lower levels of anxiety.

It is worth noting that while these patterns generally held true, there were instances where specific correlations deviated from the overall trends or the majority of other subgroups (Van der Linden et al., 2010).

Typically, when conducting research using models to predict personality traits, such as in APR research, the examination of intercorrelations is not a common practice. Traits are often computed separately, and the focus lies on implementing models on a dataset to predict personality traits and improving the performance of those models. Since the focus lies on the performance, it is not always clear if those models have also captured the intercorrelations known by established theories.

Given that the van der Linden study (Van der Linden et al., 2010) involved a meta-analysis incorporating multiple research studies, it is considered a reliable representation of the correlations observed between Big Five personality traits found in questionnaire-based research.

In APR research, the intercorrelations between personality traits are often not reported. However, understanding these intercorrelations is crucial for improving prediction accuracy. Therefore, we are interested in whether the intercorrelations between traits identified by the models align with the established correlations mentioned earlier. By assessing the consistency and validity of these intercorrelations, we can enhance our understanding and evaluation of the prediction process.

3 Methodology

The aim of this paper is to investigate the intercorrelations between traits computed by automated models and compare them with the intercorrelations observed in questionnaire-based assessments. The PANDORA dataset will serve as the dataset on which the models will be applied to compute personality traits. Subsequently, an analysis will be conducted to examine the intercorrelations between the traits identified by the models and compare them to those found in questionnaire-based research.

3.1 Intercorrelations Big Five Model

To determine the intercorrelations based on the questionnaire research, we will refer to the results obtained by van der Linden et al. (Van der Linden et al., 2010). The intercorrelations among the big five personality traits are outlined below:

	Ext	Neu	Agre	Con	Ope
Ext	1.0	-	-	-	-
Neu	-.36	1.0	-	-	-
Agre	.26	-.36	1.0	-	-
Con	.29	-.43	.43	1.0	-
Ope	.43	-.17	.21	.20	1.0

Table 1: Intercorrelations found by van der Linden et al. (2010).

Van der Linden et al. conducted a psychometric meta-analysis on the intercorrelations among the Big Five personality factors (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) (Van der Linden et al., 2010).

The researchers conducted both computerized and manual searches to identify relevant studies for their meta-analysis. They focused on studies that utilized either the Big Five model or the Five Factor Model (FFM) of personality, as these two models have significant overlap (Van der Linden et al., 2010).

To establish inclusion criteria for the meta-analysis, the researchers defined three requirements. Firstly, the personality measures used in the study had to be clearly based on either the Big Five or FFM dimensions. Secondly, the study needed to include a table presenting the ten initial Pearson correlations between the factors. Lastly, the correlation matrices had to be derived from independent samples (Van der Linden et al., 2010).

Next, the researchers employed various search methods to gather relevant data for their meta-analysis. Firstly, they conducted electronic database searches. Secondly, they manually searched specific journals in the field of personality or applied psychology. Thirdly, they examined the reference lists of retrieved articles. Additionally, the researchers contacted other experts to obtain additional intercorrelation matrices (Van der Linden et al., 2010).

For the meta-analysis, the researchers directly extracted correlation values from the original articles' matrices.

The search process resulted in 212 Big Five intercorrelation matrices from independent samples. The total sample size (N) across all matrices was 144,117. Resulting in the correlations shown in table 2, the meta-analytical intercorrelations between the Big Five dimensions (Van der Linden et al., 2010).

3.2 Models

In this section, we provide an explanation of the models we applied to estimate the personality traits.

3.2.1 Model selection

For the personality prediction, namely pre-trained transformer-based language models were used. Most of them were auto-encode transformer-based models (DistilBERT, RoBERTa, Bertweet, Albert) and one autoregressive transformer-based model(XLNet) (Ganesan, 2021).

Transformer-based language models have become the foundation for accurately approaching many tasks in natural language processing (Vaswani et al., 2017). The remarkable achievements of transformer-based language models stem from their capacity to capture both syntactic and semantic information (Tenney et al., 2019). This is accomplished through the utilization of large, deep attention-based networks, commonly known as transformers, which employ hidden state sizes in the range of 1000 across multiple layers (Ganesan, 2021).

3.3 BERT models

For the auto-encode transformer-based models, DistilBert, RoBERTa, Bertweet and Albert models were used. These models are variants of the Bidirectional Encoder Representations from Transformers (BERT) model. The BERT model is based on the "Transformer" architecture, which relies on attention mechanisms and does not have an explicit notion of word order beyond marking each word with its absolute-position embedding (Vaswani et al., 2017). BERT was trained on a huge amount of data. The original BERT model, known as "BERT Base," is trained on a dataset comprising 3.3 billion words from books and the English Wikipedia (Tenney et al., 2019). This training corpus allows the model to capture a wide range of linguistic patterns and semantic relationships. Additionally, BERT is trained on two tasks: masked language modelling (MLM) and next sentence prediction (NSP) (Aroca-Ouellette & Rudzicz, 2020). The BERT model combines bidirectional transformers and transfer learning with the objective of creating state-of-the-art models for a wide range of NLP tasks (Kici et al., 2021). Such as text classification and sentiment analysis.

BERT has several variations, including among many others, RoBERTa, ELECTRA, DistilBERT, BERTweet, and ALBERT (Kici et al., 2021). These variations differ in model size, number of parameters, pre-training data and corpora, pre-training objectives and techniques (Kumar, 2023).

3.3.1 DistilBERT

The difference between BERT and DistilBERT is that DistilBERT includes 66 million parameters, compared to the 110 million parameters of the BERT base (Mapes et al., 2019) which makes it 40% smaller and 60% faster than BERT base (Sanh et al., 2019). This model was implemented to benefit from its optimized performance without compromising accuracy.

3.3.2 RoBERTa

Following the implementation of DistilBERT, another variant of BERT called RoBERTa was introduced. RoBERTa, similar to BERT, is a language model built on the transformer architecture. However, there are significant differences between RoBERTa and BERT in terms of their training methodologies (Liu et al., 2019). RoBERTa was trained on a larger dataset using a more effective training procedure. Specifically, RoBERTa was trained on a dataset consisting of 160GB of text, which is ten time the size BERT was trained on (Liu et al., 2019). Moreover, RoBERTa incorporates a dynamic masking technique during training, enhancing the model's ability to acquire robust and generalizable word representations (Liu et al., 2019).

3.3.3 BERTweet

The next model, BERTweet, is a public BERT-based model trained using the RoBERTa pretraining procedure (Baker et al., 2022). BERTweet, is the first public large-scale pre-trained language model for English Tweets (Nguyen, 2020). BERTweet was trained on 850 million English tweets collected from 2012 to 2019. This extensive training on tweet data equips the model with the ability to perform well on various downstream classification tasks involving tweets (Baker et al., 2022).

Considering the similarities between Reddit posts and tweets, both being user-generated short texts with the ability for replies, the implementation of BERTweet was deemed suitable for this research.

3.3.4 AlBERT

Albert, another variation on BERT, consists of 12 million parameters with 768 hidden layers and 128 embedding layers (Durgia, 2021). Making it a lighter version compared to the BERT base, which consists of 110 million parameters (Durgia, 2021). ALBERT model has, as expected, the lighter model reduced the training time and inference time.

3.4 XLNet

For the autoregressive transformer-based model, XLNet was used. By maximizing the expected likelihood across various permutations of the factorization order, XLNet effectively captures bidirectional contexts. This autoregressive formulation of XLNet allows it to overcome the limitations faced by BERT (Tsang, 2022). The limitation that XLNet overcomes is Pretrain-Finetune Discrepancy. Unlike BERT, XLNet does not rely on data corruption during pretraining, which eliminates the performance differences between pretraining and fine-tuning stages (Tsang, 2022). The second limitation XLNet overcomes is the independence assumption. BERT assumes masked tokens are independent, but XLNet's autoregressive objective allows for a more flexible approach by considering all permutations of token factorization, removing this assumption (Tsang, 2022).

3.5 Experimental Settings

3.5.1 DistilBERT

To begin with, the DistilBERT model is implemented. The process starts by initializing a tokenizer using the DistilBertTokenizerFast class from the Hugging Face library ("Hugging Face", n.d.), which is loaded with the pre-trained DistilBERT tokenizer ("DistilBERT", n.d.). This tokenizer is responsible for the conversion of text data into tokenized input suitable for the model. Furthermore, the data is split into train, test, and validation sets, with the test set comprising 30% and the validation set comprising 20% of the data.

After that, a class is defined that creates a dataset that holds tweets and their corresponding targets. It prepares the data for training a machine learning model by converting it into a format that can be understood by PyTorch.

Next, the DistilBERT model, a neural network, is defined. This model processes the raw text and generates a vector representation for each input text.

After obtaining the vector representation, it undergoes a dropout layer where a portion of the input units are randomly set to 0 during each update in the training process. This dropout technique is implemented to mitigate overfitting. Here, the dropout rate is specifically set to 0.3. Finally, a linear layer is applied to map the output to the 5 personality traits.

For model Training and evaluation, an instance of the DistilBERTForPersonalityTraits model is created. The models use the Adam optimizer with a learning rate of 1e-5, and the loss function is set to mean squared error (MSE). The training loop is implemented using 10 epochs.

Finally, after training completes, the model is evaluated on the test set. The test loss and other evaluation metrics (MSE, RMSE, MAE, R^2)$arecalculatedforeachpersonalitytrait.$

Similar implementation steps were followed for the other models, with the difference being the specific model class used. They were all from the Hugging Face library ("Hugging Face", n.d.).

3.6 Error analysis

Considering the time constraints, various aspects such as exploring different parameters, loss functions, optimizers, or adjusting the number of epochs were not thoroughly examined in this study. Instead, default hyperparameters were used for all models, including the Adam optimizer with a learning rate of 1e-5, the mean squared error (MSE) loss function, and a fixed number of 10 epochs. However, conducting thorough investigations in these areas could potentially result in notable improvements.

4 Data

This chapter will introduce the dataset that serves as the foundation for the models used in the study's analysis. This section contains the details regarding the data source and preprocessing steps employed in this research. The data utilized in this study were carefully selected to address the research objectives and ensure the reliability and validity of the results.

4.1 PANDORA Dataset

In 2020, Gjurković et al. addressed the challenge of limited datasets containing both personality traits and demographic labels (Gjurković et al., 2020). To tackle this issue, they developed a dataset named PANDORA, which contains Reddit comments annotated with three personality models: MBTI, Enneagram, and the Big Five model. For the purpose of this paper, we will focus on the Big Five model. Additionally, the dataset includes demographic information such as age, gender, and location. This dataset contains over 10,000 users, offering a rich collection of Reddit comments from diverse authors (Gjurković et al., 2020). Given the extensive nature of this dataset and the variety of authors represented, it presents an ideal resource for text-based personality detection.

4.1.1 Data exploration

The PANDORA dataset consists of three zip files. The first zip file contains files related to the collected Reddit comments. The PANDORA dataset contains over 17,000,000 Reddit posts. The second zip file contains baseline code and the third contains extensive information about each author, such as their gender, age, country, MBTI, Big Five personality traits, and more (Gjurković et al., 2021).

4.2 Data cleaning

The implementation of recent models, including those utilized in this research, demand large computational power. Considering the vast size of the PANDORA dataset, we decided to only include English Reddit posts in this research. This downsizing resulted in a reduction of the dataset from 17,640,062 entries to 16,637,210, effectively reducing it by approximately 1,000,000 entries.

To complete the data, the comments and data from the 'authorprofiles' file were merged to augment each entry with the corresponding big five personality traits.

However, it is worth noting that some entries within the PANDORA dataset had missing values (NaN) for one or more Big Five personality traits. As a result, these instances were eliminated from the dataset, resulting in the removal of 13,806,899 entries.

In the dataset, the 'body' column contained the Reddit comments of varying sizes. Upon attempting to run the models on comments of both long and short sizes, it was observed that restricting the body size to 100 words led to improved performance. In order to enhance the model's effectiveness, comments exceeding 100 words were excluded from the dataset (n=2830311).

Additionally, information, such as MBTI scores, country, and age, was eliminated from the dataset since it is not relevant to the current research being conducted.

Furthermore, the personality scores were normalized to ensure they ranged from zero to one, enhancing the consistency of the dataset.

4.2.1 Sampling

The size (n = 3,006,655) of the dataset posed a significant challenge in this research. The models employed in the study required extensive training time due to the large amount of data available and the number of parameters included in the models. Unfortunately, the available time constraints did not allow for sufficient training with the complete dataset.

Therefore, we decided to undersample the dataset. During this process, particular attention was given to unique authors and their frequency of appearance in the dataset.

Upon closer examination (Figure 1), it was evident that certain authors appeared a few thousand times in the dataset (max = 52066), while others appeared only 1 time.

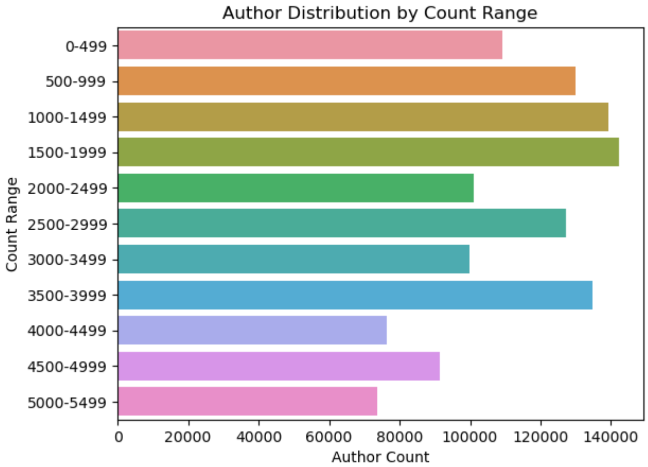

Figure 1: Author Count

This raised a concern regarding the potential dominance of a single author with a specific Big Five personality trait profile in a random sample. Such dominance is undesirable, as it can negatively affect the quality of training data for personality prediction. A random sample predominately representing one author can undermine the generalizability and diversity of the sample, which is crucial for accurate personality prediction.

To ensure diversity in the sample, we followed some steps to ensure that each author appeared only once in the random sample. This approach aimed to create a more representative and balanced dataset. The dataset consisted of 1598 unique authors, leading to a final sample size of 1568 entries.

In particular, and in order to evaluate the representativeness of the sample data, two box plots were generated, comparing the sample data to the cleaned dataset. The box plots, shown in Figures 1 and 2, provided insights into the distribution and characteristics of the sample data in relation to the larger dataset.

The boxplots clearly show that the sample data and the cleaned data have similar distributions of personality traits. This finding leads to the conclusion that the sample is indeed a strong representative of the cleaned data.

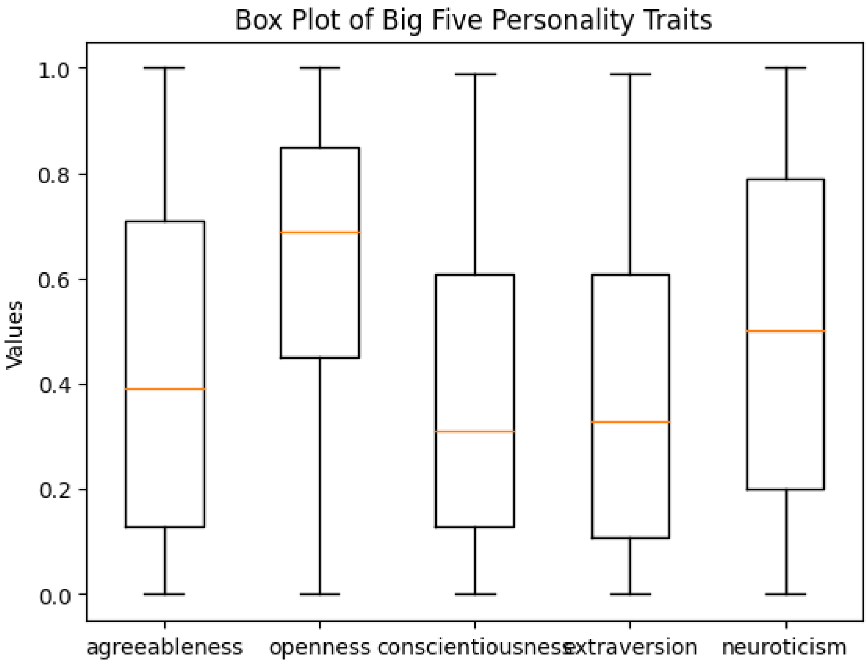

Figure 2: boxplot personality traits (n = 2733397)

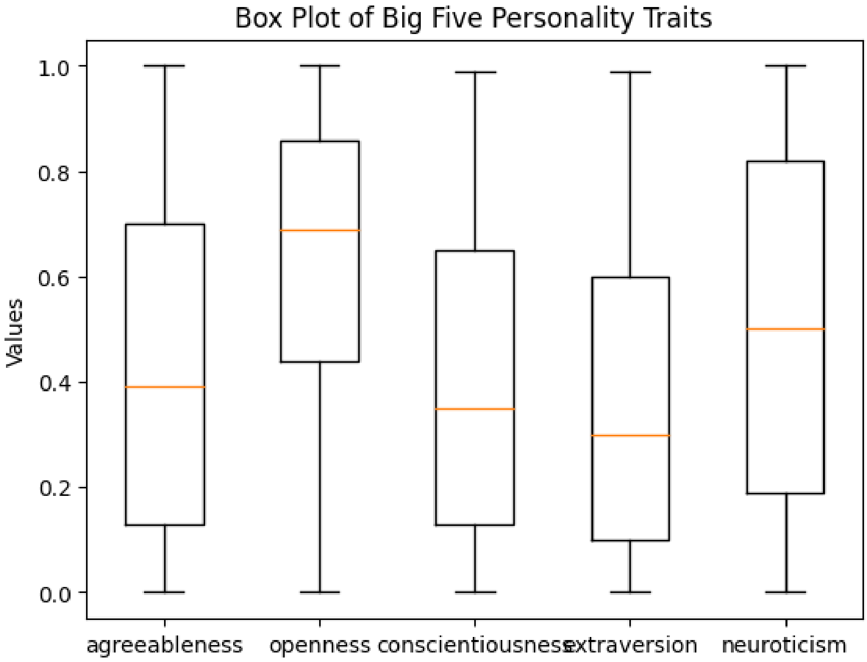

Figure 3: boxplot personality traits sample (n = 1568)

4.3 Text preprocessing

Preprocessing is a crucial step commonly taken in data science applications. In this study, the Reddit comments underwent processing. Initially, punctuation marks such as exclamation points and dollar signs were eliminated. Subsequently, stopwords were also removed from the comments.

4.3.1 Stopwords

The elimination of stopwords is an important step in text preprocessing. Stopwords are commonly used words that provide little analytical meaning and value in text mining tasks (Alshanik et al., 2020a). In English, some examples of stopwords include words like "the," "is," and "and."

The removal of stopwords during the preprocessing stage serves a few important purposes: improve the extraction of meaningful information, saving time, and reducing the document size. By eliminating stopwords, the resulting text becomes more concise and streamlined. A reduction in unnecessary words not only speeds up subsequent analysis, but also helps to decrease the overall size of the documents, making them more manageable for storage and processing (Kaur, 2018).

For the removal of stopwords, the stopword list for the English language provided by NLTK was utilized (Alshanik et al., 2020b).

These preprocessing steps help to clean and refine the data, ensuring that irrelevant elements are eliminated and improving the quality of the dataset for further analysis.

This resulted in a clean dataset containing authors with their Reddit posts and labels for their Big Five personality traits.

	author	body	agreeableness	openness	conscientiousness	extraversion	neuroticism
0	***	Tho ..	0.30	0.70	0.15	0.15	0.50
1	***	Ok ..	0.09	0.59	0.05	0.72	0.07
2	***	It s ..	0.77	0.73	0.73	0.01	0.98
3	***	ok w ..	0.09	0.61	0.13	0.04	0.72
4	***	Will ..	0.79	0.84	0.86	0.53	0.01

Table 2: Dataset used for analysis

4.3.2 Ethical considerations

To ensure responsible handling of the PANDORA dataset, which comprises actual Reddit posts from real users, specific precautions must be taken. Prior to accessing the PANDORA dataset, a person is required to adhere to a set of terms of use (Gjurković et al., 2021). To maintain the anonymity of the dataset and protect user identities, the authors' names were consistently removed whenever the dataset was shared with others or used in research.

Additionally, the following ethical considerations were addressed:

Bias and Fairness: Careful attention was given to potential biases within the dataset to ensure fair representation. Steps were taken to identify and mitigate any biases that may arise during data collection and analysis.

Transparency and Accountability: A commitment to transparency was upheld by providing detailed information about the data collection procedures and data cleaning procedures. This allows for the reproducibility of the study.

5 Results and Analysis

The following section presents the results obtained from the conducted research, shedding light on the findings that address the research questions. This part aims to provide a comprehensive analysis and interpretation of the data collected and analysed during the study.

5.1 Intercorrelations Dataset

Tables 3 and 4 display the intercorrelations observed between the Big Five personality traits within the cleaned dataset and the sample derived from that dataset.

	Ext	Neu	Agre	Con	Ope
Ext	1.0	-	-	-	-
Neu	-.29	1.0	-	-	-
Agre	-.06	.05	1.0	-	-
Con	.07	-.24	.13	1.0	-
Ope	.23	.05	.12	-.07	1.0

Table 3: Intercorrelations cleaned dataset (n=2830311)

	Ext	Neu	Agre	Con	Ope
Ext	1.0	-	-	-	-
Neu	-.25	1.0	-	-	-
Agre	.02	.05	1.0	-	-
Con	.07	-.27	.05	1.0	-
Ope	.23	.00	.12	-.03	1.0

Table 4: Intercorrelations sample dataset (n=1568)

The intercorrelations between the cleaned dataset and the sample demonstrate a high level of similarity, meaning that the sample is a good representation of the overall dataset. There are a few notable small differences observed. Specifically, in the sample, the trait pair Agreeableness (Agre) Extraversion (Ext) exhibits a change in correlation from negative to positive compared to the cleaned dataset. Additionally, the correlations between Openness (Ope) - Neuroticism (Neu) and Contentiousness

(Conn) - Agreeableness (Agre) are stronger in the sample than in the cleaned dataset. Despite these minor discrepancies, the sample remains a good representation of the dataset.

5.1.1 Comparing the Sample with Theory

Here, we aim to determine if the intercorrelations observed in our sample are consistent with the findings reported by Van der Linden et al. (2010)(table 5). By doing so, we can assess whether the patterns observed in our data align with established theories and prior research.

	S	L
O-C	-.03	.20
O-E	0.23	0.43
O-A	.12	.21
O-N	.00	-.17
C-E	.07	.29
C-A	.05	.43
C-N	-.27	-.43
E-A	.02	.26
E-N	-.25	-.36
A-N	.05	-.36

Note: S = sample, L = van der Linden

Table 5: Comparison Table Sample

We have found some disparities between the outcomes of Van der Linden et al.'s (2010) study and the sample data. Firstly, the correlation between Extraversion (Ext) and Agreeableness (Agre) was observed to be weaker compared to previous expectations. Similarly, the correlation between Contentiousness (Con) and Extraversion (Ext) exhibited a weaker association. Additionally, the correlation between Openness (Ope) and Extraversion (Ext) was also found to be weaker than anticipated.

Reddit is a platform where users engage in conversations and share content. The data mentioned is derived from Reddit posts. Reddit is a platform designed to facilitate discussion among users. This discussion-oriented nature of the platform may contribute to the discovery of less agreeable personalities in the data extracted from these posts, as the interactive and sometimes strongly opinionated nature of discussions can show diverse opinions and perspectives.

Furthermore, a significant deviation from the expected pattern was identified in the correlation between Agreeableness (Agre) and Neuroticism (Neu). Instead of the anticipated negative correlation, a positive correlation was observed. Similarly, the correlation between Openness (Ope) and Neuroticism (Neu) was close to zero, contrary to the expected negative correlation.

Moreover, the correlation between Conscientiousness (Con) and Agreeableness (Agre) displayed a weaker association, differing from previous assumptions. Lastly, the correlation between Openness (Ope) and Conscientiousness (Con) exhibited a negative correlation, contrary to the anticipated positive correlation.

Due to the discrepancies found, the intercorrelations predicted by the models will also deviate from the intercorrelations of personality traits reported in the research.

The discrepancies observed in the intercorrelations of the personality traits may be attributed to the composition of the PANDORA dataset. It has been noted that not all the labelled data for the Big Five in the dataset were derived from surveys; rather, some of the data were predicted or inferred (Gjurković et al., 2020). In contrast, Van der Linden et al. primarily utilized psychologically-based Big Five personality scores in their research (Van der Linden et al., 2010). Another possible reason is that the sample used for analysis, although it has the same distribution as the dataset, may still have inherent differences that affect the accurate representation of correlations present in the entire dataset.

5.2 Performance of Models

Table 6 displays the performance metrics of all models across different evaluation criteria. The models evaluated include RoBERTa, DistilBERT, BERTweet, Albert, and XLNet. Each model's performance is measured using various metrics such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and R-squared (R^2) score.

5.2.1 Metrics

In table 6, the metrics used to evaluate performace are explained. The first being Mean Squared Error (MSE). MSE measures the average squared difference between the predicted values and the actual values(Chugh, 2022). A lower MSE indicates better model performance, with values closer to zero indicating a better fit to the data. Next, Root Mean Squared Error (RMSE). RMSE is the square root of the MSE and provides the measure of the average magnitude of the errors in the same unit as the target variable(Chugh, 2022).Like MSE, a lower RMSE indicates better model performance, with values closer to zero indicating a better fit. The Mean Absolute Error (MAE) calculates the average absolute difference between the predicted values and the actual values (Chugh, 2022). Similar to MSE and RMSE, a lower MAE indicates better model performance. Finally, the R-squared (R^2) Score. R-squared is a statistical measure that represents the proportion of the variance in the dependent variable (target) that is predictable from the independent variables (predictions)(Chugh, 2022). It ranges from 0 to 1, where 0 indicates that the model does not explain any of the variability in the target variable, and 1 indicates that the model perfectly predicts the target variable (Chugh, 2022). Higher R^2 scores indicate better model performance.

These metrics are commonly used to evaluate the accuracy and performance of models, providing insights into how well the model fits the data and the magnitude of errors between predicted and actual values.

5.2.2 Performance

Model	Ext				Neu				Agre			
	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE	R^2
DistilBERT	0.096	0.311	0.271	-0.013	0.080	0.284	0.237	-0.026	0.099	0.315	0.276	-0.008
RoBERTa	0.096	0.310	0.271	-0.007	0.080	0.283	0.236	-0.023	0.101	0.317	0.280	-0.023
BERTweet	0.098	0.313	0.276	-0.028	0.082	0.287	0.237	-0.052	0.101	0.318	0.280	-0.025
Albert	0.096	0.310	0.273	-0.011	0.0206	0.454	0.383	-1.629	0.230	0.480	0.396	-1.333
XLNet	0.117	0.342	0.293	-0.228	0.095	0.309	0.258	-0.217	0.146	0.383	0.318	-0.487

Model	Con				Ope			
	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE	R^2
DistilBERT	0.095	0.309	0.267	-0.041	0.103	0.321	0.280	-0.023
RoBERTa	0.092	0.304	0.262	-0.008	0.100	0.316	0.278	0.007
BERTweet	0.095	0.308	0.267	-0.038	0.103	0.321	0.281	-0.025
Albert	0.218	0.467	0.380	-1.382	0.230	0.479	0.400	-1.288
XLNet	0.107	0.327	0.276	-0.166	0.153	0.391	0.325	-0.525

Table 6: Accuracy Metrics for Multiple Models

The results indicate that across all personality traits, the RoBERTa model achieved the lowest MSE, RMSE, and MAE values, suggesting better performance in predicting these traits. However, the R^2 scores for all models are negative, indicating that the models did not perform well in explaining the variance in the personality traits. Further analysis and improvement of the models may be required to enhance their predictive performance.

The models' poor performance can be attributed to several factors. Firstly, all the models were trained on the same dataset without fine-tuning the data to the specific models. It is important to fine-tune the data to better suit the requirements of each individual model architecture.

Furthermore, the presence of noise, outliers, or missing values in the training and evaluation data can have a detrimental effect on the model's performance. While missing values were handled by deleting corresponding entries, an alternative approach could have been to impute the missing data using appropriate techniques. This could potentially enhance the quality and completeness of the dataset.

Additionally, the limitations of the chosen model architectures, their complexity, or the training process itself may have contributed to the suboptimal results. It might be beneficial to explore more sophisticated models specifically designed for predicting personality traits, or experiment with different hyperparameters to improve the model's performance.

5.3 Intercorrelations of the models' predictions

Next, we were also interested in the examination of intercorrelations among personality traits when they were predicted by the utilized models. This analysis holds significance as it allows us to determine whether the individual prediction of traits, as commonly conducted in APR (Automatic Personality Recognition) research, poses any concerns, or if it eventually leads to intercorrelations that align with established psychological theories.

The tables below depict the intercorrelations identified among the Big Five model traits, determined by the employed models.

Table 7 shows the comparison between the outcome of the DistilBERT model with the van der Linden et al. (2010) outcome and the sample.

	S	L	D
O-C	-.03	.20	-.05
O-E	0.23	0.43	.07
O-A	.12	.21	.10
O-N	.00	-.17	.05
C-E	.07	.29	-.24
C-A	.05	.43	-.19
C-N	-.27	-.43	-.08
E-A	.02	.26	.02
E-N	-.25	-.36	.03
A-N	.05	-.36	.08

Note: S = sample, L = van der Linden, D = DistilBERT

Table 7: Comparison Table DistilBERT

Among the various trait pairs, the correlation found by DistilBERT between Agreeableness (Agre) and Openness (Ope) is the only one that resembles the correlations identified by Van der Linden et al. (2010).

Table 8 shows the comparison between the outcome of the RoBERTa model with the van der Linden et al. (2010) outcome and the sample.

	S	L	R
O-C	-.03	.20	-.16
O-E	0.23	0.43	.04
O-A	.12	.21	.01
O-N	.00	-.17	-.11
C-E	.07	.29	-.10
C-A	.05	.43	.07
C-N	-.27	-.43	.43
E-A	.02	.26	.44
E-N	-.25	-.36	.20
A-N	.05	-.36	.12

Note: S = sample, L = van der Linden, R = RoBERTa

Table 8: Comparison Table RoBERTa

The personality trait pairs Agreeableness (Agre) - Extraversion (Ext) and Openness (Open) - Neuroticism (Neu) show similarities to the correlations identified by Van der Linden et al. (2010). Furthermore, the pairs Openness (Open) - Extraversion (Ext), Conscientiousness (Con) - Agreeableness (Agre), and Openness (Open) - Agreeableness (Agre) exhibit positive correlations in both cases, although the correlations are stronger in the findings of Van der Linden et al. (2010).

The intercorrelations among the predicted personality traits by BERTweet have some differences compared to the research by Van der Linden et al. (2010) (table 9)

	S	L	T
O-C	-.03	.20	.11
O-E	0.23	0.43	.09
O-A	.12	.21	.04
O-N	.00	-.17	.25
C-E	.07	.29	19
C-A	.05	.43	.22
C-N	-.27	-.43	.45
E-A	.02	.26	.03
E-N	-.25	-.36	.16
A-N	.05	-.36	.37

Note: S = sample, L = van der Linden, T = BERTweet

Table 9: Comparison Table BERTweet

However, some similarities can be identified between the BERTweet and van der Linden et al. (2010) outcome:

Firstly, the correlation between Conscientiousness (Con) and Extraversion (Ext) shows a similar trend in both studies, indicating some agreement. They have a positive correlation that lays around the 0.20-0.30.

Secondly, the trait pairs Agreeableness (Agre) - Extraversion (Ext), Openness (Ope) - Extraversion (Ext), Conscientiousness (Con) - Agreeableness (Agre), Agreeableness (Agre) - Openness (Ope), and Conscientiousness (Con) - Openness (Ope) exhibit positive correlations in both the BERTTweet model and the research conducted by Van der Linden et al. (2010). However, the correlations observed in the BERTTweet outcome appear to be weaker compared to the reported correlations in Van der Linden's study.

A notable difference arises in the Neuroticism (Neu) column. While Van der Linden's research reports a negative correlation, the BERTTweet results show a positive correlation. It is interesting to note that the magnitudes of the positive correlation values in the BERTTweet outcome are quite similar to the magnitudes of the negative correlation values reported in Van der Linden's study.

The final BERT version model that was implemented in our study is the Albert model. The intercorrelations found by the Albert model have many disparities compared to the research conducted by Van der Linden et al.(2010) (table 10).

	S	L	A
O-C	-.03	.20	.96
O-E	0.23	0.43	.01
O-A	.12	.21	.97
O-N	.00	-.17	.95
C-E	.07	.29	-.02
C-A	.05	.43	.96
C-N	-.27	-.43	.98
E-A	.02	.26	-.01
E-N	-.25	-.36	-.10
A-N	.05	-.36	.95

Note: S = sample, L = van der Linden, A = Albert

Table 10: Comparison Table Albert model

According to the Albert model, the correlation between Openness (Ope) and Extraversion (Ext) is minimal, while van der Linden et al. (2010) found a correlation between these two traits. Similarly, the correlations between Openness-Neuroticism (Ope-Neu), Agreeableness-Neuroticism (Agree-Neu), and Conscientiousness-Neuroticism (Con-Neu) are negative in van der Linden et al.'s research (2010), but the Albert model shows positive correlations for these trait pairs.

In terms of Conscientiousness-Extraversion (Con-Ext) and Extraversion-Agreeableness (Ext-Agree), van der Linden et al. (2010) observed positive correlations, whereas the Albert model indicates negative correlations for these trait pairs.

However, there are some trait pairs that exhibit alignment between the ALBERT model and the findings of Van der Linden et al. (2010):

The correlation between Extraversion (Ext) and Neuroticism (Neu) is negatively correlated in both the ALBERT model and the research by Van der Linden et al. (2010). The trait pairs Conscientiousness (Con) - Agreeableness (Agre), Openness (Ope) - Agreeableness (Agre), and Openness (Ope) - Conscientiousness (Con) show positive correlations in both the ALBERT model and the research by Van der Linden et al. (2010). However, the correlations are notably stronger in the ALBERT model results.

The XLNet model was the final model that we implemented.

	S	L	XL
O-C	-.03	.20	-.09
O-E	0.23	0.43	-.58
O-A	.12	.21	-.74
O-N	.00	-.17	-.31
C-E	.07	.29	.13
C-A	.05	.43	.03
C-N	-.27	-.43	.12
E-A	.02	.26	.58
E-N	-.25	-.36	.31
A-N	.05	-.36	.35

Note: S = sample, L = van der Linden, XL = XLNet

Table 11: Comparison Table XLNet

When comparing the correlations found by XLNet (table 11) to those discovered by van der Linden et al. (2010), a few small similarities were identified:

Firstly, both the XLNet model and the research conducted by Van der Linden et al. (2010) reveal a positive correlation between Agreeableness (Agre) and Extroversion (Ext). However, the XLNet model demonstrates a higher correlation in this trait pair.

Next, for the trait pair Extroversion (Ext) and Conscientiousness (Con), van der Linden et al. (2010) identifies a stronger correlation between Extroversion and Conscientiousness compared to the XLNet model.

Additionally, both the XLNet model and Van der Linden et al. (2010) indicate a negative correlation between Openness and Neuroticism. However, the XLNet model exhibits a more pronounced negative correlation in this particular pairing.

lastly, the XLNet model and Van der Linden et al. (2010) find a positive correlation between Conscientiousness and Agreeableness. However, Van der Linden et al. (2010) reports a significantly higher correlation in this specific trait combination.

Table 12, shows the correlations of all models, the sample and the van der Linden et al. research (2010)

The substantial number of discrepancies between the intercorrelations observed in the model outcomes and those identified in the Van der Linden research (2010) are likely attributable to the negative R-squared score of the models.

	S	L	R	D	T	A	XL
O-C	-.03	.20	-.16	-.05	.11	.96	-.09
O-E	0.23	0.43	.04	.07	.09	.01	-.58
O-A	.12	.21	.01	.10	.04	.97	-.74
O-N	.00	-.17	-.11	.05	.25	.95	-.31
C-E	.07	.29	-.10	-.24	19	-.02	.13
C-A	.05	.43	.07	-.19	.22	.96	.03
C-N	-.27	-.43	.43	-.08	.45	.98	.12
E-A	.02	.26	.44	.02	.03	-.01	.58
E-N	-.25	-.36	.20	.03	.16	-.10	.31
A-N	.05	-.36	.12	.08	.37	.95	.35

Note: S = sample, L = van der Linden, R = RoBERTa, D = DistilBERT, T = BERTweet, A = Albert, XL = XLNet

Table 12: Comparison Table

Additionally, it is important to consider the process followed to create the datasets. The discrepancies could also be influenced by the dataset creation process itself, including the data collection methods and any preprocessing steps applied.

6 Discussion

In this section, we will discuss the findings and implications of our research, including the intercorrelations among personality traits when detected using computational models. In addition, we will compare those with intercorrelations found in theory. The challenges encountered in drawing conclusive results will be discussed, including below-average evaluation metrics and concerns about the reliability of the benchmark dataset used for training.

6.1 Results

6.1.1 Intercorrelations sample vs intercorrelations theory

The intercorrelations found in the sample taken from the PANDORA and those found in the meta-analysis, conducted by van der Linden et al. (2010), differed. This can mean that during the preprocessing and data cleaning, certain important factors were removed. Additionally, the unique nature of Reddit posts, characterized by discussions and opinions, could account for the disparities between the intercorrelations found in the PANDORA dataset and the research conducted by van der Linden et al. (2010).

However, another factor can be that the PANDORA dataset had some limitations:

By looking closer at the paper on the PANDORA dataset, we can see that there were multiple challenges they experienced when obtaining the Big Five labels (Gjurković et al., 2020).

The first challenge was the lack of standardized formats for Big Five test scores, unlike the MBTI. This led to a variety of scoring formats being used (Gjurković et al., 2020).

Additionally, test scores could be presented in different ways, such as raw scores, percentages, or percentiles. Numeric scores could vary in range, while descriptive scores differed for each test. Descriptive terms like "typical" and "average" could correspond to the same underlying score (Gjurković et al., 2020).

Additionally, sometimes users manually entered their results or described them in their own words. However, this introduced issues such as the misspelling of trait names or repetitive results, due to users copying and pasting their words. This negatively impacts the usability and quality of the data entries (Gjurković et al., 2020).

Lastly, in certain cases, results did not stem from inventory-based assessments but instead originate from text-based personality prediction services (Gjurković et al., 2020).

The last challenge raises questions on this research's subject. If the dataset used for predicting personality traits includes scores from personality prediction services, it becomes questionable to consider it as the "truth". The dataset itself consists of predicted personality scores, which adds complexity to the reliability of using it as a foundation for training models.

6.1.2 Model accuracy

Based on the accuracy metrics, it was observed that the RoBERTa model displayed the lowest values for MSE, RMSE, and MAE, indicating superior performance in predicting the personality traits. However, all models demonstrated negative R-squared scores, meaning their ability is limited and can not explain the variability in the personality traits.

There could be several reasons for these discrepancies. One possibility is that the preprocessing choices made during data preparation could have influenced the model's performance. Additionally, the lack of exploration and tuning of hyperparameters might have limited the models' ability to capture the complexities of the task. It is also important to consider the model architectures themselves, as some models, like Albert, may be less capable of handling the intricacies of the task compared to more sophisticated models.

6.1.3 Research Questions

In order to address the research questions posed in this study, which are "What are the intercorrelations observed among personality traits when they are detected using computational models?" and "How do these correlations compare to the correlations found in self-report questionnaires?", there are challenges in drawing definitive conclusions.

Regarding the first research question, it was difficult to arrive at a conclusive answer due to the models' low R-squared evaluation. The models' performance did not have satisfactory results for accurately determining the intercorrelations among personality traits.

Similarly, for the second research question, similar difficulties were encountered. The intercorrelations of the dataset did not entirely align with the intercorrelations found in questionnaire-based research by van der Linden at al. (2010). This inconsistency further complicated the comparison between the correlations derived from computational models and those derived from self-report questionnaires.

6.2 Limitations

In this section, the limitations of the research will be analyzed.

During this study, the only dataset used was the PANDORA dataset. We acknowledge the potential benefits of incorporating more datasets; however, due to time limitations, we were unable to incorporate additional datasets into our analysis.

Accessible large-scale datasets annotated with Big Five labels and containing text data are relatively scarce, which limited our options.

Another avenue we could have explored is modifying the preprocessing and cleaning procedures for each model to optimize the data fit. By tailoring the preprocessing steps to each model's requirements, we could potentially enhance the performance of the models. However, this approach would introduce challenges in comparing the model performances, as the preprocessing steps would vary across models.

During the data preprocessing stage, the decision was made to remove stopwords from the text bodies collected from Reddit. However, research suggests that this may not always be beneficial (Riloff, 1995). In future studies, it could be advantageous to reconsider this choice and refrain from removing stopwords, as this has the potential to improve the performance of the models.

These considerations highlight the potential avenues for future research to expand the scope of datasets and refine the preprocessing techniques to improve model performance and facilitate comparative analysis.

Due to time constraints, the hyperparameters, loss functions, and optimizers were not adjusted during the training of the models. Unfortunately, this lack of fine-tuning resulted in suboptimal predictions. However, this presents an opportunity for future research to explore and improve upon these aspects.

7 Conclusion

This research began by examining various personality trait models and discussing their strengths and limitations. We decided to use the Big Five model. Additionally, we discussed various strengths and weaknesses of the APR approach. Furthermore, the meta-analysis conducted by Van der Linden et al. (2010) was discussed, highlighting the intercorrelations discovered among the different personality traits within the Big Five model. The study also noted that current APR research fails to consider these intercorrelations, which raises concerns about the validity of APR. This served as the motivation for the present study.

Subsequently, we implemented and evaluated various models—DistilBERT, RoBERTa, BERTweet, Albert, and XLNet— on predicting the Big Five personality traits. These models were trained on the PANDORA dataset. The performance of these models was analyzed, and their predicted intercorrelations were compared to the research findings of Van der Linden et al. (2010).

Our experiments revealed disparities in specific pairs of traits between the sample dataset employed in this research (obtained from the PANDORA Dataset) and the prior study conducted by Van der Linden et al. (2010). Moreover, the models predicted the personality traits and their correlations. However, they have had limited effectiveness in explaining the variability in the personality traits.

In conclusion, this study provides insights into the implementation and evaluation of various models for predicting personality traits using text data. The findings highlight the need for further exploration and improvement in the models' performance and their ability to accurately capture the intercorrelations among personality traits. Future research should consider investigating different parameters, loss functions, and optimization techniques to enhance the models' predictive capabilities and align them more closely with existing research findings.

Lastly, we will delve into future work and potential avenues for further investigation.

7.1 Future Work

This research serves as a stepping stone in exploring novel APR techniques, particularly in assessing their validity when considering the intercorrelations among the Big Five personality traits. Additionally, this research contributes to the advancement of current APR methodologies and existing studies that aim to predict personalities based on textual data by adding knowledge to the body in the field of APR. Especially by conducting comparative analyses, and validating the findings on a new dataset. These contributions collectively contribute to the overall understanding and prediction of personalities based on textual data.

However, there is still much work to be done in the field of APR, including further improvements and development of models, as well as the creation of annotated datasets specifically designed for personality prediction. Currently, such resources remain scarce.

Bibliography

Adi, G. Y. N., Tandio, M. H., Ong, V., & Suhartono, D. (2018). Optimization for automatic personality recognition on twitter in bahasa indonesia. *Procedia Computer Science*, *135*, 473–480.

Alshanik, F., Apon, A., Herzog, A., Safro, I., & Sybrandt, J. (2020a). Accelerating text mining using domain-specific stop word lists. *2020 IEEE International Conference on Big Data (Big Data)*, 2639–2648. https://doi.org/10.1109/BigData50022.2020.9378226

Alshanik, F., Apon, A., Herzog, A., Safro, I., & Sybrandt, J. (2020b). Accelerating text mining using domain-specific stop word lists. *2020 IEEE International Conference on Big Data (Big Data)*, 2639–2648.

Aroca-Ouellette, S., & Rudzicz, F. (2020). On losses for modern language models. *arXiv preprint arXiv:2010.01694*.

Baker, W., Colditz, J. B., Dobbs, P. D., Mai, H., Visweswaran, S., Zhan, J., Primack, B. A., et al. (2022). Classification of twitter vaping discourse using bertweet: Comparative deep learning study. *JMIR Medical Informatics*, *10*(7), e33678.

Barbaranelli, C., Caprara, G. V., Rabasca, A., & Pastorelli, C. (2003). A questionnaire for measuring the big five in late childhood. *Personality and individual differences*, *34*(4), 645–664.

Behaz, A., & Djoudi, M. (2012). Adaptation of learning resources based on the mbti theory of psychological types. *International Journal Of Computer Science Issues (IJCSI)*, *9*(1), 135.

Calabrese, W. R., Rudick, M. M., Simms, L. J., & Clark, L. A. (2012). Development and validation of big four personality scales for the schedule for nonadaptive and adaptive personality—second edition (snap-2). *Psychological assessment*, *24*(3), 751.

Christian, H., Suhartono, D., Chowanda, A., & Zamli, K. Z. (2021). Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. *Journal of Big Data*, *8*(1), 1–20.

Chugh, A. (2022). Mae, mse, rmse, coefficient of determination, adjusted r squared-which metric is better? [Accessed: 06 20, 2023].

Costa, P. T., & McCrae, R. R. (2008). The revised neo personality inventory (neo-pi-r). *The SAGE handbook of personality theory and assessment*, *2*(2), 179–198.

De Raad, B. (2000). *The big five personality factors: The psycholexical approach to personality*. Hogrefe & Huber Publishers.

Distilbert [Accessed: 06 20, 2023]. (n.d.).

Durgia, C. (2021). Exploring bert variants (part 1): Albert, roberta, electra. https://towardsdatascience.com/exploring-bert-variants-albert-roberta-electra-642dfe51bc23

Elliott, K. (n.d.). Myers-briggs vs. ocean: An industrial psychologist breaks down the differences.

Entringer, T. M., Gebauer, J. E., Eck, J., Bleidorn, W., Rentfrow, P. J., Potter, J., & Gosling, S. D. (2021). Big five facets and religiosity: Three large-scale, cross-cultural, theory-driven, and process-attentive tests. *Journal of Personality and Social Psychology*, *120*(6), 1662.

Fang, Q., Giachanou, A., Bagheri, A., Boeschoten, L., van Kesteren, E.-J., Kamalabad, M. S., & Oberski, D. L. (2022). On text-based personality computing: Challenges and future directions. *arXiv preprint arXiv:2212.06711*.

Ganesan, V. A. (2021). Empirical evaluation of pre-trained transformers for human-level nlp: The role of sample size and dimensionality. *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting*. https://doi.org/10.18653/v1/2021.naacl-main.357

Gjurković, M., Karan, M., Vukojević, I., Bošnjak, M., & Snajder, J. (2021). PANDORA talks: Personality and demographics on Reddit. *Proceedings of the Ninth International Workshop on Natural Language Processing for Social Media*, 138–152. https://doi.org/10.18653/v1/2021.socialnlp-1.12

Gjurković, M., Karan, M., Vukojević, I., Bošnjak, M., & Šnajder, J. (2020). Pandora talks: Personality and demographics on reddit. *arXiv preprint arXiv:2004.04460*.

Halim, Z., Atif, M., Rashid, A., & Edwin, C. A. (2019). Profiling players using real-world datasets: Clustering the data and correlating the results with the big-five personality traits. *IEEE Transactions on Affective Computing*, *10*, 568–584.

Hugging face [Accessed: 06 20, 2023]. (n.d.).

John, O. P., Robins, R. W., & Pervin, L. A. (2010). *Handbook of personality: Theory and research*. Guilford Press.

Kaur, J. (2018). Stopwords removal and its algorithms based on different methods. *International Journal of Advanced Research in Computer Science*.

Kici, D., Malik, G., Cevik, M., Parikh, D., & Basar, A. (2021). A bert-based transfer learning approach to text classification on software requirements specifications. *Canadian Conference on AI*.

Kumar, B. (2023). Bert variants and their differences - 360digitmg. https://360digitmg.com/blog/bert-variants-and-their-differences

Lim, A. G. (n.d.). Big five personality traits: The 5-factor model of personality.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Mapes, N., White, A., Medury, R., & Dua, S. (2019). Divisive language and propaganda detection using multi-head attention transformers with deep learning bert-based language models for binary classification. *Proceedings of the second workshop on natural language processing for internet freedom: censorship, disinformation, and propaganda*, 103–106.

McCrae, R. R., & Costa Jr, P. T. (1989). Reinterpreting the myers-briggs type indicator from the perspective of the five-factor model of personality. *Journal of personality*, *57*(1), 17–40.

Mushtaq, S., & Kumar, N. (2022). Text-based automatic personality recognition: Recent developments. *Proceedings of Third International Conference on Computing, Communications, and Cyber-Security: IC4S 2021*, 537–549.

Myers, I. B., McCaulley, M. H., Quenk, N. L., & Hammer, A. L. (1998). *Mbti manual: A guide to the development and use of the myers-briggs type indicator*. Consulting Psychologists Press.

Nguyen, D. Q. (2020). Bertweet: A pre-trained language model for english tweets. *ArXiv*.

Omheni, N. (2015). Automatic recognition of personality from digital annotations. https://doi.org/10.5220/0005483002730280

Ong, V., Rahmanto, A. D., Suhartono, D., Nugroho, A. E., Andangsari, E. W., Suprayogi, M. N., et al. (2017). Personality prediction based on twitter information in bahasa indonesia. *2017 federated conference on computer science and information systems (FedCSIS)*, 367–372.

Pedregon, C. A. (2012). Social desirability, personality questionnaires, and the "better than average" effect. *Personality and Individual Differences*. https://doi.org/10.1016/J.PAID.2011.10.022

Phan, L. V., & Rauthmann, J. F. (2021). Personality computing: New frontiers in personality assessment. *Social and personality psychology compass*, *15*(7), e12624.

Radisavljević, D., Batalo, B., Rzepka, R., & Araki, K. (2022). Myers-briggs type indicator and the big five model-how our personality affects language use. *2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, 1–6.

Ramezani, M., Feizi-Derakhshi, M.-R., & Balafar, M.-A. (2022). Text-based automatic personality prediction using kgrat-net: A knowledge graph attention network classifier. *Scientific Reports*, *12*(1), 21453.

Riloff, E. (1995). Little words can make a big difference for text classification. https://doi.org/10.1145/215206.215349

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). Distilbert, a distilled version of bert: Smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Satow, L. (2021). Reliability and validity of the enhanced big five personality test (b5t).

Sharon, J. A., Christinal, A. H., Chandy, D. A., & Bajaj, C. (2023). Application of intelligent edge computing and machine learning algorithms in mbti personality prediction. In *Intelligent edge computing for cyber physical applications* (pp. 187–215). Elsevier.

Skowron, M., Tkalčič, M., Ferwerda, B., & Schedl, M. (2016). Fusing social media cues: Personality prediction from twitter and instagram. *Proceedings of the 25th international conference companion on world wide web*, 107–108.

Stachl, C. (2019). Behavioral patterns in smartphone usage predict big five personality traits.

Stricker, L. J., & Ross, J. (1964). An assessment of some structural properties of the jungian personality typology. *The Journal of Abnormal and Social Psychology*, *68*(1), 62.

Tandera, T., Suhartono, D., Wongso, R., Prasetio, Y. L., et al. (2017). Personality prediction system from facebook users. *Procedia computer science*, *116*, 604–611.

Tenney, I., Das, D., & Pavlick, E. (2019). Bert rediscovers the classical nlp pipeline. *arXiv preprint arXiv:1905.05950*.

Tsang, S.-H. (2022). Review-xlnet: Generalized autoregressive pretraining for language understanding. https://sh-tsang.medium.com/review-xlnet-generalized-autoregressive-pretraining-for-language-understanding-39e48bed2337

Van der Linden, D., te Nijenhuis, J., & Bakker, A. B. (2010). The general factor of personality: A meta-analysis of big five intercorrelations and a criterion-related validity study. *Journal of research in personality*, *44*(3), 315–327.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Vinciarelli, A., & Mohammadi, G. (2014). A survey of personality computing. *IEEE Transactions on Affective Computing*, *5*(3), 273–291.

Zhao, X., Tang, Z., & Zhang, S. (2022). Deep personality trait recognition: A survey. *Frontiers in Psychology*, 2390.

Appendix A: Intercorrelations found by models

	Ext	Neu	Agre	Con	Ope
Ext	1.0	-	-	-	-
Neu	.03	1.0	-	-	-
Agre	.02	.08	1.0	-	-
Con	-.24	-.08	-.19	1.0	-
Ope	.07	.05	.10	-.05	1.0

Table 13: Intercorrelations DistilBERT Model

	Ext	Neu	Agre	Con	Ope
Ext	1.0	-	-	-	-
Neu	.20	1.0	-	-	-
Agre	.44	.12	1.0	-	-
Con	-.10	.43	.07	1.0	-
Ope	.04	-.11	.01	-.16	1.0

Table 14: Intercorrelations RoBERTa Model

	Ext	Neu	Agre	Con	Ope
Ext	1.0	-	-	-	-
Neu	.16	1.0	-	-	-
Agre	0.03	.37	1.0	-	-
Con	.19	.45	.22	1.0	-
Ope	.09	.25	.04	.11	1.0

Table 15: Intercorrelations BERTweet Model

	Ext	Neu	Agre	Con	Ope
Ext	1.0	-	-	-	-
Neu	-.10	1.0	-	-	-
Agre	-.01	.95	1.0	-	-
Con	-.02	0.98	.96	1.0	-
Ope	.01	.95	.97	.96	1.0

Table 16: Intercorrelations Albert Model

	Ext	Neu	Agre	Con	Ope
Ext	1.0	-	-	-	-
Neu	.31	1.0	-	-	-
Agre	.58	.35	1.0	-	-
Con	.13	.12	.03	1.0	-
Ope	-.58	-.31	-.74	-.09	1.0

Table 17: Intercorrelations XLNet Model