MDPI

Article

New Approach of Measuring Human Personality Traits Using Ontology-Based Model from Social Media Data

Andry Alamsyah *[ID]**, Nidya Dudija and Sri Widiyanesti**

School of Economics and Business, Telkom University, Bandung 40257, Indonesia;
nidyadudija@telkomuniversity.ac.id (N.D.); widiyanesti@telkomuniversity.ac.id (S.W.)
* Correspondence: andrya@telkomuniversity.ac.id

Abstract: Human online activities leave digital traces that provide a perfect opportunity to understand their behavior better. Social media is an excellent place to spark conversations or state opinions. Thus, it generates large-scale textual data. In this paper, we harness those data to support the effort of personality measurement. Our first contribution is to develop the Big Five personality trait-based model to detect human personalities from their textual data in the Indonesian language. The model uses an ontology approach instead of the more famous machine learning model. The former better captures the meaning and intention of phrases and words in the domain of human personality. The legacy and more thorough ways to assess nature are by doing interviews or by giving questionnaires. Still, there are many real-life applications where we need to possess an alternative method, which is cheaper and faster than the legacy methodology to select individuals based on their personality. The second contribution is to support the model implementation by building a personality measurement platform. We use two distinct features for the model: an n-gram sorting algorithm to parse the textual data and a crowdsourcing mechanism that facilitates public involvement contributing to the ontology corpus addition and filtering.

1. Introduction

The Big Five Inventory (BFI) is formulated as a concise instrument to represent human personality [1]. Although it sounds radical to propose around 44 short questions answered in only 5 min response time, it is achievable to measure the Big Five dimensions of personality traits. In the 1990s, most instruments were much longer [2]; even the short form of the NEO-PI-R [3] has 60 questions. Today, there is a growing demand for super-short measures since the availability of large-scale data pushes toward a trend to shorter personality instruments. Many researchers that implement the BFI demand a more concise version to support faster and real-time measurement results. Several examples of the trend toward minimal measurement are the single-item self-esteem scale [4], single-item ability ratings [5], and 10-item measure of the Big Five [6,7]. Many super-short instruments produce a good psychometric characteristics result, implying that a BFI short version is feasible [8].

Personality measurement is commonly attainable in many ways, such as through interviews and questionnaires [9]. The self-administered questionnaire is widely utilized for personality measurement inside the psychological research domain [10]. This method is in vast utilization because the questionnaire shows performance with adequate reliability and is highly effective in measuring personality for the number of individuals [11]. Nevertheless, this method is usually challenged by a falsely answering respondent that causing an inaccurate result. On the other hand, an interview is another way to perform a personality measurement with the benefits of using sophisticated instruments to hinder misunderstandings [12]. Self-administered interviews generally provide better privacy

concerns and fewer sensible problems such as socially enticing responses [10]. The challenge in using interviews to measure personality is the probability of producing normative and biased results when the process is not integrated with written psycho-tests, which will be costly and time-consuming [9].

Hence, the idea to use user-generated data is logical since this approach is less obstructive than the legacy methodology of extracting personality traits. With the availability of a large volume of user-generated content through social media, we have the opportunity to turn that digital trace into user character descriptions. Some research uses machine learning to measure personality traits from social media automatically; for example, in earlier research, machine learning was used to analyze large-scale Twitter data with exceptional accuracy [13,14]. The use of machine learning provides advantages in terms of analysis time and various algorithms that can help measure extensive data [15]. Machine learning can analyze large amounts of data with the prediction algorithm to recognize personality in various data forms such as textual, speech, and graphics [16]. On the other hand, machine learning relies heavily on statistical computation; it has some weaknesses with the inability to understand the meaning and intention of some phrases and words. This weakness resulting from a machine learning method has difficulty acquiring common sense [17–19].

This study performs a rejuvenation state-of-the-art personality measurement using a new approach that combines automatic version and domain expert knowledge. The computerized version borrows machine learning ideas to classify personality traits in a fast and scalable manner. At the same time, domain expert helps us build the terms-based library we call corpus as the primary reference of terms to personality traits relations. Together, this combination is called the ontology model. We apply the ontology model to map personality traits from textual data on social media, i.e., Twitter conversational data. This research intended to distinguish the novelty and challenge of building a platform integrated with the ontology model. Social media textual data are always challenging since most posts do not follow formal language rules; more slang, street language, and occasionally temporary jargon were used.

Furthermore, there are not many studies in understanding the contextual meaning of the Indonesian language. We decipher meaning from social media textual data in the Indonesian language by two distinct features that we brought up here. First, we use the n-gram algorithm to parse textual data more accurately to the Indonesian language than our previous approach using the radix tree algorithm [20]. Second, to increase the ontology corpus quality, we invite the public to contribute to the curation process, such as adding, voting up, and voting down texts or phrases.

2. Theoretical Background

2.1. Personality Measurement

Personality measurement is a systematic method to measure several features of a person's characteristic of their interpersonal style according to specific rules. Then, one can use this measurement to predict a person's responses in a bound setting. The definition includes many different procedures, such as interviews, integrity tests, the Minnesota Multiphasic Personality Inventory (MMPI), and in-basket exercises [21]. The real issue about personality measurement is "what is a good personality measurement" instead of "what is personality measurement". A good personality measurement at least consists of two features. The first feature is the score should be stable temporarily, i.e., the score reliable over time. The second feature should be credible to measure and predict real-world performance. Even though many instruments aim to measure personality, hardly a few shares meet the two simple yet pivotal features above.

Personality scales are typically described as self-report measures, but this is misleading. The processes that govern responses to items on personality scales are formally identical to those underlying social interaction in general [22]. During the exchange, people generally try to manage how others perceive them, and they seek to control their reputations, increase

positive attention, and decrease criticism. Answering questionnaire items is like talking with a hidden interviewer. People use their item responses to tell a hidden interviewer who they are and how they would like to be seen. Thus, item endorsements are scaled samples of a person's typical interpersonal style, which build up their reputation for how others perceive them. Gough [23] states that what personality scales measure represents what they predict, and what they expect best is observers' ratings. This means that both personality scale scores and observers' ratings are rough indexes of reputation. Additionally, it is the link between scale scores and reputation that explains why well-constructed personality scales predict nontest behavior.

2.2. Social Media and Big Five Personality

The Internet has pushed online social networking to grown dramatically over the last decade. According to Twitter statistics for 2020, Twitter alone has exceeded 330 million monthly active user's members, 500 million tweets every single day, and 23 percent of the internet population are on Twitter. Users reveal many aspects about themselves when creating social networking profiles regarding what they share and how they say it. By posting self-description, status updates, photos, and interests, much of a user's personality emerges through their profile. For decades, psychology researchers have attempted to understand nature systematically. After a comprehensive effort to establish and validate a widely accepted personality model, researchers have found the connections between general personality traits and many types of behavior. Relationships have been discovered between personality and psychological disorders, job satisfaction, job performance, and even romantic accomplishment [24].

The five-factor personality model, better known as the Big Five, is acknowledged as the most comprehensive, reliable, and helpful set of personalities to date [25,26]. In this personality parameter, words and phrases are associated with the five scores corresponding to the five main personality traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism [27].

There have been few studies on how personality impacts interactions on social media, especially Twitter [28]. These studies have analyzed the impact of personality primarily on:

a. Using social media services: An extroverts character tends to find social media easy to use and valuable.
b. Selecting social contacts: Users tend to choose contacts with similar Agreeableness, Extraversion, and Openness. However, generally they prefer to stay in touch with people of high Agreeableness.
c. Keeping many contacts: As one expects, the personality trait that keeps the most with social connections is Extraversion.

People use Twitter in different ways: Zhao and Rosson [29] highlight the fact that people use Twitter for several social goals, for instance (1) staying in touch with friends and colleagues; (2) boosting the visibility of exciting things to one's social networks; (3) collecting useful information about one's profession or other personal interests; (4) seeking for help and opinions; and (5) releasing emotional stress. They also state that the way people use Twitter can be categorized into three broad classes: (1) updating personal life activities in a blog-like way of using Twitter; (2) following real-time information in a journalistic style, and (3) following people-based RSS feeds, which is a way to be informed about personal interests.

In recent years, many scholars showed interest in Twitter and from a Natural Language Processing perspective; for example, Pak and Paroubek [30] built a sentiment analysis classifier from Twitter data to automatically recognize when a post is about positive, negative, or neutral emotions. Zhao et al. [31] proposed a ranking algorithm for extracting topic key phrases from tweets. Finin et al. [32] performed Named Entity Recognition on Twitter using crowdsourcing services such as Mechanical Turk and Crowdflower to provide the first step towards semantic annotation in a Social Network Site domain [33]. Several resources regarding social media and Big Five Personality [34,35].

2.3. Ontology Model

Ontology is a formal representation of the explicit specifications of a collection of concepts. Ontology classifies the vocabulary and taxonomy that model a domain with definitions of objects, concepts, properties, and relationships [35,36]. The ontology may also be a collection of interconnected classes and subclasses, with existing classes indicating domain entities and specific interrelations between the entities [37].

We look back at some previous work on this topic to comprehensively understand the ontology approach to measure personality based on social media activity usage. Sewwandi [37] approached this methodology by designing the model using the ontology web language (OWL), a well-known ontology developing language. The research categorized the data acquired into Eysenck's three-factor personality model with help from an eminent psychologist. Several researchers work on the theoretical and practical implementation of the connection between ontology models and psychological profiles. Egloff [38] introduces an ontology model for inferring psychological profiles to capture and formally measure characteristics in digital humanity. In an earlier effort, McCrae et al. [39] created a model called Lexicon Model for Ontologies (Lemon) as the common platform to share the terminology and lexicon resources from the semantic web; however, some researchers use this general model to a specific application in psychology, but primarily for the English language. Furthermore, other research created the same ontology model, using Indonesian language and Big Five Personality Traits as the cornerstone of the study, which was also assisted and assessed by a psychologist [40]. Noy and McGuinness [41] stated that ontology modeling in personality measurement is mainly used to share a common understanding of the information structure, reuse, and analyze the domain knowledge.

The research proposes classifying social media textual data into one of the five personality traits in the Big Five model. Big Five traits are significantly associated with users' behaviors on social media [42]. For instance, an individual with high Extraversion has been identified as having high social media activity [43]. On the other hand, a higher Neuroticism individual has shown the opposite behavior; they tend to self-disclose hidden aspects of themselves and use social media to learn about other people in a submissive way [44].

3. Methodology

3.1. Participants

We utilize Twitter as our primary textual data source for ease of use, provide large-scale data, and open access. Thus, our participants are all Twitter members. We investigate how the Indonesian language is used as a daily means of communication. We collect as many language variations as possible in informal situations setting. Thus, our ontology model corpus is constructed from thousands of tweets, regardless of the individual personality. Our previous research [20] is based on several prominent individuals as our sample with the qualifications as follows: (1) a verified account, (2) having more than 1000 tweets or 500,000 followers, (3) post variation of topics in their tweet, (4) show many interactions or conversations to others, and (5) not a protected account. We have successfully created an ontology model corpus based on the tweets of three famous figures. Those tweets have been verified and mapped to the correct personality traits by a panel of psychology experts. This paper enriches the previous corpus with new tweets from a more extensive collection of individuals. By enlarging the scope of individuals, our objective is not to limit the spectrum of various types of Indonesian language as the input for the ontology model corpus.

3.2. Procedure

This research proposes constructing a tool or platform to recognize a user's personality according to their online activity (e.g., Tweets status). The ontology model is then built using free, open-source tools for logic description known as Protégé. Further, the ontology model is reconstructed using the n-grams language model algorithm to undertake the

phrase parsing process into words, allowing it to be executed inside a shell platform. Utilizing an n-gram based approach is an upgrade from the previous parsing process using the radix-tree method [20]. We have found that the radix-tree has several issues in parsing phrases of the Indonesian language structural language. The architectural research diagram is shown in Figure 1. Each phrase from the input text is parsed using the n-gram algorithm, and hereafter, they are checked to the ontology model. Inside the ontology model, an ontology process maps the word dictionary into the personality trait list from the Big Five model. After successfully mapping the input text to the corresponding personality trait, the model computes the probability of a person who expresses via the text into several personality traits. The input process, followed by the parsing process, and then the mapping process, and at last compute probability of personality trait is part of the personality measurement platform workflow.
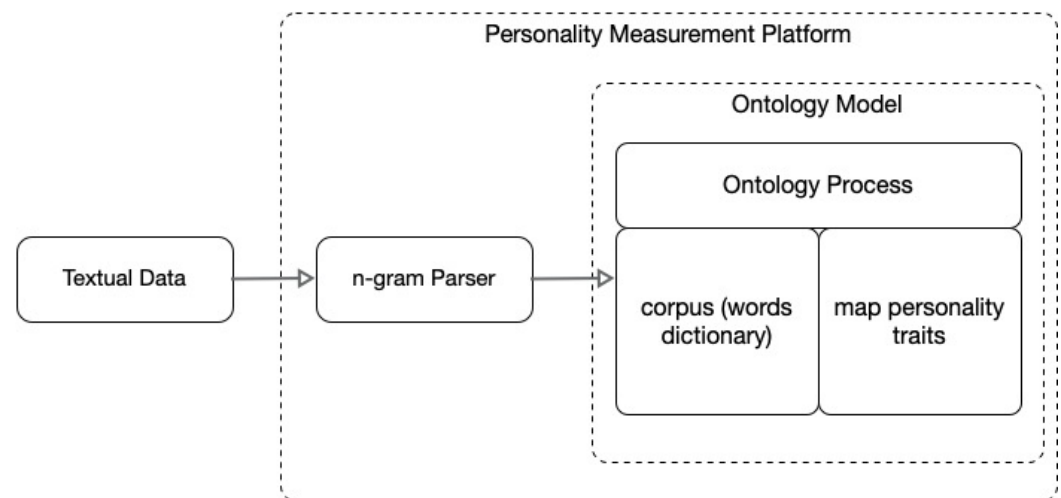

Figure 1. Ontology model workflow under Personality Measurement Platform.

3.3. Datasets

We collect Twitter data following participants' guidelines mention in 3.1. Twitter allows us to collect their data through their Application Programming Interface (API) protocol. The Twitter platform enables researchers to simplify their work by mining conversational data by accessing large-scale data in a specific range of time. The datasets used in this research are available to other researchers via our open platform. Therefore, it could be used as the benchmark of new research methods and approaches.

Our previous research successfully measures human personality with the ontology model that consists of 2125 words that correspond to personality traits in the Big Five model. To have better accuracy in the ontology model, we enrich the ontology corpus by adding words and phrases from several four Twitter account. The Twitter account sample requirements are as follows:

1. Public figure's account.
2. Actively interacting with other users.
3. Giving opinions.
4. Share a lot of daily activities.

Those criteria are used in our previous research to consider having datasets that can resemble human personality. We observe the real character by observing their interactions, so the sample must interact with other users. Using the Twitter API, we acquired 7328 tweets from four different Twitter accounts, filtered into 4389 tweets classified into 6889 words to represent the user's personality traits.

We filter all the data collected using pre-processing steps to gain more accuracy and relevance in personality measurement. Pre-processing is one of the most critical steps before performing data analysis [45]. Conducting data pre-processing is one way to achieve

more significant meaning and information [46]. In this research, the pre-processing steps are divided into two phases. The first step is to remove unnecessary features from data retrieved via Twitter API, to have datasets that include only the tweets themselves and the retweet status. The second step is to remove retweeted tweets so that the data consist only of tweets generated from the user as an individual. Currently, in total, the personality corpus has 10,265 words/phrases categorized into 5 traits and 30 sub traits/facets. The number expectedly grows when the public on crowdsourcing mechanism has fully participated in corpus enrichment.

3.4. Ontology Model Development

The ontology model development is intended to classify words and phrases from the acquired textual data into facets and traits available on the Big Five Personality Theory. After undergoing pre-processing steps, the acquired tweets are classified into 6889 words and phrases corresponding to each personality trait in the Big Five model. After successfully creating an ontology model containing mapping words and phrases into corresponding personality traits, we continue to the next step to deploy the ontology model in the form of an application platform.

Generally, there are two ways to classify linguistic-featured data into personality traits: using domain expert judgment or using machine learning classification [35]. We use the domain expert judgment in this research since this approach has higher accuracy than the machine learning approach in human personality prediction. The domain expert judgment scenario is formalized into an ontology model. The domain expert role in the ontology model is

1. To ensure the correctness of the mapping.
2. To measure model performance or the accuracy of the personality class decision. The model is validated by two domain experts in the psychology discipline. The validation process requires the experts to validate every single keyword in the ontology model that corresponded to the available traits in the Big Five Personality model.

The role of ontology is to map all previously classified words into an enormous knowledge domain. In this research, the ontology model is designed using an ontology modeling language (OWL), using Protégé software with the aid of the OWL-DL package. The results of ontology construction in Protégé can be seen in Figure 2. There are some advantages of using protégé to develop a model of an ontology according to Sewwandi et al. [37], the benefits are as follows:

a. Protégé OWL provides multiuser support for synchronous knowledge entry.
b. Protégé OWL can be extended with back-ends for alternative file formats. Currents formats include Clips, XML, RDF, and OWL.

The categories in the ontology model are divided into main categories, subcategories, and individual levels. The main categories rely on the personality traits of the Big Five Personality model, which are Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. The subcategories are at the same levels as the personality facet, which is the sub traits in the Big Five Personality model that can be seen in Table 1. Words and phrases are placed on the leaf levels. An example is shown in Figure 3, considering the main category as Extraversion. Warmth and Gregariousness's subcategories are the facets of Extraversion, and words related to those subcategories are displayed.
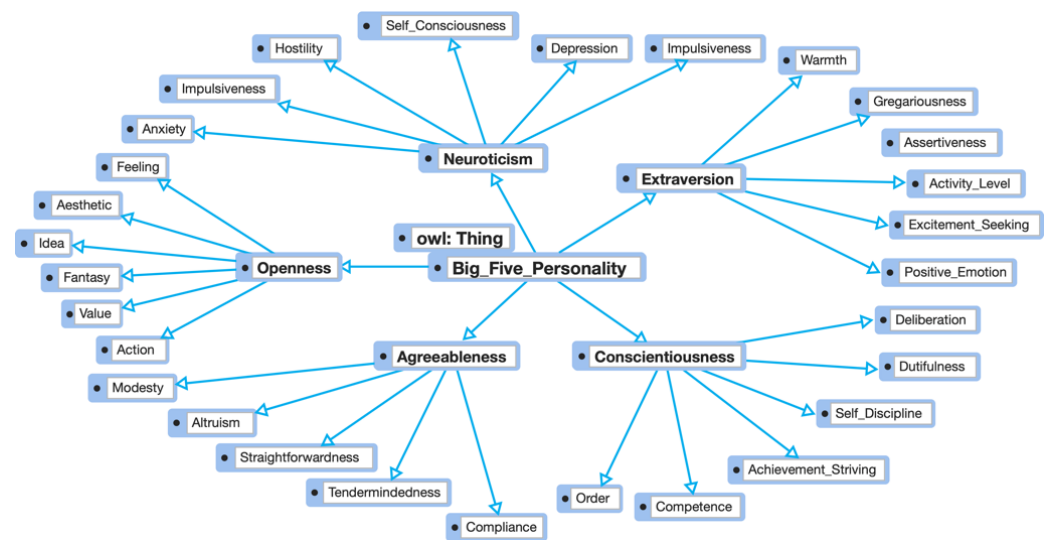
Figure 2. OWL Protégé Visualization.

Table 1. Big Five Personality Traits.

| Personality Traits	Definition	Sub-Trait/Facet
Openness	The openness to experience: the degree to which an individual exhibits intellectual curiosity, self-awareness, and nonconformance.	Aesthetic, Fantasy, Action, Idea, Feeling, Value.
Conscientiousness	The degree to which individuals value planning, acquire the tenacity quality, and achievement oriented.	Competence, Order, Dutifulness, Achievement-Striving, Self-Discipline, Deliberation.
Extraversion	The degree to which individuals involved with the external world, encounter enthusiasm and other positive emotions.	Warmth, Gregariousness, Assertiveness, Activity-Level, Excitement-Seeking, Positive Emotion.
Agreeableness	The degree to which individuals value mutual effort and social harmony, modesty, dignity, and trustworthiness.	Trust, Compliance, Altruism, Straightforwardness, Modesty, Tendermindedness.
Neuroticism	The degree to which individuals deal with negative feelings and their propensity to overreact emotionally.	Anxiety, Depression, Hostility, Self-Consciousness, Impulsiveness, Vulnerability.
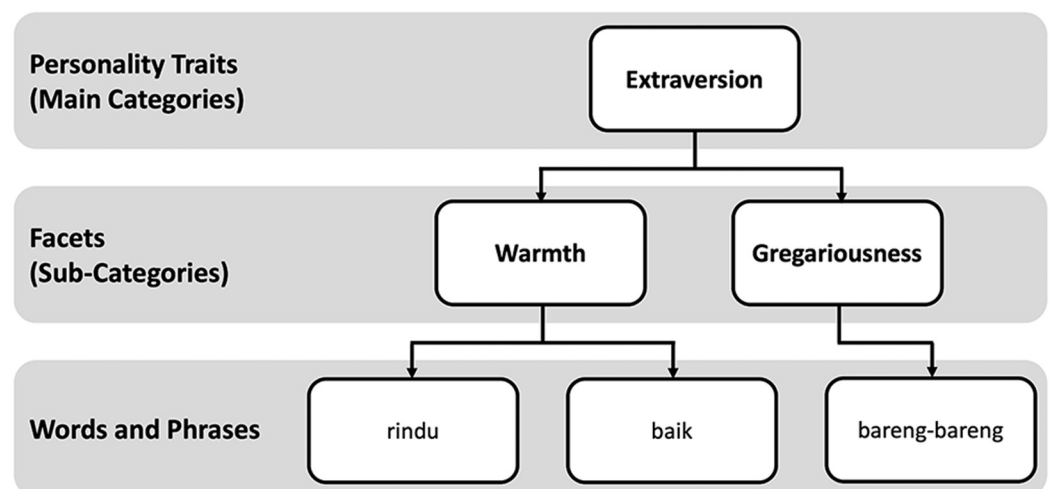
Figure 3. Main Categories and Subcategories in the Ontology Model.

At the final stage, the validated ontology model is classified into main categories and subcategories based on the Big Five Personality model, as shown in Figure 4. The example keywords shown are "rindu" and "bareng-bareng", two commonly used words in the Indonesian language. The word "Rindu" refers to the phrase "missing someone" in English. This word corresponds to the traits of Extraversion owing to the word's intention to express their feeling in a friendly way. The word "bareng-bareng" also corresponds to the Extraversion traits inconsequent to the word meaning in English, which is "together" that mainly expresses the feeling of having pleasure in being a part of a crowd or group of people. This ontology model will be the construction base of our proposed platform, which is intended to automate the process of words-to-personality mapping shown in Figure 4.
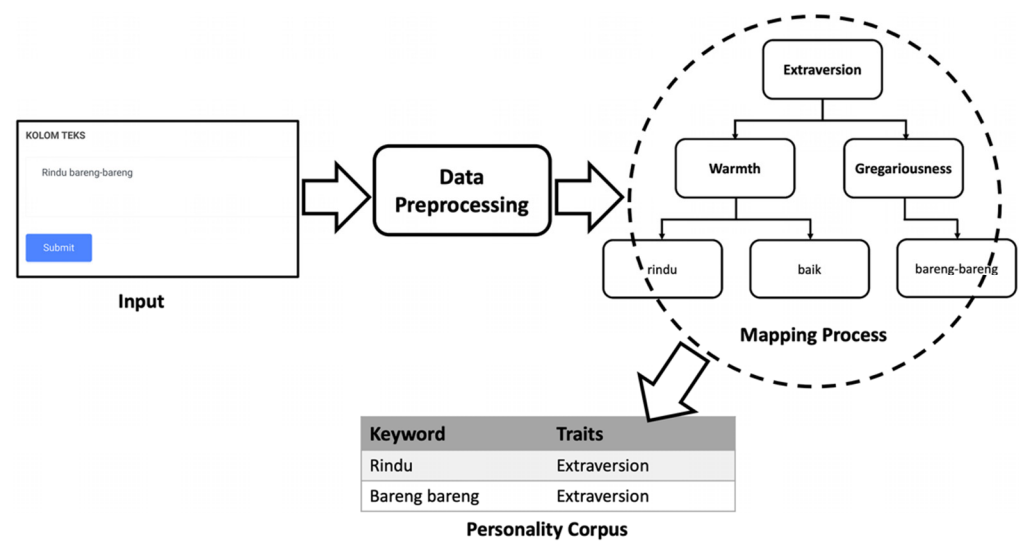

Figure 4. Personality Measurement Process.

3.5. Proposed Platform

The main objective of this research is to deploy the ontology model to the public. Public accessibility is essential for maintaining the model's high quality and facilitating public feedback by voting words or phrases into correct traits. By involving the public in the trait classification process, we use a crowdsourcing mechanism, which yields much better performance than domain expert judgment and improves the scalability and parallelization process. We provide the options of upvote or downvote on each word or phrase. The higher upvote represents the majority of the public agree that the particular word or phrase belongs to the destined personality trait. Those objectives are materialized in the form of an application platform as a Personality Measurement Platform or simply called the platform. However, we would periodically invite domain experts to check the corpus integrity and correctness from the crowdsourcing mechanism.

The platform measure personality only by inserting words or text in a short length of time. To achieve that, we need to build a versatile model that is highly adaptive and immensely flexible. Creating a web-based platform initiates an idea to have a long-term applicable platform with a low maintenance cost, ease of facilitation, and flexible management system. The platform constructed was made to simplify personality measurement for many uses, such as marketing intelligence purposes, social media influencer selection, and talent recruitment.

Our previous proposed platform can significantly measure human personality via words and phrases generated from their online activity [20]. It can parse every word that corresponds to available traits in the ontology model. The main challenge in our previous platform is the inability to distinguish phrases from words. In the ontology model that was previously built, some terms existed that reflect a human personality (e.g., really_hate, never_cry). Our previous platform could not yet differentiate two separate words and a single phrase; an example is shown in Table 2.

**Table 2.** Keyword and Traits Example.

Tweets	Keyword	Traits
Jelek banget	Jelek	Neuroticism
	Banget	Agreeableness
Jelek banget	Jelek_banget	Neuroticism

This challenge is one of our main works to improve from our previous platform. The work is instrumental in having better accuracy in measuring personality. By this enhanced work, the personality measurement platform can understand the human intention of speaking through the context of words, which is conceived into separate words or a single phrase. The way we can overcome this deficiency is by changing the main construction of our platform. Previously, our platform worked under the radix tree parsing algorithm. The way the radix tree works is unlike the regular way: the key at each node is compared chunk-of-bits by chunk-of-bits, where the quantity of bits in that chunk at that node is the radix r of the radix tree [47].

Despite the advantages that the radix tree had, the algorithm faced a considerable challenge. The radix tree works not optimally when it is confronting a phrase-reflected trait. The radix tree works well in parsing the English language but not the Indonesian language; thus, constructing an applicable engine needs to be done precisely to represent our model conveniently. We rearrange our platform algorithm to work under the n-gram language model, which is better to parse phrases in the Indonesian language. An n-gram is an n character slice of a longer string when a new document arrives for classification.

The system first computes its n-gram frequency profile. In the model we created, the n-gram language model was used to compare this profile against each category's profiles using an easily calculated distance measure [48,49]. We illustrate how the radix tree parsing algorithm works differently than n-gram in a brief illustration in Table 3.

Table 3. The radix tree and n-gram mechanism comparison.

Radix Tree	n-Gram
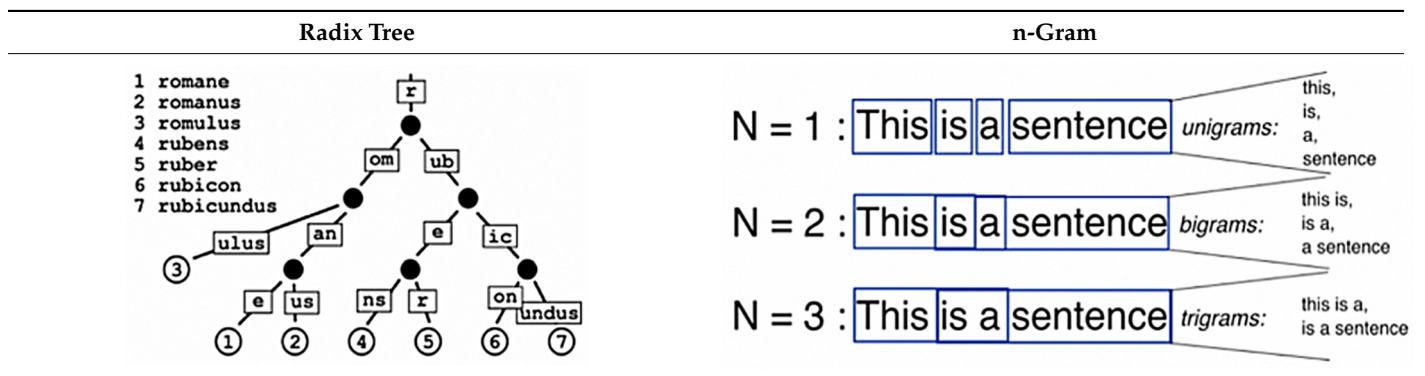
The n-gram parsing algorithm supports a more robust platform to undertake a personality measurement based on textual data. With the help of the n-gram language model, the platform can differentiate whether it is a phrase or a word from the document inserted, based on the words and phrases that reflected personality traits in the ontology model. The n-gram algorithm pseudo-code is shown in Table 4.

We explain the algorithm in Table 4. There are two main functions: the looping function and the if function. The looping function for j aims to examine the input of words or phrases in the ontology model that has been constructed. Then, the looping function for *k* is operating to do a comparative assessment of the input of words or phrases that will drive through the process of personality measurement and compare them with words or phrases found in the ontology model made previously. Furthermore, the if function of the algorithm is employed to check the words or phrases with the highest array of words. The types of words or phrases in the n-gram language model can be described as unigram

(one word), bigram (two words), trigram (three words), and n-gram (n-words). The if function works by detecting words with the largest n-gram values and discarding words with n-gram values smaller than those of other words or phrases. The model is constructed under a complex algorithm, with the pseudo-code shown in Table 4.

Table 4. The n-gram parsing algorithm.

Function	Algorithm

The looping for j function	while i < len(token):
	tmp = []
	tmp_trait = []
	for j in range(len(phrase)):
	if token[i] in phrase[j]:
	tmp.append(phrase[j])
	tmp_trait.append(traits[j])
	max = 0
	trait = ' '

The looping for k function	for k in range(len(tmp)):
	if re.sub('_', ' ', tmp[k].lower()) in sent:
	if len(tmp[k].split('_')) > max:
	trait = tmp_trait[k]
	max = len(tmp[k].split('_'))

The if function	if max > 0:
	list_freq[list_trait.index(trait)] += 1
	i += max
	else:
	i += 1

The platform architectural diagram is shown in Figure 5. Our platform architectural diagram and framework representation consist of user test data input, personality corpus, and the n-gram language model parser. The results of textual personality measurement will be displayed in the form of a spider plot or radar diagram. We built the platform using the flask framework of the python programming language as the main engine and a SQL database mainly to manage the personality corpus. The interface for input and output is shown as a web platform. The personality corpus provides data feeds to the personality measurement platform. Together both map the word/phrases from the n-gram parsing process and show the result as personality radar. The employed technologies to support the mechanism are the searching and matching process, the input data and the personality corpus in the database.
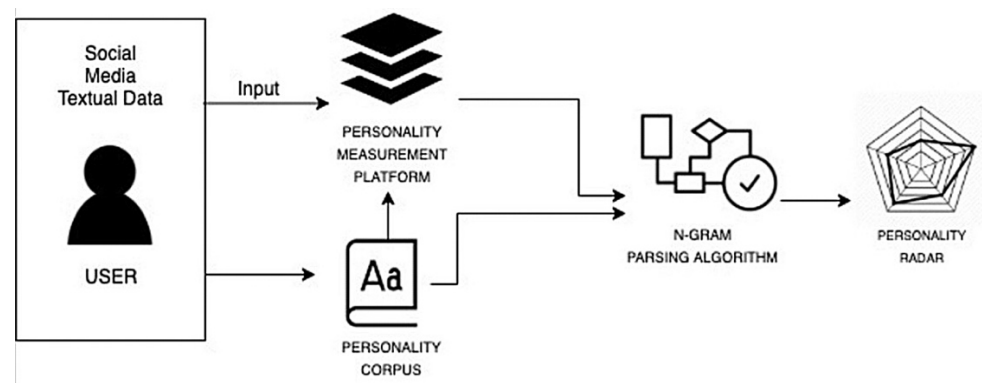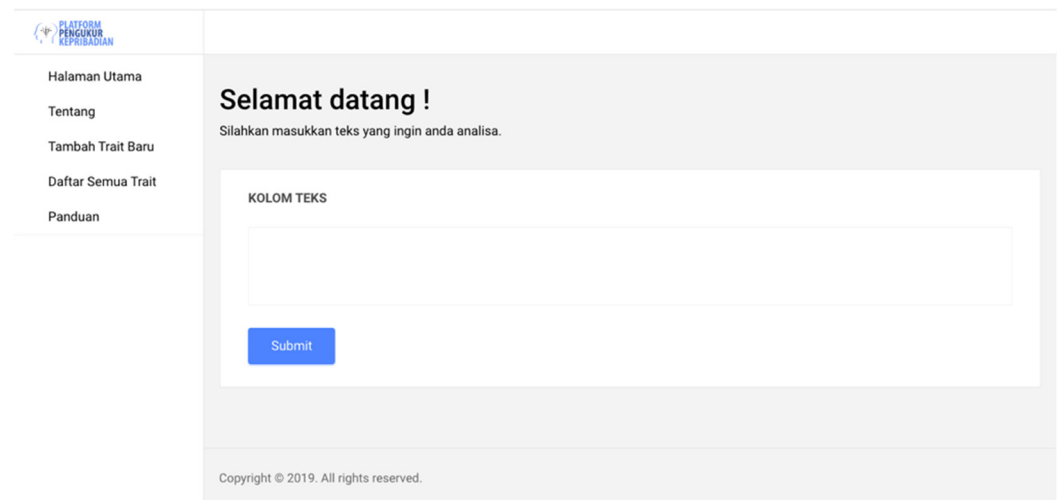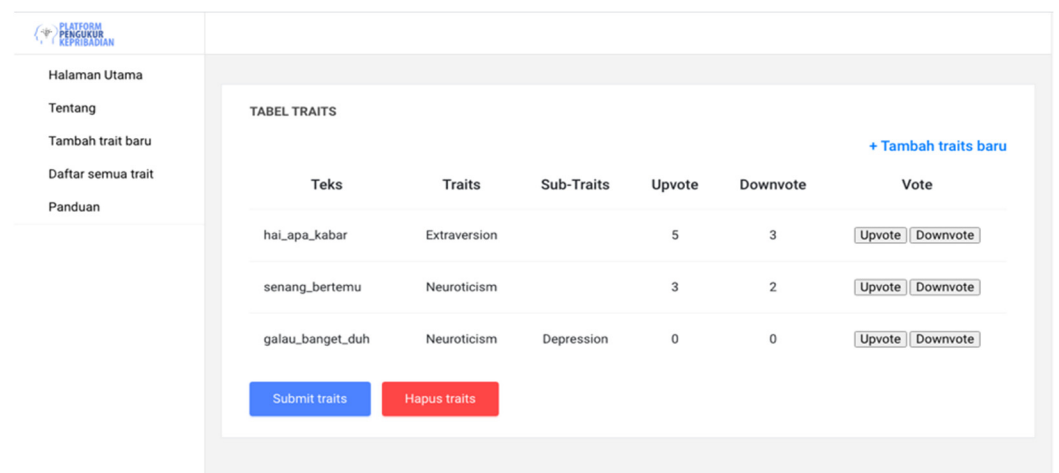

Figure 5. Platform architectural diagram.

The platform is called Platform Pengukur Kepribadian; this is the Indonesian language version of the Personality Measurement Platform. The platform can be accessed via the address http://kepribadian.labscbd.id (accessed on 7 October 2021). The platform interface

shown in Figure 6a consists of input text or phrase and submit button, while the voting page for the voting mechanism on each word or phrase is shown in Figure 6b.

(**a**)

(**b**)

Figure 6. (**a**) The platform interface, (**b**) The interface for voting mechanism in crowdsourcing.

3.6. Personality Measurement

We test the constructed personality measurement platform using several samples of a famous Indonesian Twitter account that the domain expert can easily verify. This step is essential to evaluate the reliability and accuracy of our personality mapping prediction. The most important is eliminating the parsing mechanism problem in the Indonesian language in our previous platform version. The sample accounts are @faldomaldini, @benakribo, @shitlicious, and @fajarnugros. Table 5 shows the measurement results in the form of a spider plot.

We see the personality traits of each account in Table 5. The personality traits of each account can be easily captured and discussed depending on our needs. To measure the actor personality consistency based on their tweets, we can also dynamically measure their personality over the number of tweets or over the designated time frame. Table 6 shows the result of account personality consistency based on the number of tweets measured. We consistently read the personality of @faldomaldini and @benakribo over 30 tweets, while *@fajarnugros* and @shitlicious gave different results over 30 tweets. We may find the same phenomena during our measurement; thus, we need to consider the complexity of human nature. The given text reading depends on behavior or personality at the measurement

time. We frequently measured the textual data over a more extended period or a more significant volume of tweets to give a more conclusive or convergent result into one or several dominant personality traits.

Table 5. Personality measurement results.

Account	Result
@faldomaldini	
@benakribo	
@shitlicious	

Table 5. *Cont.*

Account	Result
@fajarnugros	

Table 6. Personality measurement test.

Twitter Account	First 10 Tweets	First 20 Tweets	First 30 Tweets
@faldomaldini	Agreeableness	Agreeableness	Agreeableness
@benakribo	Openness	Openness	Openness
@fajarnugros	Openness	Agreeableness	Agreeableness
@shitlicious	Agreeableness	Agreeableness	Openness

4. Analysis and Conclusions

The burgeoning of brief personality measurement success is commonly depicted as two main factors: time-effectiveness and cost-effectiveness [50]. Nevertheless, time and cost are essential factors behind developing short measurement instruments in personality psychology in the last decade. The need for fast and cost-effective personality measurement has been growing to overcome rapid knowledge advancement [51]. Recent research seeks to develop a recourse to have a dashing and cost-effective method to measure human personality [52]. Language usage analysis is the most common way to have a fast and inexpensive personality determiner [9]. The last decade, over a hundred studies have linked linguistic feature usage to a wide range of psychological research [53].

Encouraged by the growing evidence of the connection between personalities and online behavior, researchers have begun to explore the use of digital footprints left by people on social media to derive the characteristics of the Big Five model [40]. Recent studies in this field have led to a typical research design. However, some studies vary in terms of the social media platform they used to gain textual data. For example, Park et al. [52] investigated the feasibility of predicting personality traits based on text features extracted from Facebook status updates using topic modeling techniques. Likewise, Liu et al. [17] and Qiu et al. [54] analyzed the language of text used on Twitter to create a predictive model for the Big Five feature. While Gao et al. [55], Li et al. [56], and Wei et al. [57] identified the characteristics of the Big Five theory sampled from the Sina Weibo microblog, and different combinations of digital footprints (activity vs. activity + language vs. activity + speech + image) were used in their analysis.

Our previous research uses knowledge-based representation known as ontology. Ontology provides a better way of performing accurate results based on human expertise [41] compared to other approaches such as machine learning. Most machine learning algorithms perform faster for simple language patterns at the current stage and in day-to-day implementation. Still, they have difficulty extracting complex language patterns, thus

failing to extract the contextual meaning of texts. Sharing a common understanding of the research by reusing the model and reanalyzing the model is another reason for using the ontology model [37]. The ontology model that we built in our previous research successfully measured human personality from social media textual data with high velocity.

The model can map human personality by classifying every single word posted by a person in the Indonesian language into a group of traits provided by the Big Five model.

Our research proposes a platform architecture that can efficiently run and execute personality measurements based on our ontology model. A similar approach should be conveniently implemented to other languages, with some considerations on how complex to map the words/phrases to the personality corpus, a suitable parsing algorithm, and the most important is the effort to build the corpus itself.

The Internet has provided a borderless world to share information, opinions, and interactions with others. Social media consumption has been a part of daily human life. Social media activity may expose human behavior and personality that is beneficial for many areas, including psychology, human resources management, and business management. By measuring personality traits based on the Big Five model, a person's personality can be depicted by the linguistic usage of that person. This research provided a technique to detect the personality of a person by ontology-based personality measurement.

This research proves that the ontology model is one rapid model that can be used in many areas. In this case, it is used to detect the personality of a person. We integrated our ontology model with the automation platform constructed to create a more significant implication for the community or organization by providing a faster and easy-to-use model.

Author Contributions: Conceptualization, A.A.; methodology, A.A. and N.D.; software, A.A. and S.W.; validation, N.D.; formal analysis, A.A. and N.D.; investigation, N.D.; resources, S.W.; data curation, S.W.; writing—original draft preparation, A.A. and N.D.; writing—review and editing, A.A. and N.D.; visualization, S.W.; supervision, A.A.; project administration, S.W.; All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not Applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. John, O.P.; Donahue, E.M.; Kentle, R.L. *Big Five Inventory (BFI)*; American Psychological Association: Washington, DC, USA, 1991.
2. Goldberg, L.R. The development of markers for the Big Five factor structure. *Psychol. Assess.* **1992**, *4*, 26–42. [CrossRef]
3. Costa, P.T.; McCrae, R.R. Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychol. Assess.* **1992**, *4*, 5–13. [CrossRef]
4. Robins, R.W.; Hendin, H.M.; Trzesniewski, K.H. Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg self-esteem scale. *Personal. Soc. Psychol. Bull.* **2001**, *27*, 151–161. [CrossRef]
5. Rammstedt, B.; Rammsayer, T. Gender differences in self-estimated intelligence and their relation to gender-role orientation. *Eur. J. Personal.* **2002**, *16*, 369–382. [CrossRef]
6. Gosling, S.D.; Rentfrow, P.J.; Swann, W.B., Jr. A very brief measure of the Big-Five personality domains. *J. Res. Personal.* **2003**, *37*, 504–528. [CrossRef]
7. Rammstedt, B.; John, O.P. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *J. Res. Personal.* **2007**, *41*, 203–212. [CrossRef]
8. Burisch, M. Test length and validity revisited. *Eur. J. Personal.* **1997**, *11*, 303–315. [CrossRef]
9. Farr, J.L.; Tippins, N.T. (Eds.) *Handbook of Employee Selection*; Taylor & Francis Group: New York, NY, USA, 2017.
10. Hilgert, L.; Kroh, M.; Richter, D. The effect of face-to-face interviewing on personality measurement. *J. Res. Personal.* **2016**, *63*, 133–136. [CrossRef]
11. Hunt, C.; Andrews, G. Measuring personality disorder: The use of self-report questionnaires. *J. Personal. Disord.* **1992**, *6*, 125–133. [CrossRef]

12. Dwivedula, R.; Bredillet, C.N.; Müller, R. Personality and work motivation as determinants of project success: The mediating role of organizational and professional commitment. *Int. J. Manag. Dev.* **2016**, *1*, 229–245. [CrossRef]

13. Golbeck, J.; Robles, C.; Edmondson, M.; Turner, K. Predicting Personality from Twitter. In Proceedings of the IEEE 3rd International Conference on Privacy, Security, Risk, and Trust and the 3rd International Conference on Social Computing, Boston, MA, USA, 9–11 October 2011.

14. Setyawan, M.A. Ontological Search Engine on Twitter to Collect Data for Bandung Happiness Index Measurement. In Proceedings of the Indonesia Symposium on Computing (IndoSC), Bandung, Indonesia, 24–25 September 2016.

15. Pratama, B.Y.; Sarno, R. Personality classification based on Twitter text using Naive Bayes, KNN, and SVM. In Proceedings of the IEEE International Conference on Data and Software Engineering (ICoDSE), Yogyakarta, Indonesia, 25–26 November 2015; pp. 170–174.

16. Stachl, C.; Au, Q.; Schoedel, R.; Gosling, S.D.; Harari, G.M.; Buschek, D.; Volkel, S.T.; Schuwerk, T.; Oldemeier, M.; Ullman, T.; et al. Predicting personality from patterns of behavior collected with smartphones. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 17680–17687. [CrossRef] [PubMed]

17. Farnadi, G.; Sitaraman, G.; Sushmita, S.; Celli, F.; Kosinski, M.; Stillwell, D.; De Cock, M. Computational personality recognition in social media. *User Modeling User-Adapt. Interact.* **2016**, *26*, 109–142. [CrossRef]

18. Bleidorn, W.; Hopwood, C.J.; Wright, A.G. Using big data to advance personality theory. *Curr. Opin. Behav. Sci.* **2017**, *18*, 79–82. [CrossRef]

19. Alamsyah, A.; Widiyanesti, S.; Putra, M.R.D.; Sari, P.K. Personality Measurement Design for Ontology-Based Platform using Social Media Text. *Adv. Sci. Technol. Eng. Syst. J.* **2020**, *5*, 100–107. [CrossRef]

20. Hathaway, S.R.; McKinley, J.C. *The Minnesota Multiphasic Personality Inventory*; American Psychological Association: Washington, DC, USA, 2016.

21. Hogan, R.; Hogan, J. Personality, and status. In *Personality, Social Skills, and Psychopathology: An Individual Differences Approach*; Gilbert, D.G., Connolly, J.J., Eds.; Plenum Press: New York, NY, USA, 1991; pp. 137–154. [CrossRef]

22. Gough, H.G. Cross-Cultural Validation a Measure of Asocial Behavior. *Psychol. Rep.* **1965**, *7*, 379–387. [CrossRef]

23. Conway, M.; O'Connor, D. Social media, big data, and mental health: Current advances and ethical implications. *Curr. Opin. Psychol.* **2016**, *9*, 77–82. [CrossRef] [PubMed]

24. Goldberg, L.R.; Johnson, J.A.; Eber, H.W.; Hogan, R.; Ashton, M.C.; Cloninger, C.R.; Gough, H.G. The international personality item pool and the future of public-domain personality measures. *J. Res. Personal.* **2006**, *40*, 84–96. [CrossRef]

25. Costa, P.T.; McCrae, R.R. The five-factor model of personality and its relevance to personality disorders. *J. Personal. Disord.* **1992**, *6*, 343–359. [CrossRef]

26. Cieciuch, J.; Łaguna, M. The Big Five and beyond: Personality traits and their measurement. *Rocz. Psychol.* **2014**, *17*, 249–257.

27. Quercia, D.; Kosinski, M.; Stillwell, D.; Crowcroft, J. Our Twitter Profiles, Our Selves: Predicting Personality with Twitter. In Proceedings of the 2011 IEEE 3rd International Conference on Privacy, Security, Risk and Trust and 2011 IEEE 3rd International Conference on Social Computing, Security, Boston, MA, USA, 9–11 October 2011; pp. 180–185. [CrossRef]

28. Zhao, D.; Rosson, M.B. How and why people Twitter: The role that micro-blogging plays in informal communication at work. In Proceedings of the ACM 2009 International Conference on Supporting Group Work (GROUP '09), Sanibel, FL, USA, 10–13 May 2009; pp. 243–252. [CrossRef]

29. Pak, A.; Paroubek, P. Twitter-based system: Using Twitter for disambiguating sentiment ambiguous adjectives. In Proceedings of the ACL 2010 5th International Workshop on Semantic Evaluation, Uppsala, Sweden, 15–16 July 2010; pp. 436–439.

30. Zhao, W.X.; Jiang, J.; He, J.; Song, Y.; Achananuparp, P.; Lim, E.P.; Li, X. Topical keyphrase extraction from Twitter. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA, 19–24 June 2011; pp. 379–388.

31. Finin, T.; Murnane, W.; Karandikar, A.; Keller, N.; Martineau, J.; Dredze, M. Annotating named entities in Twitter data with crowdsourcing. In Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, Los Angeles, CA, USA, 6 June 2010; pp. 80–88.

32. Barbier, G.; Liu, H. Data mining in social media. In *Social Network Data Analytics*; Springer: Boston, MA, USA, 2011; pp. 327–352.

33. Farnadi, G.; Zoghbi, S.; Moens, M.F.; De Cock, M. Recognizing personality traits using Facebook status updates. In Proceedings of the 7th International AAAI conference on weblogs and social (WCPR13), Bostan, MA, USA, 11 July 2013.

34. Lambiotte, R.; Kosinski, M. Tracking the digital footprints of personality. *Proc. IEEE* **2014**, *102*, 1934–1939. [CrossRef]

35. Wu, W.; Li, H.; Wang, H.; Zhu, K.Q. Probase: A probabilistic taxonomy for text understanding. In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, Scottsdale, AR, USA, 20–24 May 2012; pp. 481–492.

36. Aaberge, T.; Akerkar, R. Ontology and Ontology Construction: Background and Practices. *Int. J. Comput. Sci. Appl.* **2012**, *9*, 32–41.

37. Sewwandi, D.; Perera, K.; Sandaruwan, S.; Lakchani, O.; Nugaliyadde, A.; Thelijjagoda, S. Linguistic features-based personality recognition using social media data. In Proceedings of the 6th National Conference on Technology and Management (NCTM), Malabe, Sri Lanka, 27 January 2017; pp. 63–68.

38. Egloff, M.; Lieto, A.; Picca, D. An Ontological Model for Inferring Psychological Profiles and Narratives Roles of Character. In Proceedings of the Digital Humanities Conference, Mexico City, Mexico, 26–29 June 2018.

39. McCrae, J.; Spohr, D.; Cimiano, P. Linking Lexical Resources and Ontologies on Semantic Web with Lemon. In Proceedings of the 8th Extended Semantic Web Conference (ESWC), Heraklion, Greece, 29 May–2 June 2011; pp. 245–259.

40. Alamsyah, A.; Putra, M.R.D.; Fadhilah, D.D.; Nurwianti, F.; Ningsih, E. Ontology Modelling Approach for Personality Measurement Based on Social Media Activity. In Proceedings of the 6th International Conference on Information and Communication Technology (ICoICT), Bandung, Indonesia, 3–5 May 2018; pp. 507–513. [CrossRef]

41. Noy, N.F.; McGuinness, D.L. *Ontology Development 101: A Guide to Creating Your First Ontology*; Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880; Stanford University: Stanford, CA, USA, 2001.

42. Azucar, D.; Marengo, D.; Settanni, M. Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personal. Individ. Differ.* **2018**, *124*, 150–159. [CrossRef]

43. Kuss, D.J.; Griffiths, M.D. Online social networking and addiction—A review of the psychological literature. *Int. J. Environ. Res. Public Health* **2011**, *8*, 3528–3552. [CrossRef]

44. Seidman, G. Self-presentation and belonging on Facebook: How personality influences social media use and motivations. *Personal. Individ. Differ.* **2013**, *54*, 402–407. [CrossRef]

45. Arusada, M.D.N.; Putri, N.A.S.; Alamsyah, A. Training Data Optimization Strategy for Multiclass Text Classification. In Proceedings of the 5th International Conference on Information and Communication Technology (ICOICT), Melaka, Malaysia, 17–19 May 2017; pp. 1–5. [CrossRef]

46. Zheng, H.; Wu, C. Predicting Personality Using Facebook Status Based on Semi-supervised Learning. In Proceedings of the 11th International Conference on Machine Learning and Computing, Zhuhai, China, 22–24 February 2019; pp. 59–64.

47. Morin, P. Data Structures for Strings. Available online: https://cglab.ca/~morin/teaching/5408/notes/strings.pdf (accessed on 15 April 2012).

48. Cavnar, W.B.; Trenkle, J.M. N-gram-based text categorization. In Proceedings of the SDAIR-94, 3rd Anual Symposium on Document Analysis and Information Retrieval, Las Vegas, NV, USA, 11 April 1994.

49. Pauls, A.; Klein, D. Faster and smaller n-gram language models. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA, 19–24 June 2011; pp. 258–267.

50. Konstabel, K.; Lönnqvist, J.E.; Walkowitz, G.; Konstabel, K.; Verkasalo, M. The 'Short Five'(S5): Measuring personality traits using comprehensive single items. *Eur. J. Personal.* **2012**, *26*, 13–29. [CrossRef]

51. Vazire, S. Informant reports A cheap, fast, and easy method for personality assessment. *J. Res. Personal.* **2006**, *40*, 472–481. [CrossRef]

52. Park, G.; Schwartz, H.A.; Eichstaedt, J.C.; Kern, M.L.; Kosinski, M.; Stillwell, D.J.; Seligman, M.E. Automatic personality assessment through social media language. *J. Personal. Soc. Psychol.* **2015**, *108*, 934. [CrossRef]

53. Tausczik, Y.R.; Pennebaker, J.W. The psychological meaning of words: LIWC and computerized text analysis methods. *J. Lang. Soc. Psychol.* **2010**, *29*, 24–54. [CrossRef]

54. Qiu, L.; Lin, H.; Ramsay, J.; Yang, F. You are what you tweet: Personality expression and perception on Twitter. *J. Res. Personal.* **2012**, *46*, 710–718. [CrossRef]

55. Gao, R.; Hao, B.; Bai, S.; Li, L.; Li, A.; Zhu, T. Improving user profile with personality traits predicted from social media content. In Proceedings of the 7th ACM Conference on Recommender Systems, Hong Kong, China, 12–16 October 2013; pp. 355–358.

56. Li, L.; Li, A.; Hao, B.; Guan, Z.; Zhu, T. Predicting active users' personalities based on micro-blogging behaviors. *PLoS ONE* **2014**, *9*, e84997. [CrossRef]

57. Wei, H.; Zhang, F.; Yuan, N.J.; Cao, C.; Fu, H.; Xie, X.; Ma, W.Y. Beyond the words: Predicting user personality from heterogeneous information. In Proceedings of the of 10th ACM international conference on web search and data mining, Cambridge, UK, 6–10 February 2017; pp. 305–314.