IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 5, 2020, accepted June 15, 2020, date of publication June 22, 2020, date of current version July 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3004002

Predicting Personality Using Answers to Open-Ended Interview Questions

MADHURA JAYARATNE[1,2], (Member, IEEE), AND BUDDHI JAYATILLEKE[1]
[1]PredictiveHire Pty Ltd., Cremorne, VIC 3121, Australia
[2]Centre for Data Analytics and Cognition, La Trobe University, Bundoora, VIC 3083, Australia

Corresponding author: Buddhi Jayatilleke (buddhi@predictivehire.com)

ABSTRACT One's personality is widely accepted as an indicator of job performance, job satisfaction and tenure intention. The ability to measure an applicant's personality in the selection process helps recruiters, hiring managers and the applicant make better hiring decisions. Our work shows that textual content of answers to standard interview questions related to past behaviour and situational judgement can be used to reliably infer personality traits. We used data from over 46,000 job applicants who completed an online chat interview that also included a personality questionnaire based on the six-factor HEXACO personality model to self-rate their personality. Using natural language processing (NLP) and machine learning methods we built a regression model to infer HEXACO trait values from textual content. We compared the performance of five different text representation methods and found that term frequency-inverse document frequency (TF-IDF) with Latent Dirichlet Allocation (LDA) topics performed the best with an average correlation of r = 0.39. As a comparison, a large study of Facebook messages based inference of Big 5 personality found an average correlation of r = 0.35 and IBM's Personality Insights service built using twitter text data reports an average correlation of r = 0.31. We further validated our model with a group of 117 volunteers who used an agreement scale of yes/no/maybe to rate the individual trait descriptors generated based on the model outcomes. On average, 87.83% of the participants agreed with the personality description given for each of the six traits. The ability of algorithms to objectively infer a candidate's personality using only the textual content of interview answers presents significant opportunities to remove the subjective biases involved in human interviewer judgement of candidate personality.

INDEX TERMS HEXACO personality traits, linguistic analysis, personality prediction.

I. INTRODUCTION

Organizational psychologists have long hypothesised that one's personality is closely related to his job performance [1]–[3], job satisfaction [4], [5] and tenure intention [3], [6], [7]. The research outcomes suggest that the work is more enjoyable and thus engaging to the individual and beneficial to the employer and the society at large when there is congruence between one's personality and career. The widely accepted approach to testing for personality in the recruitment process is to use a personality inventory such as NEO-PI-R [8], HEXACO-PI-R [9] or a similar personality questionnaire based on either the five or six factor model of personality. However, conducting a personality test adds an extra cost to the recruitment process and also tend to diminish candidate experience as personality tests are less favoured by candidates compared to other methods such as job interviews [10]. Hence personality tests are not as ubiquitous as the employment interviews, the most widely used selection method in the past 100 years [11]. However, a strong criticism of the job interview is the likelihood of bias introduced by the prejudices of the interviewer. Structured interviews where the same questions are asked from every candidate, in a controlled conversation flow and evaluated using a well-defined rubric have shown to reduce bias [12] and also increase the ability to predict future job performance [13]. The questions asked in a structured interview are derived using a job analysis as opposed to interviewer preference and are typically based on past behaviour and situational judgement.

The ability to infer personality from interview responses could replace lengthy and less favoured personality tests while providing objective outcomes from interviews. Numerous studies have demonstrated that one's informal languages use, such as those used in blogs and social media

The associate editor coordinating the review of this manuscript and approving it for publication was Tallha Akram.

can reveal personality insights [14]–[16] and those insights have been extended to build predictive models to infer personality traits with success [17]–[22]. While encouraging, such correlations and predictability have not been demonstrated in more formal contexts such as job interviews.

We hypothesized that the textual content of answers given by candidates to questions of past behaviour and situational judgement to include patterns that are correlated with their personality (Figure 1). We envisage that such a finding could lead to algorithmic measurement of personality and further reduce human bias and increase the fairness of the structured interview approach. Here we discuss the methodology and the outcomes of the study we conducted to test the aforementioned hypothesis.

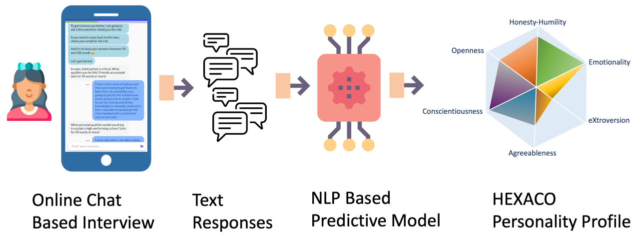

FIGURE 1. Using interview responses to predict candidate personality.

In this work, we make the following contributions to the crossroads of computational linguistics and organizational psychology domains.

1) We demonstrate that responses to typical interview questions related to past behaviour and situational judgement can be used to reliably infer one's personality characteristics.
2) With an average correlation of r = 0.39 between text and personality, the results achieved represent the state-of-the-art for text to personality in any domain, surpassing the previously reported highest correlation of r = 0.35. The best results were achieved by using terms with TF-IDF weighting scheme and topics generated with LDA as the features to the predictive model.
3) We demonstrate that open-vocabulary approach to modelling language use outperforms dictionary-based closed-vocabulary approaches by a great margin. The results further demonstrate that LIWC [23], a popular lexicon used in previous studies in informal contexts (e.g. social media), extract little to no information in formal contexts such as job interview responses.

The rest of the paper is organised as follows. In section II we explore related work in personality in the employment context and different approaches used in language-based personality inference. Section III details the methodology including the data used and the five different text representation methods we evaluated, namely terms (TF-IDF), topics (LDA), Word2Vec, Doc2Vec and LIWC. We also briefly outline the chat-based structured interview tool (called FirstInterview) that was used to conduct the online interviews. A comparison of accuracies of the regression models built using the different text representations and the inter-correlations between the six HEXACO traits are listed in section IV. In sections V and VI we discuss the results comparing them to other related research findings and present a further validation we conducted with a group of 117 volunteer applicants. We conclude and suggest future directions in section VII.

II. BACKGROUND AND RELATED WORK

The congruence of personality and career has been linked to important life outcomes of the employees [24], [25]. Denissen *et al.* [24] examined the effect of congruence between one's personality and the expert-rated demand of the job, on income. The outcomes suggest the fit between the person and job demand is a good indicator of income. Thus, they conclude that "economic success depends not only on having a 'successful personality' but also, in part, on finding the best niche for one's personality". Specifically, in service-oriented roles such as flight attendants, personality has shown to predict training performance, a leading indicator of job performance, supporting the efficacy of using personality tests in candidate selection [26]. The same study found that personality-based measures were able to predict aspects of performance not explained by cognitive measures.

Personality has also been linked to employee well-being [27]. Friedman and Kern [27] note that someone who naturally strives for accomplishments and is dedicated towards their career, employer or society enjoys sizeable health benefits. This is in line with other research where conscientiousness has been shown to play a significant role in health, with implications across the lifespan [28]–[30]. Similarly, Roberts *et al.* [25] have demonstrated specific personality traits predict important life outcomes such as longevity, marital success, and occupational attainment.

Given the link between one's personality and career, employers have utilized personality testing to assess the match between candidates and job/organizational profile. These personality tests are administered using personality inventories whereby candidates rate various statements from the inventory linked to personality traits. A downside of using self-report personality tests as a candidate selection tool is that many are quite long, in some cases taking several hours to complete. This is mainly due to each statement in the test contributing a single data point in measuring a facet of the underlying trait. In order to obtain a reliable measure of a facet, a test taker is required to answer several statements (e.g. 3-5 items) that are similar sounding and often showing no direct relevance to the role a candidate is applying to, making respondents bored and frustrated. Hence applicant reactions to personality tests have shown to be less favourable than interviews [10], [31]. Based on a meta-analysis of multiple studies on application reaction to selection methods, Anderson *et al.* [10] found that compared to job interviews and work sample tests, personality tests fall short of making a positive impression with candidates in areas of face validity, opportunity to perform, interpersonal warmth and respectful of privacy. These indicate candidates' preference to express themselves and not be restricted to self-rating themselves on a pre-defined set of multiple-choice questions (typically over 100 items) as found in standard personality tests.

A. HEXACO PERSONALITY MODEL

In this work, we use the HEXACO model of personality as the underlying personality representation model. HEXACO is a six-factor model of personality introduced by Ashton and Lee [32], [33], following a number of lexical studies. The six factors are Honesty-humility (H), Emotionality (E), eXtraversion (X), Agreeableness (A), Conscientiousness (C) and Openness [to experience] (O). HEXACO personality model is closely related to the Big Five model [34] (also known as Five-factor or OCEAN model) consisting of five factors, Openness [to experience] (O), Conscientiousness (C), Extraversion (E), Agreeableness (A) and Neuroticism (N). Both Big Five and HEXACO models are grounded in the ''Lexical Hypothesis'' that claims personality characteristics are encoded in language and factor analysis of verbal descriptors of human behaviour are used to derive the factors and the underlying facets in each model. While there are subtle differences in the five comparable factors, the unique sixth factor in the HEXACO model is the ''honesty-humility'' factor represented by the facets sincerity, fairness, greed avoidance, and modesty.

HEXACO model of personality has been proposed as a better alternative to the Big Five model as it explains a number of personality phenomena that are not explained by the Big Five model such as patterns of gender differences in personality traits [32]. The H-factor is specifically important in an organizational context given it represents characteristics desired in a workplace environment such as honesty, fairness, integrity and modesty. Various studies have shown how the H-factor can help explain and predict workplace deviance [35], delinquency [36], [37], integrity [38], counterproductive work behaviour and organizational citizenship [39] and job performance [40]. It has also shown to be positively correlated with many desirable traits while negatively correlated with undesirable traits such as the Dark Triad (narcissism, psychopathy, and Machiavellianism) [33], [41].

B. INFERRING PERSONALITY FROM LANGUAGE

Pennebaker and King [14] conducted one of the earliest research in the area analyzing the written content from daily diaries, assignments and journal abstracts and concluded that despite the differences in domains, language is an ''independent and meaningful way of exploring personality''. Most of the early research relied on the closed-vocabulary approach, where counting the number of words over predefined categories were used to identify personality attributes. Linguistic Inquiry and Word Count (LIWC) [23] is one such lexicon and the 2007 version includes 64 categories such as pronouns, present, family, sad, health, etc, while the 2015 version is extended to 76 categories. A given word may belong to one or more categories and word counts under these human-engineered categories are used to classify and analyze the language use. Using the LIWC categories, researchers have analyzed and found interesting correlations among language patterns and personality. For example, compared to introverts, extroverts have been identified as producing less complex writings, and more social and positive emotional words [14]. Amongst other correlations between word use and personality, Fast and Funder [15] found several personality and behavioural correlates of article usage (e.g. words such as 'the', 'an' and 'a'), indicating that those who use more articles are highly intellectual (philosophical, verbally fluent, and sceptical) and open to experience (wide-ranging interests and aesthetic). Similarly, people who are high on conscientiousness have been found to discuss achievements more compared to people who are low on conscientiousness in self-narratives [16].

With advances in machine learning, predictive models have been developed using closed vocabulary approaches to infer personality from one's language use. Presence of certain words and word categories have been highly predictive of personality types and researchers have been able to develop machine learning models using language content from blogs [17], essays [42], microblogs (Twitter, Sina Weibo) [18]–[20], social media posts [21], [43] etc. However, most of these closed-vocabulary attempts are either based on small datasets (<1000 users) or classifies users as high/low on personality traits. Small datasets limit the accuracy and validity of the trained models [21] while personality is a continuum rather than a binary variable.

Facilitated by large volumes, recent attempts on predicting personality traits from textual contents have utilized the open-vocabulary approach. In the open-vocabulary approach, single words and multi-word phrases (also known as n-grams) are used to model language use compared to human-engineered categories in closed-vocabulary approaches. Large dimensionality produced by the open-vocabulary approach requires a large amount of training data to infer the relationship between language representation and personality. Such predictive models have been demonstrated on textual data from social media [22] and blogs [17] with success.

The accuracy of open-vocabulary approach has been further enhanced by the recent successes in Natural Language Processing (NLP) in word/document embedding techniques where words or documents are converted to a vector representation by a language model trained on a large text corpus such as Wikipedia. Doc2Vec (Document to Vector) models [44] convert a document to a vector representation directly while Word2Vec (Word to Vector) [45] based approached use some form of aggregation - usually averaging - over vector representations of words to obtain a document vector. For example, IBM's Personality Insights Service [46] uses the GloVe word embedding method [47] in their text to personality inference model. IBM also claims that ''Earlier versions of the service used the Linguistic Inquiry and Word Count (LIWC) psycholinguistic dictionary with its machine-learning model. However, the open-vocabulary approach outperforms the LIWC-based model.'', which aligns with our own findings (see Results section).

Use of language as input for inferring one's personality has significant merits over using self-rated statements as used in standard personality tests. Firstly, widely accepted personality models such as Big Five and HEXACO are grounded on the "Lexical Hypothesis" that claims personality characteristics are encoded in language, showing the foundational impact of language in defining identifiable personality traits [48]. Moreover, the wealth of research in psycholinguistics demonstrates the relationship between thought, language and human nature [49], [50]. Following quote from a pioneer in the field is worth stating:

> "Language is a mirror of mind in a deep and significant sense. It is a product of human intelligence ... By studying the properties of natural languages, their structure, organization, and use, we may hope to learn something about human nature; something significant, ..."
> - Noam Chomsky [51]

Both lexical hypothesis and psycholinguistics validate why language analysis is a first-principles approach to understanding personality and human nature. For example, the amount of information captured in syntax and semantics of a language-based open interaction such as an interview is far greater than from a test inventory of a limited set of multiple-choice questions found in a standard personality test. The advent of advanced NLP, large public text corpora and machine learning technology has made the analysis and computational understanding of language scalable and far more effective. Moreover, the constant advancements in NLP driven by academia and industry giants such as Google and Facebook, allow language content to be analysed in ever more newer ways, further increasing the insights language data can generate [52]. The data value of language in a personality assessment setting is further enhanced by the engaging nature of the conversational approach compared to a lengthy personality test. For example, asking test-takers to express themselves by answering a few open-ended questions that are relevant and context-sensitive makes the experience enjoyable and empowering.

III. METHODOLOGY

In order to test the correlation between language use and personality, we built a regression model that is able to infer a rating for each of the six personality traits in the HEXACO model using textual answers given to open-ended interview questions. Given the importance of numerical representation of language in building a machine learning model, we compared the performance of five different text representation methods namely, terms (TF-IDF), topics (LDA), Word2Vec, Doc2Vec and LIWC. In this section, we describe the training dataset, the five different text representation methods and the regression model building approach.

A. DATASET

The training dataset for our experiment comes from the PredictiveHire[1] FirstInterview (TM) product, an online chat-based interview tool. Job applicants answer 5-7 open-ended questions and 40 self-rating questions based on a proprietary personality inventory that examines HEXACO personality traits. FirstInterview (TM) is typically the very first engagement the applicant has with the hiring organisation, placed at the top of the recruitment funnel and close to 40% of applicants complete it on a mobile.

The online interview questionnaire includes open-ended free-text questions on past experience, situational judgement and values. The questions are customisable by role family (e.g. sales, retail, call centre etc.) and specific customer value requirements. The questions are rotated regularly to address gaming risk. Following are some example questions.

- *What motivates you? What are you passionate about?*
- *Not everyone agrees all the time. Have you had a peer, teammate or friend disagree with you? What did you do?*
- *Give an example of a time you have gone over and above to achieve something. Why was it important for you to achieve this?*
- *Sometimes things don't always go to plan. Describe a time when you failed to meet a deadline or personal commitment. What did you do? How did that make you feel?*
- *In sales, thinking fast is critical. What qualifies you for this? Provide an example.*

Following are two example answers to two of the free text questions.

- What motivates you? What are you passionate about?
  - *Money and success are my motivators, however I am an extremely competitive person and enjoy simply competing with colleagues or even just aiming to beat whatever good work I have previously done.*
  - *Generally enjoy helping others and helping others makes me feel iv done something good for someone else is very rewarding.*
- Sometimes things don't always go to plan. Describe a time when you failed to meet a deadline or personal commitment. What did you do? How did that make you feel?
  - *I had an uni assignment that I had left to the last minute that was due. I still managed to complete it and submit a few days after it was due with a penalty of 5% deduction for each day taken after the due date. You know the saying, better late than never. It made me feel that I wished I had started working on it a bit earlier and not start on the week it was due. A lesson learnt to not repeat with future assignments.*
  - *I was managing an interstate project where I had to prepare a site for shut down. When I got to the site I very quickly realised that it was going to take a lot longer than I had anticipated and that I was going to be away from home a lot longer than I thought. I was frustrated in myself for not doing the correct preparation and I then requested a shift in*

[1] https://www.predictivehire.com/

> *the deadline and set about engaging others to assist me. I learned to be more careful and not rush my planning.*

Each candidate wrote on average 286 words (a minimum of 150 to a maximum of 6196 words) answering the open-ended questions.

The HEXACO based statements are rated on a 5-point scale from *Strongly agree* to *Strongly disagree*. Each candidate responded to a 40 item personality test with statements similar to below (the measuring trait is listed in parenthesis. This is not shown to the candidate):

- I pretend to be concerned for others to get what I want. (Honesty-Humility)
- I am deeply moved by other people's misfortunes. (Emotionality)
- I am good at making unplanned speeches. (Extraversion)
- I find that it takes a lot to make me angry at someone. (Agreeableness)
- I do things without thinking of the consequences. (Conscientiousness)
- I like to think up lots of ideas. (Openness)

The HEXACO inventory responses were coded 1 to 5 (accounting for any reverse-coding) and each candidate rated 6-8 statements on each trait. The variance in the number of statements per trait is due to each role family assessment carrying the most predictive 40 statements for the job fitness criterion. The final individual trait scores were calculated by averaging over all the responses for each trait the candidate rated.

We analysed free-text responses from 46,888 candidates who used the FirstInterview (^{TM}) platform. The demographics of the candidates in terms of gender and the job family they applied to are shown in Table 1.

TABLE 1. Demographic breakdown of the participants.

Attribute	Group	Count
Gender	Female	11,859
	Male	12,504
	Not specified	22,525
Job family	Airline	6,115
	Call centre	1,762
	Healthcare	18,212
	Retail	2,319
	Sales	17,900
	Other	580

Table 2 shows the number of candidates included in the study under each trait where HEXACO trait data were available. These trait scores formed the ground-truth for building predictive models for each trait. The total number of unique individuals is 46,888. The mean and standard deviation of each trait are also listed in Table 2.

B. LANGUAGE TO TRAIT MODEL BUILDING

We experimented with four open-vocabulary representations of textual information and one closed-vocabulary approach for comparison. The open-vocabulary approaches do not rely on a priori word or category judgment compared to closed-vocabulary approaches. See below for the different representation approaches we compared:

TABLE 2. Total number of candidates for each traits who rated at least 6 statements and wrote at least 150 words in their free-text responses and the mean and standard deviation of the trait scores.

Trait	Number of records	Mean	SD
Honesty-Humility	6,195	4.10	0.62
Emotionality	11,659	2.89	0.48
Extraversion	7,865	3.98	0.52
Agreeableness	14,151	3.87	0.49
Conscientiousness	16,165	4.24	0.44
Openness	14,868	3.40	0.49

1) TERMS

In this approach, we first tokenized the text response from interview questions and developed a vectorized representation in n-dimensional space using unigrams, bigrams and trigrams of tokens. The vectorized representation uses the term frequency-inverse document frequency (TF-IDF) [53] scheme where the value for a response-term combination increases with the number of times the word is used in the response while offsetting for the overall usage of the term in the corpus. We experimented three dimensionality values of 500, 1000 and 2000 using the most frequent n-grams (n = 1,2,3) used in the representation. We found 2000 most frequent n-grams to yield the best results. We denote this approach as TF-IDF.

With t, r, and R denoting term, response and the set of all responses respectively, $n_{t,r}$ is the number of times term t occurring in response r, $N = |D|$ and $n_t = |r \in R : t \in r|$

$$tfidf(t, r, R) = tf(t, r) \times idf(t, R); \quad t \in r, r \in R \tag{1}$$

where

$$tf(t, r) = \frac{n_{t,r}}{\sum_{\acute{t} \in d} n_{\acute{t},r}} \tag{2}$$

$$idf(t, R) = \log(\frac{N}{n_t} + 1) + 1 \tag{3}$$

2) TOPICS

We used the Latent Dirichlet Allocation (LDA) [54] topic modelling algorithm to derive 100 topics from the text data. LDA assumes the presence of latent topics in a given set of text documents and attempts to probabilistically uncover these topics. Topics are generated over the vocabulary with each term having a particular affinity to each topic. These affinities can be used to comprehend the machine-generated topic by using terms with high affinities to describe the topic. Once uncovered, each document (in our case each candidate response) can be represented with a one-dimensional vector of 100 elements. We used the Gensim software package[2] for topic modelling. We denote this approach as LDA.

[2]https://radimrehurek.com/gensim/

Using the notation defined in (1) and θ denoting a LDA topic, the probability of a topic given a candidate's response,

$$P(\theta|r) = \sum_{t \in \theta} P(\theta|t) \times P(t|r) \tag{4}$$

3) Word2Vec

The Word2Vec [45] approach assigns n-dimensional vectors, called word embeddings, for terms in the candidate responses. A word embedding for a term represents the linguistic context of the term and due to this, similar words are placed closer to each other in the vector space. The Skip-gram models of Word2Vec are trained with the objective of finding word representations that can predict the surrounding words in a sentence. Formally, given a sequence of training words $w_1, w_2, w_3, \ldots, w_n$, and k, the size of the training context, the objective is to maximize the average log-probability

$$\frac{1}{n}\sum_{i=1}^{n} \sum_{-k \leq j \leq k, j \neq 0} \log P(w_{i+j}|w_i) \tag{5}$$

Word2Vec models are usually trained on large corpora such as Wikipedia or web pages gathered by a web crawler. Word2Vec based textual representations have been used in solutions that have achieved state-of-the-art results in many NLP tasks in recent times. To achieve a vector representation for a given response, we averaged across word embeddings of terms in that response.

4) Doc2Vec

Similarly, Doc2Vec is an embedding technique but unlike Word2Vec, it operates directly on variable-length textual content, such as sentences, paragraphs, and documents rather than on terms. Doc2Vec approach has shown to perform better on capturing context and semantics compared to averaged Word2Vec representations [44]. Le and Mikolov [44] proposed two Doc2Vec models, a distributed memory (Doc2Vec-DM) model and a distributed bag of words (Doc2Vec-DBOW) model. We trained both Doc2Vec-DM and Doc2Vec-DBOW models on content from Wikipedia.

5) LIWC

For the comparison purpose, we used word categories from LIWC [23], a dictionary-based closed-vocabulary approach. We used the LIWC 2015 version which consists of 76 categories such as pronoun, present, family, sad, health, etc. In LIWC, a given word may belong to one or more categories and word counts under these human-engineered categories are used to model textual content.

Using the notation defined in (1) and c denoting a category in LIWC lexicon, category frequency,

$$cf(c, r) = \frac{\sum_{t \in c} n_{t,r}}{\sum_{\acute{t} \in d} n_{\acute{t},r}} \tag{6}$$

C. PREPROCESSING AND MODEL TRAINING

Candidate responses were preprocessed by removing all non-informative content such as special characters, numbers and stop words. Moreover, the responses were converted to lower case and lemmatized to remove multiple forms of the same terms.

We used the Random Forest algorithm [55] to train regression models for each trait using the above text representations as input features and the corresponding trait scores as the target. We selected Random Forest for its known superior performance compared to many other regression algorithms. We see as future work to compare the outcomes using different algorithms. We find it as sufficient to show the outcomes on a single algorithm in order to establish the correlation between language use and personality as any improvement made over our findings using a different algorithm would only make the case for language-based inference of personality stronger.

We used 80% of the data to train the models while the rest of the data was used to validate the accuracy of the trained models.

D. TOPIC ANALYSIS

To better understand the content of the candidate responses and their relationship to HEXACO personality traits, we employed LDA-based topic analysis of the textual data. We used the 100 LDA-based topics derived from the entire text corpus (see Section III-B2 for details) and calculated the correlations between individual topics and personality traits. Noting that not all 100 topics are informative in terms of correlations with personality traits, we selected the top 2 positively correlated topics for each trait, resulting in 9 unique topics, for further analysis. Two topics were selected for each trait as we noticed that the same topic was calculated as the highest positively correlated topic for more than one trait in some instances. Figure 2 presents the correlation between these 9 topics and the HEXACO traits in a heat map. The topic correlations reveal inherent patterns in word usage and personality. For example, the topic with the highest correlation to conscientiousness and openness (topic 23) shows the highest negative correlation to emotionality. This aligns well with the inter-correlation between conscientiousness, openness and emotionality as found in the norm group shown in Table 4. Moreover, the word cloud for topic 23 (see Figure 3-(e)) shows words related to passion, results and sales focus, reference to life as indicative of someone who is conscientiousness, open to experience and less anxious (low in emotionality).

Topic #	H	E	X	A	C	O
6	**-0.047**	**-0.054**	**0.048**	-0.019	**0.037**	**0.051**
20	0.022	**0.064**	-0.017	0.004	-0.001	**-0.151**
23	0.010	**-0.124**	**0.054**	0.020	**0.234**	**0.235**
29	**-0.065**	0.004	-0.025	**-0.059**	0.004	**0.134**
35	0.015	**0.065**	-0.001	**0.032**	**0.056**	**0.091**
49	0.029	0.024	-0.019	**0.046**	**0.036**	**-0.112**
52	**0.154**	**0.054**	**-0.044**	**0.063**	**0.060**	**0.116**
60	**-0.135**	**-0.111**	0.020	**-0.076**	**0.160**	**0.131**
62	**0.050**	**0.053**	**-0.046**	**0.030**	**0.048**	**0.108**

FIGURE 2. **Correlations between the selected LDA topics and HEXACO traits. Bold indicates correlations significant at p < 0.001 level.**

FIGURE 3. Word clouds of significant LDA topics. (a) Topic 52: The highest positively correlated topic with honesty-humility. (b) Topic 35: The highest positively correlated topic with emotionality. (c) Topic 6: The second highest positively correlated topic with extraversion. See (e), topic 23, for the highest correlated. (d) Topic 49: The second highest positively correlated topic with agreeableness. See (a), topic 52, for the highest correlated. (e) Topic 23: The highest positively correlated topic with conscientiousness. (f) Topic 29: The second highest positively correlated topic with openness. See (e), topic 23, for the highest correlated. All the above correlations are significant at $p < 0.001$ level.

As LDA topics are distributed over the entire vocabulary with term-topic affinities representing the composition of a given topic, word clouds can be used to visually represent the topics by using affinities as the weights determining the font size of each term. Figure 3 presents the word cloud-based representation of the most positively correlated topics (the second most positively correlated in case the most positively correlated topic is already included) of each trait. Positively correlated topics were chosen for each trait as the terms in word clouds indicate the high usage of those terms by candidates who were high on the corresponding trait.

IV. RESULTS

We evaluated the trained 'language use to personality trait' regression models on the left out 20% of the data. The models are evaluated using the correlation coefficient between the actual trait scores (ground truth) and the model predicted scores. The correlation coefficient (r) is a statistical measure of the strength of the relationship between two interdependent variables. In this case, it is the strength of the relationship between actual scores and the inferred scores.

Table 3 presents the accuracy of the models trained on LIWC and other selected open-vocabulary features.

TABLE 3. Comparison of accuracies in terms of correlation across different language modelling approaches. All correlations, but LIWC, are significant at the $p < 0.001$ level. Bold indicates the best correlation for each trait.

Features	H	E	X	A	C	O	Average
LIWC	0.02	0.01	0.08	-0.03	0.05	0.06	0.032
TF-IDF	**0.44**	0.32	**0.34**	**0.28**	**0.44**	0.48	0.383
LDA	0.40	0.28	0.29	0.22	0.42	0.47	0.347
TF-IDF + LDA	0.43	**0.33**	**0.34**	**0.28**	**0.44**	**0.50**	**0.387**
Doc2Vec-DM	0.31	0.21	0.27	0.16	0.27	0.21	0.238
Doc2Vec-DBOW	0.32	0.20	0.23	0.17	0.27	0.23	0.237
Averaged Word2Vec	0.39	**0.33**	0.33	**0.28**	0.43	0.49	0.375

In research literature related to personality research, especially with linguistics based ones, correlations greater than 0.2 are considered acceptable and it is rare to see correlations that exceed 0.4. This result demonstrates that the language used in responding to typical interview questions are predictive of one's personality.

Table 4 presents the intercorrelations between the personality traits inferred with the TF-IDF + LDA model (overall best performing model) on an independent group of over 12,000 candidates who only answered the free-text questions and hence were not included in the model training. Composition of this group is shown in Table 5.

TABLE 4. Intercorrelations between personality traits inferred with the TF-IDF + LDA model on the norm group ($n = 12{,}183$). * $p < 0.01$.

	H	E	X	A	C	O
Honesty-Humility (H)	1.00					
Emotionality (E)	0.01	1.00				
Extraversion (X)	-0.05*	-0.07*	1.00			
Agreeableness (A)	0.27*	0.01	0.03*	1.00		
Conscientiousness (C)	0.15*	-0.01	0.24*	0.13*	1.00	
Openness (O)	0.04*	-0.09*	-0.01	-0.08*	0.12*	1.00

TABLE 5. Demographic breakdown of the norm group.

Attribute	Group	Count
Gender	Female	4,117
	Male	5,214
	Not specified	2,852
Job family	Airline	5,223
	Customer service	1,590
	Graduate	567
	Retail	3,741
	Other	762

V. DISCUSSION

Textual content representation using terms and topics (TF-IDF + LDA) achieved the best accuracy in terms of average correlation ($r = 0.387$) over other representation methods. It also achieved the best accuracies across 5/6 traits (emotionality, extraversion, agreeableness, conscientiousness, openness) and 2nd best accuracy for honesty-humility. Correlations for honesty-humility, conscientiousness and openness at 0.43, 0.44, 0.50 respectively, fell above the expected standard "correlation upper-limit", a correlation coefficient of 0.4, for predicting personality with behaviour [25], [56]. These results exceed the best average correlation of 0.354 achieved by Schwartz *et al.* [22] on

Facebook messages and 0.31 achieved by IBM Watson Personality Insights [46] on Tweets while using the Big-5 personality model. It is important to note that the textual content used in our experiment comes from a set of targeted questions where the user is presented with a scope compared to open content from a source like social media. This high vocabulary variation can be one reason for the lower correlations in the studies using social media data. It is encouraging to see the stronger correlations in TF-IDF + LDA, which indicates the presence of word combinations related to personality, given both TF-IDF and LDA are vectoring the responses only using the total vocabulary found within the responses as a whole (as opposed to word embedding methods based on external corpora).

As shown in Table 4, the HEXACO dimensions inferred via the TF-IDF + LDA regression model were weakly related overall, with nearly all intercorrelations falling below 0.2. Weak inter-correlations are desired as it indicates the fix personality factors to be independent. The exceptions were correlations between honesty-humility (H) and agreeableness (A) of 0.27 and extraversion (X) and conscientiousness (C) of 0.24. The high correlation between H-A has been reported elsewhere [9], [57]. In [9] the authors, who are also the creators of the HEXACO model, found the H-A correlation to range between 0.28 - 0.42 across three groups (N = 2,868; 100,318; 8,233) which was also the highest inter-correlation in each group. In the meta-analysis of HEXACO studies listed in [57], authors also found H-A to be the highest intercorrelation at 0.35 (attenuated) and the correlation between X-C to be the second highest at 0.19. These findings are in-line with the correlations we found.

Averaging over Word2Vec word embeddings achieved a very encouraging accuracy of 0.375. The Word2Vec model we used was trained on very generic content from web pages crawled from the Internet. Moreover, simply averaging over terms can lead to information loss from the representation. Hence, the accuracy achieved here is significant and demonstrates the power of language use in predicting one's personality. It is also worth noting that given the generic nature of the content used in training Word2Vec and the ability of embedding models to assess word similarities, this model is more generalizable to words unseen by the predictive model. Doc2Vec-DM and Doc2Vec-DBOW approaches achieved decent results but fell short of the results achieved by the Word2Vec model. While further analysis is required to understand the variations, one reason can be the content and the styles of text used to train the embedding models. Doc2Vec models were trained on Wikipedia content which is more formal and small compared to Word2Vec model, which uses content from common web crawl.

The closed-vocabulary approach, LIWC, failed to achieve significant results for any of the personality traits which is in contrast to results reported by Schwartz *et al.* [22]. We believe this to be due to the domain differences; while the social media content is generic, the responses from candidates contain more work-related terms. It also highlights the limitations of closed-vocabulary approaches where a tediously developed lexicon for a certain domain cannot be applied to a different domain or generic content and vice versa. As stated earlier in section II, similar findings were reported by IBM in [46].

VI. FURTHER VALIDATION

To further validate the accuracy of the 'text to personality' models developed, we carried out a survey involving 117 volunteers (male: 50, female: 67). These volunteers answered a similar set of open-ended questions on past experience and situational judgement on FirstInterview (^{TM}) platform. Based on their text responses, we derived the scores for each of the six personality traits using the trained models. We then placed the derived trait values relative to the representative population to calculate the percentile values for each trait which were then used to provide volunteers with a trait level description of their personality. We used a quintile scale to provide a description of the personality under each trait (e.g. 0%-20%, 80%-100%).

Below are two examples of insights provided to the volunteers.

- Volunteers with their 'agreeableness' percentile between 0 and 20: "You are likely to defend and champion an opinion. You do not back down without evidence, and you are keen to hold others firmly to account. You are happy to solve problems independently and usually don't require much hand-holding. This may come across as you being less of a team player."
- Volunteers with their 'agreeableness' percentile between 80 and 100: "You are likable and people feel comfortable coming to you for help. You are naturally compassionate, and put the interests of others before your own. You value social harmony and so you may defer to others to gain consensus. Some may even call you a 'pleaser'."

Volunteers then provided us with trait level feedback on the personality insights they received. They graded the six descriptors they received (for the six HEXACO traits) on a 'Yes'/'Maybe'/'No' scale indicating whether they agreed on each descriptor. Table 6 presents the outcome of this validation.

TABLE 6. The number of volunteers rating 'Yes'/'Maybe'/'No' on whether they agree with trait level personality insights. Aggregated accuracy is derived by assigning 1, 0.5 and 0 weights to 'Yes', 'Maybe' and 'No' ratings respectively.

Trait	'Yes'	'Maybe'	'No'	Aggregated accuracy
Honesty-Humility	86	14	9	85.32%
Emotionality	87	16	7	86.36%
Extraversion	83	14	14	81.82%
Agreeableness	93	14	3	90.91%
Conscientiousness	95	7	8	89.55%
Openness	84	5	3	94.02%

By assigning 1, 0.5 and 0 weights to 'Yes', 'Maybe' and 'No' ratings respectively, we derived trait level aggregated accuracies which varied from 81.82% - 94.02%. With the same weights, the overall accuracy was 87.83%. It is also worth noting that the aggregated accuracy for "Openness" was the highest, which aligns well with the highest test accuracy for any trait from the model training outcomes (see Table 3).

VII. CONCLUSION AND FUTURE WORK

In this paper, we presented how open-vocabulary approaches in natural language processing alongside machine learning can be utilized to infer one's personality from their language use in a recruitment interview setting. Using data from over 46,000 individuals who answered open-ended interview questions and a HEXACO based personality assessment, we built regression models for each trait using four open-vocabulary text representation methods (namely terms, topics, word embeddings and document embeddings) and one closed-vocabulary method (LIWC). Terms and topics based text representation achieved the best accuracy, an average correlation of 0.387 over other representation methods. As a comparison, the average correlation reported by IBM Watson Personality Insights service is 0.31. Openness, honesty-humility and conscientiousness saw correlations over 0.4. Moreover, intercorrelation among the inferred trait scores remained weak, except for two known intercorrelations (between agreeableness and honesty/humility), strengthening the quality claims of the trained models. In a further study involving 117 volunteers, the participants agreed with the trait-level personality descriptions based on the inferred trait scores at an accuracy of 87.83%.

Given the well-established relationship between one's personality and job satisfaction, performance and tenure intention, we find the above outcome to be significant in at least two ways. Firstly, the ability of algorithms to objectively infer a candidate's personality using only the textual content of interview answers could remove the subjective biases involved in human interviewer judgement of personality. Secondly, algorithmic inference of personality can enable every job applicant to respond to open-ended interview questions, regardless of the applicant volume, and their personality profiles generated at scale. This is near impossible if a human is to interview each applicant in roles where large volumes apply. Giving every candidate an opportunity to express themselves and being assessed as equals by the same algorithm significantly elevate the individual fairness of the selection process. Numerous studies have shown that widely used personality tests are less favoured by candidates compared to interviews. The ability to infer personality from interview answers displaces the need for a separate personality test, increasing candidate satisfaction and engagement.

Further work is required in increasing the clarity of the outcomes. One area of further research is exploring the term and topic patterns discovered by the algorithm as correlated with each trait to help answer questions such as "what specific terms and topics are used by highly Extroverted applicants?". In our current study, we used only the semantic level features (terms, topics etc). Exploring whether other types of features, such as the use of parts of speech (POS), readability, formality, use of emojis etc. can further increase the accuracy, is another useful future extension of this work. Similarly testing the performance of other available regression algorithms, including neural network approaches, may help increase the accuracy of the regression models. Moreover, multi-modal information such as the audio and video signals captured while candidates answer the questions can also be explored as signals to enhance the text-based inference of personality.

REFERENCES

[1] M. R. Barrick and M. K. Mount, "The big five personality dimensions and job performance: A meta-analysis," *Personnel Psychol.*, vol. 44, no. 1, pp. 1–26, Mar. 1991, doi: 10.1111/j.1744-6570.1991.tb00688.x.

[2] S. Rothmann and E. P. Coetzer, "The big five personality dimensions and job performance," *SA J. Ind. Psychol.*, vol. 29, no. 1, pp. 68–74, Oct. 2003. [Online]. Available: https://journals.co.za/content/psyc/29/1/EJC88938

[3] J. F. Salgado, "The big five personality dimensions and counterproductive behaviors," *Int. J. Selection Assessment*, vol. 10, nos. 1–2, pp. 117–125, Mar. 2002, doi: 10.1111/1468-2389.00198.

[4] J. W. Lounsbury, R. P. Steel, L. W. Gibson, and A. W. Drost, "Personality traits and career satisfaction of human resource professionals," *Hum. Resource Develop. Int.*, vol. 11, no. 4, pp. 351–366, Sep. 2008, doi: 10.1080/13678860802261215.

[5] J. W. Lounsbury, N. Foster, H. Patel, P. Carmody, L. W. Gibson, and D. R. Stairs, "An investigation of the personality traits of scientists versus nonscientists and their relationship with career satisfaction," *R&D Manage.*, vol. 42, no. 1, pp. 47–59, Jan. 2012, doi: 10.1111/j.1467-9310.2011.00665.x.

[6] V. Ariyabuddhiphongs and S. Marican, "Big five personality traits and turnover intention among thai hotel employees," *Int. J. Hospitality Tourism Admin.*, vol. 16, no. 4, pp. 355–374, Oct. 2015, doi: 10.1080/15256480.2015.1090257.

[7] T. A. Timmerman, "Predicting turnover with broad and narrow personality traits," *Int. J. Selection Assessment*, vol. 14, no. 4, pp. 392–399, Nov. 2006, doi: 10.1111/j.1468-2389.2006.00361.x.

[8] P. T. Costa, Jr., and R. R. McCrae, "The Revised NEO Personality Inventory (NEO-PI-R)," in *The SAGE Handbook of Personality Theory and Assessment* (Personality Measurement and Testing), Vol. 2. Thousand Oaks, CA, US: Sage Publications, 2008, pp. 179–198.

[9] K. Lee and M. C. Ashton, "Psychometric properties of the HEXACO-100," *Assessment*, vol. 25, no. 5, pp. 543–556, Jul. 2018, doi: 10.1177/1073191116659134.

[10] N. Anderson, J. F. Salgado, and U. R. Hülsheger, "Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity," *Int. J. Selection Assessment*, vol. 18, no. 3, pp. 291–304, Aug. 2010.

[11] T. Macan, "The employment interview: A review of current studies and directions for future research," *Hum. Resource Manage. Rev.*, vol. 19, no. 3, pp. 203–218, Sep. 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053482209000382

[12] J. Levashina, C. J. Hartwell, F. P. Morgeson, and M. A. Campion, "The structured employment interview: Narrative and quantitative review of the research literature," *Personnel Psychol.*, vol. 67, no. 1, pp. 241–293, Mar. 2014. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/peps.12052

[13] M. Mcdaniel, D. Whetzel, F. Schmidt, and S. Maurer, "The validity of employment interviews: A comprehensive review and meta-analysis," *J. Appl. Psychol.*, vol. 79, pp. 599–616, Aug. 1994.

[14] J. W. Pennebaker and L. A. King, "Linguistic styles: Language use as an individual difference," *J. Personality Social Psychol.*, vol. 77, no. 6, p. 1296, 1999.

[15] L. A. Fast and D. C. Funder, "Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior," *J. Personality Social Psychol.*, vol. 94, no. 2, pp. 46–334, 2008.

[16] J. B. Hirsh and J. B. Peterson, "Personality and language use in self-narratives," *J. Res. Personality*, vol. 43, no. 3, pp. 524–527, Jun. 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0092656609000439

[17] F. Iacobelli, A. J. Gill, S. Nowson, and J. Oberlander, "Large scale personality classification of bloggers," in *Affective Computing and Intelligent Interaction* (Lecture Notes in Computer Science), S. D'Mello, A. Graesser, B. Schuller, and J.-C. Martin, Eds. Memphis, TN, USA: Springer, 2011, pp. 568–577.

[18] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from Twitter," in *Proc. IEEE 3rd Int. Conf. Privacy, Secur., Risk Trust IEEE 3rd Int. Conf. Social Comput.*, Oct. 2011, pp. 149–156.

[19] C. Sumner, A. Byers, R. Boochever, and G. J. Park, "Predicting dark triad personality traits from Twitter usage and a linguistic analysis of tweets," in *Proc. 11th Int. Conf. Mach. Learn. Appl.*, Dec. 2012, pp. 386–393.

[20] D. Xue, Z. Hong, S. Guo, L. Gao, L. Wu, J. Zheng, and N. Zhao, "Personality recognition on social media with label distribution learning," *IEEE Access*, vol. 5, pp. 13478–13488, 2017.

[21] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Personality predictions based on user behavior on the facebook social media platform," *IEEE Access*, vol. 6, pp. 61959–61969, 2018.

[22] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. P. Seligman, and L. H. Ungar, "Personality, gender, and age in the language of social media: The open-vocabulary approach," *PLoS ONE*, vol. 8, no. 9, Sep. 2013, Art. no. e73791. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0073791

[23] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn. (2005). *The Development and Psychometric Properties of LIWC2015*. Accepted: Sep. 2015. [Online]. Available: https://repositories.lib.utexas.edu/handle/2152/31333

[24] J. J. A. Denissen, W. Bleidorn, M. Hennecke, M. Luhmann, U. Orth, J. Specht, and J. Zimmermann, "Uncovering the power of personality to shape income," *Psychol. Sci.*, vol. 29, no. 1, pp. 3–13, Jan. 2018, doi: 10.1177/0956797617724435.

[25] B. W. Roberts, N. R. Kuncel, R. Shiner, A. Caspi, and L. R. Goldberg, "The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes," *Perspect. Psychol. Sci.*, vol. 2, no. 4, pp. 313–345, Dec. 2007, doi: 10.1111/j.1745-6916.2007.00047.x.

[26] D. F. Cellar, D. J. G. De Grendel, J. D. Klawsky, and M. L. Miller, "The validity of personality, service orientation, and reading comprehension measures as predictors of flight attendant training performance," *J. Bus. Psychol.*, vol. 11, no. 1, pp. 43–54, Sep. 1996. [Online]. Available: www.jstor.org/stable/25092532

[27] H. S. Friedman and M. L. Kern, "Personality, well-being, and health," *Annu. Rev. Psychol.*, vol. 65, no. 1, pp. 719–742, Jan. 2014, doi: 10.1146/annurev-psych-010213-115123.

[28] B. P. Chapman, K. Fiscella, I. Kawachi, and P. R. Duberstein, "Personality, socioeconomic status, and all-cause mortality in the united states," *Amer. J. Epidemiology*, vol. 171, no. 1, pp. 83–92, Jan. 2010. [Online]. Available: https://academic.oup.com/aje/article/171/1/83/85373

[29] H. S. Friedman, M. L. Kern, S. E. Hampson, and A. L. Duckworth, "A new life-span approach to conscientiousness and health: Combining the pieces of the causal puzzle.," *Develop. Psychol.*, vol. 50, no. 5, pp. 1377–1389, 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3651756/

[30] G. Hagger-Johnson, S. Sabia, H. Nabi, E. Brunner, M. Kivimaki, M. Shipley, and A. Singh-Manoux, "Low conscientiousness and risk of all-cause, cardiovascular and cancer mortality over 17years: Whitehall II cohort study," *J. Psychosomatic Res.*, vol. 73, no. 2, pp. 98–103, Aug. 2012.

[31] J. P. Hausknecht, D. V. Day, and S. C. Thomas, "Applicant reactions to selection procedures: An updated model and meta-analysis," *Personnel Psychol.*, vol. 57, no. 3, pp. 639–683, 2004, doi: 10.1111/j.1744-6570.2004.00003.x.

[32] M. C. Ashton and K. Lee, "Empirical, theoretical, and practical advantages of the HEXACO model of personality structure," *Personality Social Psychol. Rev.*, vol. 11, no. 2, pp. 150–166, May 2007, doi: 10.1177/1088868306294907.

[33] K. Lee and M. Ashton, *The H Factor of Personality: Why Some People are Manipulative, Self-Entitled, Materialistic, and Exploitive—And Why It Matters for Everyone*. Waterloo, ON, Canada: Wilfrid Laurier Univ. Press, 2013.

[34] L. R. Goldberg, "The structure of phenotypic personality traits," *Amer. Psycholog.*, vol. 48, no. 1, pp. 26–34, 1993.

[35] J. L. Pletzer, M. Bentvelzen, J. K. Oostrom, and R. E. de Vries, "A meta-analysis of the relations between personality and workplace deviance: Big five versus HEXACO," *J. Vocational Behav.*, vol. 112, pp. 369–383, Jun. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0001879119300582

[36] K. Lee, M. C. Ashton, and R. E. de Vries, "Predicting workplace delinquency and integrity with the HEXACO and five-factor models of personality structure," *Human Perform.*, vol. 18, no. 2, pp. 179–197, Apr. 2005.

[37] R. E. de Vries and J.-L. van Gelder, "Explaining workplace delinquency: The role of Honesty-Humility, ethical culture, and employee surveillance," *Personality Individual Differences*, vol. 86, pp. 112–116, Nov. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S019188691500389X

[38] K. Lee, M. C. Ashton, D. L. Morrison, J. Cordery, and P. D. Dunlop, "Predicting integrity with the HEXACO personality model: Use of self- and observer reports," *J. Occupational Organizational Psychol.*, vol. 81, no. 1, pp. 147–167, Mar. 2008. [Online]. Available: https://onlinelibrary.wiley.com/doi/pdf/10.1348/096317907X195175, doi: 10.1348/096317907X195175.

[39] J. Anglim, F. Lievens, L. Everton, S. L. Grant, and A. Marty, "HEXACO personality predicts counterproductive work behavior and organizational citizenship behavior in low-stakes and job applicant contexts," *J. Res. Personality*, vol. 77, pp. 11–20, Dec. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0092656618302769

[40] M. K. Johnson, W. C. Rowatt, and L. Petrini, "A new trait on the market: Honesty-Humility as a unique predictor of job performance ratings," *Personality Individual Differences*, vol. 50, no. 6, pp. 857–862, Apr. 2011.

[41] K. Lee and M. C. Ashton, "Psychopathy, machiavellianism, and narcissism in the five-factor model and the HEXACO model of personality structure," *Personality Individual Differences*, vol. 38, no. 7, pp. 1571–1582, May 2005.

[42] Y. Neuman and Y. Cohen, "A vectorial semantics approach to personality assessment," *Sci. Rep.*, vol. 4, no. 1, pp. 1–6, May 2015. [Online]. Available: https://www.nature.com/articles/srep04761

[43] Z. Wang, C. Wu, K. Zheng, X. Niu, and X. Wang, "SMOTETomek-based resampling for personality recognition," *IEEE Access*, vol. 7, pp. 129678–129689, 2019.

[44] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proc. 31st Int. Conf. Int. Conf. Mach. Learn.*, vol. 32, 2014, pp. II-1188–II-1196.

[45] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. 26th Int. Conf. Neural Inf. Process. Syst.*, Lake Tahoe, NV, USA, vol. 2, 2013, pp. 3111–3119.

[46] Watson Personality Insights. (2020). *The Science Behind the Service*. [Online]. Available: https://www.ibm.com/watson/services/personality-insights/

[47] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543. [Online]. Available: https://www.aclweb.org/anthology/D14-1162

[48] G. Saucier and R. L. Goldberg, "The language of personality: Lexical perspectives on the five-factor model," in *The Five-Factor Model of Personality: Theoretical Perspectives*. New York, NY, USA: Guilford Press, 1996, pp. 21–50.

[49] L. Gleitman and P. Papafragou, "New perspectives on language and thought," in *The Oxford Handbook of Thinking and Reasoning.*, 2nd ed. London, U.K.: Oxford Univ. Press, 2012.

[50] S. Pinker, *The Stuff of Thought: Language As a Window Into Human Nature*. London, U.K.: Penguin Group, 2007.

[51] N. Chomsky, *Reflections on Language*. San Francisco, CA, USA: Pantheon, 1975.

[52] A. Torfi, R. A. Shirvani, Y. Keneshloo, N. Tavvaf, and E. A. Fox, "Natural language processing advancements by deep learning: A survey," 2020, *arXiv:2003.01200*. [Online]. Available: http://arxiv.org/abs/2003.01200

[53] M. Christopher, R. Prabhakar, and S. Hinrich, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[54] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003. [Online]. Available: http://www.jmlr.org/papers/v3/blei03a

[55] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[56] G. J. Meyer, S. E. Finn, L. D. Eyde, G. G. Kay, K. L. Moreland, R. R. Dies, E. J. Eisman, T. W. Kubiszyn, and G. M. Reed, ''Psychological testing and psychological assessment. a review of evidence and issues,'' *Amer. Psychologist*, vol. 56, no. 2, pp. 128–165, 2001.

[57] M. Moshagen, I. Thielmann, B. E. Hilbig, and I. Zettler, ''Meta-analytic investigations of the HEXACO personality Inventory(-Revised): Reliability generalization, self–observer agreement, intercorrelations, and relations to demographic variables,'' *Zeitschrift Für Psychologie*, vol. 227, no. 3, pp. 186–194, Jul. 2019.

MADHURA JAYARATNE (Member, IEEE) received the B.Sc. degree in engineering and the M.Sc. degree in computer science from the University of Moratuwa, Sri Lanka, in 2011 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Research Center for Data Analytics and Cognition, La Trobe University, Melbourne, VIC, Australia. From 2011 to 2013, he was a Senior Engineer with the Research and Development Division, Codegen Ltd., and he has been working as a Data Scientist at PredictiveHire, Melbourne, VIC, Australia, since 2019. He is a Continuous Contributor to open-source software. His research interests include natural language processing, multimodal data fusion, and cognitive computing.

BUDDHI JAYATILLEKE received the B.Sc. degree in computer science and engineering from the University of Moratuwa, Sri Lanka, the master's degree in software systems engineering from The University of Melbourne, Australia, and the Ph.D. degree in computer science from RMIT University, Australia. Since 2018, he has been the Principal Data Scientist at PredictiveHire, Melbourne, Australia, where he leads the Data Science Team developing machine learning products to increase fairness and efficiency of the recruitment process. Prior to joining PredictiveHire, he was the Lead Data Scientist at Culture Amp, where he led various data science projects focused on people analytics. His academic career includes Senior Research Fellow positions at La Trobe University and Deakin University, Australia, and a Research Fellow with RMIT University. His research interests include people analytics, natural language processing, and intelligent software agents.