Exploring the Possibility of Predicting Intelligence and Personality Traits of an Individual Using Transcript of Their Speech

Aryan Ahadinia, Saba Nasiri, Shahrzad Javidi; *supervised by* Prof. Vita Akstinaitė and Prof. Phil Garner
{aryan.ahadinia, saba.nasiri, shahrzad.javidi}@epfl.ch
{vita.akstinaite, pgarner}@idiap.ch
EPFL, IDIAP

***Abstract*—Intelligence and personality have always been considered two important dimensions for characterizing an individual. Recent advancements in Natural Language Processing methods have initiated an interdisciplinary field of research aimed at inferring these variables. This report explores the possibility of inferring such variables. In this project, the goal is to investigate this potential using a new dataset (IDIAP). Our studies shed light on the possibility of accurately inferring intelligence and, to some extent, personality.**

I. Introduction

Inferring the intelligence and personality traits of an individual has attracted the attention of the psychology and psychoanalysis community for a long time [1], as they can explain and predict the behavior of a person in different circumstances [2]. This may have potential application in domains such as social network analysis [3], recommendation and personalization [4], mental diagnosis [5], and application screening [6]. The objective of this project is to explore the possibility of predicting aspects of intelligence, ICAR, and personality, Big $5+1$, using a reflection of their thoughts in their writings or speeches.

Although there is no uniform agreement on measurement for intelligence or personality, *International Cognitive Ability Resource (ICAR)* score [7] is widely considered as a good indication of individual intelligence, which is measuring a person's ability in abstract, spatial, verbal and logical reasoning. On the other hand, *Big 5* [8] and MBTI [9] are measurements for aspects of an individual personality. While MBTI classifies people into 16 classes, the Big 5 is made up of 5 pseudo-orthogonal values for characterizing an individual personality. These measurement are (1) *openness*, (2) *conscientiousness*, (3) *extroversion*, (4) *agreeableness* and (5) *neuroticism*. In addition, the Big Five can be enriched as the Big $5+1$ with (6) *honesty* as the sixth dimensionality.

The prior literature in psychology suggests that the *semantic load of words* can be a reflection of a person's way of thinking, and thus words said by people, e.g., career interviews, can be used by researchers to extract information on the personality traits of each individual. In this project, the aim is to find intelligence and personality indications using a dataset of interview transcripts, answering our research question as follows: **(RQ:)** How can career interview transcripts of individuals reveal insights on their intelligence and/or personality? While inference of personality is well-explored in the literature, inferring intelligence has been omitted due to the lack of a suitable dataset.

Our brand-new dataset (IDIAP), which is introduced in [10], comprises 1,998 participants recruited through *Prolific* [11] with a short video describing their leadership and ability skills, the ICAR intelligence scores with performance incentives, along with side information, such as level of education, marital status, etc. In addition, we have also used a subset of the *My Personality* dataset [12], which includes anonymized data on social network users' personality traits and behaviors collected via online surveys, containing Big Five scores, for comparison. However, it is worth mentioning that My Personality dataset which contains tens of thousands of records is not available anymore due to privacy protection rules, so, only a subset of it is available for access.

A category of the challenges of this project is due to the nature of the problem. Training a machine learning model for predicting aspects of personality and intelligence is considered a difficult problem in the literature due to (1) sparsity and limitation of the data and (2) the abstraction of the target variables. Another challenge during this project was the limitation of using any foreign API-based model like ChatGPT or Google Translate or any foreign computational resource due to the confidentiality of the data according to the *Swiss Federal Act of Data Protection* [13], so, we were restricted to use models that we are able to run locally.

In this project, different classical and neural methodologies for inferring intelligence and personality along with methodologies for data augmentation have been explored and compared to the baseline models. To this end, a literature review has been done in section II, and then the proposed methodologies for data augmentation, feature extraction, and predicting intelligence and personality are explained in detail in section III. Afterward, the empirical results are illustrated in section IV. Then, in V, there is a conclusion and a discussion for the future directions.

II. Literature Review

Research by [14], [15], and [16] has approached personality prediction as a classification problem, using models such as Decision Trees and Support Vector Machines to identify linguistic patterns linked to psychological traits. Similarly, [17] and [18] have demonstrated the effectiveness of machine learning techniques in analyzing social media text for personality prediction.

On the other hand, [19], [20], and [21] have utilized deep learning architectures, including Recurrent Units in Neural Networks to capture contextual relationships in textual datasets like MBTI and Big Five trait assessments. Hierarchical models using recurrent layers have also been effective in analyzing social media text for detecting personality traits [22], [23], and [24]. These neural methods have demonstrated notable success in extracting semantic information and enhancing prediction accuracy.

Recent approaches, such as stylometry for text quality and vocabulary analysis, show promise in estimating cognitive abilities [25]. Expanding on this, our research examines video transcriptions to explore intelligence through natural, spontaneous language in real-world contexts.

III. Proposed Methodology

In this section, the methodologies for data augmentation (III-A), feature extraction (III-B), and predicting the target variables (III-C) are explained in detail.

A. Data Augmentation

Data augmentation techniques can be used to overcome the problem of limited data quantity. In this section, the details of methods that have been used or explored for data augmentation are illustrated.

1) Chunking: Intelligence and personality traits are considered to be constant within an individual [26]. Therefore, it can be immediately derived that these variables also remain constant during a short speech In this regard, a possible approach for data augmentation is to split long speeches into chunks and consider each of them as a separate data point. The transcripts of the lectures in our main dataset are approximately consisted of 1000 tokens (words) on average, and are long enough to apply this methodology. To this end, each speech is divided into chunks with a maximum length of 250 tokens. The start and the ends of the chunks are aligned with the start and end of sentences to preserve the semantics and sentiment.

2) Sentiment-Preserving Text Augmentation: Several other methods for data augmentation have been explored such as noise injection [27], masking [28], synonym replacement [27], back translation [29], which are considered as sentiment preserving textual data augmentations in the literature. However, in the final methodologies used for this project, the models are resilient to small changes in the input, so, these methods cannot act efficiently. To overcome this issue, back translation using state-of-the-art models like ChatGPT and Google Translates is promising, however, The usage of these API-based models was prohibited in this project due to the reasons explained in section I.

B. Feature Extraction

A combination of the National Research Council Emotion Lexicon (NRC), NRC + Valence, Arousal, Dominance Lexicon (NRC-VAD), and Linguistic Inquiry and Word Count (LIWC) dictionary are highly effective in extracting features from the text for psychological purposes. Moreover, some linguistic features can be extracted using readability measurements. The pre-trained models on the Go-Emotion [30] dataset have also been explored using pre-trained model [31]. Also, pre-trained RoBERTa [32] and DeBERTa [33] has been leverage to generate text embedding, for classical models.

1) NRC Emotion Lexicon: NRC Emotion Lexicon is a comprehensive lexicon that maps words to eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (positive and negative) [34]. It is useful for sentiment analysis and understanding the emotional tone of the text. this implementation is done by applying [35].

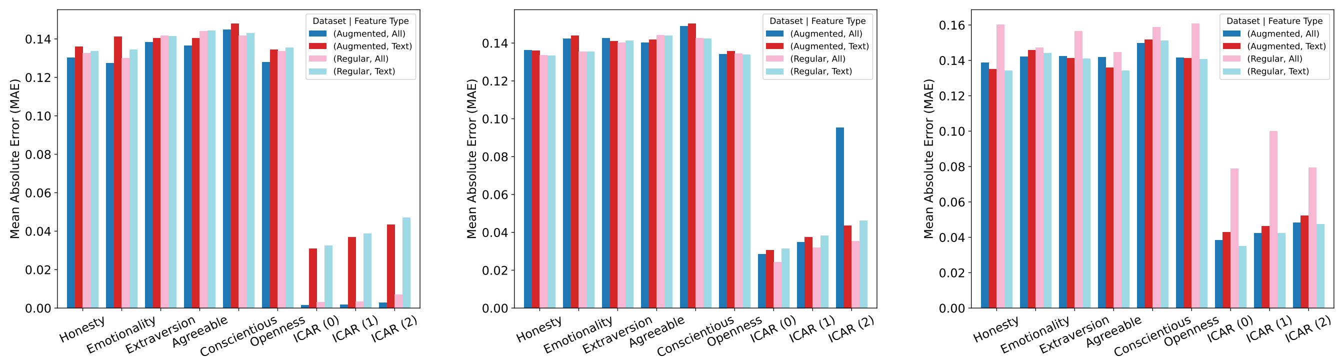
2) (NRC + VAD) Lexicon: NRC + VAD Lexicon extends the NRC lexicon by assigning each word a score on three dimensions valence (positive or negative emotion), arousal (level of excitement or calmness), and dominance (degree of control or influence) [36]. It provides an emotional profile of the text, enabling more sophisticated sentiment and emotion analyses. implementation is done by applying [35].

3) LIWC: LIWC is a text analysis tool that categorizes words into psychologically meaningful categories, such as emotional tones, cognitive processes, or social processes [37]. It is ideal for understanding underlying psychological states. implementation is done by applying [38].

4) Readability Features: Readability features are metrics that assess how easy or difficult a text is to read. Common measures include Flesch-Kincaid Grade Level which indicates the U.S. school grade level required to understand the text, or SMOG Index which predicts the years of education needed to understand a piece of writing [39]. It is useful for evaluating the complexity of the text. implementation is done by applying [40].

C. Models

The details of the models are outlined as follows. Mean Absolute Error is used (MAE) as the loss function for training and tuning the parameters in all of our models because, unlike other loss functions, MAE is less sensitive to outliers, which helps in building more robust models. Additionally, MAE ensures that each error contributes equally to the final loss, promoting fairness in the model's performance. Due to these reasons, MAE is chosen as the criterion which has improved the results. To enhance the stability of training procedures, the target variables have been normalized throughout this report.



(a) Model: LightGBM (b) Model: SVR (c) Model: MLP

Figure 1: Comparison of MAE across different models, datasets, and feature types.

1) Epsilon-Support Vector Regression (SVR): The problem of predicting personality and intelligence is mostly solved as a classification problem in the literature, in which Support Vector Machines are one of the top-performing models. To extend this model to the regression setting, we have used the Epsilon-Support Vector Regression (SVR) model which is a regression extension of the SVM model. While SVM is designed to find a hyperplane that best separates classes, SVR aims to fit the best possible hyperplane (in higher dimensions) that approximates the relationship between input features and the target variable. The hyperparameters of this model are epsilon (margin of tolerance), kernel, and c (complexity of the model and regularization) which have been tuned using a grid-search strategy combined with k-fold cross-validation. this implementation is done by applying [41].

2) Light Gradient-Boosting Machine (LightGBM): Decision Trees (DT) also have been widely used for the classification variant of this problem, however, DT cannot be naively applied to this setting since the dimensionality of the input is too high for a DT which results in low performance and long training time. To overcome this problem, *Light Gradient-Boosting Machine* is leveraged which is designed to be efficient and scalable, and it is able to handle large-scale datasets with high dimensionality and offer faster training speeds. LightGBM is based on Gradient Boosting DTs, a powerful ensemble learning technique that builds models sequentially, with each new model attempting to correct the errors of the previous ones. The key innovations in LightGBM are around its tree-growing strategy and leaf-wise algorithm. To regularize the model, Lasso regularization and bagging techniques are applied. The learning rate, Lasso regularization coefficient, and the fraction of data used for bootstrap sampling are tuned by grid search combined by k-fold cross-validation. this implementation is done by applying [41], and [42].

3) Recurrent Models: Although language models such as BERT or LLaMa are considered as state-of-the-art models for embedding the texts, the large size of their outputs results in an increasing need for trainable parameters in the prediction model, and, it results in to lack of generalization of these models. Our approach to overcome this problem is to use smaller embeddings of a pre-trained FastText [43]. Instead of simple averaging, to derive an embedding for the whole text, we've utilized a bi-directional GRU recurrent unit for aggregating fast-text embeddings. The embedding of this model is then fed into a multi-layer perceptron for final prediction. A multi-task regression model with MAE loss along with a scheduled Adam optimizer has been used for training. "Optuna" [44] library has been leveraged for efficient hyper-parameter tuning for all tunable hyper-parameters. Training a multi-task inference model instead of multiple models for each with one output implicitly enforces a regularization to the model to learn meaningful, intermediate representations.

IV. Experiments

Table I presents MAE and RMSE, along with their standard deviations, across 3 targets based on ICAR. The analysis includes two datasets: IDIAP and Augmented IDIAP. For each dataset, three types of features are utilized: (1) embeddings, (2) psychological features derived from both the original data and extracted features from the text, and (3) a combination of embeddings and psychological features. Additionally, three distinct models, as described in the preceding sections, were implemented to evaluate performance. The same thing is done for the results of regression on the personality labels in table II.

Table I shows that LGBM outperforms both SVR and MLP models in predicting ICAR. LGBM usually beats SVR and MLP because it uses efficient algorithms, can handle large and high-dimensional datasets, avoids overfitting, and is easy to adjust. Furthermore, LGBM achieves better results on the Augmented IDIAP dataset, demonstrating the effectiveness of our data augmentation method. Moreover,

Table I: Performance of different methodologies on predicting ICAR.

Dataset	Model	Text Embedding MAE	Text Embedding RMSE	Psychological Features MAE	Psychological Features RMSE	All Features MAE	All Features RMSE
IDIAP	NN	$.0414 \pm .0062$	$.2661 \pm .3558$	$.0496 \pm .0045$	$.0715 \pm .0043$	$.0860 \pm .0121$	$.0989 \pm .0110$
	SVR	$.0386 \pm .0074$	$.0550 \pm .0069$	$.0304 \pm .0056$	$.0407 \pm .0053$	$.0305 \pm .0057$	$.0409 \pm .0052$
	LGBM	$.0394 \pm .0073$	$.0558 \pm .0069$	$.0037 \pm .0019$	$.0085 \pm .0030$	$.1372 \pm .0021$	$.0096 \pm .0032$
Aug. IDIAP	NN	$.0471 \pm .0047$	$.0685 \pm .0040$	$.0433 \pm .0041$	$.0433 \pm .0041$	$.043 \pm .0050$	$.0615 \pm .0047$
	SVR	$.0372 \pm .0065$	$.0555 \pm .0057$	$.0299 \pm .0053$	$.1803 \pm .2431$	$.0529 \pm .0369$	$.0512 \pm .0043$
	LGBM	$.0371 \pm .0062$	$.0033 \pm .0024$	$.0017 \pm .0005$	$.0058 \pm .0003$	$.0020 \pm .0007$	$.0060 \pm .0005$

Table II: Performance of different methodologies on predicting Big 5 (+ 1).

Dataset	Model	Text Embedding MAE	Text Embedding RMSE	Psychological Features MAE	Psychological Features RMSE	All Features MAE	All Features RMSE
IDIAP	NN	$.1409 \pm .0064$	$.1770 \pm .0083$	$.1443 \pm .0066$	$.1792 \pm .0081$	$.1546 \pm .0069$	$.1911 \pm .0085$
	SVR	$.1383 \pm .0046$	$.1732 \pm .0054$	$.1383 \pm .0044$	$.1731 \pm .0052$	$.1384 \pm .0044$	$.1731 \pm .0052$
	LGBM	$.1386 \pm .0046$	$.1727 \pm .0055$	$.1352 \pm .0067$	$.1695 \pm .0090$	$.1372 \pm .0059$	$.1706 \pm .0062$
Aug. IDIAP	NN	$.1417 \pm .0062$	$.1793 \pm .0083$	$.1417 \pm .0035$	$.1756 \pm .0058$	$.1426 \pm .0037$	$.1784 \pm .0066$
	SVR	$.1414 \pm .0053$	$.1767 \pm .0085$	$.1394 \pm .0054$	$.1751 \pm .0081$	$.1407 \pm .0052$	$.1763 \pm .0082$
	LGBM	$.1400 \pm .0047$	$.1422 \pm .0081$	$.1131 \pm .0060$	$.1422 \pm .0081$	$.1341 \pm .0068$	$.1674 \pm .0101$

the highest accuracy is achieved by using only psychological features, indicating that incorporating embeddings and higher-dimensional learning may lead to overfitting on the training set. The same thing applies for table II for predicting personality. LGBM has the best performance on the augmented dataset using just psychological features. It is worth mentioning that better results can be achieved in predicting ICAR compared to the personality.

Table III illustrates the comparability of our model with state-of-the-art methods [45]. However, please note that this comparison is not quite fair, since they are on different samples of my personality dataset, and there is no access to the same dataset for fair comparison. The results provided for our models in this table are based on using both embeddings and psychological features as the feature set. The results of rival methods have been translated to similar normalization.

Table III: Comparing the Result of Our Models with Other Models on My Personality Dataset in Term of RMSE.

Method	E	N	A	C	O
AdaWalk	0.215	0.200	0.183	0.187	0.180
Node2Vec	0.225	0.208	0.189	0.201	0.185
DeepWalk	0.225	0.195	0.193	0.202	0.185
2CLSTMs	0.218	0.215	0.192	0.198	0.199
Doc2Vec	0.229	0.193	0.188	0.197	0.197
SVR	0.234	0.221	0.205	0.209	0.210
LGBM	0.231	0.218	0.205	0.206	0.205
NN	0.2365	0.2202	0.2093	0.2076	0.2178

Figure 1 displays three plots for each model, illustrating the Mean Absolute Error (MAE) for each target across two datasets: IDIAP and Augmented IDIAP. The analysis is conducted using two feature sets: embedding features alone and a combination of embedding and psychological features.

V. Discussion and Conclusion

This project highlights the potential of using machine learning models for accurately predicting intelligence based on the speech of an individual. However, our results suggest that intelligence traits can only be well-predicted if (1) data augmentation techniques are conducted to mitigate the data limitation problem, (2) a parameter-efficient model, e.g., LightGBM, is used, and (3) meaningful features are extracted before the training procedure and fed into the network. This shows the importance of the quality and quantity of data in dealing with the optimization-generalization tradeoff. While the initial hypothesis was that a neural network-based architecture should perform better than smaller models due to their expressive abilities, it was later found that the abundance of data did not enable the neural network to generalize on this task, necessitating further data collection. To this end, this report calls for future researchers to enhance the performance of the models in a data-centric way by working on methodologies for gathering/generating large-scale datasets, as well as to find additional approaches to similarly improve the performance for personality traits. Also, using other modalities of the data such as embedding the body language of the individual into a meaningful dense numerical space has the potential to help the model make more precise predictions. Finally, in such problems, ground-truths are also subjected to noises; thus, using more accurate methods to create ground-truths can also enhance the applicability of the model and the stability of the training procedure.

Ethical Risk

While artificial intelligence (AI) technology offers significant potential, it also brings ethical challenges, such as algorithmic discrimination, data bias, and ambiguous accountability. These risks could have tangible consequences if such algorithms are misused.[46].

One ethical concern in this project is *algorithmic bias*, particularly because the dataset have been used, *IDIAP*, includes only native English speakers [10]. This means the model may not work as well for non-native speakers, as it might struggle to interpret language patterns that differ from those of the training data. The group most affected by this bias would be non-native English speakers, especially in situations where the model is used to assess things like cognitive ability or personality. Mistakes in these cases could worsen inequities and lead to unfair decisions. Research shows that AI systems trained mostly on English data often marginalize non-native speakers, reinforcing existing social and cultural inequalities [47].

While the models used in this project achieve good accuracy in predicting traits like ICAR, the complexity of these traits and the limitations of textual data mean that human oversight remains essential. By incorporating a "human-in-the-loop" approach, we can significantly reduce the severity of risks by ensuring that predictions are reviewed and interpreted by humans before any decisions are made. This approach not only mitigates ethical risks but also promotes accountability and fairness in the system's use.

Canvas is available in GitHub repository.

References

[1] G. W. Allport, *Personality: A Psychological Interpretation*. Holt, Rinehart & Winston, 1937.

[2] P. T. Costa and R. R. McCrae, *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional Manual*. Psychological Assessment Resources, 1992.

[3] D. Karanatsiou, K. Kalimeri, and M. G. Beiró, "My tweets bring all the traits to the yard: Predicting personality and relational traits in online social networks," *arXiv preprint arXiv:2009.10802*, 2020. [Online]. Available: https://arxiv.org/abs/2009.10802

[4] L. Gou, J. Li, and J. Tang, "Improving recommendation systems by incorporating personality traits," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, pp. 5937–5951, 2022. [Online]. Available: https://link.springer.com/article/10.1007/s12652-022-04200-5

[5] R. Smith and M. Jones, "Applications of personality assessment in mental health diagnosis," *Journal of Clinical Psychology*, vol. 77, pp. 345–360, 2021. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/10731911211029104

[6] T. A. Judge and S. P. Robbins, *Personality Traits and Their Role in Job Screening*. Cambridge University Press, 2013, pp. 321–344. [Online]. Available: https://www.cambridge.org/core/books/personality-traits/applications-of-personality-assessment/41B2389B22B770FCB321A42475D848BF

[7] D. M. Condon and W. Revelle, "The international cognitive ability resource: Development and initial validation of a public-domain measure," *Intelligence*, vol. 43, pp. 52–64, 2014. [Online]. Available: https://doi.org/10.1016/j.intell.2014.01.004

[8] L. R. Goldberg, "The development of markers for the big-five factor structure," *Psychological Assessment*, vol. 4, no. 1, pp. 26–42, 1992. [Online]. Available: https://doi.org/10.1037/1040-3590.4.1.26

[9] I. B. Myers and P. B. Briggs, *MBTI Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator*, 3rd ed. Consulting Psychologists Press, 1998.

[10] V. Akstinaite, U. T. Jensen, M. Vlachos, A. Erne, and J. Antonakis, "Charisma is a costly signal," *The Leadership Quarterly*, vol. 35, p. 101810, 2024. [Online]. Available: https://doi.org/10.1016/j.leaqua.2024.101810

[11] "Prolific," 2024, accessed: 2024-12-15. [Online]. Available: https://www.prolific.com

[12] D. Stillwell and M. Kosinski, "mypersonality project website," 2015. [Online]. Available: https://www.psychometrics.cam.ac.uk/productsservices/mypersonality

[13] S. F. Government, "Swiss federal act on data protection (fadp)," 2023, revised Act effective September 1, 2023. [Online]. Available: https://www.bj.admin.ch/bj/en/home/staat/datenschutz/rechtsgrundlagen.html

[14] J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer, "Psychological aspects of natural language use: Our words, our selves," *Annual review of psychology*, vol. 54, no. 1, pp. 547–577, 2003.

[15] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. Seligman *et al.*, "Personality, gender, and age in the language of social media: The open-vocabulary approach," *PloS one*, vol. 8, no. 9, p. e73791, 2013.

[16] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from twitter," in *2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing*. IEEE, 2011, pp. 149–156.

[17] F. Celli, E. Bruni, and B. Lepri, "Personality traits and social network analysis in facebook," *Proceedings of the ACM Web Science Conference*, 2014.

[18] W. Youyou, M. Kosinski, and D. Stillwell, "Computer-based personality judgments are more accurate than those made by humans," *Proceedings of the National Academy of Sciences*, vol. 112, no. 4, pp. 1036–1040, 2015.

[19] N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, “Deep learning-based document modeling for personality detection from text,” *IEEE Intelligent Systems*, vol. 32, no. 2, pp. 74–79, 2017.

[20] D. Hernandez, S. Knight, S. Kirschenbaum, and G. Little, “Using lstm networks to predict personality traits from text,” in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2017, pp. 471–477.

[21] R. Hernandez and I. Scott, “Predicting myers–briggs type indicator with text,” in *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS)*, 2017, pp. 4–9.

[22] D. Xue, L. Wu, Z. Hong, S. Guo, L. Gao, Z. Wu *et al.*, “Deep learning-based personality recognition from text posts of online social networks,” *Applied Intelligence*, vol. 48, pp. 4232–4246, 2018.

[23] M. Kosinski, D. Stillwell, and T. Graepel, “Private traits and attributes are predictable from digital records of human behavior,” *Proceedings of the National Academy of Sciences*, vol. 110, no. 15, pp. 5802–5805, 2013.

[24] G. Park, H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, M. Kosinski, D. J. Stillwell, L. H. Ungar, and M. E. Seligman, “Automatic personality assessment through social media language,” *Journal of personality and social psychology*, vol. 108, no. 6, p. 934, 2015.

[25] A. Hendrix and R. Yampolskiy, “Automated iq estimation from writing samples,” pp. 3–7, 2017.

[26] Y. Mehta, S. Fatehi, A. Kazameini, C. Stachl, E. Cambria, and S. Eetemadi, “Bottom-up and top-down: Predicting personality with psycholinguistic and language model features,” in *2020 IEEE International Conference on Data Mining (ICDM)*, 2020, pp. 1184–1189.

[27] G. Chao, J. Liu, M. Wang, and D. Chu, “Data augmentation for sentiment classification with semantic preservation and diversity,” *Knowledge-Based Systems*, vol. 280, p. 111038, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705123007888

[28] B. Hollander, F. Frasincar, and F. van der Knaap, “Data augmentation using bert-based models for aspect-based sentiment analysis,” in *Web Engineering*, K. Stefanidis, K. Systä, M. Matera, S. Heil, H. Kondylakis, and E. Quintarelli, Eds. Cham: Springer Nature Switzerland, 2024, pp. 115–122.

[29] T. Body, X. Tao, Y. Li, L. Li, and N. Zhong, “Using back-and-forth translation to create artificial augmented textual data for sentiment analysis models,” *Expert Systems with Applications*, vol. 178, p. 115033, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421004747

[30] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, “Goemotions: A dataset of fine-grained emotions,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2020, pp. 4040–4054.

[31] B. Savani, “Bert base goemotions,” https://huggingface.co/bhadresh-savani/bert-base-go-emotion, n.d., accessed: 2024-12-19.

[32] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, “Roberta: A robustly optimized bert pretraining approach,” *arXiv preprint arXiv:1907.11692*, 2019.

[33] P. He, X. Liu, J. Gao, and W. Chen, “Deberta: Decoding-enhanced bert with disentangled attention,” *arXiv preprint arXiv:2006.03654*, 2020.

[34] S. M. Mohammad and P. D. Turney, “Crowdsourcing a word-emotion association lexicon,” *Computational Intelligence*, vol. 29, no. 3, pp. 436–465, 2013.

[35] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. Sebastopol, CA: O’Reilly Media, Inc., 2009. [Online]. Available: http://www.nltk.org/book/

[36] S. M. Mohammad, “Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words,” in *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*, Melbourne, Australia, 2018.

[37] M. L. Newman, J. W. Pennebaker, D. S. Berry, and J. M. Richards, “Lying words: Predicting deception from linguistic styles,” *Personality and Social Psychology Bulletin*, vol. 29, no. 5, pp. 665–675, 2003, pMID: 15272998. [Online]. Available: https://doi.org/10.1177/0146167203029005010

[38] E. Fast, M. Yatskar, L. Zettlemoyer, and E. Hovy, “Empath: Exploring and building lexicons for text analysis,” in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2018, pp. 2534–2544. [Online]. Available: https://aclanthology.org/D18-1350/

[39] T. Deutsch, M. Jasbi, and S. Shieber, “Linguistic features for readability assessment,” in *Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications*, J. Burstein, E. Kochmar, C. Leacock, N. Madnani, I. Pilán, H. Yannakoudakis, and T. Zesch, Eds. Seattle, WA, USA → Online: Association for Computational Linguistics, Jul. 2020, pp. 1–17. [Online]. Available: https://aclanthology.org/2020.bea-1.1

[40] D. Bevan, “Readability: Python library for text readability analysis,” https://github.com/domdfcoding/readability, 2023, accessed: 2024-04-27.

[41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. [Online]. Available: https://scikit-learn.org/

[42] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017, pp. 3146–3154. [Online]. Available: https://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf

[43] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.

[44] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2019, pp. 2623–2631.

[45] X. Sun, B. Liu, Q. Meng, J. Cao, J. Luo, and H. Yin, "Group-level personality detection based on text generated networks," *World Wide Web*, vol. 23, no. 3, pp. 1887–1906, 2020. [Online]. Available: https://doi.org/10.1007/s11280-019-00729-2

[46] M. Kozlow *et al.*, "Ethical risk factors and mechanisms in artificial intelligence decision making," *Journal of Artificial Intelligence Ethics*, 2022. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9495402/

[47] A. Mukherjee, C. Raj, Z. Zhu, and A. Anastasopoulos, "Global voices, local biases: Socio-cultural prejudices across languages," *arXiv preprint arXiv:2310.17586*, 2023. [Online]. Available: https://arxiv.org/abs/2310.17586