

# COINs - Predicting personality from text

Justus Cappel <sup>a</sup>, Esma Hüner <sup>b</sup>, Fation Selimllari <sup>c</sup>

a University of Cologne, jcappel@smail.uni-koeln.de

b Bamberg University, esma.huener@stud.uni-bamberg.de

c Bamberg University, fation\_selimllari@stud.uni-bamberg.de

**Abstract** In the field of Natural Language Processing (NLP), there is a great deal of interest in predicting personality traits from text data. This research addresses the relationship between linguistic cues in texts and individual personality traits using the Big Five model and Schwartz's value categories. The goal is to develop a robust machine learning model capable of accurately predicting personality traits from text data extracted from various videos on YouTube. The study aims to categorize people into four different tribes: Nerds, Fatherlands, Treehuggers, and Spiritualists, based on their personality profiles. It also aims to create a user-friendly communication interface that will allow individuals to interact with the trained model and provide them with valuable insights into their own personality traits based on their written text. This web application provides a convenient and accessible way for individuals to explore and gain self-knowledge about their personality traits. This paper presents a comprehensive methodology that includes data processing, feature extraction, and model training. It describes the data set used for training and evaluation, and describes the experiments conducted to accurately assess the performance of the model. The study also discusses the results obtained, showing the model's precision, recovery rate, and F1 score, which indicate balanced performance in predicting personality traits. A key aspect of the study is to explore the implications of the results, particularly with regard to the fuzzy boundaries between tribes and linguistic similarities between specific groups. The study highlights the challenges and nuances associated with categorizing individuals into different tribes based on linguistic patterns and sheds light on the complexity of understanding language-based personality predictions. The paper also suggests several avenues for future research to improve the understanding of language-based personality prediction. It explores possibilities such as using larger and more diverse datasets, comparing the performance of different pre-trained models, enabling personality analysis using audio files or dialogs, and providing additional insights and visualizations to improve the user experience. In summary, this study provides valuable insights into the field of personality trait prediction from text data by leveraging NLP and advanced machine learning techniques. The study demonstrates the potential of these methods in accurately predicting personality traits and provides individuals with meaningful self-knowledge through an intuitive web application. The limitations of the study are discussed, and future directions are proposed to improve and expand the application of language-based personality prediction.