Francisco José de Caldas District University

Faculty of Engineering

Systemic Analysis and Design

Psychopathy Prediction Based on Twitter Usage

Authors: Juan David Avila - 20232020154

Raul Andres Diaz Losada - 20232020058

Juan David Bejarano Cristancho - 20232020056

David Sanchez Acero - 20232020049

Professor: Carlos Andrés Sierra Virguez

October 25, 2025

Contents

List of Figures

List of Tables

Abstract

Predicting psychopathy through the analysis of Twitter usage and language represents an innovative application of data science and computational psychology. Through text processing techniques and the study of variables such as word frequency, interaction, and behavioral patterns, it is possible to identify psychopathic profiles on social media. This approach offers potential for early detection and prevention in mental health, although methodological challenges remain, such as data imbalance and the risk of overfitting. It also raises important ethical challenges regarding responsible use and privacy.

Introduction

Online platforms generate vast amounts of user data that can reveal patterns of behavior and personality. The Online Privacy Foundation's Kaggle competition on psychopathy prediction addresses a central challenge: identifying behavioral markers of psychopathy from social media activity without relying on invasive or unethical profiling. The dataset, built from 2,927 anonymized Twitter users across 80 countries, captures linguistic, temporal, and interaction features. Current models struggle to balance predictive accuracy with fairness and privacy concerns, often overfitting to specific linguistic or cultural contexts. This project seeks to explore how a systemic and modular approach can enhance the interpretability and ethical reliability of psychopathy prediction models, ensuring that the analysis remains both statistically sound and socially responsible.

Literature review

Different studies have explored the relationship between linguistic patterns and digital behavior in relation to personality traits, particularly those associated with the Dark Triad: narcissism, Machiavellianism, and psychopathy. Research has shown correlations between specific lexical choices, syntactic markers, and levels of psychopathy, as well as the influence of message length, posting frequency, and interaction style on computational predictions. These findings suggest that online communication may provide measurable indicators of underlying personality dimensions, which can be leveraged for behavioral modeling under ethical and privacy constraints.

Background

The competition was organized in 2012 by the Online Privacy Foundation and Kaggle, with more than 3,000 volunteers completing the SD3 instrument. These data made it possible to evaluate the ability of the models to identify the small fraction of users (around 3%) with high scores, based on 337 variables extracted from Twitter activity.

Objectives

The main objectives were to develop models capable of predicting psychopathy from public Twitter data, analyze relative accuracy compared to other personality traits, and encourage discussion about the ethical and responsible use of psychological data mining.

Scope

The scope of this study focuses on analyzing the relationship between linguistic behavior and psychopathy indicators among Twitter users, using publicly available datasets and computational modeling techniques. The research emphasizes feature extraction, data preprocessing, and the evaluation of predictive models capable of identifying behavioral patterns associated with psychopathic traits. It does not aim to provide clinical diagnoses or psychological assessments but rather to explore how digital language and user activity can serve as indirect indicators of personality tendencies. Ethical considerations, while

acknowledged, are addressed at a general level without delving into medical or therapeutic implications.

Assumptions

The report is based on several assumptions: that self-assessments of the trait reflect actual values, that language and interactions on Twitter are representative of personality, and that the 337 variables selected contain sufficient information to build efficient predictive models.

Limitations

The main obstacle was class imbalance, with few positive cases. In addition, the self-assessed nature of the trait may incorporate personal biases, the limited amount of text per tweet restricts analysis, and the risk of overfitting is high in competitive contexts.

Methodology

In this first stage, basic analysis and classification methods were used to explore the relationship between language and possible personality traits. Initial data cleaning and organization was performed, followed by testing with simple prediction models to observe general trends. Evaluations were made using basic accuracy measures, with the aim of identifying preliminary patterns and establishing a basis for future model improvements.

Results

The work focused on exploring general trends within the dataset rather than achieving high predictive performance. The preliminary analyses revealed differences in word choice, message length, and posting frequency between user groups.

Discussion

The analysis supports the usefulness of digital data for psychological profiling, although it warns of technical and ethical limitations. The study confirms that psychopathy is

the most predictable trait in the context of Twitter but emphasizes the need for external validation and rigorous ethical controls.

Conclusion

The study demonstrated the viability of computational analysis for detecting psychopathy, although it highlighted the importance of robust validation and ethical reflection in the interpretation and use of predictive models built with public data.

References

Bibliography

[1] Park, Greg. "The dangers of overfitting: a Kaggle postmortem." (2012).

[2] KDnuggets. "Competitions: Predict personality traits, Psychopathy based on Twitter usage." (2011).

[3] Online Privacy Foundation. "Are you what you Tweet?" (2015).

Appendices

- **Appendix A:** Summary of dataset variables: word frequency, use of punctuation, message length, social network metrics, and interaction patterns.
- **Appendix B:** Practical example of the average precision calculation used in the competition.

Glossary

Psychopathy Personality trait characterized by low empathy and tendencies toward antisocial behavior.

Dark Triad A group composed of three traits: psychopathy, narcissism, and Machiavellianism.

Overfitting Loss of generalization ability in a predictive model due to excessive adaptation to the training data.

Leaderboard Ranking table used to compare results in modeling competitions like Kaggle.