# Systemic Analysis and Design for Psychopathy Prediction on Twitter

Juan David Bejarano Cristancho 20232020056
Raúl Andrés Díaz Losada 20232020058
Juan David Ávila 20232020154
David Sánchez Acero 20232020049

***Abstract*—The Kaggle competition *"Predicting Psychopathy Based on Twitter Usage"* presents a unique intersection between data science, psychology, and systems engineering. This work proposes a systemic architecture to model psychopathic traits from linguistic, temporal, and behavioral indicators in social networks. The methodology integrates principles of chaos theory, modular design, and computational efficiency for high-dimensional, unbalanced data. Key components such as the *BehaviorPatternAnalyzer* and *LinguisticFeatureExtractor* were developed to manage textual complexity and user variability. The proposed workflow aims to improve prediction consistency, manage sensitivity to input changes, and ensure reproducibility in real-world machine learning contexts.**

***Index Terms*—Psychopathy prediction, Systems engineering, Chaos theory, Natural language processing, Kaggle competition, Twitter analysis**

## I. Introduction

### A. Background and Problem Statement

Online behavior analysis provides new insights into human personality through digital footprints. The Kaggle competition organized by the Online Privacy Foundation explores the feasibility of predicting psychopathy levels using tweets from anonymized users. However, the task is inherently complex due to the sensitivity of linguistic cues, feature diversity, and the chaotic behavior of machine learning models facing unbalanced data.

The competition's dataset contains linguistic features, usage statistics, and inferred personality labels for approximately 3,000 users. Predicting psychopathy from text therefore involves modeling subtle emotional signals embedded in high-dimensional language data—a challenge that demands a systemic, transparent, and adaptive approach.

### B. Aims and Objectives

The objectives of this research are:

- Design a modular and interpretable system architecture for psychopathy score prediction.
- Implement control strategies based on chaos theory to mitigate instability in model performance.
- Validate model accuracy, sensitivity, and robustness under real-world data variability.

### C. Solution Approach

The system was developed using a modular design pattern to ensure clarity and maintainability. Three key principles guided the process:

- **Functional decomposition:** Separation of linguistic, behavioral, and psychometric modules.
- **Chaos management:** Application of sensitivity control techniques during preprocessing and training.
- **Efficiency:** Implementation of lightweight models capable of running under limited computational resources.

### D. Summary of Contributions and Achievements

This work presents:

1) A modular architecture for psychopathy detection combining feature extraction, classification, and chaos mitigation.
2) Implementation of interpretability tools to analyze linguistic relevance.
3) A proof of concept in Python using datasets derived from Twitter activity.

### E. Organization of the Report

The report is organized as follows: Chapter 2 presents the methodology and design. Chapter 3 details the proposed architecture. Chapter 4 describes system dynamics and experimental insights. Chapter 5 includes discussions and conclusions.

## II. Methodology

### A. Systemic Approach

Following engineering principles, the analysis decomposes the system into three interacting components:

- **Data Interpretation Layer:** Focused on understanding linguistic and behavioral patterns from users.
- **Learning Engine:** Responsible for building predictive models capable of correlating features with psychopathy scores.
- **Stability Controller:** Manages sensitivity issues through parameter regularization and stability monitoring.

### B. Identified Challenges

Two critical difficulties were identified:

1) **Feature noise and ambiguity:** Certain language patterns may appear psychopathic but represent harmless irony or humor.
2) **Systemic chaos:** Small changes (e.g., tokenization or hyperparameter tuning) produce large performance variations across folds.

To address this, subsystems were designed with specific roles, aiming to isolate interference and reduce unpredictability.

### C. System Architecture (Modular Design)

The proposed system is composed of the following modules:

- **TweetLoader:** Imports and validates raw data from Twitter archives.
- **DataSanitizer:** Removes emojis, links, and special characters; standardizes casing.
- **LinguisticFeatureExtractor:** Computes syntactic and semantic features using SpaCy and LIWC.
- **BehaviorPatternAnalyzer:** Derives metrics such as posting frequency, activity time, and interaction entropy.
- **PsychopathyPredictor:** Applies ensemble models (XGBoost, LightGBM) for psychopathy score estimation.
- **ChaosController:** Adjusts model sensitivity through variance analysis and weight stabilization.
- **OutputFormatter:** Generates Kaggle-compatible CSV submissions and summary reports.

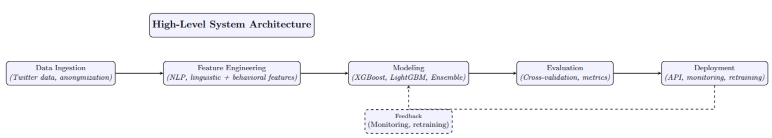


Fig. 1: High-level schematic of the proposed architecture for psychopathy prediction.

### D. System Dynamics

The interaction between modules exhibits features typical of complex adaptive systems. Nonlinear dependencies exist between preprocessing quality and model generalization. Chaos manifests when small variations in feature selection or initialization propagate through the ensemble.

Three strategies are used to stabilize this system:

- Feedback loops that adjust learning rates based on validation error variance.
- Ensemble averaging to minimize extreme deviations.
- Anomaly detection to track overfitting and divergence across iterations.

### E. Summary

This approach treats psychopathy prediction as a dynamic engineering system rather than a simple classification task. The modular design supports traceability, interpretability, and iterative improvement aligned with systems engineering principles.

## III. Discussion and Conclusion

The presented system demonstrates that applying systems thinking and chaos control principles to psychological prediction tasks enhances both reliability and interpretability. By modularizing the processing pipeline, the methodology mitigates instability common to unbalanced textual datasets.

Future work includes integrating transformer-based embeddings for deeper linguistic representation, incorporating ethics-aware evaluation metrics, and deploying real-time inference architectures under federated constraints.

## Acknowledgment

This work was carried out as part of the course *Systems Analysis and Design*, directed by Eng. Carlos Andrés Sierra, M.Sc., Universidad Distrital Francisco José de Caldas, during the 2025–III academic period.

## References

[1] Online Privacy Foundation, "Predicting Psychopathy Based on Twitter Usage," Kaggle, 2012.
[2] D. L. Paulhus and D. N. Jones, "Introducing a Short Measure of the Dark Triad," Society for Personality and Social Psychology, 2011.
[3] A. Kossiakoff et al., *Systems Engineering Principles and Practice*, 2nd ed., Wiley, 2011.
[4] E. Borgonovo and E. Plischke, "Sensitivity analysis: A review of recent advances," *Eur. J. of Operational Research*, vol. 248, no. 3, 2016.