



SYSTEM DESIGN FOR PSYCHOPATHY PREDICTION ON TWITTER

JUAN DAVID BEJARANO C. - RAÚL ANDRÉS DÍAZ L. - JUAN DAVID ÁVILA A. - DAVID SÁNCHEZ A.
UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS - INGENIERÍA DE SISTEMAS
SYSTEMS ANALYSIS AND DESIGN

Introduction

This paper presents the design of a system to predict psychopathy levels based on Twitter usage. The study explores how linguistic and behavioral patterns on social media can indicate psychopathic traits.

The main challenges involve high-dimensional and unbalanced data, along with the chaotic nature of social interactions. These factors demand a robust and adaptable architecture capable of managing noise and randomness.

To address previous issues such as overfitting and low reproducibility, the proposed system emphasizes modularity, scalability, and reliability for future real-world applications.

Goal

Develop a robust, modular, and scalable system architecture to predict psychopathy levels based on Twitter usage data, integrating chaos-aware principles to ensure stability and reproducibility during model training and evaluation.

Proposed Solution

The system is designed in five main parts to make it easier to build, test, and improve over time.

1. Data Ingestion: load and check the Twitter dataset from the Kaggle competition.
2. Feature Engineering: get linguistic and behavioral features using tools like NLTK, SpaCy, and LIWC-type analyses.
3. Modeling: train and compare several machine learning models such as Logistic Regression, XGBoost, and LightGBM.
4. Evaluation: test the models with metrics like accuracy, AUC, and F1-score, using fixed random seeds to reduce randomness.
5. Deployment: plan a simple API for real-time predictions in future versions.

Each part is independent but connected through data pipelines, making the system modular, scalable, and easier to maintain or improve later on.

Experiments

Results

Conclusions