

INTELLISHIELD: AI FOR SUBSTANCE MISUSE PREVENTION

Department of Data Science and Artificial Intelligence
Faculty of Information Technology and Computer Sciences

Duaa Sawalmeh | Sarah Bani Khalaf
Supervised by Dr.Enas Khashashneh



ABSTRACT

IntelliShield is an AI-based system designed to predict the risk of cannabis misuse among youth using behavioral and psychological traits. It applies machine learning techniques such as XGBoost, LightGBM, and SMOTE for data balancing. The system supports early prevention and intervention efforts through accurate prediction and interpretable models. IntelliShield contributes to public health by enabling data-driven strategies to reduce substance misuse.

INTRODUCTION

Cannabis misuse among youth is a growing concern worldwide. Traditional prevention strategies often lack precision and early detection capabilities. IntelliShield aims to bridge this gap by using artificial intelligence to analyze personal and behavioral traits and identify individuals at risk. The system enhances early intervention strategies and supports educational and health sectors in data-driven decision-making.

LR OR SIMILAR PROJECTS

Several studies have explored AI in substance misuse prediction.

- Park et al. (2022): Multi-country XGBoost analysis
- Koirala et al. (2021): Systematic review on ML in substance use
- Zafari et al. (2022): LSTM + SHAP for interpretable prediction
- Uddin et al. (2021): PCA + Logistic Regression on local data
- Langton et al. (2023): Predicting early adulthood drug use

Unlike these studies, IntelliShield focuses on cannabis misuse in youth, using scalable tree-based models with preprocessing and class balancing, aiming for real-time prediction and preventive intervention.

PROJECT DESIGN

Dataset:

NSDUH (National Survey on Drug Use and Health), 56,276 records × 12 features

Preprocessing Steps:

- Target cleaning
- Feature scaling using MinMaxScaler
- Outlier capping (Winsorization)
- Class balancing using SMOTE

Feature Engineering:

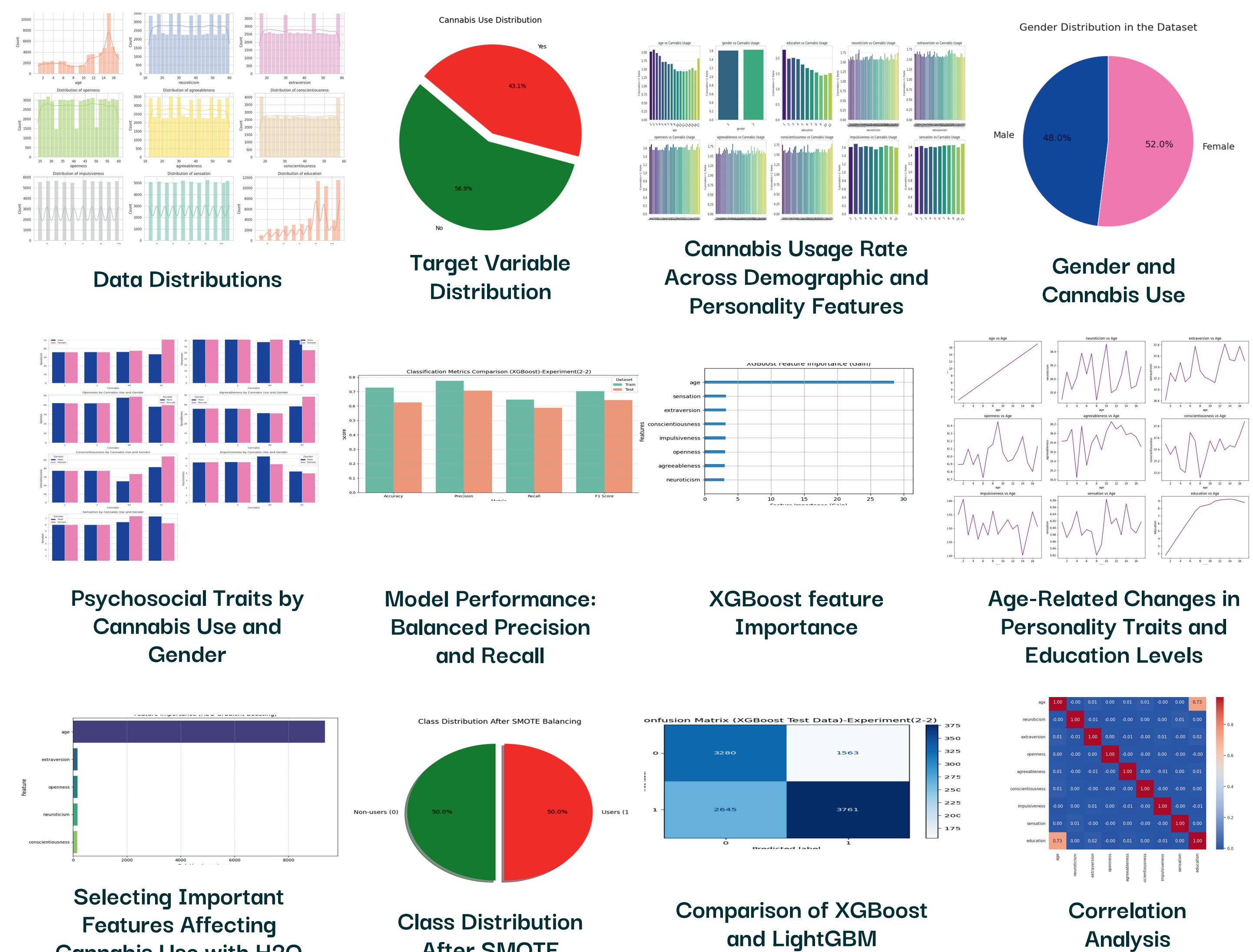
Important features selected using XGBoost's feature importance

Machine Learning Models:

- XGBoost
- LightGBM
- CatBoost
- Random Forest
- Support Vector Machine (SVM)

Best Model:

- XGBoost
- F1-score: 0.7639, Accuracy: 0.7728



SAMPLE SCREENS

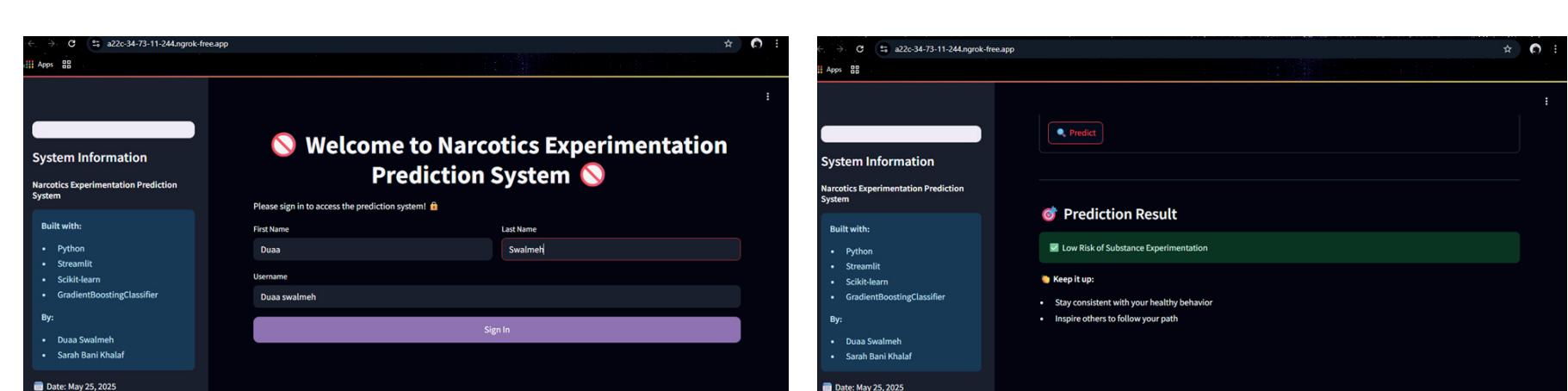
Web Interface - Streamlit Prototype

A basic web interface was developed using Streamlit. Users can input features such as age, neuroticism, and openness.

The system provides real-time predictions of cannabis misuse risk.

The prototype was deployed via Ngrok for testing purposes.

A full-featured version is planned for future deployment.



CONCLUSION

IntelliShield demonstrates the effectiveness of machine learning in predicting youth susceptibility to cannabis misuse using behavioral and psychological traits. By identifying high-risk individuals early, the system enables timely and targeted preventive actions for educators, families, and health authorities. The achieved results—such as an F1-score of 0.66 using XGBoost—highlight the potential of AI as a reliable tool in public health risk assessment. With further enhancements and real-world deployment, IntelliShield can play a vital role in reducing substance misuse and supporting data-driven prevention strategies.

FUTURE PATHS

- Expand dataset with newer and more diverse records
- Deploy a full web application for public use
- Integrate explainable AI tools (e.g., SHAP, LIME)
- Improve classification of minority (user) class
- Collaborate with schools or NGOs for real-world testing



SCAN TO ACCESS SOURCE
CODE & DATASET