

MarijuanaAnalyzer: an Intelligent Visualization and Analysis Web App

Muzi Bi, Yang Lu, Jing Ren, Fangxia Tian, Jian Wang, and Yutong Wu

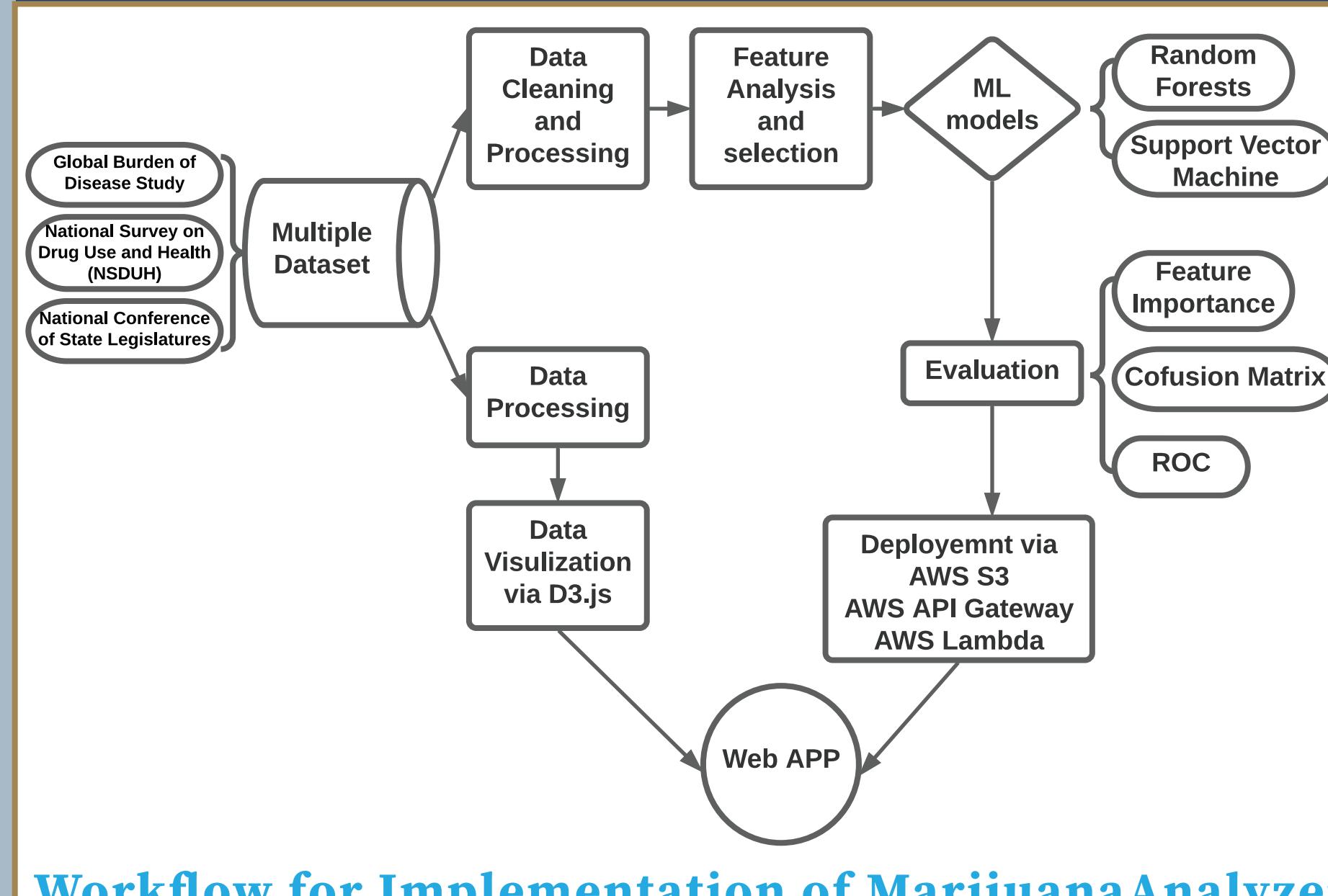
College of Computing, Georgia Institute of Technology, Atlanta, GA

INTRODUCTION

Since California legalizing medical marijuana in 1996, 42 states in U.S. have already legalized medical marijuana. Moreover, 15 states have fully legalized recreational marijuana. Marijuana can be beneficial to health and stimulate the brain with a euphoric feeling. Nonetheless, it might lead to the use of other illicit drugs and mental disorders. **A deeper understanding of marijuana's impact on illicit drug use and mental health is still lacking.**

In this study, we made an interactive web app to provide an overview of marijuana. Also, we performed statistical analysis and applied machine learning algorithms to evaluate the impact of marijuana. The goal is to provide insights into policymakers' decision making as well as providing more informative education to the general public.

STRATEGY



Workflow for Implementation of MarijuanaAnalyzer

Intuition, Innovations, and Approaches:

- Merging multiple large public datasets
- Overview of marijuana through interactive data visualization
- Statistical analysis by correlation matrices
- Feature selection among 2691 features
- Machine learning model selection and training
- A predictor tool to help user identify potential drug disorder and mental disorder risks
- Host product on interactive&responsive web app

Tools: Web app: HTML, CSS, JavaScript, Bootstrap 4, GitHub Pages; Data viz: D3.js, Matplotlib, Seaborn; Data analysis/Machine learning: Python, Scikit-learn, Numpy, Scipy, Pandas; Model deployment: AWS API Gateway, AWS Lambda, AWS S3.

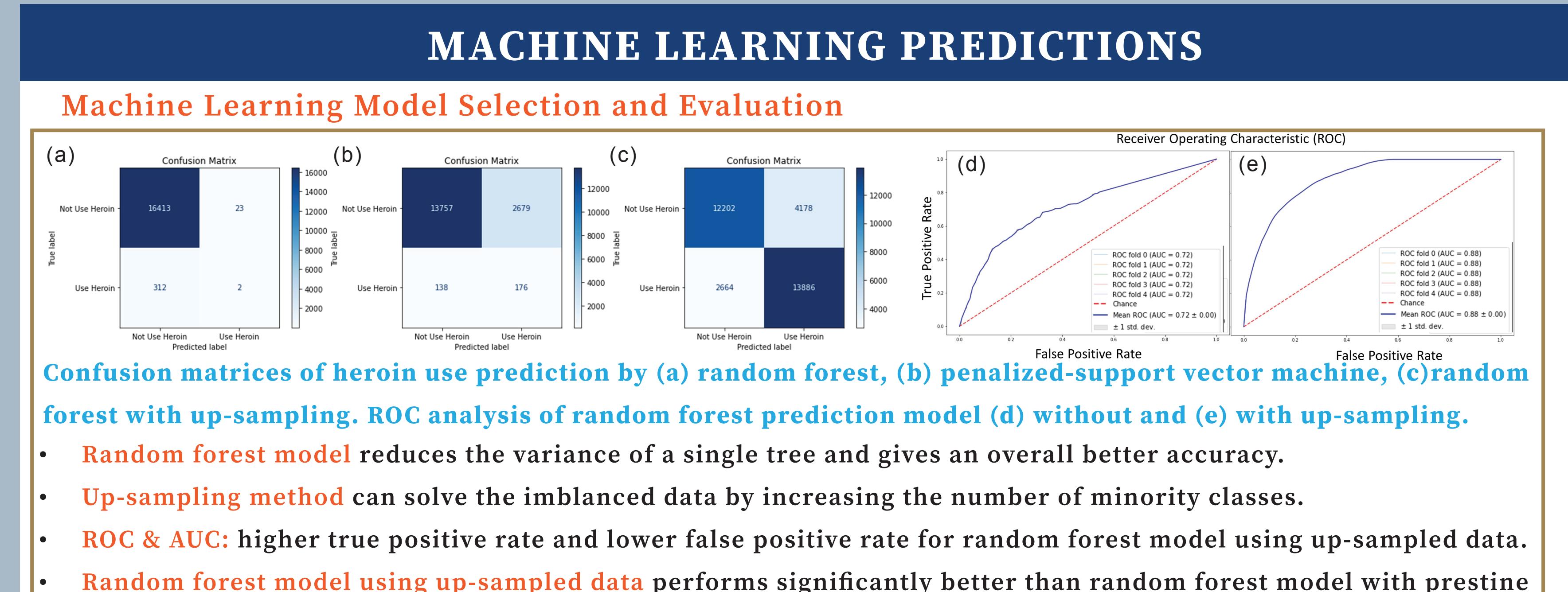
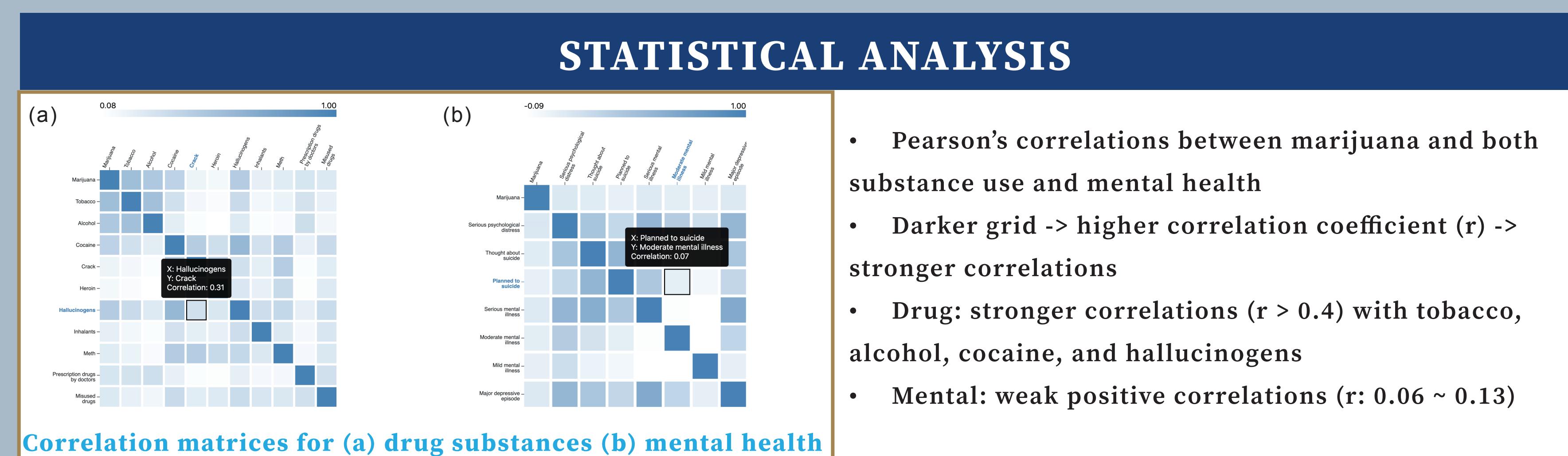
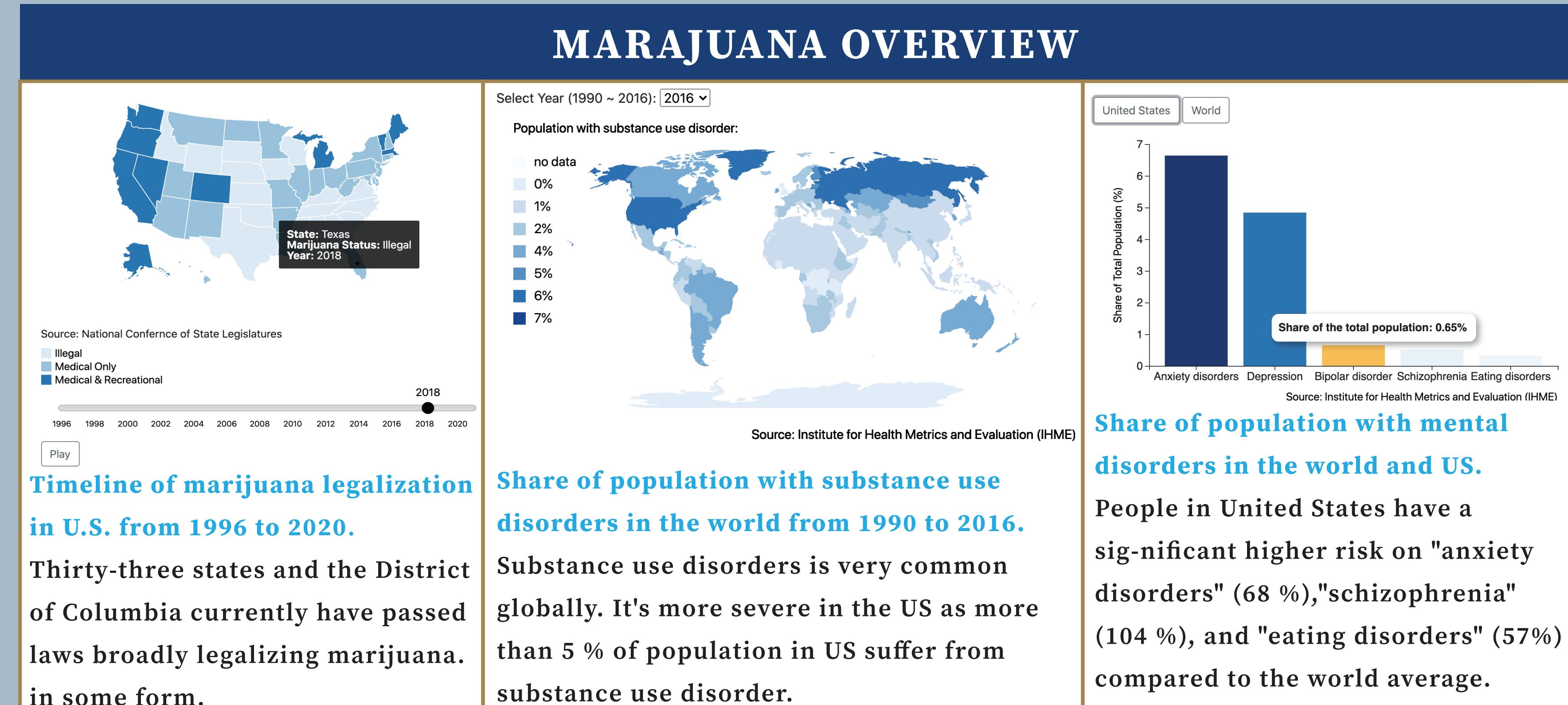
DATA

Main dataset (Downloaded, 394.7 MB):

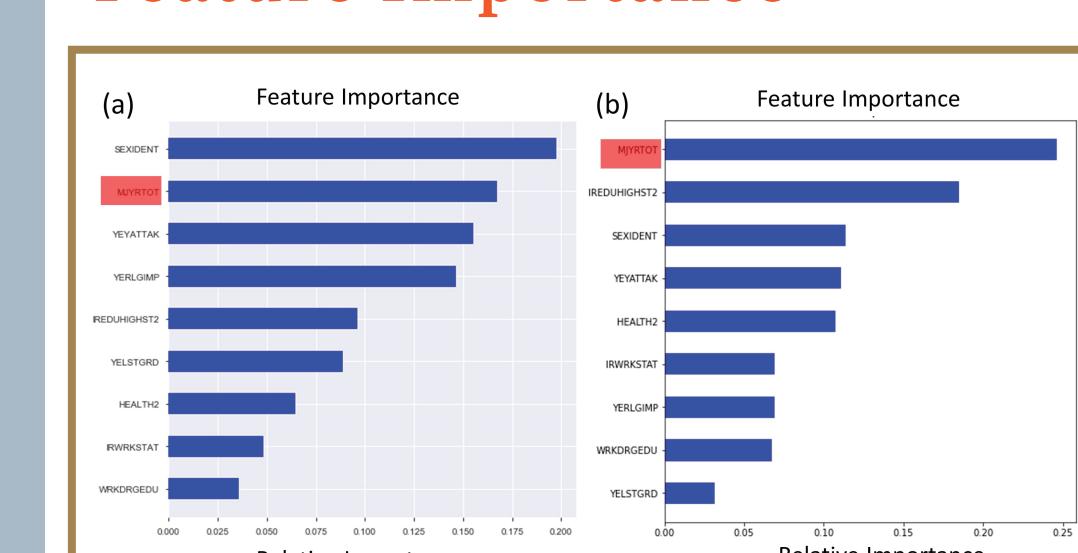
- National Survey on Drug Use and Health (NSDUH): 67,791 data points, 2691 attributes.

Additional data source (Results tool):

- Global Burden of Disease Study (GBD)
- National Conference of State Legislatures (NCLS)



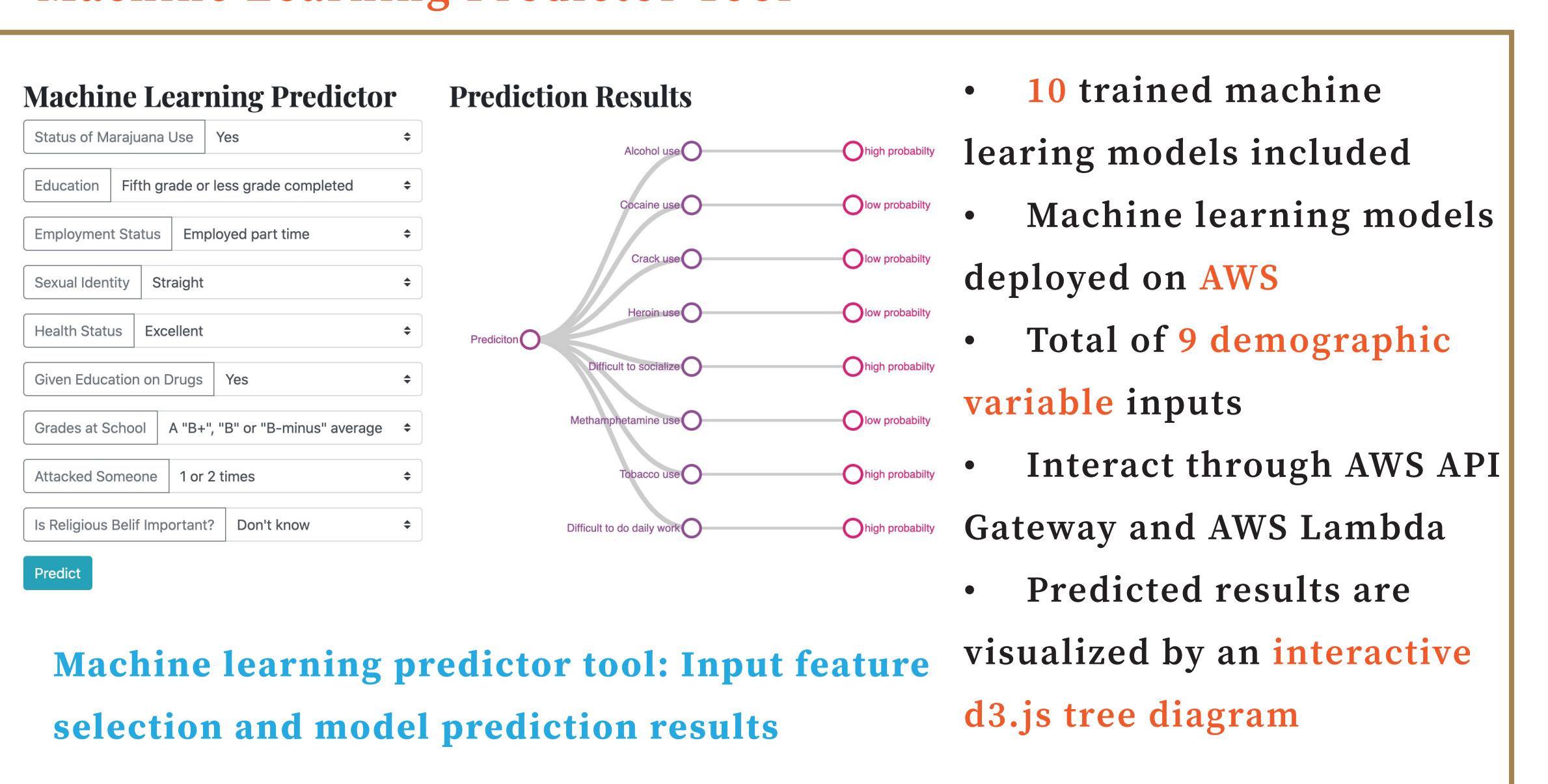
Feature Importance



Feature importance of (a) tobacco use prediction and (b) heroin use prediction

- Marijuana highlighted in red
- Marijuana has an obvious impact on the predicting results compared to other input features.

Machine Learning Predictor Tool



Machine learning predictor tool: Input feature selection and model prediction results

SUMMARY

We developed a web app using data visualization techniques, machine learning models, and cloud platforms to give an overview of marijuana legalization, correlations between marijuana and both drug disorders and mental disorders. In addition, the web app utilized best performed trained machine learning models to help users predict their risks on drug disorder and mental disorder.