

# Data Nutrition Project Label (Team 1 | Project Title: Prediction of Pharmaceutical Removal from Wastewater with Machine Learning)

## Prediction of Pharmaceutical Removal from Wastewater with Machine Learning

### Description

The dataset is used for a group project for I-GUIDE Summer School 2024. The dataset contains information extracted from scholarly literature, searched and extracted from Google Scholar, Web of Science, Scopus and PubMed. 'Pharmaceutical Waste Removal' has been used as the major keyword while searching for articles. The collected information has been inserted into an excel file across and these information included - pore volume, pyrolysis temperature, types of pharmaceuticals, excess molar refraction, solute dipolarity/polarizability, Hydrogen bond acidity, Hydrogen bond basicity, McGowan's characteristics volume, gas-to-hexadecane partition, pH of the solution and point of zero charge of sorbents.

#### Keywords

Pharmaceutical Biochar Machine Learning Wastewater

### How to use it?

#### Intended Use

- Intended Domain.** Machine Learning & Pharmaceutical Waste Removal
- Intended Use.** Intended use was to predict pharmaceutical removal from wastewater using machine learning technique
- Other Responsible Uses.** Federal agencies, academicians and industrialists may use this dataset in their respective field and can ensure water security.

#### Known Uses

- The dataset is used for a group project in I-GUIDE Summer School 2024..

#### Restrictions on Use

- The dataset has been made available for public view. However, its usage is restricted and dependent on the owner's consent.

#### Do Not Use

- Domain.** This dataset should not be used for any research or industrial activity without the consent of the dataset owner.
- This dataset should not be used for any research or industrial activity without the consent of the dataset owner.** The owner of the dataset initiated a rigorous searching of relevant literature with his initiative. And this dataset is intended to be used only for the Summer School's project activity. Thus, it is suggested not to be used in other purposes to avoid conflict of interest.

### About the dataset

#### People

Owned by Jude Okolie  
Created by Jude Okolie  
Maintained by Jude Okolie  
Funding No funding was needed since data was collected through online literature search.

Management No management funding was needed.

#### Technical information

Publish Date 2024-04-30  
Format .xlsx  
Instances 86

Version  
The original dataset has been revised through a systematic exploratory data analysis (EDA) stage. The dataset being described is the original dataset.

License  
The dataset hasn't been made available so licensing was not required.

Collection timeframe January 2024 to April 2024

Collection process  
Scientific literatures on the relevant topic were searched from 04 key websites - Google Scholar, Web of Science, Scopus and PubMed. 'Pharmaceutical Waste Removal' was used as the major keyword while searching for articles.

#### Useful links

Dataset access point  
<https://github.com/J4RELY/I-GUIDEpharmDATASET>  
This metadata repository describes the data.  
<https://github.com/J4RELY/I-GUIDEpharmDATASET>

### Inference risks

#### At a Glance



About humans

No



Upstream sources

Yes



Technical review

Yes



Ethical review

No



Update frequency

No

#### Data values

What values are in each column?

Collection and Labeling Protocols  
Data Manipulation Protocols  
Raw Data

Data Imputation Protocols  
Missing Data

#### Number of issues

Risky 1  
Safe 1  
Unknown 3

#### Feature selection

Which columns were chosen and why?

Cultural or Domain Assumptions  
Planning Representation

Proxy Characteristics  
Domain Knowledge

#### Number of issues

Risky 0  
Safe 2  
Unknown 2

#### Representation

Which rows were included and why?

Subpopulation Information  
Individual Inferences  
Collection Representation

Representation  
Individual Inferences - Mitigation  
Other Representation Issues

#### Number of issues

Risky 1  
Safe 1  
Unknown 4

#### Upstream sources

Are there known risks in datasets upstream?

Intended Use Familiarity  
Data Processing Familiarity

Data Collection Familiarity  
Other Known Issues

#### Number of issues

Risky 0  
Safe 0  
Unknown 4

#### General risks

Any additional risks?

Individual Information  
Generalized Inferences  
Sensitive Content  
Other Known Issues

Consent  
Generalized Inferences - Mitigation  
Documented Known Issues

#### Number of issues

Risky 0  
Safe 1  
Unknown 6