Data Nutrition Project (DNP) for Team 1’s Project - Prediction of Pharmaceutical Removal from Wastewater with Machine Learning

The tool, DNP, interrogates the dataset for a variety of imbalances or problems it could have, ascertains if it is the right dataset for a model and represents key components presenting dataset quality resembling the format of FDA’s nutrition label. For this project, DNP label has been created to understand multiple insights on - data usage, data values, risk, data collection process and so on.

In brief, since our project doesn’t deal with data collection on human, many of the questions present in the tool don't align with our dataset. The inference risks’ summary shows positive result regarding - human engagement, technical review and data update frequency while neutral observations are noted for parameters such as upstream sources and ethical review. However, no ethical review was needed, since the project doesn’t concern data collection from human. The tool also identified few risks on data collection protocol and data representation. For many of the parameters, the tool was unable to identify any risk, since the associated questions were not relevant with the project.

Overall, the tool provided quite positive feedback on the dataset’s GeoEthics component. The output of DNP label is as followed.









