

Dataset Questions

1. Describe the dataset. (ex: Information about Traffic violations in Montgomery County, MD; or The number of microaneurysms found in a patient's eye and whether or not they have diabetic retinopathy.)

Information regarding mental health statistics, education rates, and unemployment within the counties of each state, in order to determine if a county is deficient in health care professionals.

2. How many records does the dataset have?

3,142 records.

3. How many features does the dataset have? List or describe a few of them.

579. Examples of some of these features include the number of individuals with major depression within each county, the number of individuals who lack a high school diploma, the number of individuals who are unemployed, the number of individuals who are severely work disabled, and the number of people who have participated in recent drug use. We also plan on using another dataset as a supplement to assist with feature engineering.

4. What can you try to predict in this dataset? (ex: We can try using the features, including age, race, gender, car make and model, etc, to predict the type of traffic violation; or We can use the number of microaneurysms measured in the patient's eye to predict whether or not they have diabetic retinopathy.)

We can attempt to predict whether or not a county is deficient in health professionals based on features related to drug use, unemployment, mental health disorders, and disability.

5. Is this a **labeled** dataset, appropriate for a supervised learning classification problem? (In other words, if you are trying to predict whether or not someone has a disease, does your dataset contain whether or not each record has the disease?)

This is a labeled dataset. One thing to note is that the dataset does not use the common indicator features of 0 and 1, but instead uses 1 and 2.

6. Provide a link to the dataset, if there is one. If you are getting your data from somewhere other than a link, where are you getting it from?

Main Dataset:

ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/CHDI/chsi_dataset.zip

Supplementary Data:

<https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/>