Description of the Datasets

**Input Data**

* **File Name**: ‘Phosphopeptide\_abundance.csv’
* **Data Description:** Mass spec data indicating phosphopeptide abundance for each of the 30 samples in the dataset. The higher the value, the more abundant the phosphopeptide is in that sample.
* **Index labels:** Identifies which phosphopeptide that row is referring to. In the form of (“Protein ID\_SiteInProtein”). These are the features.
* **Column labels**: Each column holds data for a different patient sample with a unique ID. These are the observations.

**Output Data**

* **File Name**: ‘drug\_sensitivity.csv’
* **Data Description**: Contains data describing how sensitive each patient sample was to a drug. Five different drugs were tested for each patient sample. In this case, the lower the value, the more sensitive the sample was to that drug.
* **Index Labels**: Each row holds data for a different patient sample with a unique ID. These are the observations, and correspond to the columns of the input data.
* **Column Labels**: Each column contains drug sensitivity data for a single drug. Column name is in the form “TargetedKinase (drug name)”. Choose one of the drugs to use the output of the machine learning model.