**Dataset:** Shuttle dataset

**Number of instances:** 49,098

**Attributes Info:** The shuttle dataset contains 9 attributes all of which are numerical. The first one being time.

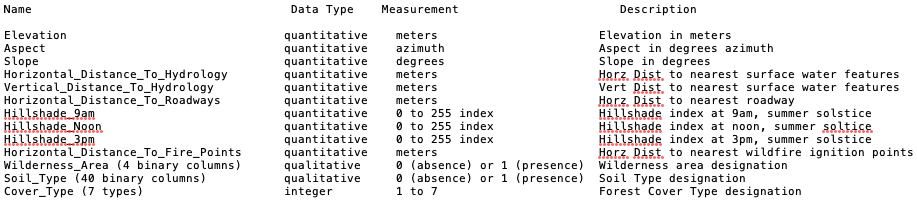
**Class distribution:** Approximately 80% of the data is normal.

**Comments:** It is a simple anomaly detection dataset with time stamp. However, it shuffle the data so the items are not in time consequence. We are not sure whether the time relation is important or not.

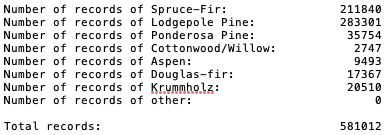
**Dataset:** ForestCover dataset

**Number of instances:** 581,012

**Attributes Info:**



**Class distribution:**



**Comments:** It is more like a classification dataset. No time stamp. I don’t think it suitable to our project.

**Dataset:** SMTP dataset

**Number of instances:** 95156

**Attributes Info:** This dataset is a subset of KDD CUP 99, which reduced to 4 attributes (service, duration, src\_bytes, dst\_bytes) as these attributes are regarded as the most basic attributes. Only “SMTP” in service is used so only 3 attributes in this dataset.

**Class distribution:** 0.03% anomaly rate

**Comments:** It is a simple sub-dataset of KDD CUP 99. We can use this dataset to test the performance on simple dataset. For other purpose, I suggest use the original KDD CUP 99 dataset.

**Dataset:** HTTP dataset

**Number of instances:** 567497

**Attributes Info:** This dataset is a subset of KDD CUP 99, which reduced to 4 attributes (service, duration, src\_bytes, dst\_bytes) as these attributes are regarded as the most basic attributes. Only “HTTP” in service is used so only 3 attributes in this dataset.

**Class distribution:** 0.4% anomaly rate

**Comments:** It is a simple sub-dataset of KDD CUP 99. We can use this dataset to test the performance on simple dataset. For other purpose, I suggest use the original KDD CUP 99 dataset.