****

**CSE 523 Machine Learning**

**Weekly Report Progress**

**(Week - 2)**

**Topic - Password strength checker**

**Group name - Predictors\_4.0**

| **NAME** | **ENROLLMENT NUMBER** |
| --- | --- |
| Zalak Shah | AU2040217 |
| Freyal Shah | AU2040228 |
| Freya Modi | AU2040222 |
| Avinash Raval | AU2040031 |

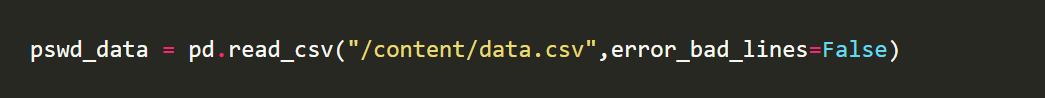
As researched in the previous week and as mentioned in the last weekly report, there are a variety of approaches to our topic of password strength detection. Of them, we are using logistic regression and decision tree approach to train our model. Firstly, we will look at the decision tree approach which uses a classifier and then we will look into the logistic regression approach which will also develop a classifier using the sigmoidal function and Bernoulli’s probability.

**Tasks performed this week:**

1. **Loading of the dataset:**

As we had already found the accurate data, the first task we did was to load the data in Google collab. For that we imported the pandas library and used its function read\_csv to read a comma separated value in a DataFrame.



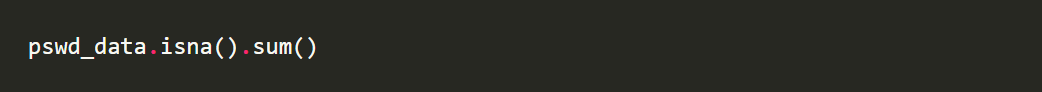


Writing “error\_bad\_lines=False” ensures that only complete data samples are loaded into the DataFrame.

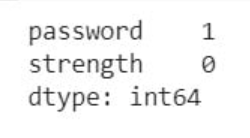
1. **Checking for missing values:**

Whenever datasets are imported using Pandas, there is always a possibility that many data samples have a missing value (Not Applicable-NA) in the DataFrame. This may happen when the data is not captured properly.

Code:



Output:



This shows that there is one password in the entire dataset which has a value NA and therefore it is a missing value.

1. **Removing the missing values:**

Missing values can lead to erroneous predictions and therefore it is very important to remove them. For this we simply use the dropna() functions of Pandas which removes all the rows with missing values.



This removes all the missing values and hence our dataset becomes complete and error free.

1. **Conversion of Dataset into an array:**

Since arrays are much easier to work with, we convert the DataFrames into arrays. We did this by first importing the NumPy library and using its function array() to get an array of data samples.



The “pswd” variable contains the dataset as an array of data samples.