

**CSE 523 Machine Learning**

**Weekly Report Progress**

**(Week - 1)**

**Topic - Password strength checker**

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

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

**Task performed this week:**

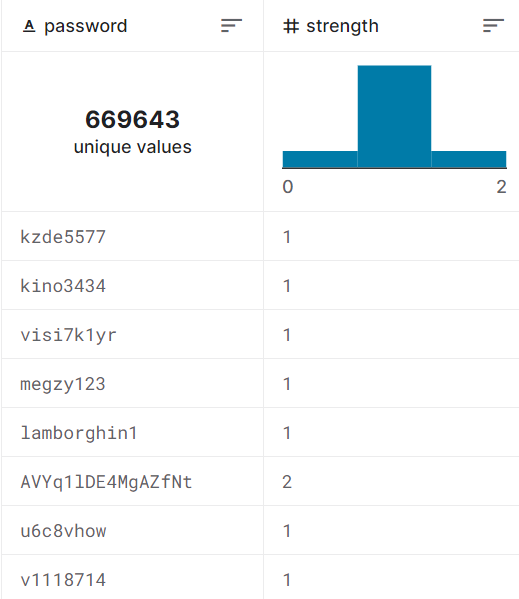
* **Finding relevant research report:**

The source of the research report we submitted was not credible. So, we went through a few reports and finalized the one. We are attaching the link below:

<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9936117&casa_token=KgFph5aUikYAAAAA:fBF68tIdDKWtM6mriH2j9rCfAMFIPh01qMH7UAWwKK3VwFYzvJfGXND7HYFb2UUwEo3egJlN6FLVwQ>

* **Finding Accurate Data from a reliable source:**

<https://www.kaggle.com/datasets/bhavikbb/password-strength-classifier-dataset>



After doing the data analysis, things which were concluded:

1. Our model would give output in the form of numbers ranging from 0-2, where 0 is the weak strength password and 2 is the strongest password.
2. The reason for the above is that we would be training our model whose labels are numbers ranging from 0-2.
3. Also, we checked for any data uncertainties in our data set. We looked for missing values and did the changes accordingly.

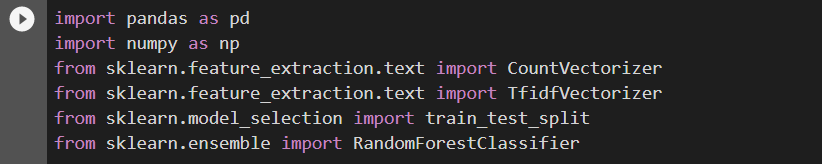
* **Understanding the topic:**

Passwords are exposed to various types of vulnerabilities due to the predictable patterns which human employs for setting passwords likely dictionary words, commonly used people and places names, keyboards patterns, date of birth, known phrases etc. and thereby, pave a route for online or offline muggers to intrude into the system by assuming the passwords. For guessing the passwords, various password cracking tools are available either online or offline and most of the tools can readily crack such accounts possessing passwords with weak strengths or passwords with common patterns. Therefore, the organizations should enforce unassailable strategies governing the use of strong passwords such that those vulnerabilities can be addressed.

* **Approaches of machine learning used in our topic:**

The following machine learning classifiers have been employed for password strength classification into different categories:

1. **Logistic Regression**: It is employed for the prediction purpose of the variable which is categorically dependent on the independent variables
2. **Decision Tree**: It is a tree based classifier where the internal node denotes features of a dataset, the branches signifies decision rules and the leaf node predicts the outcomes.
3. **Random Forest**: The concept of ensemble learning is implemented in this algorithm, which is used to solve a complex problem by a process of mixture of multiple classifiers.
4. **Naive Bayes theorem**: It is used for resolving text classification problems possessing a high-dimensional training dataset.
5. **XGBoost**: It is primarily implemented in prediction problems involving unstructured data like images, text, etc.
6. **Support Vector Machine**: The ultimate goal of the SVM is the creation of a decision boundary such that the n-dimensional space can be segregated into different classes such that it can easily adopt the new data point in the accurate class in future.
7. **Multilayer Perceptron**: In MLP, the multiple layers contain various units in an organized manner where each unit comprises general purpose, flexible, nonlinear models.

****

From datasets made up of formats like text and image, the **sklearn.feature\_extraction** module can be used to extract features in a format that machine learning algorithms can understand. Sklearn is a Python library possessing a number of data processing capabilities, including classification, clustering, and model selection.