Machine Learning-based Password Strength Checker

Project Overview:
For this project, the aim is to develop a Machine Learning-based password strength checker. The primary goal of this project is to assess the strength of passwords using machine learning techniques rather than traditional rule-based algorithms. We will utilize a dataset of passwords from the 000webhost leak, and our approach will be based on the Tf-idf scores of individual characters within the passwords. Additionally, we will use machine learning algorithms for classification to predict whether a password is weak, medium, or strong.

Project Description:
In this project, a dataset of passwords will be collected from the 000webhost leak available online. To determine password strength, we will utilize a tool called PARS, which integrates commercial password strength meters (Twitter, Microsoft, and Battle). The passwords will be classified into weak, medium, or strong based on the consensus of these three commercial algorithms.

Then the data will be preprocessed, splitting it into training (80%) and test (20%) sets. For feature extraction, Tf-idf scores will be used, treating each character within the passwords as a token. This approach allows us to consider the significance of individual characters in determining password strength.

Next, a machine learning algorithm will be implemented, specifically Logistic Regression with multi-class classification, to predict the strength of passwords. This algorithm will be trained on the training dataset and evaluated on the test dataset. To ensure statistical robustness and reduce overfitting, k-fold cross-validation will also be employed during the model training process if time permits.

In summary, the project aims to create a machine learning-based password strength checker that goes beyond rule-based algorithms. It combines the strengths of multiple existing algorithms and offers robust password assessment capabilities. While this project is an initial exploration, it has the potential to be further extended and refined into a comprehensive password strength checker.