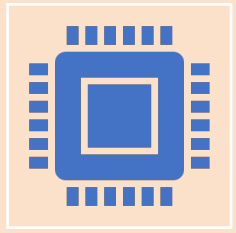
HYBRID PASSWORD STRENGTH ANALYSIS TOOL BASED ON MACHINE LEARNING

Supervisor:

Dr. Shaima Qureshi

INTRODUCTION

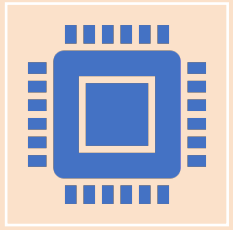
With the advent of digital age, passwords have become the most prevalent authentication scheme because of their superior versatility and deployability in comparison to other authentication mechanisms.



Unfortunately, humans still struggles with creating strong passwords that are hard to crack. Users in general end up creating passwords that contain predictable patterns or reuse the same password across multiple platforms which in turn reduces the work of hackers.

AIM

Since with passing of each year hackers become more creative with their attacks, therefore to counter them new pattern tests have to be added to the traditional password analyzers. This makes them computationally intensive.



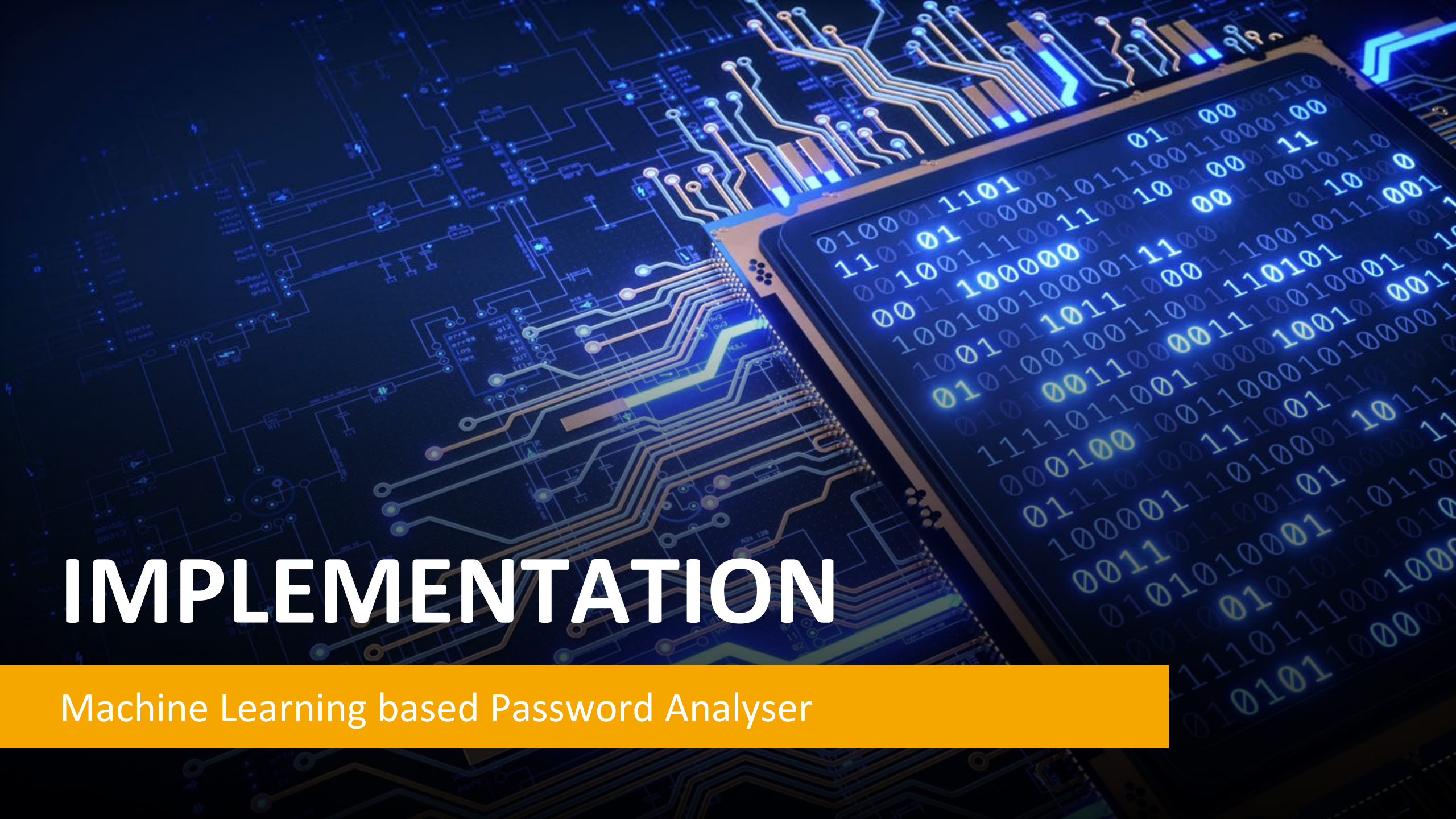
To explore the use of machine learning algorithms to reduce the ever increasing computational needs of traditional rule-based password analysis tools and to make them more flexible without compromising their accuracy.

Background and Motivation

While accurate traditional password analysers are highly computation intensive and inflexible.

Current Machine Learning based password analyzers produces classification results. Any attempts at regression based model overfits the data and produces bad prediction results.

We are trying to create a password analyzer that utilizes a machine learning component but gives password strength measure in terms of real numbers or integers.



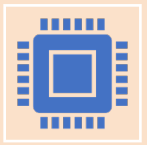
IMPLEMENTATION

Machine Learning based Password Analyser

Approach

A Machine Learning component utilizing classification algorithms.

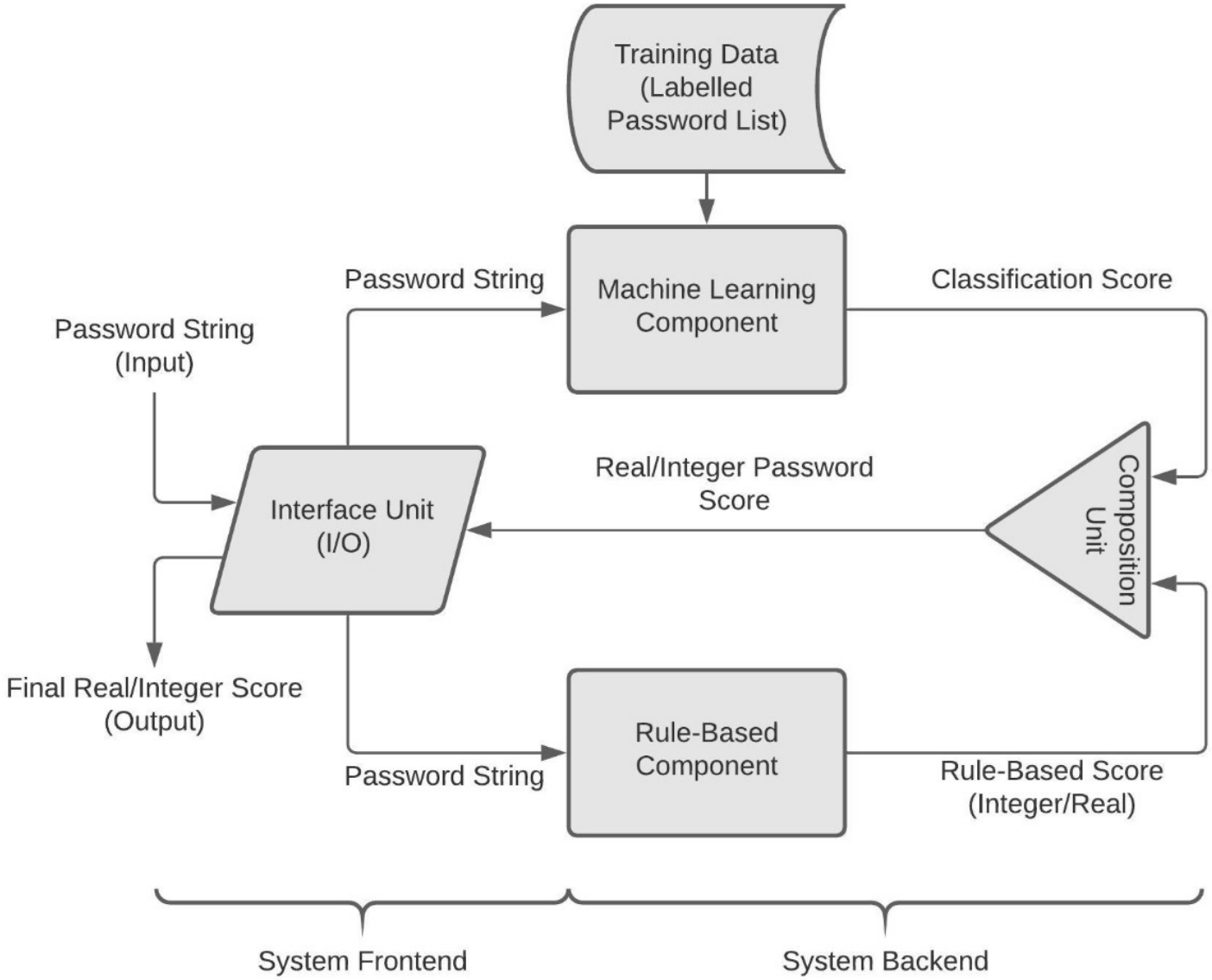
A light-weight Rule-based component using specific pattern testing algorithms to assign a real-value/integer score to a password.



A Composition/Integration unit that combines the scores of the aforementioned two components using a specific strategy to generate the final normalized score allocated to the password.

An Interface unit for password input and score output.

Block Diagram



1. Machine Learning Component

It consists of prediction models created by training various classification algorithms e.g., Logistic Regression, KNN, SVM, Naïve Bayes, Decision Tree etc. on a specific data set.

Different strategies are used by different algorithms based on some parameters.

1.1 Dataset Description

The data-sets we are using are leaked password lists from hashkiller, 000webhost and rockyou etc.

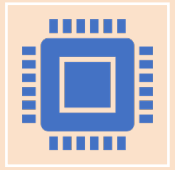
To classify these passwords in our data by current standards we have used a tool called PARS(Password Analysis and Research System) developed at Georgia Tech University.

Dataset Description Format

password	strength
kzde5577	1
kino3434	1
visi7k1yr	1
megzy123	1
lamborghin1	1
AVYq1lDE4MgAZfNt	2
u6c8vhow	1
v1118714	1
universe2908	1
as326159	1
asv5o9yu	1
612035180tok	1
jytifok873	1
WUt9IZzE0OQ7PkNE	2
jerusalem393	1
g067057895	1
52558000aaa	1
idofo673	1
6975038lp	1
sbl571017	1
elyass15@ajilent-ci	2
intel1	0
klara-tershina3H	2
czuodhj972	1
faranumar91	1
cigicigi123	1

1.2 Feature Matrix Modification

As of now if we were to use this data set to train our classification model, we would get bad prediction results.



Now instead of taking the entire string of password as a token, we take individual characters inside the password string as tokens.Thereby reducing its domain of values.

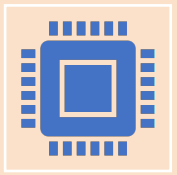
1.2 Feature Matrix Modification

After carrying out this modification a row in our dataset looks as follows:

Password								Strength
k	z	d	e	5	5	7	7	1

2. Rule Based Component

It is a light-weight unit that contains a collection of tests that makes the password immune from specialized attacks like Mask Attacks, Dictionary Attacks, Combinator attacks etc.



The password string is evaluated by these tests and then assigned a real/integer score based on the number and type of tests in which it gets the desired results.

3.Composition/Integration Unit

Training the model often takes the longest amount of time. Hence pickling the machine model can save us time to train the model once and reload it if and when it is required.

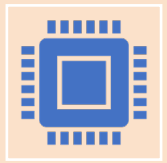
.It takes the classification score of a given password allocated by the machine learning unit and the integer/real score of the rule-based unit as its inputs.

It combines these scores using a specific strategy to generate the final real/integer normalised score assigned to the given password.

4. Interface Unit

The function of the interface unit is to take password inputs from users and output final scores assigned to the password.

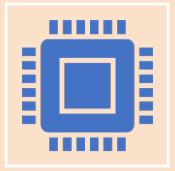
Additions assigns a score to the password based on some patterns and deduction reduces the score based on some patterns. These scores are used for the calculation of final score.



Hosting the analysis tool on a cloud platform Heroku.

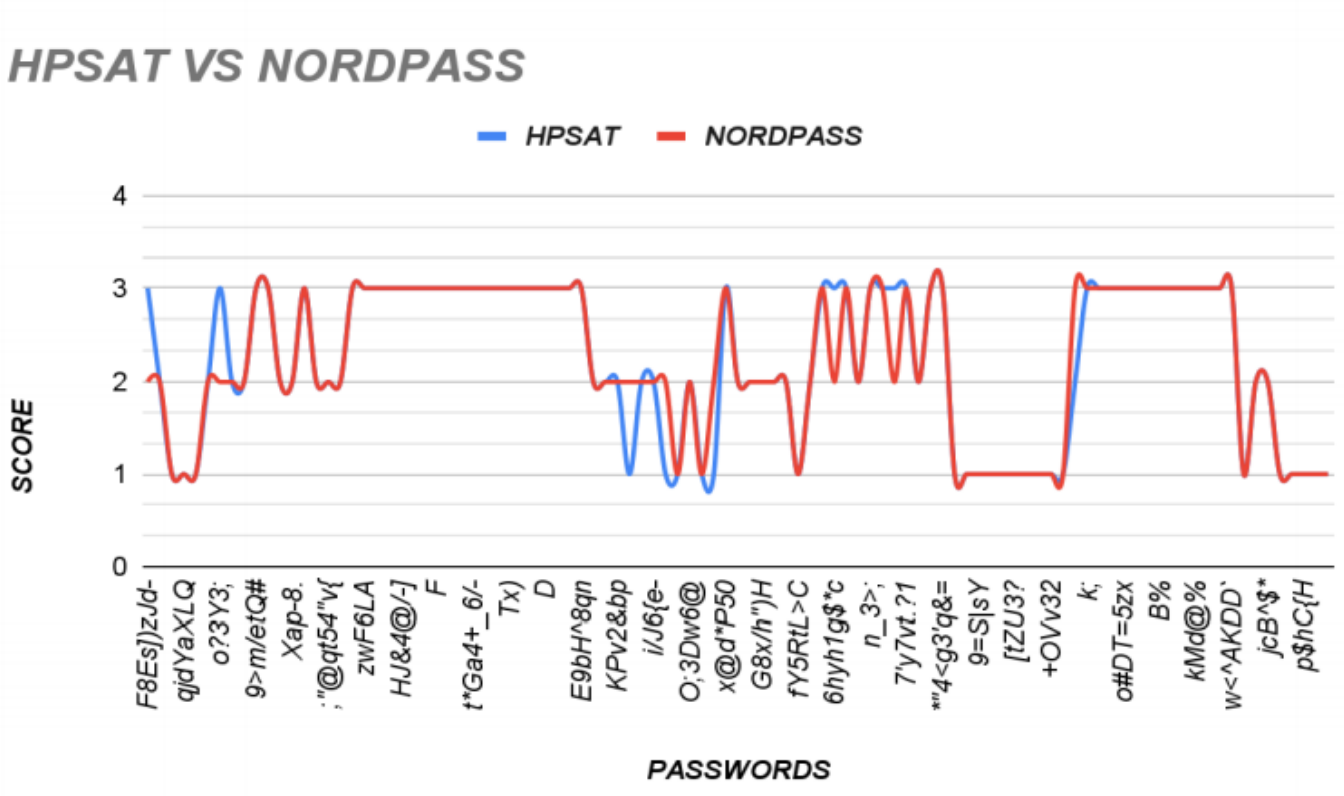
5.Project Result

At the end of this project, we will try to analyse the results of our project along three major axes:



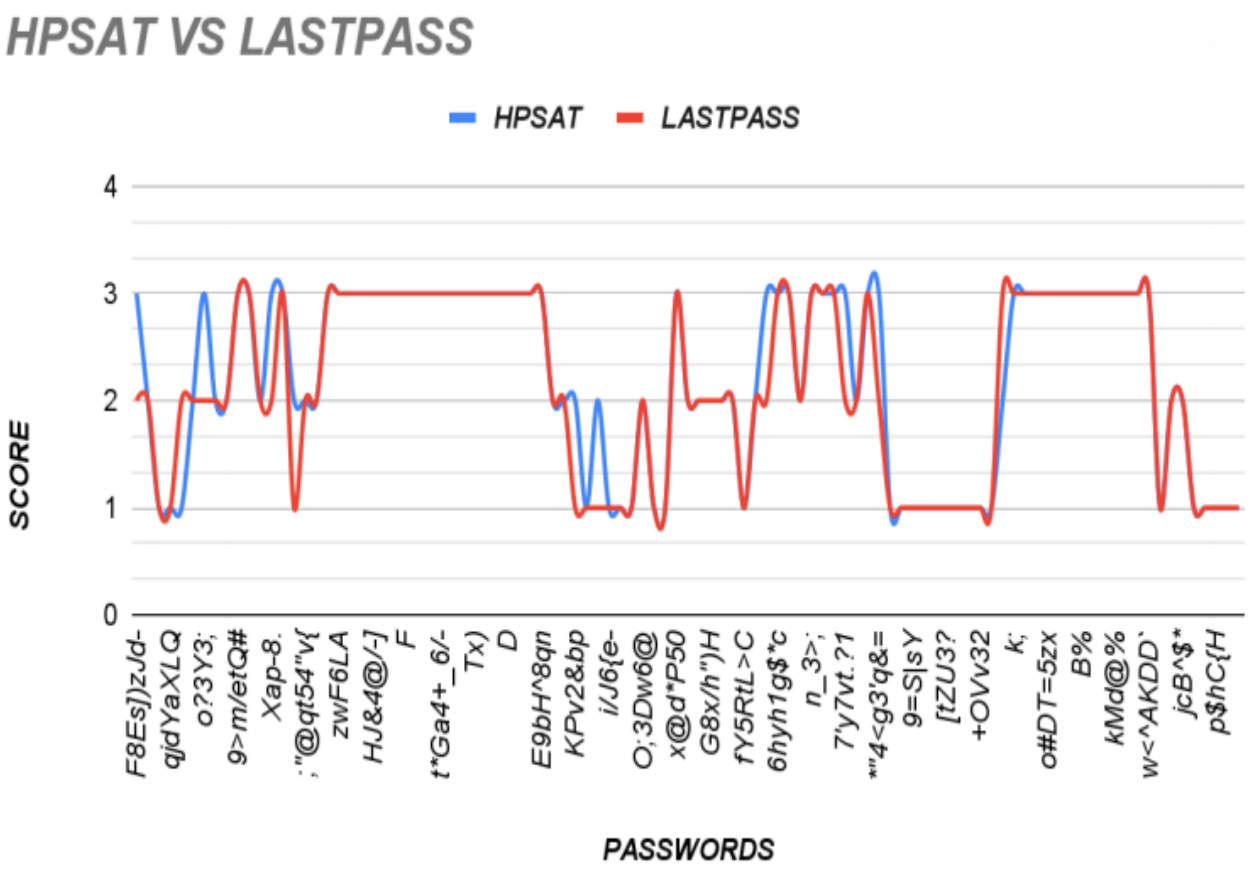
I. Accuracy ii. Flexibility iii. Processing Time

5.1 HPSAT vs NordPass



- Upon analysis, we found out that our password analyser had an accuracy of about 92% when compared against the NordPass password analysis tool.

5.2 HPSAT vs LastPass



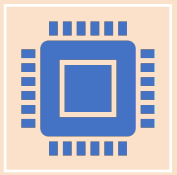
- Upon analysis, we found out that our password analyser had an accuracy of about 89% when compared against the NordPass password analysis tool.

5.3 Result

Thus overall accuracy of our password strength analysis tool based on these trials comes out to be 90.5%.

6. Flexibility

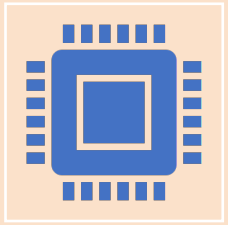
Our password analyser is far more flexible than any traditional password analyser.



Because all we need to do to modify our tool is to train the ML component with the updated knowledge data.

7. Processing Time

Since we have replaced most of the computation intensive pattern tests in traditional password analysers with a Machine Learning component which only carries out class based predictions while carrying out password testing.



This cuts short most of the processing time to evaluate the strength of passwords.

8. Future Scope

In the domain of this project there is an active chance of future work in terms of further increasing the accuracy of the analyser. This can be done by further experimenting with various classification and clustering techniques to create efficient machine learning components which gives better results on their own and rely less on rule-based component.

Project Timeline

	October	November	December	January	February	March	April	May	June
Project Research & Planning									
Machine Learning Component Development									
Rule-Based Component Development									
Composition Unit Development & Testing									
Interface Unit Develpment & Unit Testing									
Interface Integration									
Overall Project Testing & Debugging									

Legend: JAMYANG · AQUIF · COLLABORATIVE WORK

REFERENCES:

1. Ahmed, Faizan (2016). Machine Learning based Password Strength Classification. Available from World Wide Web: https://medium.com/@faizann20/machine-learning-based-password-strength-cl assification-7b2a3c84b1a6

2. E.g of PARS processed dataset. Available from World Wide Web: https://www.kaggle.com/bhavikbb/password-strength-classifier-dataset

3. Ji, S., Yang, S., Wang, T., Liu, C., Lee, W., & Beyah, R. (2015). PARS: A Uniform and Open-source Password Analysis and Research System. ACSAC 2015. Available from World Wide Web: https://www.semanticscholar.org/paper/PARS%3A-A-Uniform-and-Open-sourcePassword-Analysis-Ji-Yang/99b9bff925fdab9c4680f5667e7e6bc93cde2980

4. Mackay, W. (2016). How to Perform a Rule-Based Attack Using hashcat. Available from World Wide Web: https://www.4armed.com/blog/hashcat-rule-based-attack/

5. Todd, M. (2016). An Investigation of Machine Learning for Password Evaluation. Available from World Wide Web: https://www.semanticscholar.org/paper/An-Investigation-of-Machine-Learning-f or-Password-Todd/d40dee322446d8a3e45a622e1cf80f990976627c

6. Flask Documentation https://flask.palletsprojects.com/

7. Corey Schafer Flask Playlist on Youtube. https://www.youtube.com/watch?v=MwZwr5Tvyxo&list=PL-osiE80TeTs4UjLw5MM6OjgkjFeUxCYH

8. NordPass password strength checker. Available from World Wide Web: https://nordpass.com/secure-password/

THANKS!

End of Slides...

“People worry that computers will get too smart and take over the world, but the real problem is that they're too stupid and they've already taken over the world.”
— Pedro Domingos

Project url: lotusaquifhpsat.herokuapp.com

lotusjamyang@gmail.com

rezaaquif11@gmail.com

Presented by:
Jamyang Lotus
Aquif Reza Mir