INTRODUCTION

Password Strength Predictor Using Machine Learning and Artificial Intelligence

Mentored By :
Prof. D.S Yadav

Submitted By :
Aman Kumar Singh (1805210007)
Kishan Rana (1805210025)
Pushpa Devi (1805210040)

CONTENTS

1 Introduction

2 Motivation

3 Objectives

4 Components of Project

5 Model Architecture

6 Data Set Preparation

7 Algorithm Overview

8 Feature Selection

9 Decision Tree Diagram

10 Working Model Results

11 Tool and Technology

12 Challenges and Learnings

13 References

INTRODUCTION

Life these days has become largely dependent on passwords. A typical computer user may require passwords for many purposes such as logging in to computer accounts, retrieving email from servers, transferring funds, shopping online, accessing programs, databases, web sites.
Our goal is to create a machine learning model which can predict the strength of passwords so that weak passwords can be detected and avoided to secure our social and personal information on internet.

MOTIVATION

Cyber security and machine learning are one of the top trending tech stacks in today's world almost everything which is connected to internet somehow connects with these two domains of computer science technology.
Wheather we talk about our you tube , social media feed or doing online transactions everyday through Paytm, Gpay etc. we are somehow using features of cyber security and machine learning.
These two technology are core of this project so having a good understanding of these two will help us to understand and operate software industry in well manner.

objective

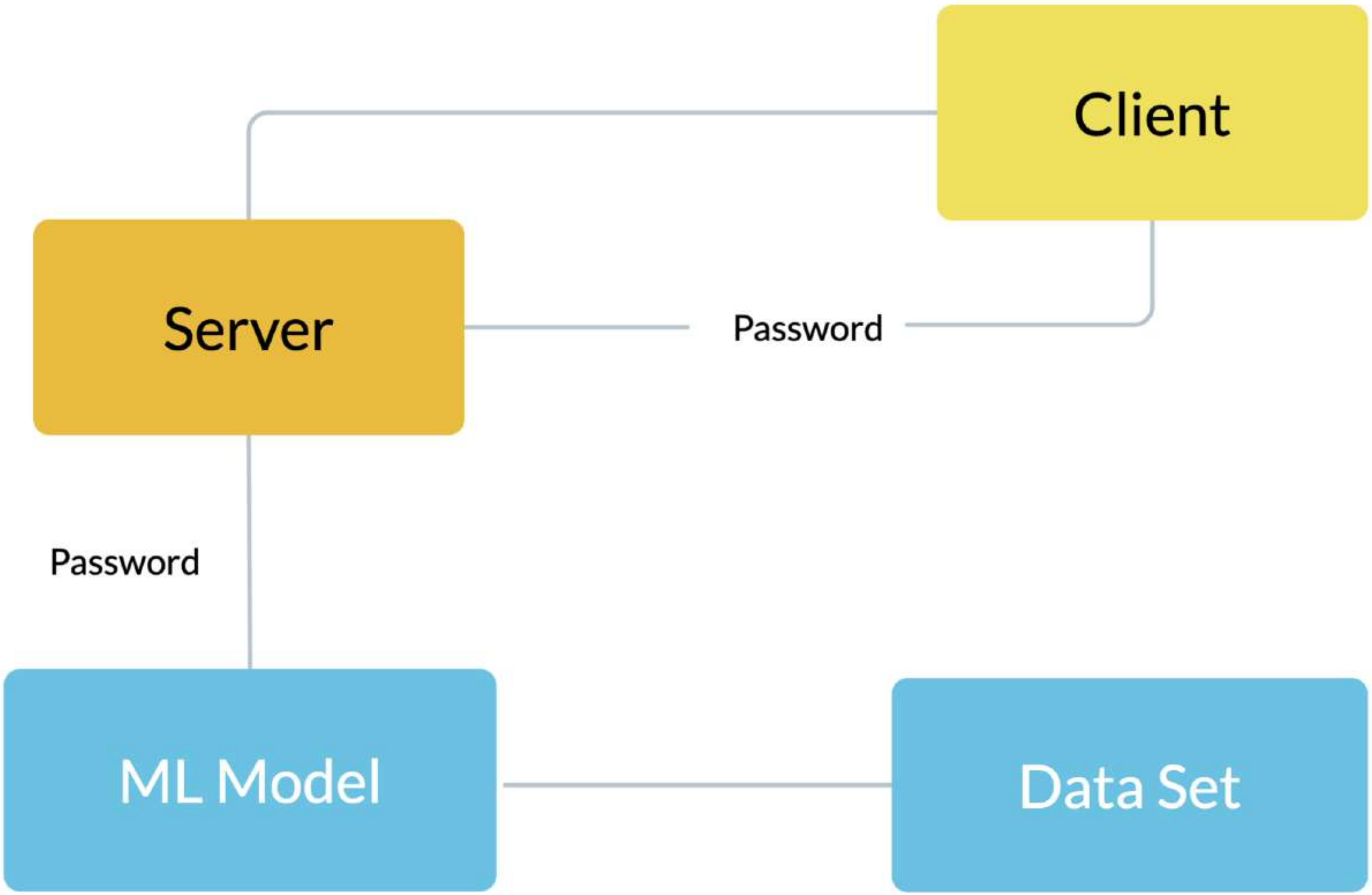
Our objective is to create a machine learning model which can predict strength of passwords with high accuracy.
And also we are objected to takeaway as much as learnings we can take from this project.

Components of Project

Our whole project is devided in below components

- Research
- Dataset preparation
- Writing Algorithm
- Training Model
- Testing Model
- Evaluation

SYSTEM ARCHITECTURE



Dataset Preparation

01 Data collection

Resources :

1. Open source data sets , Machine Learning dev communities
2. Writing scripts to create data sets

02 Data Cleaning

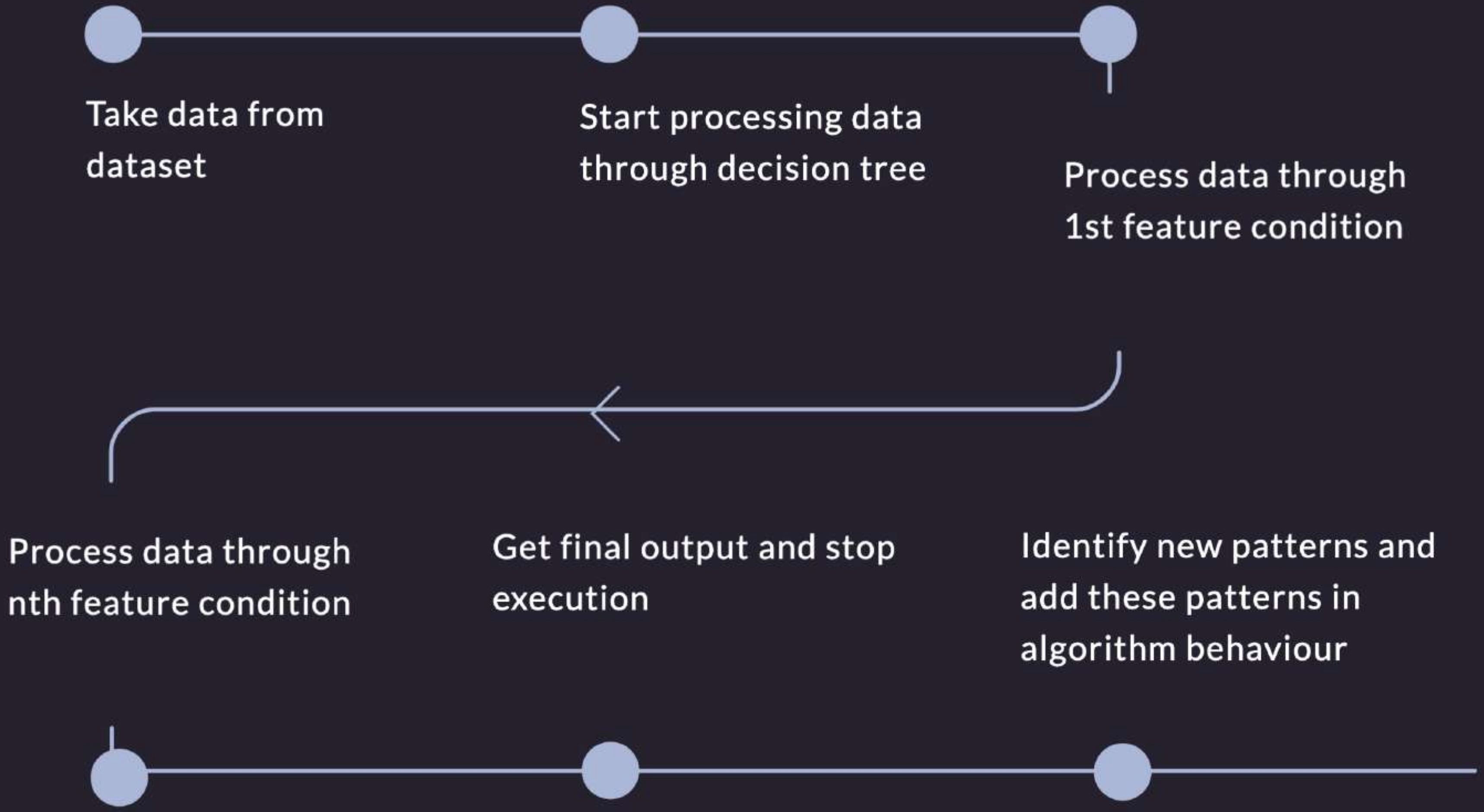
We have written python scripts to format and clean data to give the format to the data which our model can understand

03 Data Structure

For our project our data needs to have two attributes.

1. Password
2. Expected strength of password

Algorithm Overview



Take data from dataset

Start processing data through decision tree

Process data through 1st feature condition

Process data through nth feature condition

Get final output and stop execution

Identify new patterns and add these patterns in algorithm behaviour

FEATURE SELECTION

SELECTING CORRECT FEATURES IS ONE OF THE MOST CRITICAL STEPS IN ML MODEL TRAINING BECAUSE WHOLE BEHAVIOUR OF OUR MODEL DEPENDS ON HOW OUR FEATURES INFORCE OUR MODEL TO TAKE DECISIONS

01. Length of password

02. Number of Distinct characters in password

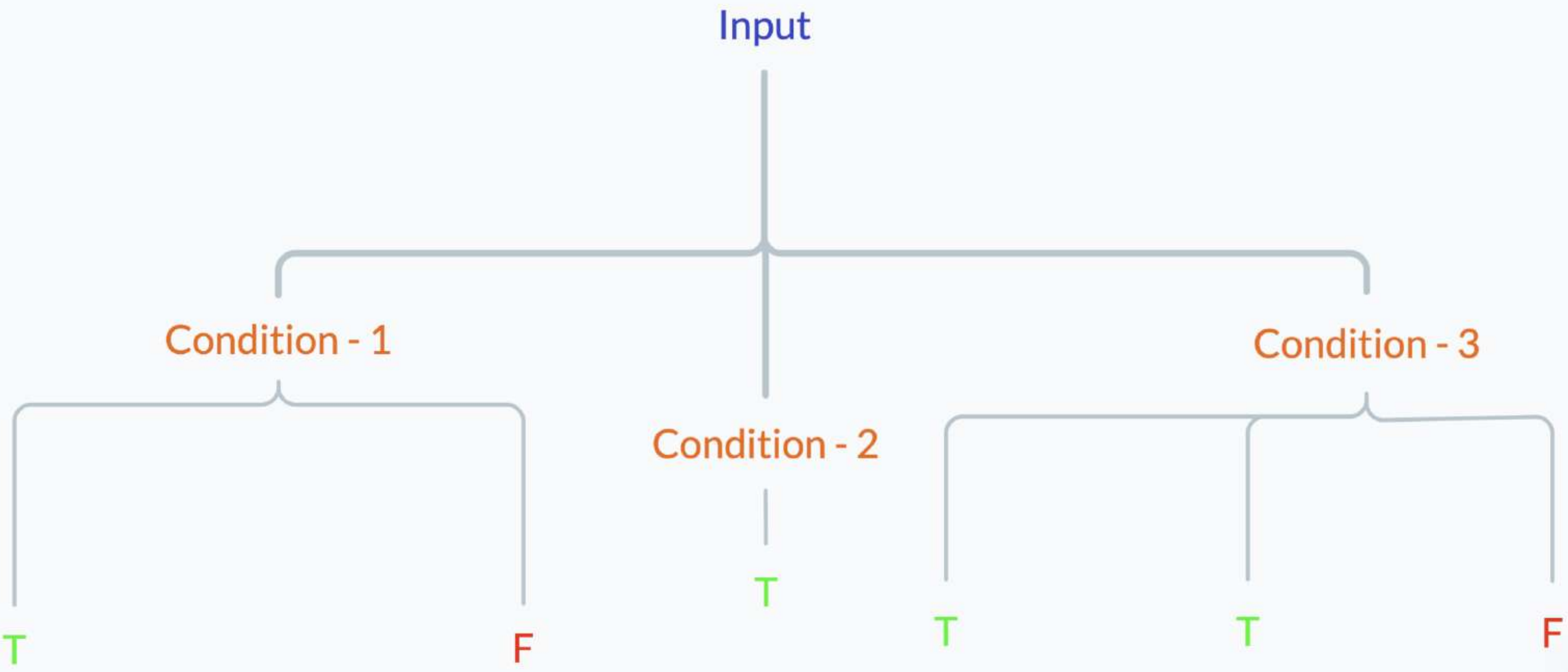
03. Number of numeric , alphabetic and other characters in password

04. Position of characters with respect to all other characters

05. Number of uppercase and lowercase alphabatic characters in password

06. Naive and commonly used passwords

DECISION TREE DIAGRAM



Input

Condition - 1

Condition - 2

Condition - 3

T

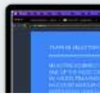
F

T

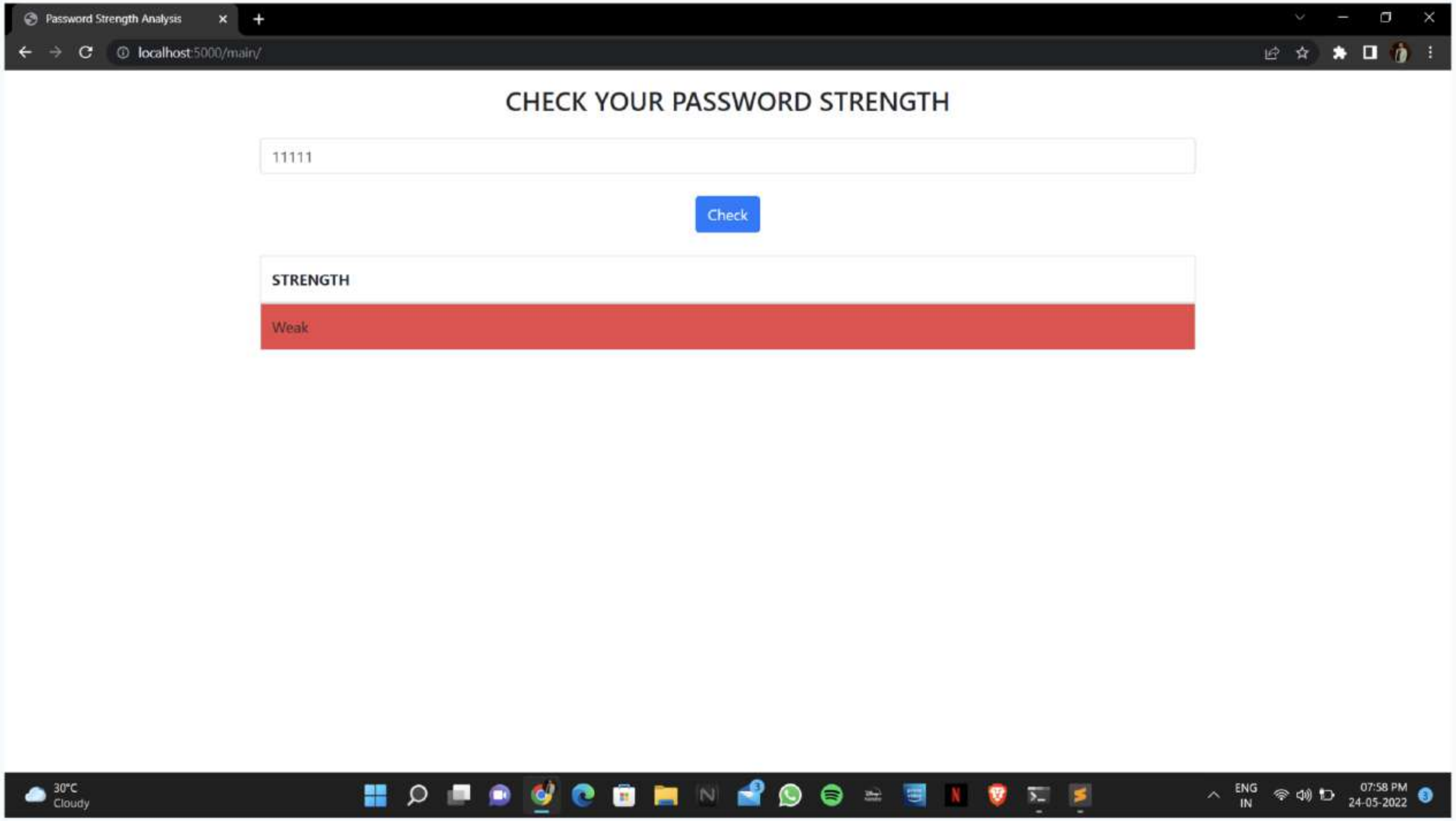
T

T

F



WORKING MODEL RESULTS : WEAK PASSWORD

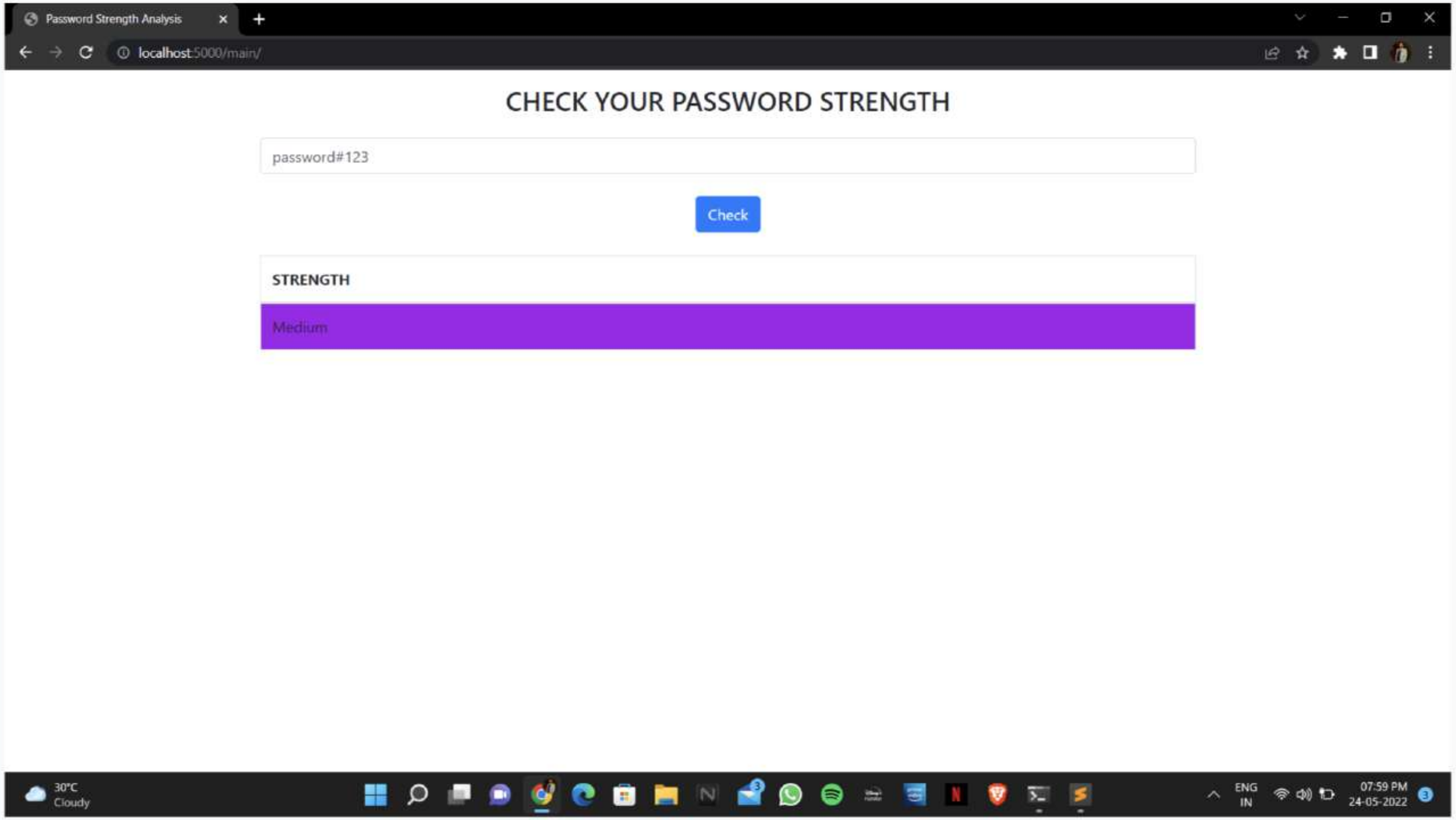


CHECK YOUR PASSWORD STRENGTH

11111

STRENGTH
Weak

WORKING MODEL RESULTS : MEDIUM PASSWORD

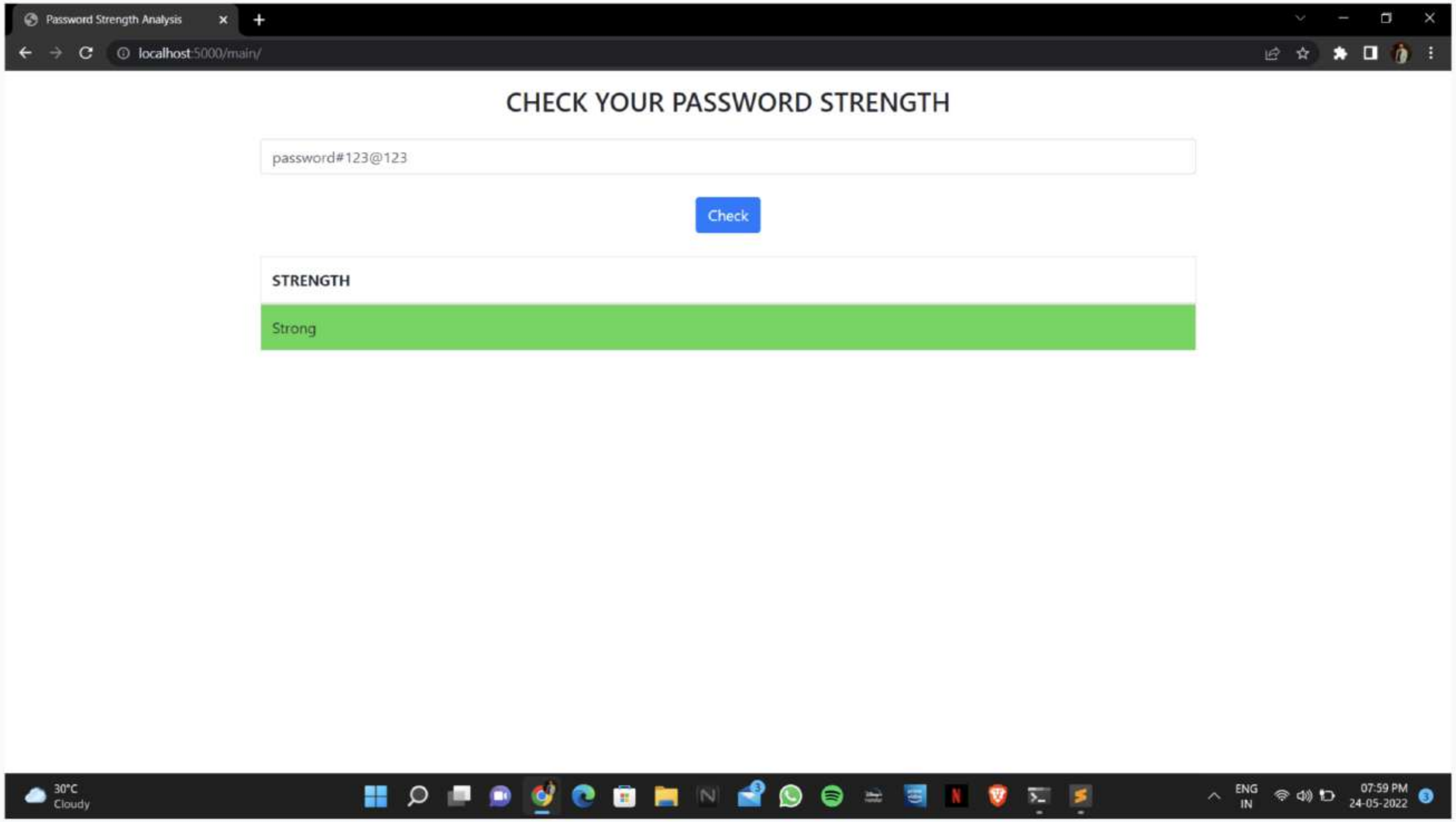


CHECK YOUR PASSWORD STRENGTH

password#123

STRENGTH
Medium

WORKING MODEL RESULTS : STRONG PASSWORD



CHECK YOUR PASSWORD STRENGTH

password#123@123

STRENGTH
Strong

Tools and Technology

Our whole project utilizes various tech stacks for various purposes which are listed below

Client

HTML
CSS
JS
Jinja2

Server

Flask
Python
Gunicorn

Model

Werkzeug
Numpy
Scipy
Scikit-Learn
Pandas

Dataset

CSV
TXT

CHALLENGES AND LEARNINGS

Challenges

Dataset for training model is tough to find and refactor

Building a model with a high accuracy will be quite challenging for us

Finding good feature patterns who will train model in better way for getting a good accuracy will be challenging task for us

Learnings

This project will be a good opportunity for us to get a good introduction of ML

Through this project we will get overview of cyber security

Through this project we will learn how ML is integrated in real life projects.

REFERENCES

[1] Giancarlo Ruffo, Francesco Bergadano, “EnFilter : A Password Enforcement and Filter Tool Based on Pattern Recognition Techniques”, Springer Berlin / Heidelberg, 1611-3349 (Online), Volume 3617/2005.

[2] John Shawe-Taylor, Nello Cristianini, “Support Vector Machines and other kernel-based learning methods”, 2000, Cambridge University Press, UK.

[3] F.Bergadano, B.Crispo, G.Ruffo, “Proactive password checking with decision trees”,Proc. of the 4 th ACM conference on computer and communications security, Zurich, Switzerland, 1997, pp 67-77.

[4] Soman K.P, Loganathan R, Ajay V, “ Machine Learning with SVM and other Kernel Methods”, 2009, PHI, India.

[5] https://en.wikipedia.org/wiki/Password_strength

[6] https://en.wikipedia.org/wiki/Brute-force_attack

[7] https://it.ufl.edu/it-policies

[8] https://medium.com/rangeforce/password-cracking-6d9612915f03

CHALLENGES AND LEARNINGS

Challenges

THANK YOU

REFERENCES