



SHRI VISHNU ENGINEERING COLLEGE FOR WOMEN

DEPARTMENT OF CSE-CYBERSECURITY
CS-12

PHISHX

D E T E C T . P R O T E C T . P R E V E N T

Presented By:

A.Jhansi - 23B01A4601

A.Aishwarya - 23B01A4605

K.Amrutha Valli - 23B01A4631

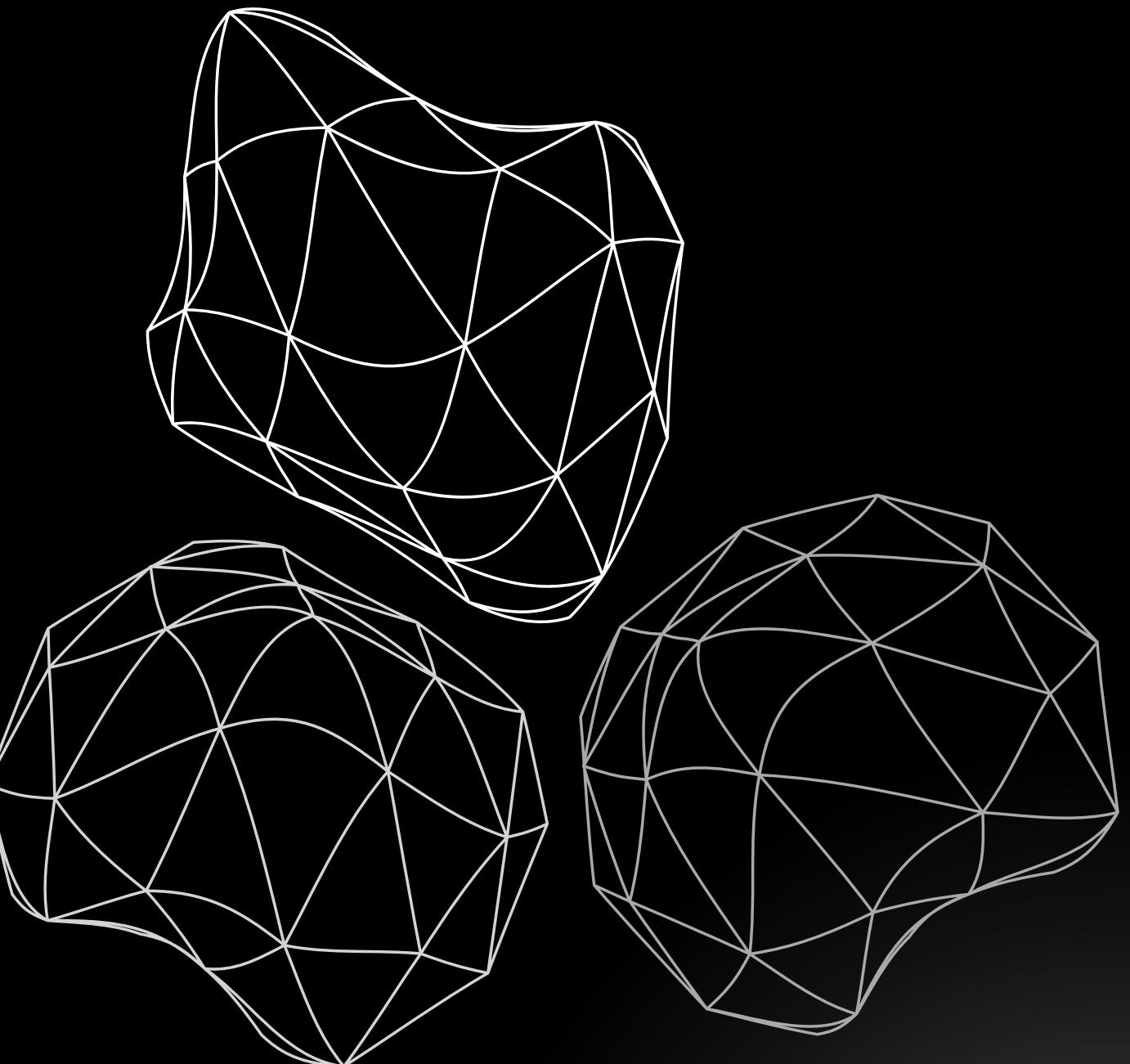
N.Nikhitha - 23B01A4642

V.Deepika - 23B01A4660

Guided By: Mrs.N.Durga
Asst.Professor

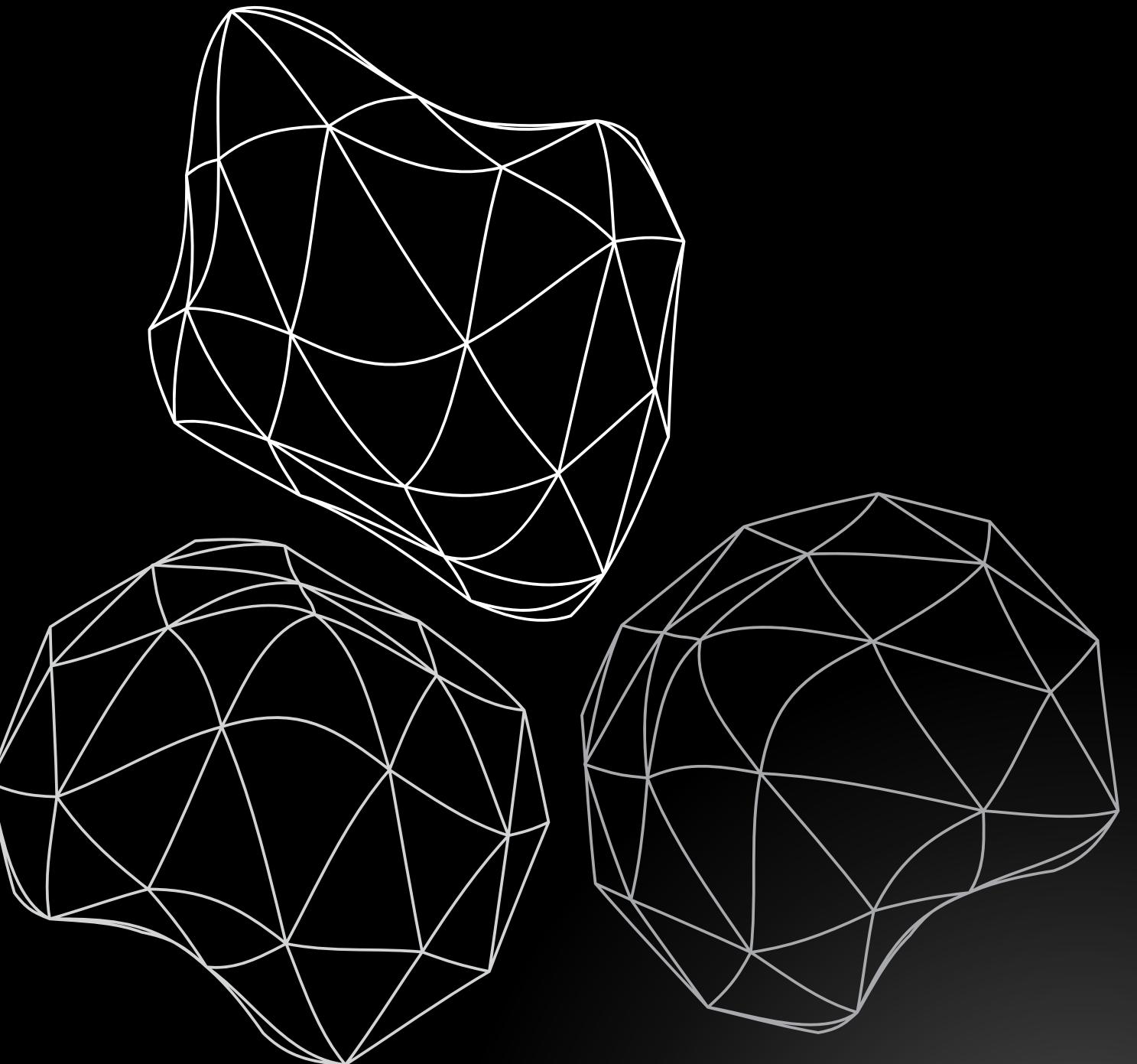
INTRODUCTION

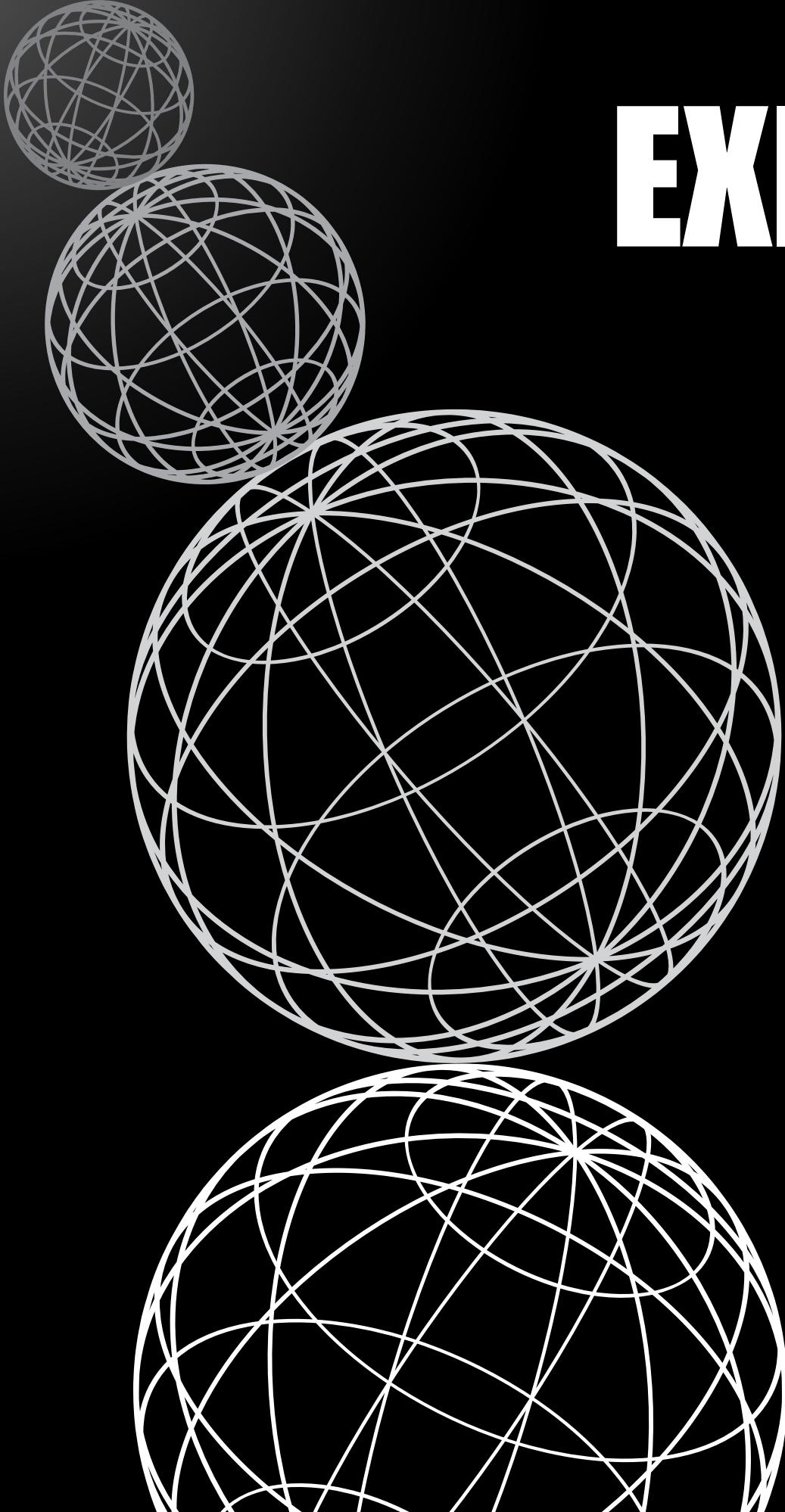
- Emails are vital for communication but are often targeted by phishing attacks.
- These attacks aim to steal sensitive information like passwords and credit card details.
- PhishX uses machine learning to detect and block such phishing threats.
- It analyzes email content, structure, and patterns to identify malicious messages.
- This helps distinguish between legitimate and harmful emails.
- The system enhances security and keeps users' inboxes safe from cyber threats.



ABSTRACT

Our project is about building a system to detect phishing emails. We used Natural Language Processing (NLP) to clean and process the email text. Two machine learning models were trained – SVC and Random Forest. We tested these models using a dataset of phishing and legitimate emails. The models showed good accuracy in spotting phishing emails. This shows how ML can help improve email security and protect users from scams.





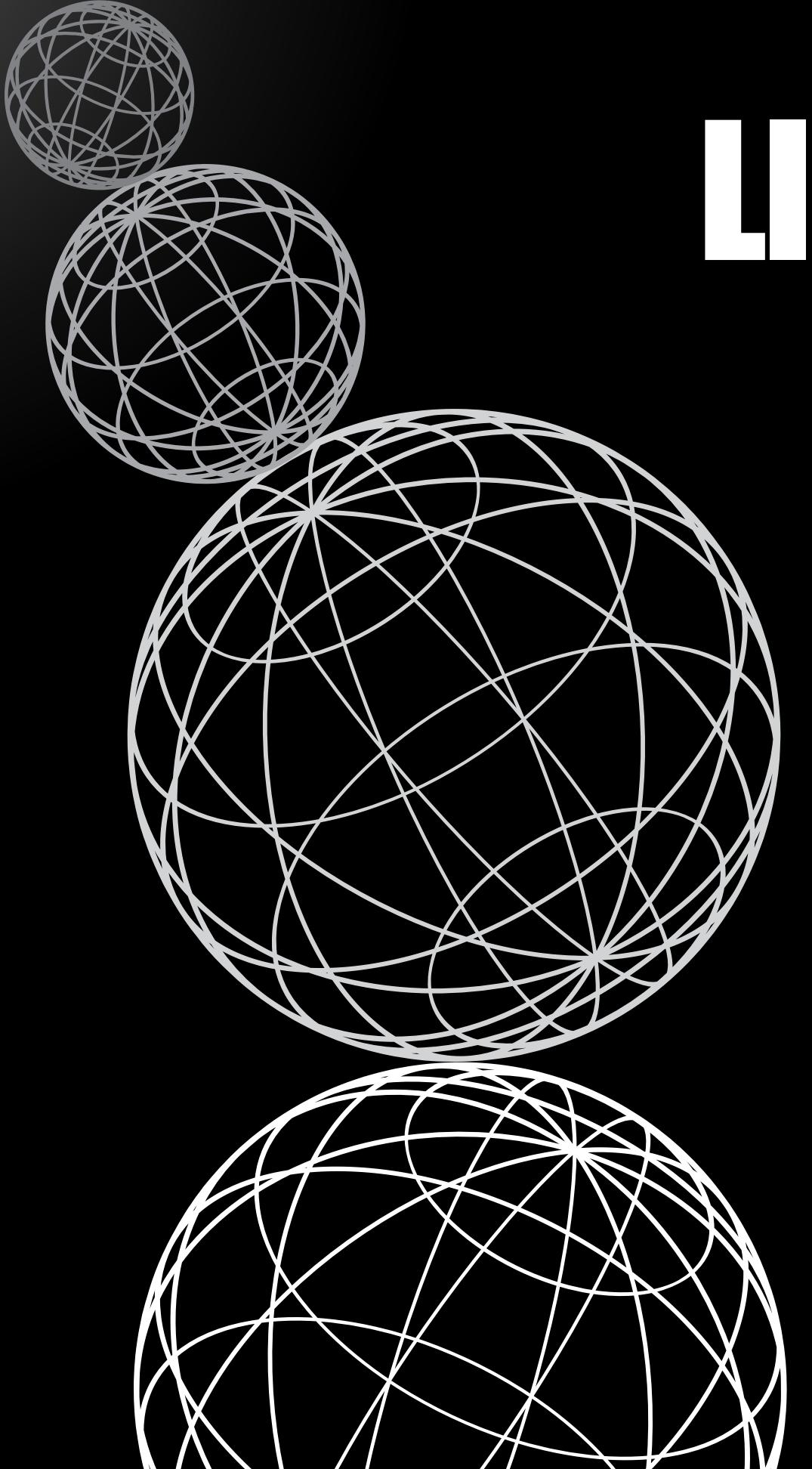
EXISTING SOLUTIONS

Proofpoint Email Protection

- Protects company inboxes from spam, phishing & malware
- Ideal for securing enterprise communication at scale

Zoho Mail

- Filters junk and spam for clean, focused inboxes
- Supports secure business email with custom domains
- Perfect for startups, freelancers, and small teams



LIMITATIONS

- Proofpoint Limitations
Expensive, complex setup, meant for big enterprises, can block legit emails, not user-friendly.
- Zoho Mail Limitations
Spam filter isn't perfect, advanced features cost extra, admin tools tricky, UI can lag, slow free-user support.

TECHNOLOGY STACK

python

The main programming language used for building and training ML model

NLTK

NLTK(Natural Language Tool Kit) is a python library that helps process text by cleaning and extracting key patterns for phishing detection

SCIKIT-LEARN

A library that provides ready to use machine learning algorithms like SVC & random forest for classification

DESIGN

Data Collection : Gather necessary datasets for model training.

Preprocessing : Clean and transform raw data for analysis.

Data Splitting : Divide data into training and testing sets.

Model Training : Train both Random Forest and SVC models.

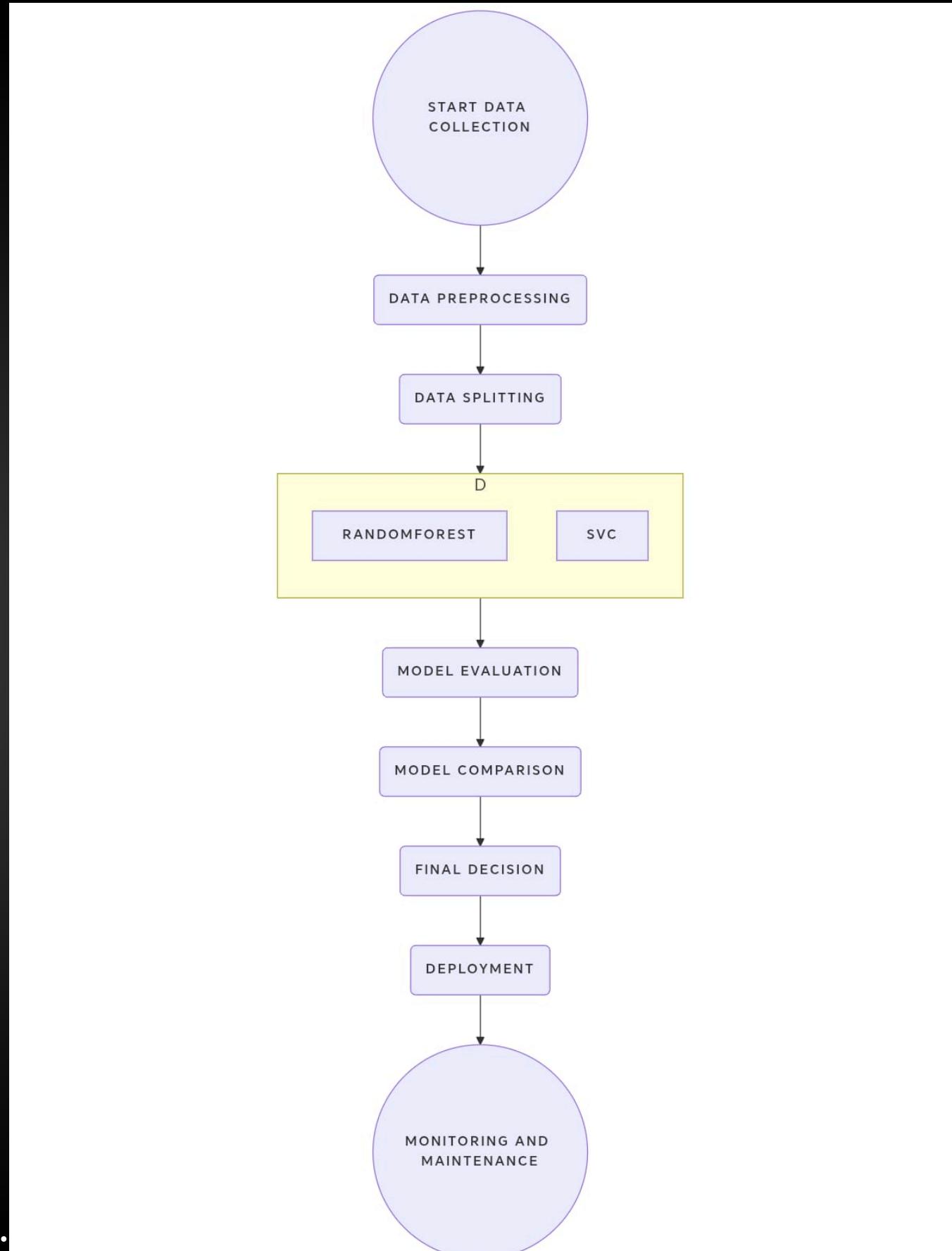
Evaluation : Assess model performance using key metrics.

Comparison : Compare models to choose the best one.

Final Decision : Select the most effective model.

Deployment : Launch the chosen model into production.

Monitoring : Continuously track and maintain model performance.



Run Terminal Help ← → 🔍 SMD

index.html X spam.tsv bg.jpg # Style.css

```
PhishX > templates > index.html > html > body > section > header
1  <!DOCTYPE html>
2  <html lang="en">
3  {% load static %}
4  <head>
5      <meta charset="UTF-8">
6      <meta name="viewport" content="width=device-width, initial-scale=1.0">
7      <meta http-equiv="X-UA-Compatible" content="ie=edge">
8      <link rel="stylesheet" href="{% static 'Css/bootstrap.min.css' %}">
9      <link rel="stylesheet" href="{% static 'Css/Style.css' %}">
10     <link rel="stylesheet" href="{% static 'Css/radio.css' %}">
11     <title>PhishX - Detect Phishing Smarter</title>
12 </head>
13
14 <body>
15     <section>
16         <header>
17             <nav class="navbar">
18                 <div><a href="#" class="brand">PHISHX</a></div>
19                 <ul>
20                     <li><a href="#" class="active">Web App</a></li>
21                 </ul>
22             </nav>
23         </header>
24
25         <div class="fluid-container content">
26             <div class="row m-auto">
27                 <div class=" p-5 m-0 form-content align-right">
28                     <h1>Welcome to PhishX</h1>
29                     <p>Analyze suspicious text and detect phishing attempts in real time.</p>
30                 </div>
31
32                 <div class="col-md-6 p-5 m-0 form-content">
33                     <form method="post" action="check" class="main-form">
34                         {% csrf_token %}
35                         <textarea name="rawdata" placeholder="Enter your suspicious message here..." id="rawdata"></textarea>
36                         <div class="algo-choice">
37                             <label class="radiocontainer">Algo - 1
38                                 <input type="radio" name="algo" value="Algo-1">
39                                 <span class="checkmark"></span>
40                             </label>
41                             <label class="radiocontainer">Algo - 2
42                                 <input type="radio" name="algo" value="Algo-2">
```

Ln 16, Col 17 Space

Terminal Help ← → 🔍 SMD

index.html mySVCModel.pkl phishx.py 9+ spam.tsv # Style.css

```
C: > Users > mrudu > OneDrive > Desktop > mini 2 > phishx > phishx.py ...
1 import numpy as np
2 import pandas as pd
3 import joblib
4 import matplotlib.pyplot as plt
5 from sklearn.model_selection import train_test_split
6 from sklearn.ensemble import RandomForestClassifier
7 from sklearn.svm import SVC
8 from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
9 from sklearn.pipeline import Pipeline
10 from sklearn.feature_extraction.text import TfidfVectorizer
11
12 df = pd.read_csv('spam.tsv', sep='\t')
13
14
15 hamDf = df[df['label'] == "ham"]
16 spamDf = df[df['label'] == "spam"]
17
18 hamDf = hamDf.sample(spamDf.shape[0])
19
20 finalDf = pd.concat([hamDf, spamDf], ignore_index = True)
21
22 X_train, X_test, Y_train, Y_test = train_test_split(finalDf['message'], finalDf['label'], test_size = 0.2, random_state=42)
23
24 #model = Pipeline([('tfidf', TfidfVectorizer()), ('clf', RandomForestClassifier(n_estimators=100, n_jobs=-1))])
25
26 model = Pipeline([('tfidf', TfidfVectorizer()), ('model', SVC(C=1000, gamma='auto'))])
27 model.fit(X_train, Y_train)
28 Y_pred = model.predict(X_test)
29
30 joblib.dump(model, "mySVCModel.pkl")
31
32 print(accuracy_score(Y_test, Y_pred))
33
```

WELCOME TO PHISHX

Analyze suspicious text and detect phishing attempts in real time.

Enter your suspicious message here...

Algo - 1

Algo - 2

Analyze

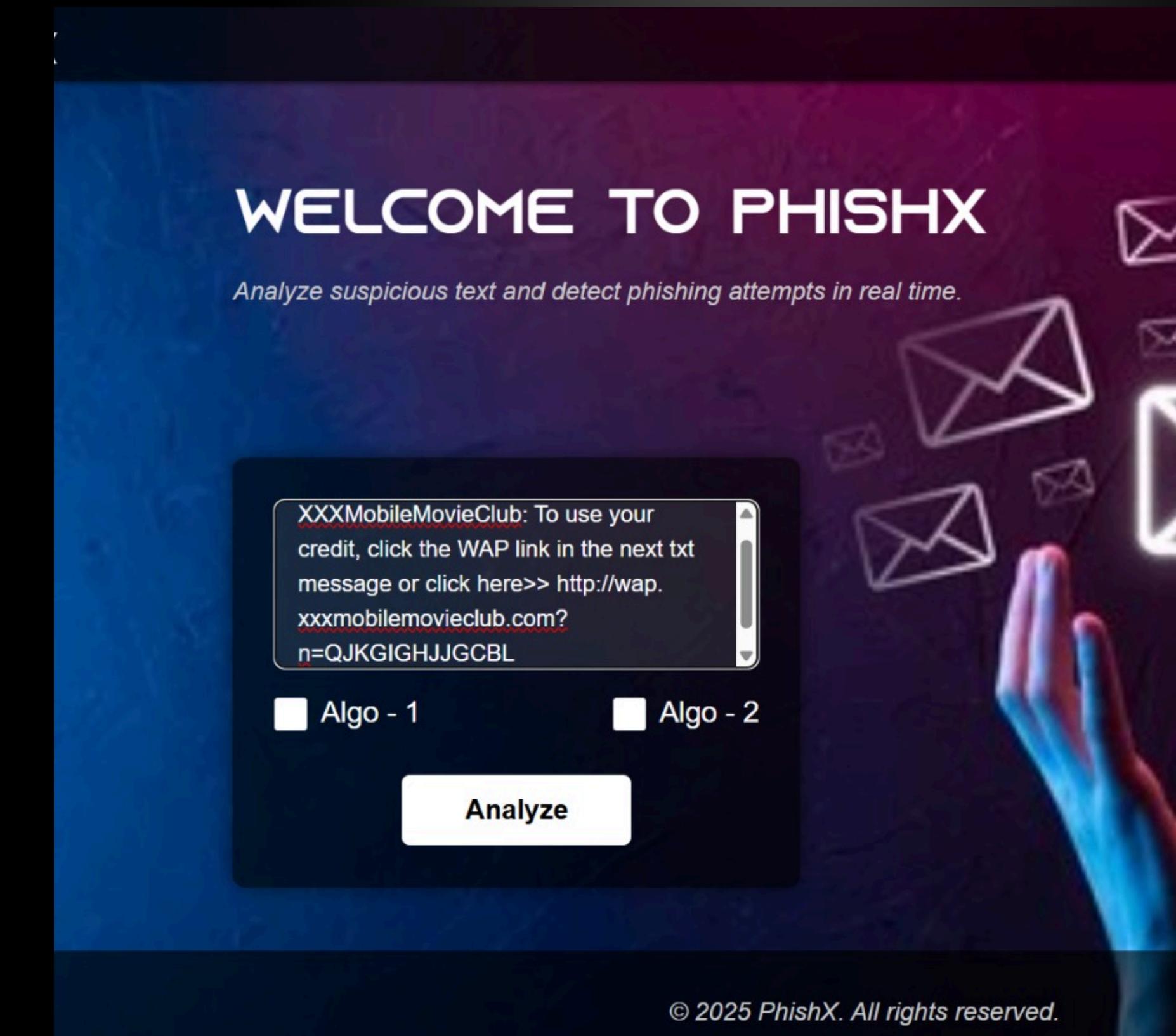


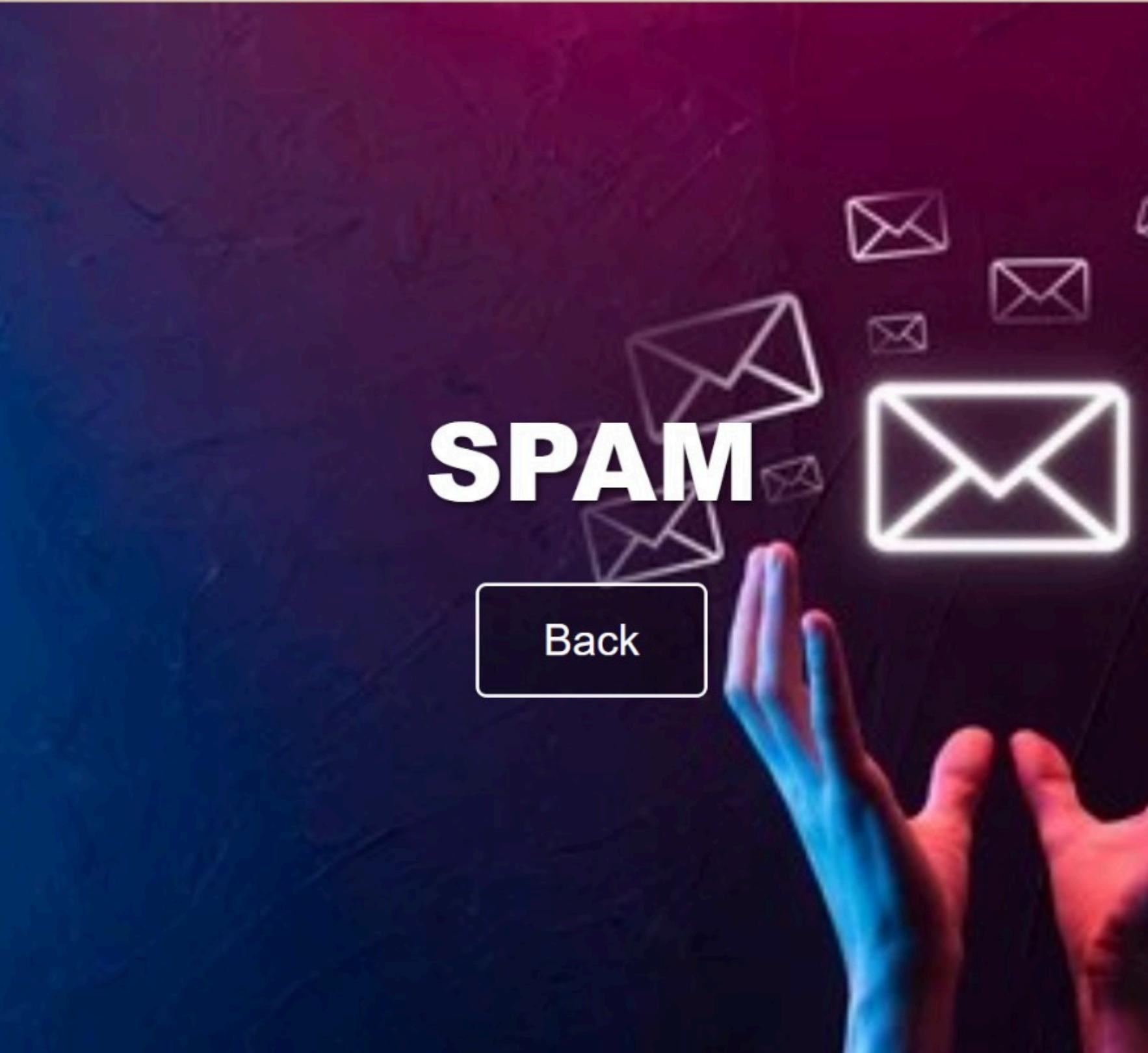
sun Terminal Help ⌘ ⌘ SMD

index.html phishx.py 9+ spam.tsv # Style.css

```
C: > Users > mrudu > OneDrive > Desktop > mini 2 > phishx > spam.tsv
1 label message length punct
2 ham Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat... 1
3 ham Ok lar... Joking wif u oni... 29 6
4 spam Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question
5 ham U dun say so early hor... U c already then say... 49 6
6 ham Nah I don't think he goes to usf, he lives around here though 61 2
7 spam FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb
8 ham Even my brother is not like to speak with me. They treat me like aids patient. 77 2
9 ham As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertune for all ca
10 spam WINNER!! As a valued network customer you have been selected to receivea £900 prize reward! To claim call 090
11 spam Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free!
12 ham I'm gonna be home soon and i don't want to talk about this stuff anymore tonight, k? I've cried enough today. 1
13 spam SIX chances to win CASH! From 100 to 20,000 pounds txt> CSH11 and send to 87575. Cost 150p/day, 6days, 16+ Ts
14 spam URGENT! You have won a 1 week FREE membership in our £100,000 Prize Jackpot! Txt the word: CLAIM to No: 81010
15 ham I've been searching for the right words to thank you for this breather. I promise i wont take your help for grant
16 ham I HAVE A DATE ON SUNDAY WITH WILL!! 35 2
17 spam XXXMobileMovieClub: To use your credit, click the WAP link in the next txt message or click here>> http://wap
18 ham Oh k...i'm watching here:) 26 6
19 ham Eh u remember how 2 spell his name... Yes i did. He v naughty make until i v wet. 81 5
20 ham Fine if that's the way u feel. That's the way its gotta b 56 1
21 spam England v Macedonia - dont miss the goals/team news. Txt ur national team to 87077 eg ENGLAND to 87077 Try:WA
22 ham Is that seriously how you spell his name? 41 1
23 ham I'm going to try for 2 months ha ha only joking 47 0
24 ham So ü pay first lar... Then when is da stock comin... 52 6
25 ham Aft i finish my lunch then i go str down lor. Ard 3 smth lor. U finish ur lunch already? 88 3
26 ham FFFFFFFF. Alright no way I can meet up with you sooner? 57 2
27 ham Just forced myself to eat a slice. I'm really not hungry tho. This sucks. Mark is getting worried. He knows I'm s
28 ham Lol your always so convincing. 30 1
29 ham Did you catch the bus ? Are you frying an egg ? Did you make a tea? Are you eating your mom's left over dinner ?
30 ham I'm back & we're packing the car now, I'll let you know if there's room 75 7
31 ham Ahhh. Work. I vaguely remember that! What does it feel like? Lol 64 4
32 ham Wait that's still not all that clear, were you not sure about me being sarcastic or that that's why x doesn't wan
33 ham Yeah he got in at 2 and was v apologetic. n had fallen out and she was actin like spoilt child and he got caught
34 ham K tell me anything about you. 29 1
35 ham For fear of fainting with the of all that housework you just did? Quick have a cuppa 84 1
36 spam Thanks for your subscription to Ringtone UK your mobile will be charged £5/month Please confirm by replying Y
37 ham Yup... Ok i go home look at the timings then i msg ü again... Xuhui going to learn on 2nd may too but her lesson
38 ham Oops, I'll let you know when my roommate's done 47 3
39 ham I see the letter B on my car 28 0
40 ham Anything lor... U decide... 27 6
41 ham Hello! How's you and how did saturday go? I was just texting to see if you'd decided to do anything tomo. Not tha
42 ham Pls go ahead with watts. I just wanted to be sure. Do have a great weekend. Abiola 82 3
```

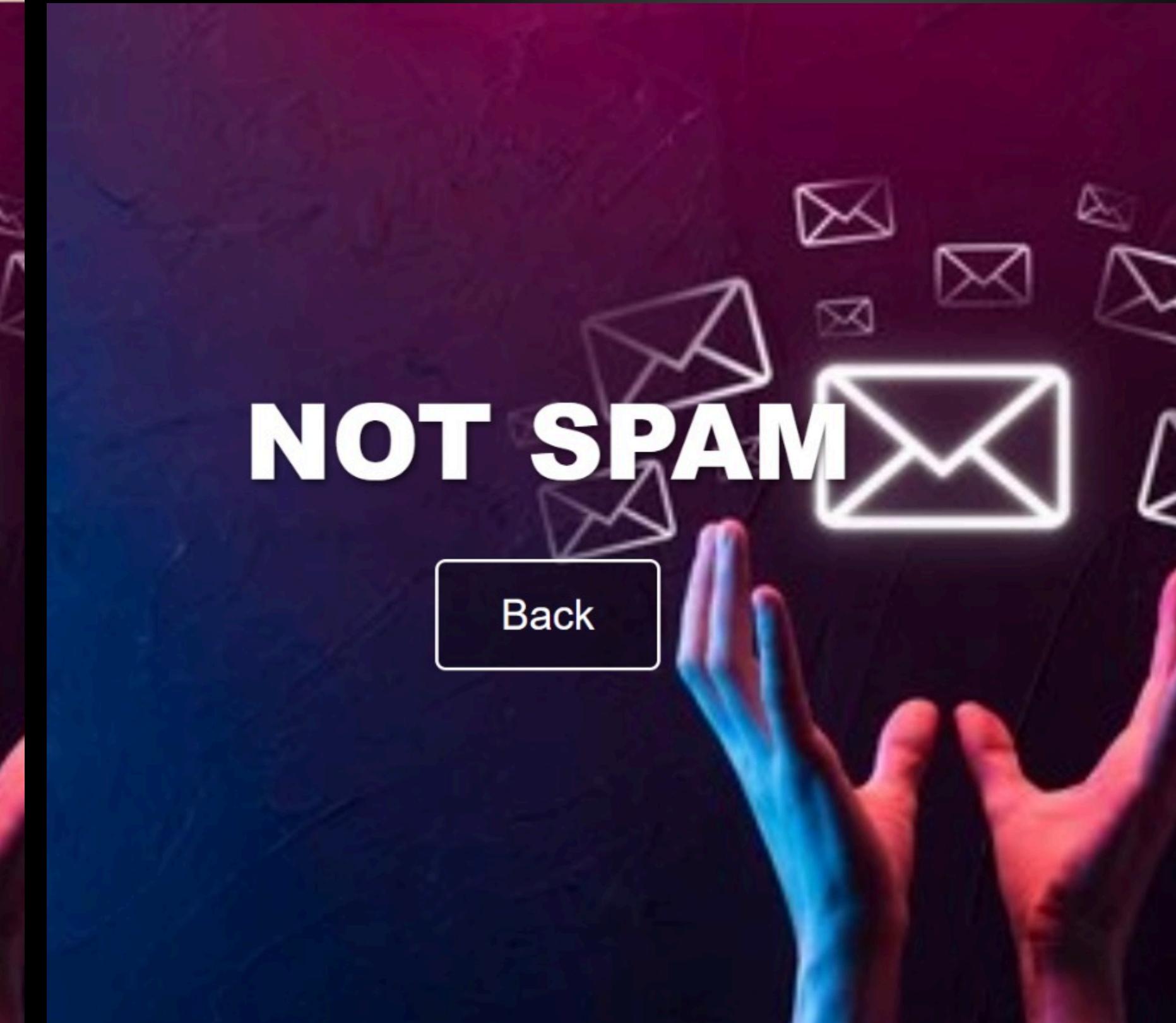
In 237, Col 166 (157 selected) Spaces:





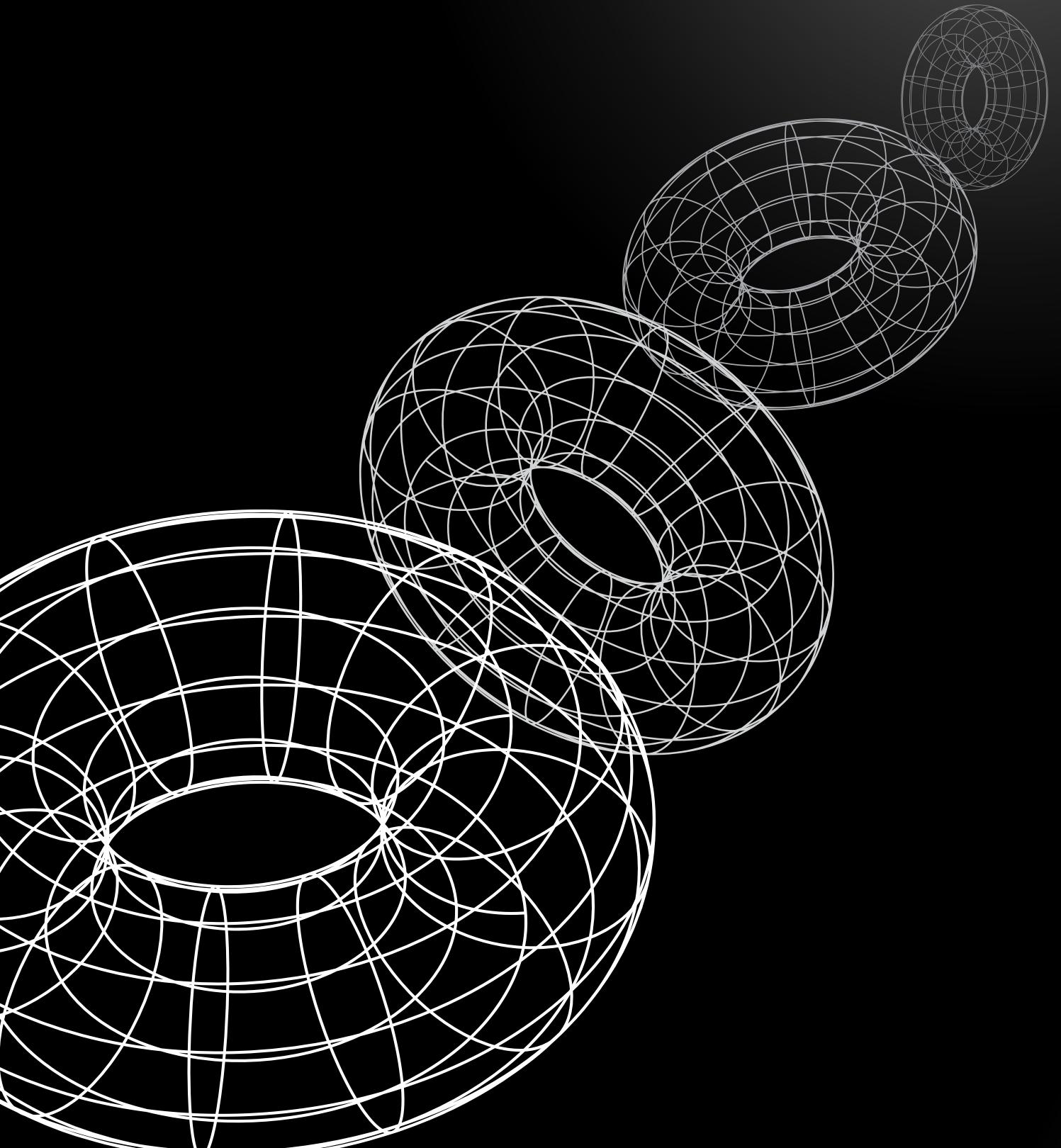
SPAM

Back



NOT SPAM

Back

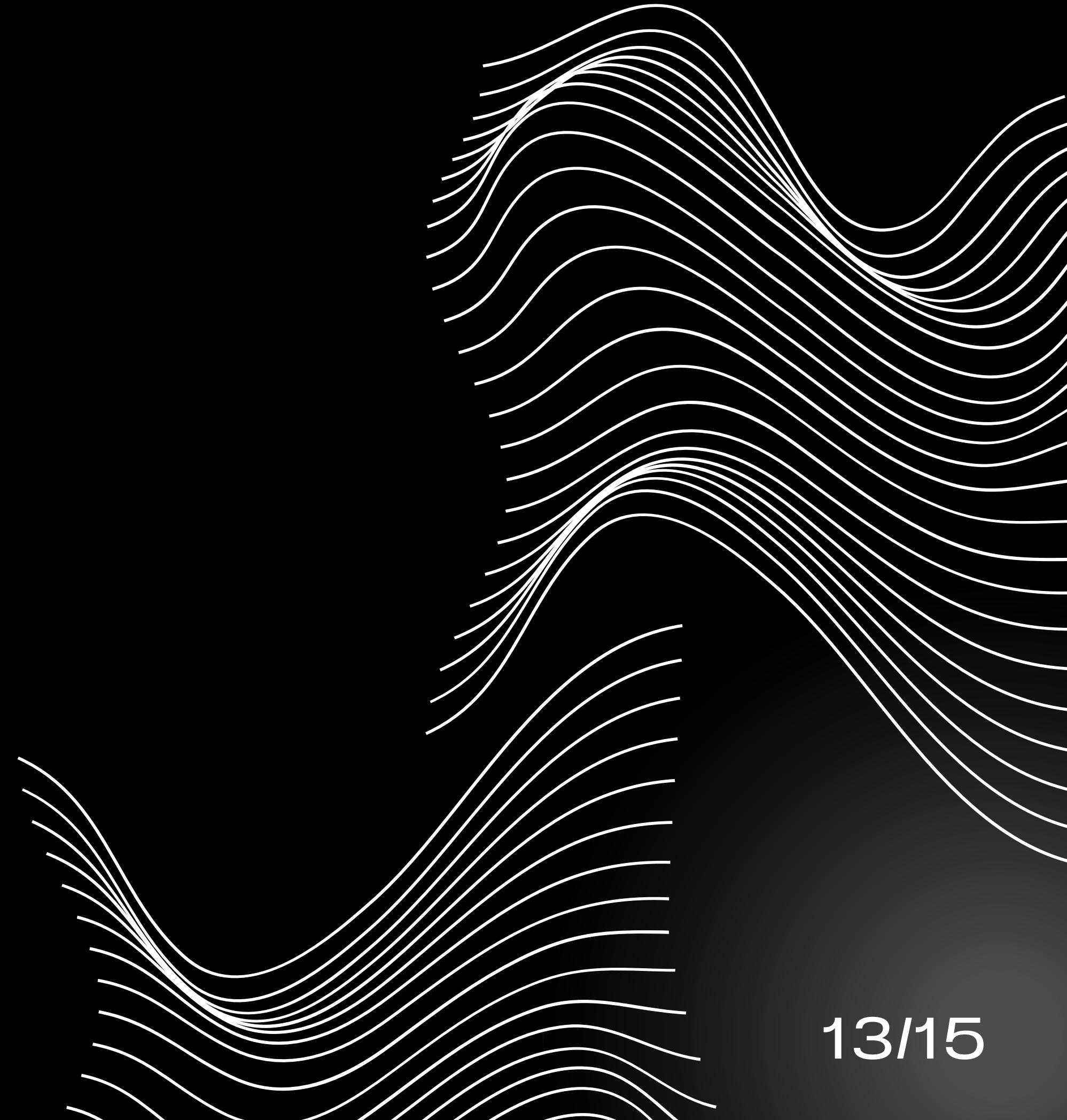


UNIQUENESS

We enhance accuracy and reliability by using two different algorithms for spam detection. This hybrid approach allows for better identification of malicious emails and reduces false positives.

CONCLUSION

- Protect users from phishing emails using machine learning.
- Improve email safety through intelligent content analysis.
- Enhance cybersecurity and build user trust in digital communication.



REFERENCES



<https://ieeexplore.ieee.org/document/10575110>



<https://www.youtube.com/watch?v=nkPNQk4-3UE>



https://www.researchgate.net/publication/320187204_Spam_Email_Detection_Using_Deep_Support_Vector_Machine_Support_Vector_Machine_and_Artificial_Neural_Network

THANK YOU

FOR GIVING THIS
OPPORTUNITY