

Institiúid Teicneolaíochta Cheatharlach



INSTITUTE *of*
TECHNOLOGY

CARLOW

At the Heart of South Leinster

Email Spam Filter using Machine Learning

Usability Manual

Author: Hazel Murphy

Student ID: C00230058

Project Supervisor: James Egan

Date: Friday 30th April 2021

Abstract

The purpose of this project is to develop a web application where users can securely view their emails in the correct folders. Each email is classified as spam or ham using machine learning algorithms. The purpose of this document is to describe how and use the email spam filter tool known as “ByeHackers”.

Table of Contents

Abstract	2
Introduction.....	4
System requirements	4
Tools.....	4
Technologies	4
Versions.....	4
Installing instructions.....	5
Pre-requirements	5
Setup Backend.....	5
Setup Web Application	6
System usage.....	8
Registration.....	8
Login	9
General Usage	9
Inbox mail	10
Sent mail.....	10
Spam mail.....	10
Account information	11
Spam report	11
Registered Users.....	12
Back-end configuration.....	12
Troubleshooting	12

Introduction

This document is targeted at individuals and organizations who use emails regularly. The aim of this document is to provide details on the general usage of this email spam tool known as “ByeHackers” and how this tool is installed and setup. The document also describes how the backend is configured.

System requirements

This application will run on any modern desktop operating system. The web application can be executed via any browser for example chrome, Firefox etc. You will need enough system storage to download the required tools for this tool.

The following tools and technologies were used to create the application:

Tools:

- MySQL Workbench (<https://www.mysql.com/products/workbench/>)
- hMailServer (<https://www.hmailserver.com/download>)
- Thunderbird (<https://www.thunderbird.net/en-US/>)

Technologies:

- MySQL server
- Flask
- Python
- SQLAlchemy
- Scikit-learn
- Pickle
- Matplotlib

Versions

The versions of the languages and tools used are as follows:

- MySQL Workbench Version 8.0.23
- hMailServer 5.6.7-B2425
- Python 3.9.1
- Thunderbird 78.9.0

- MySQL server 8.0.24

Installing instructions

Pre-requirements

1. Install MySQL server to your computer. <https://dev.mysql.com/downloads/mysql/>
2. Create a local database

Setup Backend

1. Install hMailServer

Setup a domain called “@test.com”.

Setup test account:

- test@test.com
- test2@test.com
- test3@test.com
- admin@test.com

2. Install Thunderbird

Thunderbird will be used as the email client.

You need to login to thunderbird using the following settings for each account:

Set Up Your Existing Email Address

Use your current email address

Your name: (Student) - Hazel Murphy

Email address: test@test.com

Password: ••••••••

Remember password

	INCOMING	OUTGOING
Protocol:	POP3	SMTP
Server:	127.0.0.1	127.0.0.1
Port:	110	25
SSL:	None	None
Authentication:	Normal password	Normal password
Username:	test@test.com	test@test.com

[Advanced config](#)

Cancel Re-test Done

3. Install MySQL Workbench

This tool is used to host and display the database for the application.

Setup Web Application

1. Download code

You can download all the code needed on GitHub: <https://github.com/HazelMur/Spam-filter-tool.git>

2. Create a virtual environment (<https://docs.python.org/3/library/venv.html>)

Enter the following commands in the command line to create and activate your virtual environment:

- `py -3 -m venv venv`
- `venv\Scripts\activate`

3. Navigate to the directory.

In the command line navigate to the directory where the “requirement.txt” file is located.

Enter “pip install requirements.txt”

4. Edit the `_init_.py` file

In the user's folder edit the `_init_.py` file

Edit the following line to suit your database you have created.

```
app = Flask(__name__)
app.config['SECRET_KEY'] = '35f6d593ea98517a248e1e14f958b2f6'
app.config['SQLALCHEMY_DATABASE_URI'] = 'mysql://root:hazel19@127.0.0.1/pythonlogin' #setting up database
db = SQLAlchemy(app) #database
migrate = Migrate(app, db)
```

To update these changes, in the command line enter the following commands:

- `flask db migrate -m` (Migrating models to local database)
- `flask db upgrade`

5. Run the flask application

Finally, run the flask application.

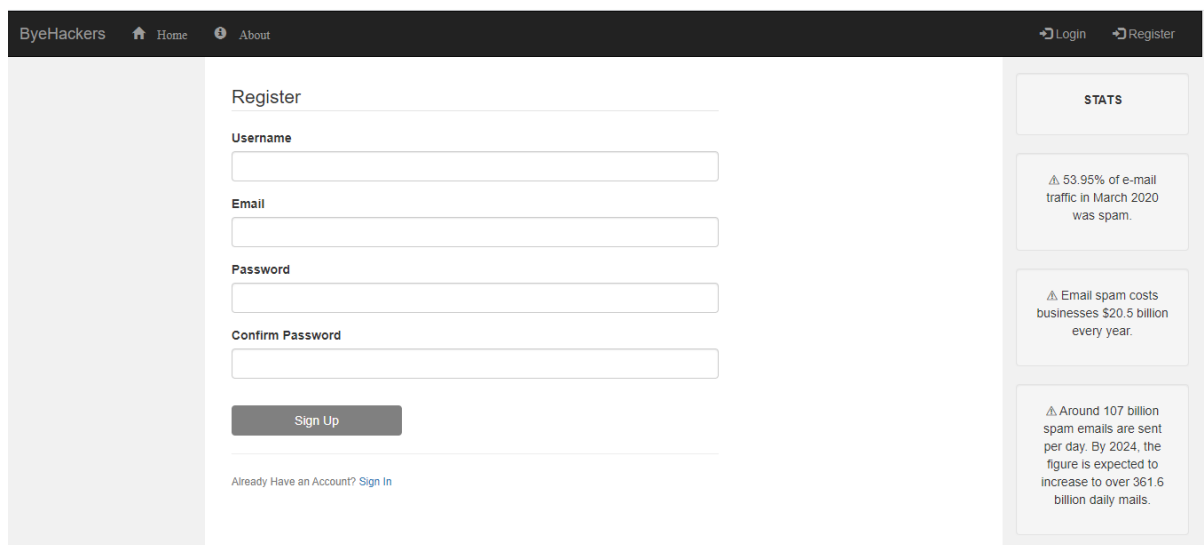
Enter “`flask run`” in the command line and when you enter <http://127.0.0.1:5000/> on your web browser the flask app should appear.

System usage

The below section details all the steps required to begin using the application.

Registration

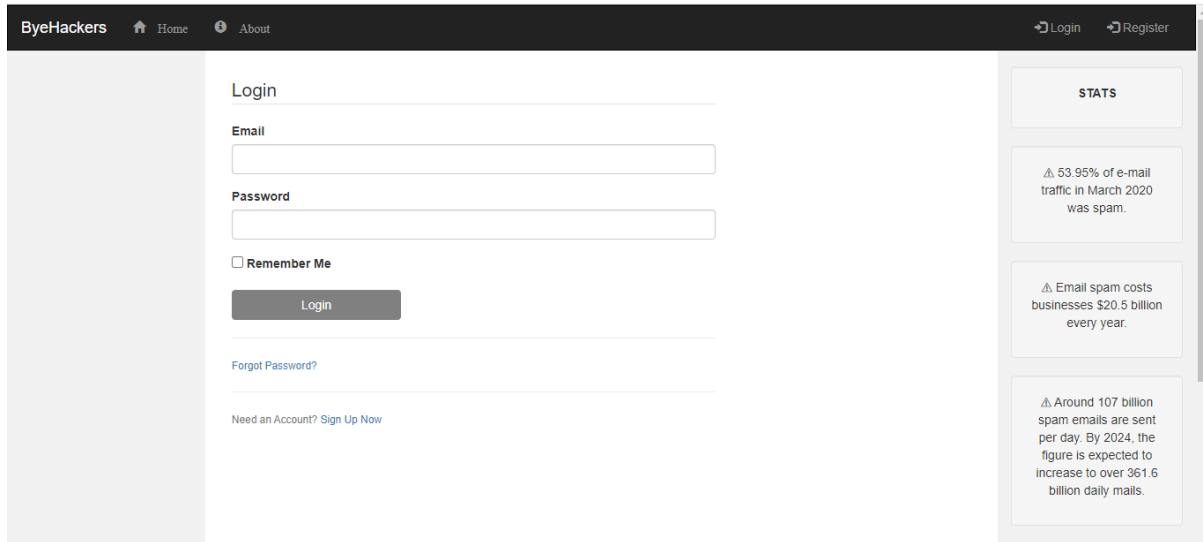
Before you can receive and view your emails you must register with the application. You must complete the registration form as seen below. You will have to provide a username, email address and password. If the username and/or email address is already in use an error message will display to notify you.



The screenshot shows a web application interface for registration. The header includes the site name 'ByeHackers' and navigation links for 'Home' and 'About'. On the right side of the header, there are links for 'Login' and 'Register'. The main content area is divided into two sections. The left section is a registration form with the following fields: 'Username', 'Email', 'Password', and 'Confirm Password'. Below these fields is a 'Sign Up' button. At the bottom of the form, there is a link: 'Already Have an Account? [Sign In](#)'. The right section is titled 'STATS' and contains three informational boxes. The first box states: '53.95% of e-mail traffic in March 2020 was spam.' The second box states: 'Email spam costs businesses \$20.5 billion every year.' The third box states: 'Around 107 billion spam emails are sent per day. By 2024, the figure is expected to increase to over 361.6 billion daily mails.'

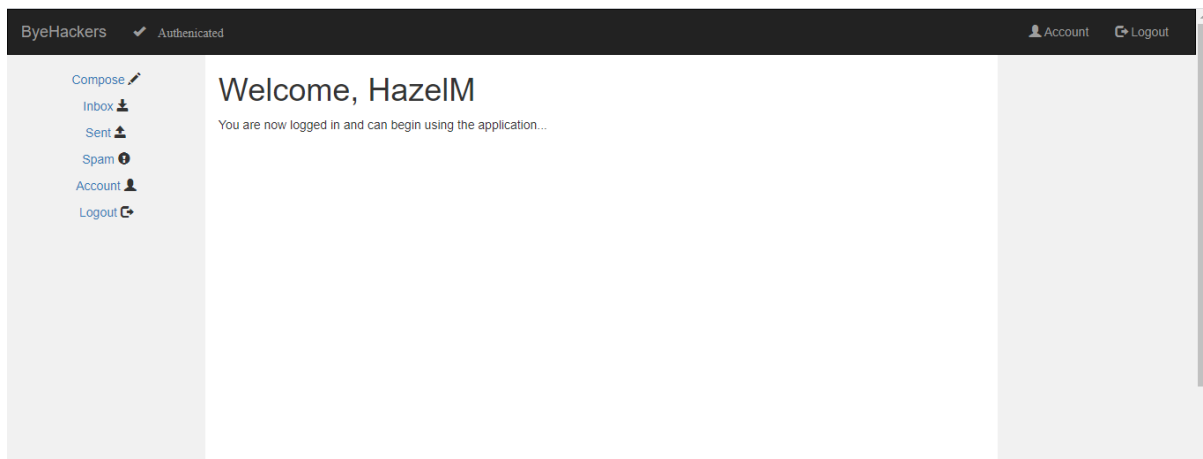
Login

Once you have registered successfully, you can then login to the application. To login you will need to provide the email address and password used when registering.



The screenshot shows the login page of the ByeHackers application. The header includes the site name 'ByeHackers' and navigation links for 'Home' and 'About'. On the right side of the header, there are links for 'Login' and 'Register'. The main content area is titled 'Login' and contains two input fields for 'Email' and 'Password'. Below these fields is a checkbox labeled 'Remember Me' and a 'Login' button. There are also links for 'Forgot Password?' and 'Need an Account? Sign Up Now'. On the right side of the page, there is a 'STATS' section with three data points: '53.95% of e-mail traffic in March 2020 was spam.', 'Email spam costs businesses \$20.5 billion every year.', and 'Around 107 billion spam emails are sent per day. By 2024, the figure is expected to increase to over 361.6 billion daily mails.'

Once you have logged in successfully you will be greeted with the below screen.



The screenshot shows the user dashboard after a successful login. The header includes the site name 'ByeHackers' and a status indicator 'Authenticated'. On the right side of the header, there are links for 'Account' and 'Logout'. The main content area is titled 'Welcome, HazelM' and contains a message: 'You are now logged in and can begin using the application...'. On the left side of the page, there is a sidebar with navigation links: 'Compose', 'Inbox', 'Sent', 'Spam', 'Account', and 'Logout'.

General Usage

Once you have registered and logged into the application you have access to your emails. You can view your inbox, sent, spam mails and account information. However, if you have logged into the application as the administrator you have additional features such as view spam reports and view all the registered users of the application.

Inbox mail

The screenshot shows an email client interface for 'ByeHackers' with a navigation sidebar on the left containing 'Compose', 'Inbox', 'Sent', 'Spam', 'Account', and 'Logout'. The main area is titled 'Inbox' and displays a table of five emails. The table has columns for Date & Time, Sender, Subject, and Body.

Date & Time	Sender	Subject	Body
2021-04-08 11:06:11	test2@test.com	HAZELS TEST	HAZELS TEST HAZELS TEST HAZELS TEST
2021-04-06 18:26:10	test2@test.com	another test	testing for spam
2021-03-31 14:58:31	test2@test.com	Checking if this one gets marked off	Spam score test
2021-03-31 11:50:43	test2@test.com	trying again	one more test
2021-03-31 11:00:34	test2@test.com	hi	just another test, really this should come up as just ham, not spam

Sent mail

The screenshot shows the 'Sent' section of the email client. The table lists six sent emails with columns for Date & Time, Recipient, Subject, and Body.

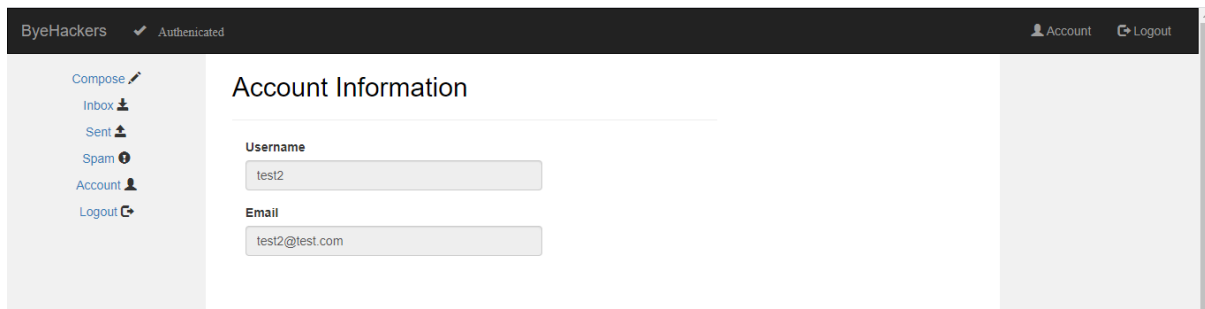
Date & Time	Recipient	Subject	Body
2021-04-15 15:33:49	test2@test.com	Final Project	Kate jackson rec center before 7ish, right?
2021-04-15 15:33:49	test2@test.com	Final Project	Some of them told accenture is not confirm. Is it true.
2021-04-15 15:33:49	test2@test.com	Final Project	Dear i have reache room
2021-04-15 15:33:49	test2@test.com	Final Project	
2021-04-15 15:33:49	test2@test.com	Final Project	Nope... Think i will go for it on monday... Sorry i replied so late
2021-04-15 15:33:48	test2@test.com	Final Project	Sorry,in meeting I'll call later

Spam mail

The screenshot shows the 'Spam Mail' section of the email client. The table lists four spam emails, including a 'Naive Bayes Score' column where all scores are 'True'. The table has columns for Date & Time, Sender, Subject, Body, and Naive Bayes Score.

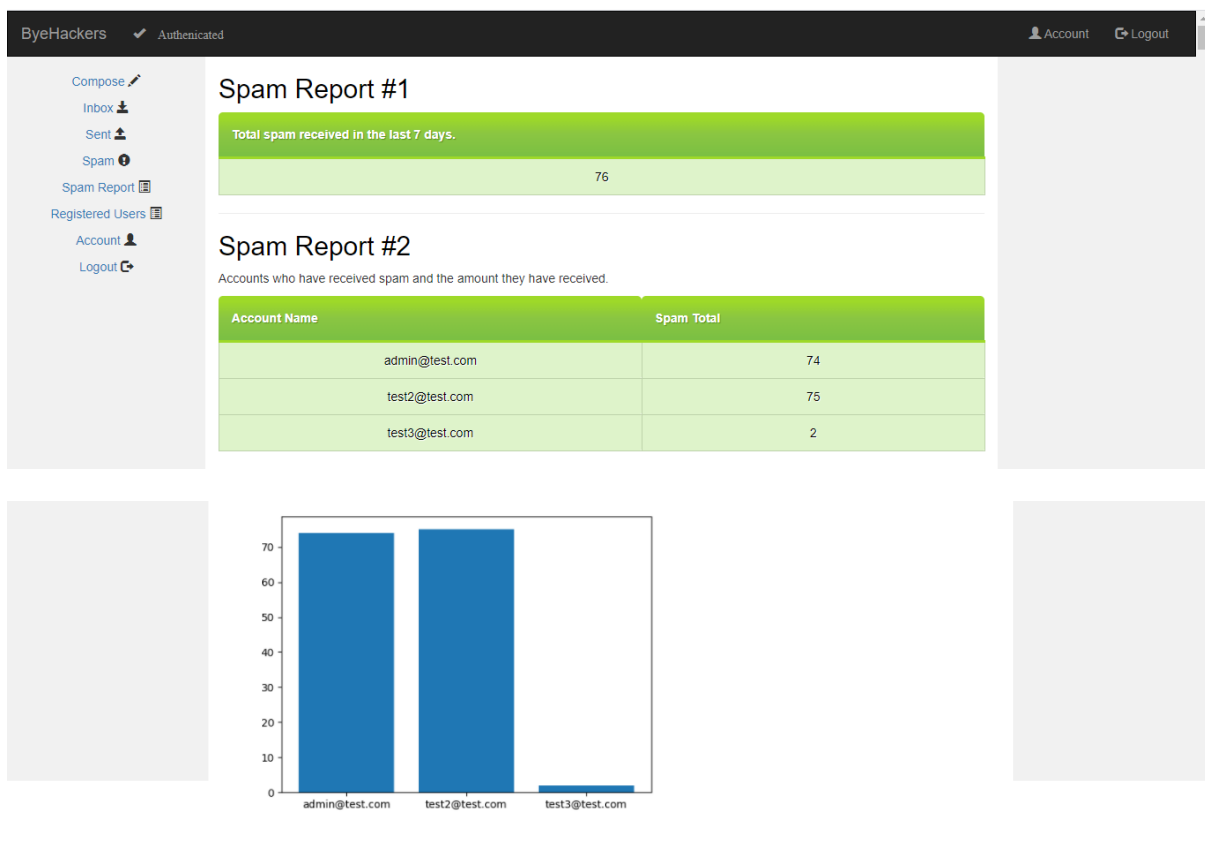
Date & Time	Sender	Subject	Body	Naive Bayes Score
2021-04-15 15:33:48	test@test.com	Final Project	Are you free now?can i call now?	True
2021-04-15 15:33:48	test@test.com	Final Project	Congrats! 1 year special cinema pass for 2 is yours. call 09061209465 now! C Suprman V, Matrix3, StarWars3, etc all 4 FREE! bx420-lp4-5we. 150pm. Dont miss out!	True
2021-04-15 15:33:47	test@test.com	Final Project	FREE MESSAGE Activate your 500 FREE Text Messages by replying to this message with the word FREE For terms & conditions, visit www.07781482378.com	True
2021-04-15 15:33:45	test@test.com	Final Project	Want 2 get laid tonight? Want real Dogging locations sent direct 2 ur Mob? Join the UK's largest Dogging Network by txtng MOAN to 69688Nyt. ec2a. 31p.msg@150p	True

Account information



Spam report

The admin user only has access to this page.



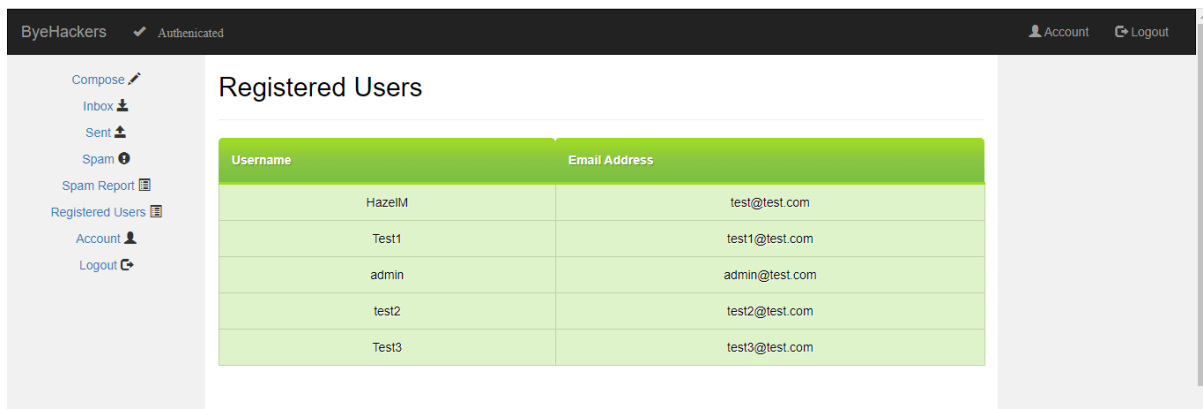
Spam Report #3

Emails received that have been classified as spam.

Date received	Subject	Body	Naive Bayes Score
2021-04-20 12:15:53	an invitation to advertise on freightmart !	an invitation to advertise on freightmart ! freightmart . com - ship anything . . . anywhere . . . anytime . . . auction style ! need more info ? if you ' re a shipper if	True

Registered Users

The admin user only has access to this page.



Username	Email Address
HazellM	test@test.com
Test1	test1@test.com
admin	admin@test.com
test2	test2@test.com
Test3	test3@test.com

Back-end configuration

For these emails to be stored in the SQL database I had to implement a scrape function. This function scans through the directories of the local mail server, and for each email which has not already been stored it transforms the emails data for storage and then calls upon each of the model's functions created and inserts the spam classification boolean data for each model into the database. The four models, Naïve Bayes, SVM, Random Forest and Logistic Regression were trained using the following dataset:

<https://www.kaggle.com/uciml/sms-spam-collection-dataset>.

Troubleshooting

If you experience any issues while setting up this project, please contact me:

hazelm742@gmail.com