

Project Report Template

Title of Project: fake email classifier

Name of the Innovator: Shashank.R

Start Date:13-10-2025

End Date: 17-10-2025

Day 1: Empathise & Define

Step 1: Understanding the Need

- Which problem am I trying to solve?

I'm solving solving the problem of **detecting and filtering fake or fraudulent emails automatically** using **machine learning or AI techniques** — helping users stay safe from scams and ensuring that only genuine emails reach their inbox.

Step 2: What is the problem?

In today's digital world, people receive hundreds of emails daily, including **fraudulent or fake emails** that attempt to **steal personal information, spread malware, or deceive users** through **phishing scams**.

Why is this problem important to solve?

The most important problem in the *Fake Email Classifier* project is the **increasing difficulty in detecting fake or fraudulent emails** that are designed to look like genuine ones.

Take-home task

Ask 2-3 people what they think about the project:

- **1. Student (Rural College Student):**
“find it **interesting and educational**, as it helps them understand **online safety and cybersecurity** — topics that are becoming more important even in rural areas.”
- **2. Teacher (Career Guidance Teacher):**
“**valuable learning experience**, because it helps students understand **real-world cybersecurity problems, data analysis, and machine learning concepts** in a practical way.”
- **3. Parent (From a Rural Area):**
“appreciate that the project focuses on **protecting people from online fraud and scams**, which are becoming common. They'd see it as a **useful and responsible project** that helps their children understand **how to stay safe online**.”

• ”

AI Tools you can use for Step 1 and 2:

AI Tools Used:

1. Meta MGX

- Used Built-in datasets, CSV file.
- Machine Learning Classifier.

2. ChatGPT

- objectives, scope, and requirements.
- dataset sources flowcharts, algorithms.

Day 2: Ideate

Step 3: Brainstorming solutions

- List **at least 5 different solutions** (wild or realistic):
- **Machine Learning:** traditional machine learning algorithms to classify emails as “Fake”
- **Deep Learning:** Feed sequences into an LSTM model to learn context patterns
- **Transformer-Based Solution:** pre-trained transformer model like for advanced text understanding
- **Rule-Based and Keyword Detection System:** predefined rules and keyword patterns to identify fake or spam email
- **Hybrid Model:** Combine ML-based detection with keyword-based filtering for better accuracy

Step 4: My favourite solution:

Machine Learning approach using the Naïve Bayes algorithm. This solution uses artificial intelligence to automatically identify whether an email is fake (spam) or genuine based on the words and patterns present in the message.

Step 5: Why am I choosing this solution?

It's simple, efficient, and accurate for text classification ,It requires less computing power compared to deep learning models.It's easy to explain and implement in Python

AI Tools you can use for Step 3-5:

AI Tools for Step 3–5

1. Meta MGX

- Used to design and build the fake email classifier without coding.
- simplifies the entire process of building a Fake Email Classifier.

2. ChatGPT

- designed **to** automatically detect and classify emails as either fake or genuine.
- Prevents users from falling victim to phishing emails that try to steal sensitive information like passwords, bank details, or personal data.
- Educates users about the risks of fake emails and how to identify them.

AI Tools you can use for the take-home task:

Canva AI/CoPilot AI/Meta AI: Use these mobile-based tools to generate images for the solution they want to design

Day 3: Prototype & Test

Step 6: Prototype – Building my first version

What will my solution look like?

- Gather a small set of emails labelled as Fake/Spam and Genuine.
- Clean text by removing punctuation, numbers.
- Use a Naïve Bayes Classifier for your first version. It is easy and fast to implement.
- Train the model on the sample dataset.
- Check basic accuracy and note any misclassifications.

Design Style:

- Colour Theme.
- Typography.
- Visual Elements.

Prototype Tools:

- Programming & ML Tools, Dataset Tools, No-Code.

AI Tools Needed to Build Career Path

1. **Meta MGX**
 - Data Management Tools ,Data Preprocessing Tools.
2. **ChatGPT**
 - Import your email dataset.
 - Automatically split data into training and testing sets for model development.
 - To analyze user interactions and improve recommendations over time.

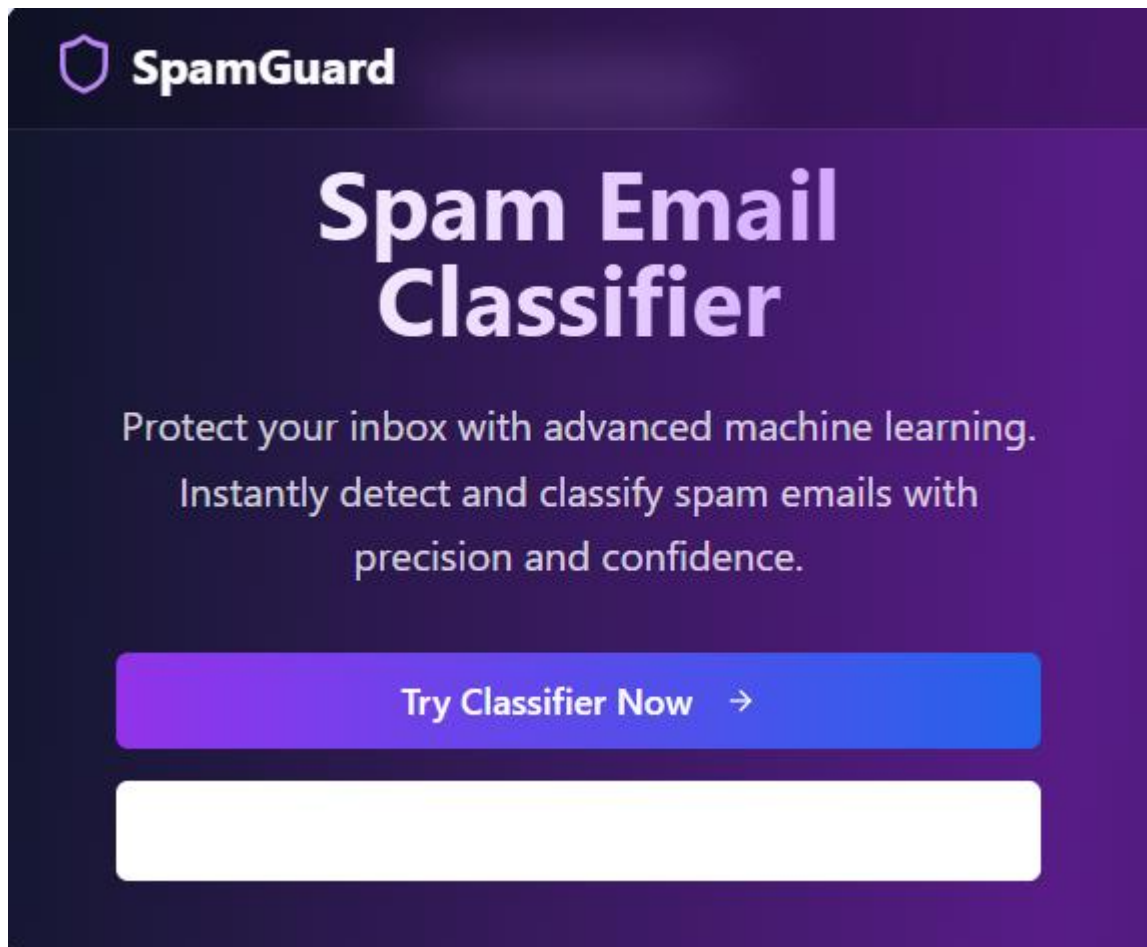
What AI tools I finally selected to build this solution?

1. **Chat GPT**
2. **Metamgx**

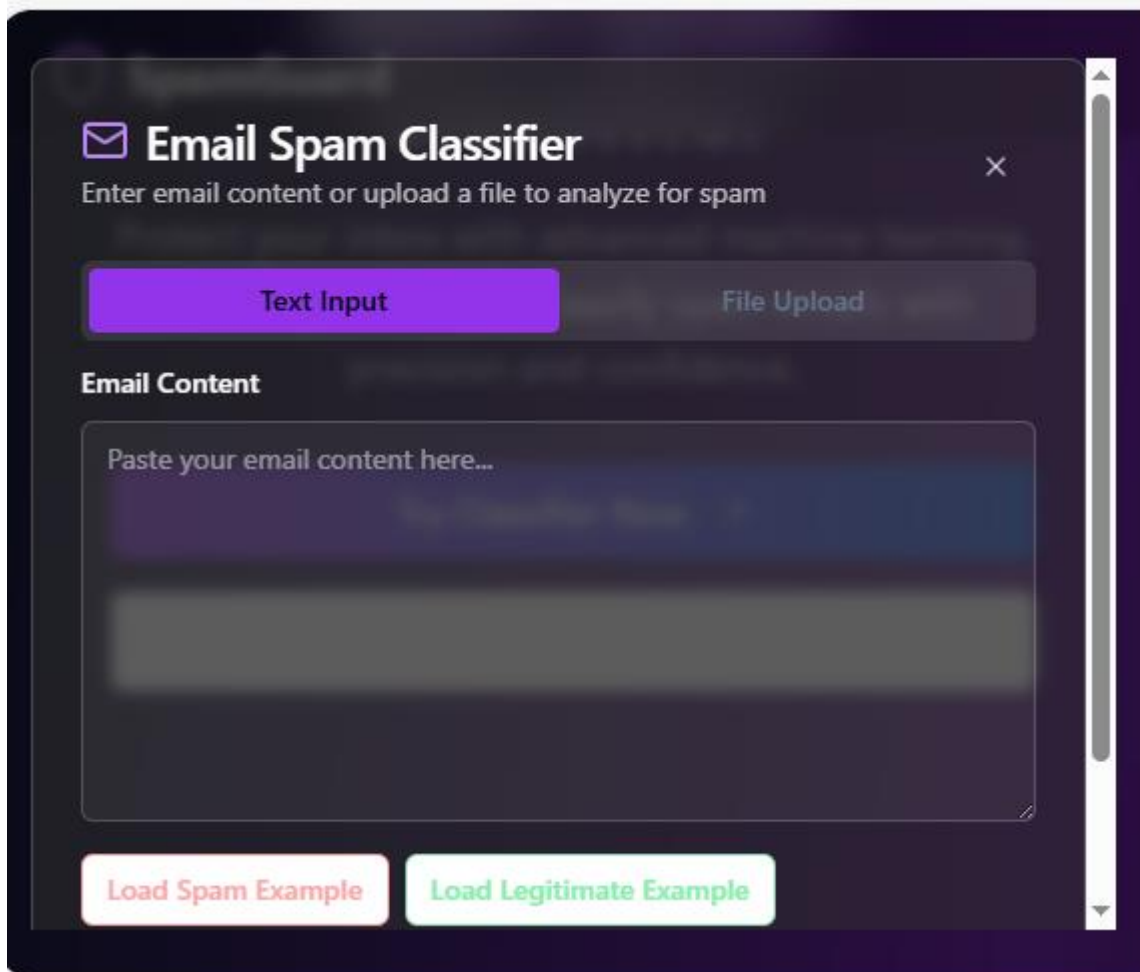
< Build The Innovation >

<DASHBOAD OF THE TOOL>

Tool Link:<https://mgx.dev/app/c1f6b3a6e15343fc8d74b76bf256c5f8>



Internal Working of tool:

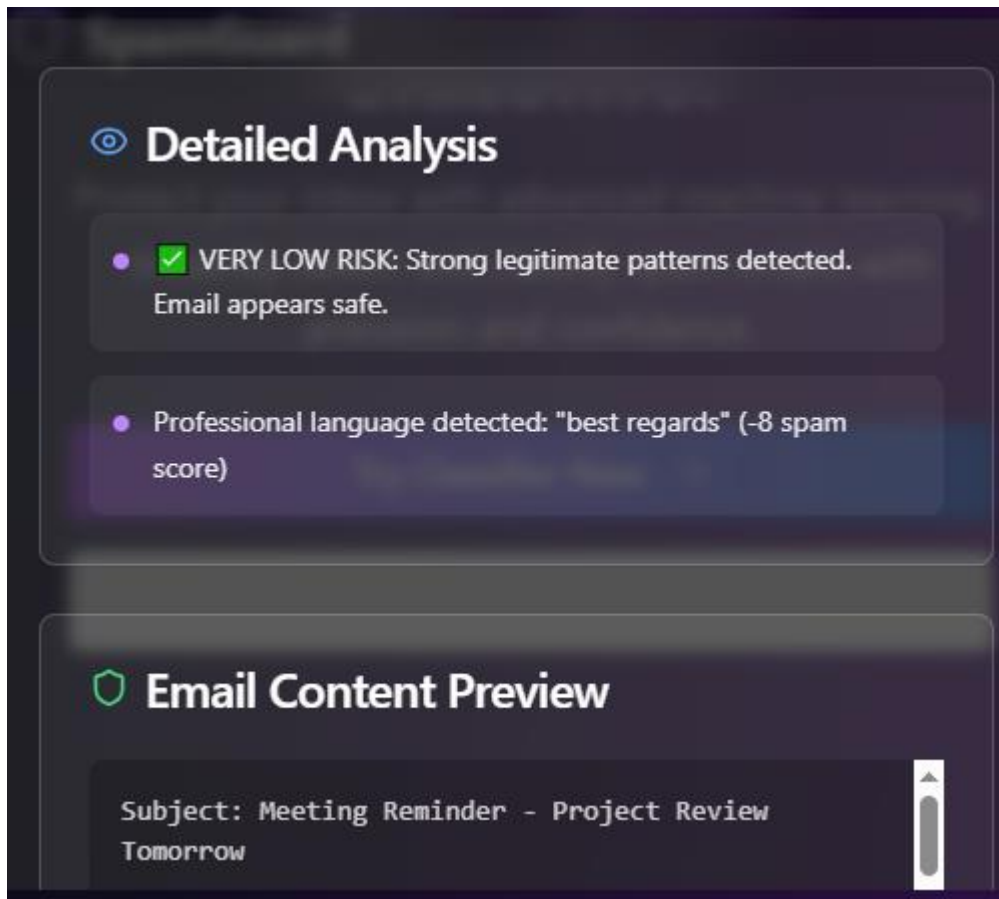


The screenshot shows a web application titled "Email Spam Classifier" with a close button (X) in the top right corner. Below the title is a subtitle: "Enter email content or upload a file to analyze for spam". There are two main input methods: "Text Input" (highlighted in blue) and "File Upload" (highlighted in light blue). Under the "Text Input" section, there is a large text area with the placeholder text "Paste your email content here...". At the bottom of the interface, there are two buttons: "Load Spam Example" (highlighted in red) and "Load Legitimate Example" (highlighted in green). A vertical scrollbar is visible on the right side of the interface.

Spam probability analysis:



Detailed analtsis:



Step 7: Test – Getting Feedback

- Who did I share my solution with?

I shared my **fake email classification** solution with:

- **Students from rural areas** – The project helps them understand fake emails, phishing attacks, and online safety.
- **Teachers and career guidance counselors** - Help with understanding algorithms, text preprocessing, and evaluation metrics.
- **Parents of rural students** – It helps protect both students and family members from digital fraud and financial loss.
- **Peers and mentors** – Help overcome challenges such as data imbalance, model errors, or low accuracy.

What feedback did I receive?

Feedback: Pros and Cons

Pros (Positive Insights from Feedback):

1. Saves time by automatically classifying emails as fake.
2. Helps students learn AI, machine learning, and text analysis.

3. Can be improved with larger datasets or advanced models for better accuracy.

Cons (Areas to Improve Noted in Feedback):

1. Accuracy Limitations.
2. Dataset Dependency.
3. Handling Complex Emails.

My Response for The Feedback: I will use a larger and more diverse dataset and experiment with advanced models like LSTM or BERT to reduce false positives and false negatives, I will explore techniques to process attachments, HTML content, and images in emails to make the classifier more robust, I plan to implement periodic retraining and a continuous learning approach so the model can adapt to evolving phishing and spam patterns.

👍 What works well:

What Works Well

- ✚ **Automated Email Detection:** The system effectively classifies emails as Fake (spam/phishing) or Genuine, reducing the need for manual checking.
- ✚ **Use of Machine Learning Algorithms:** Models like Naïve Bayes or Logistic Regression provide good accuracy and are easy to implement for text classification.
- ✚ **Text Preprocessing and Feature Extraction:** Techniques like cleaning, tokenization, and TF-IDF vectorization help the model understand email content efficiently.
- ✚ **User-Friendly Prototype:** The initial prototype allows testing new emails and gives clear output, making it easy to demonstrate functionality.

🔧 What needs improvement:

• Model Accuracy

- Reduce **false positives and false negatives** by using a **larger dataset** and experimenting with **advanced models** like LSTM or BERT.

• Handling Complex Emails

- Improve the model to handle **attachments, HTML content, and images** in emails, not just plain text.

• Adaptability to New Fake Patterns

- Implement **periodic retraining** so the model stays updated with evolving phishing and spam tactics.

• User Interface Enhancements

- Make the prototype **more interactive and user-friendly**, with **real-time predictions, alerts, and confidence scores**.

AI Tools you can use for Step 6-7:

ChatGPT/Perplexity AI/Claude AI/Canva AI/Chatling AI/Figma AI/Metamgix/Gamma AI: You can use these tools to build solutions/models or mock-up dummy prototypes

Day 4: Showcase

Step 8: Presenting my Innovation:

1. Project Title

Fake Email Classifier – An AI-powered system to detect and classify emails as **Fake**.

2. Problem Statement

- Fake emails, phishing attacks, and scams are **increasing rapidly**, targeting individuals and organizations.
- Manual checking of emails is **time-consuming and error-prone**.
- There is a **need for an automated system** to detect and prevent email fraud.

3. Objective / Innovation

- Develop a **machine learning model** to automatically classify emails as fake or genuine.
- Raise **cybersecurity awareness** and protect users from online scams.
- Introduce **AI/ML techniques** to a practical, real-world problem.

4. Methodology

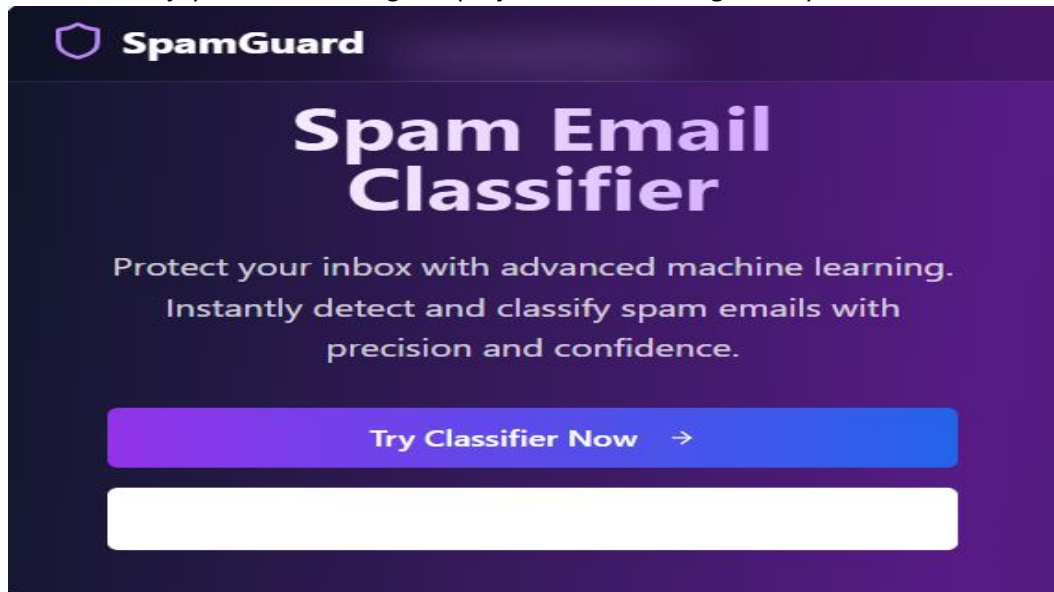
1. **Data Collection:** Gather emails labeled as Fake or Genuine.
2. **Preprocessing:** Clean text, remove punctuation/stopwords, tokenize words.
3. **Feature Extraction:** Use **TF-IDF or Count Vectorizer** to convert text to numeric data.
4. **Model Selection:** Train models like **Naïve Bayes, Logistic Regression, or Random Forest**.
5. **Evaluation:** Measure **accuracy, precision, recall**, and use a **confusion matrix**.
6. **Prototype:** Create a **user-friendly interface** to test new emails.

Impact: fake email classifier is Protects Users from Online Threats, Promotes Cybersecurity Awareness,

<SHOWCASE YOUR INNOVATION TO YOUR PEERS>

Step 9: Reflections

- What did I enjoy the most during this project-based learning activity?



I enjoyed learning how AI and machine learning can solve real-world problems, especially detecting fake emails automatically. I found it exciting to write code, preprocess data, train the model, and see it correctly classify emails.

What was my biggest challenge during this project-based learning activity?

The biggest challenge I faced during this project was handling and preprocessing the email data effectively. Emails often contain special characters, HTML content, links, and inconsistent formatting.

Take-home task

https://github.com/shashank181733-svg/Fake-email-classifier_project-report

AI Tools you can use for Step 8:

Canva AI: You can use this to design your pitch document. Download your pitch document as a PDF file and upload on GitHub