A
Project Presentation On

SPAM EMAIL DETECTION

Guru Ghasidas Vishwavidyalaya, Koni
Bilaspur, Chhattisgarh
May-2025

PRESENTED BY :
Vijaya L. Singh
B.Sc (Hons) CSIT
VI Sem

GUIDED BY :
Mr. Vivek Sarathe
Asst. Professor
Dept. of CS & IT

INTRODUCTION

In today’s digital world, spam emails are a growing concern, affecting both individuals and organizations.

They range from annoying advertisements to serious threats like phishing and malware.This project focuses on developing a **machine learning-based system** to detect and classify spam messages using the **Naive Bayes algorithm**.

By training the model on real-world data and integrating it with an interactive web interface using **Streamlit**, we can efficiently identify and filter spam messages.

This enhances email security and improves user productivity.

REQUIREMENTS & SPEIFICATIONS

Software Requirements:

1. Programming Environment

Python 3.7 or higher

2. Python Libraries

pandas - for data manipulation

scikit-learn - for machine learning (train-test split, CountVectorizer, Naive Bayes)

streamlit - to create the web interface

streamlit-lottie - to display Lottie animations

requests - for fetching animation files from the web

3. Dataset

spam.csv - labeled dataset containing email/text messages classified as "spam" or "ham'

4. Web Interface

Streamlit-compatible browser (e.g., Chrome, Firefox)

Internet connection (for Lottie animation loading)

5. Operating System

Works on **Windows**, **macOS**, or **Linux**

Hardware Requirements:

Minimun Requirement:

Processor: Dual-Core 2.0 GHz or higher
RAM: 4 GB
Storage: 500 MB of free disk space
Display: 1024x768 resolution
Internet: Required (for loading Lottie animations)

OBJECTIVE

The main objective of this project is to develop a **machine learning-based application** that can accurately classify email or text messages as **Spam** or **Not Spam.**Using the **Naive Bayes algorithm** and **natural language processing (NLP)** techniques, the system aims to:

- Detect spam messages with high accuracy.
- Minimize false positives and negatives.
- Provide a simple and interactive **Streamlit-based web interface** for real-time testing.
- Enhance user safety and productivity by filtering unwanted messages.

DATA FLOW – Spam Email Detection

1. Input

User provides a message through the Streamlit web interface.
Alternatively, labeled data (spam.csv) is used for training.

2. Data Preprocessing

Remove duplicate messages.
Label categories as “Spam” or “Not Spam”.
Split the dataset into training and testing sets.

3. Feature Extraction

Use CountVectorizer to convert messages into numerical features (bag-of-words model).
Apply stop word removal for better accuracy.

4. Model Training

Train a **Multinomial Naive Bayes** classifier using the processed features and labels.

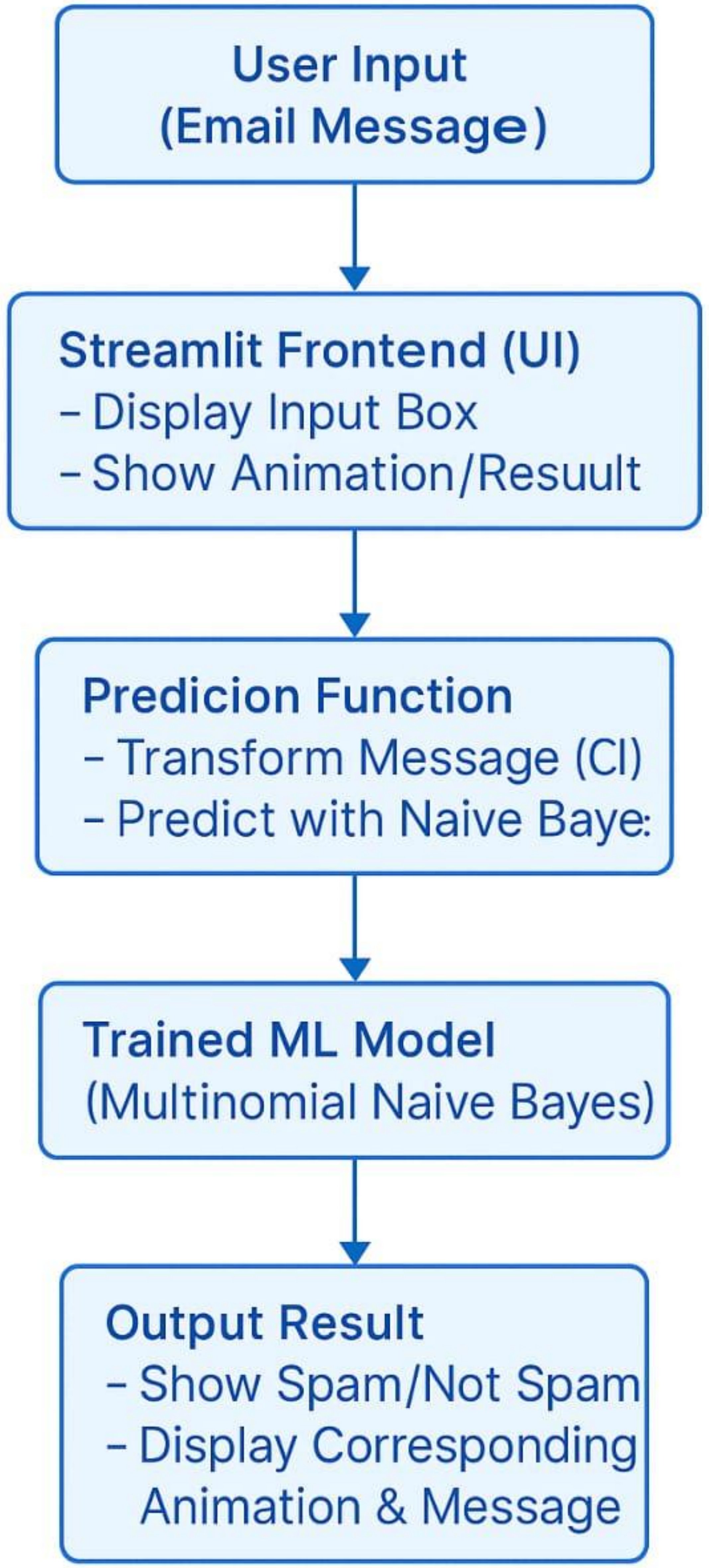
5. Prediction

Convert user input into the same feature format.
Predict using the trained Naive Bayes model.

6. Output

Display prediction result (Spam or Not Spam).
Visual feedback provided via Lottie animations.

Data Flow Diagram:- Level 0



User Input
(Email Message)

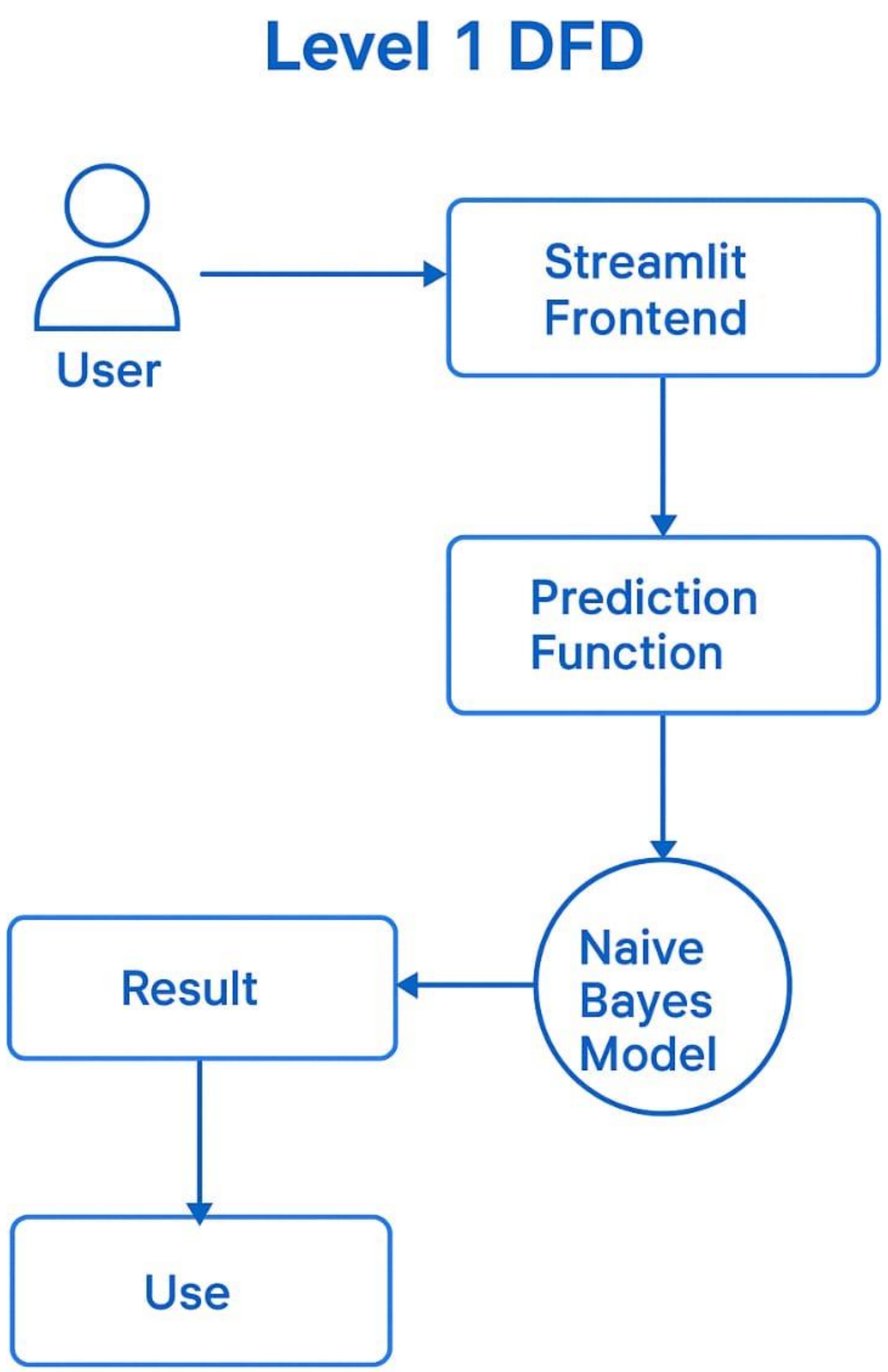
Streamlit Frontend (UI)
- Display Input Box
- Show Animation/Resuult

Predicion Function
- Transform Message (Cl)
- Predict with Naive Baye:

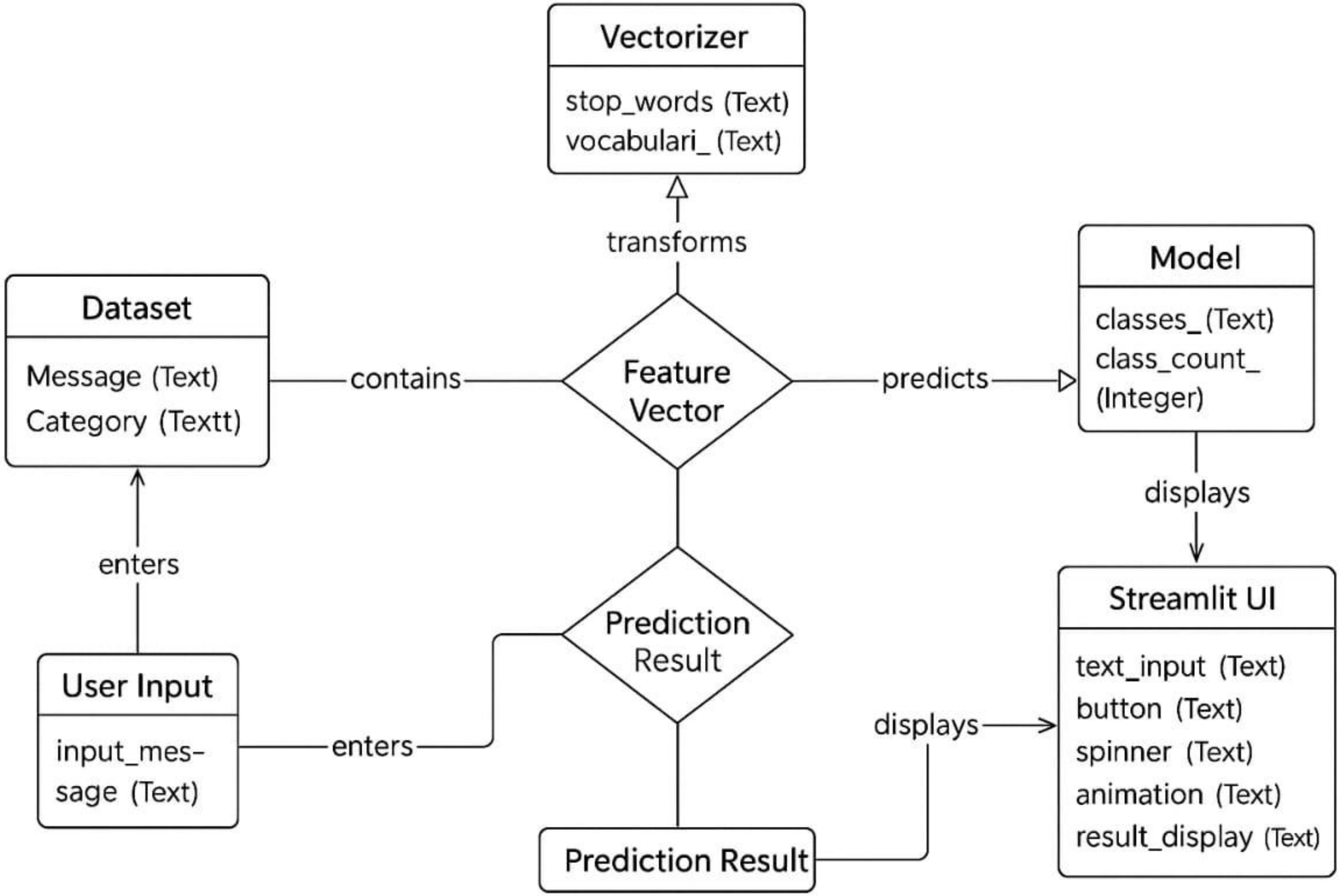
Trained ML Model
(Multinomial Naive Bayes)

Output Result
- Show Spam/Not Spam
- Display Corresponding
Animation & Message

Data Flow Diagram:- Level 1

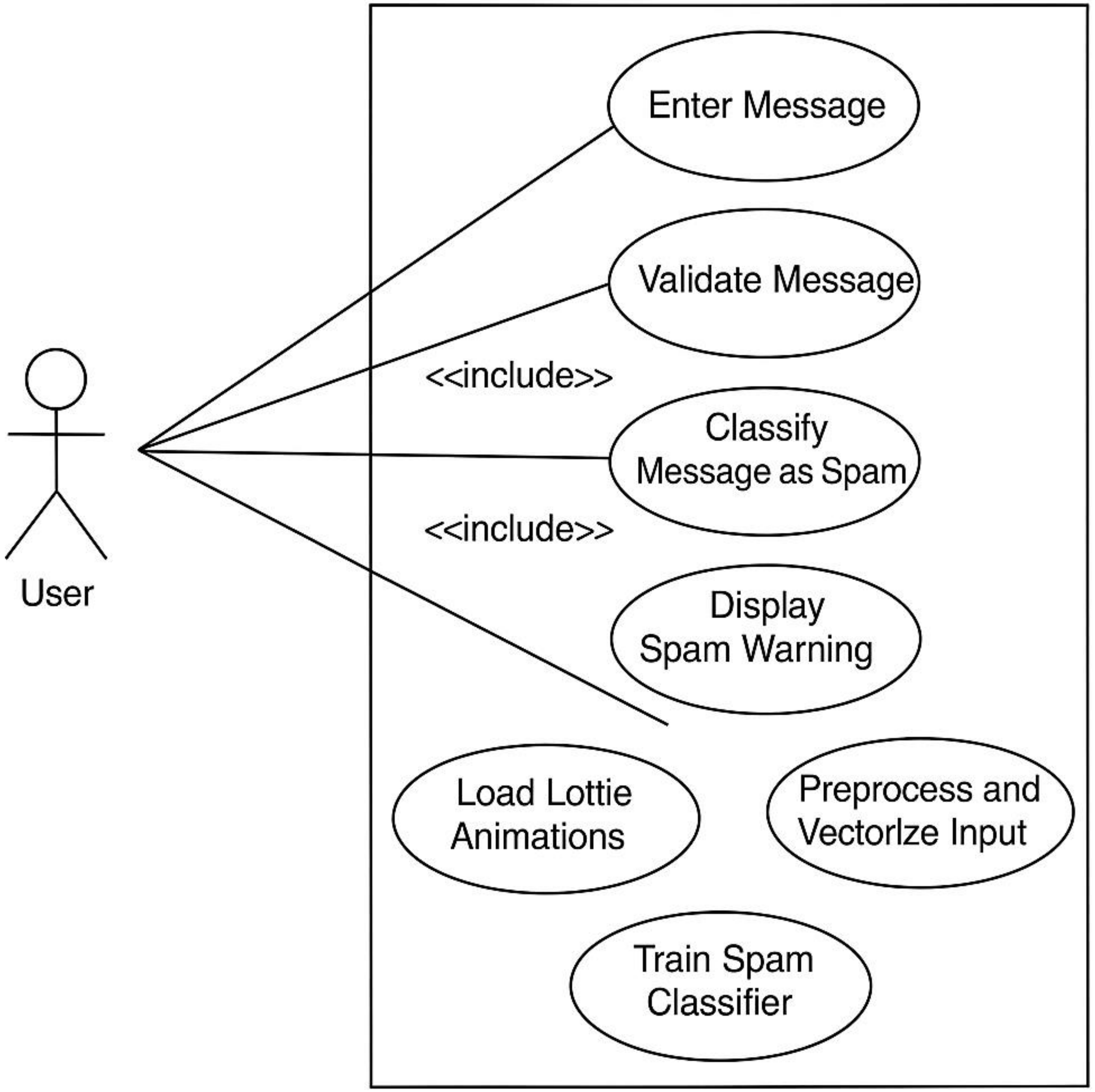


ER Diagram:-

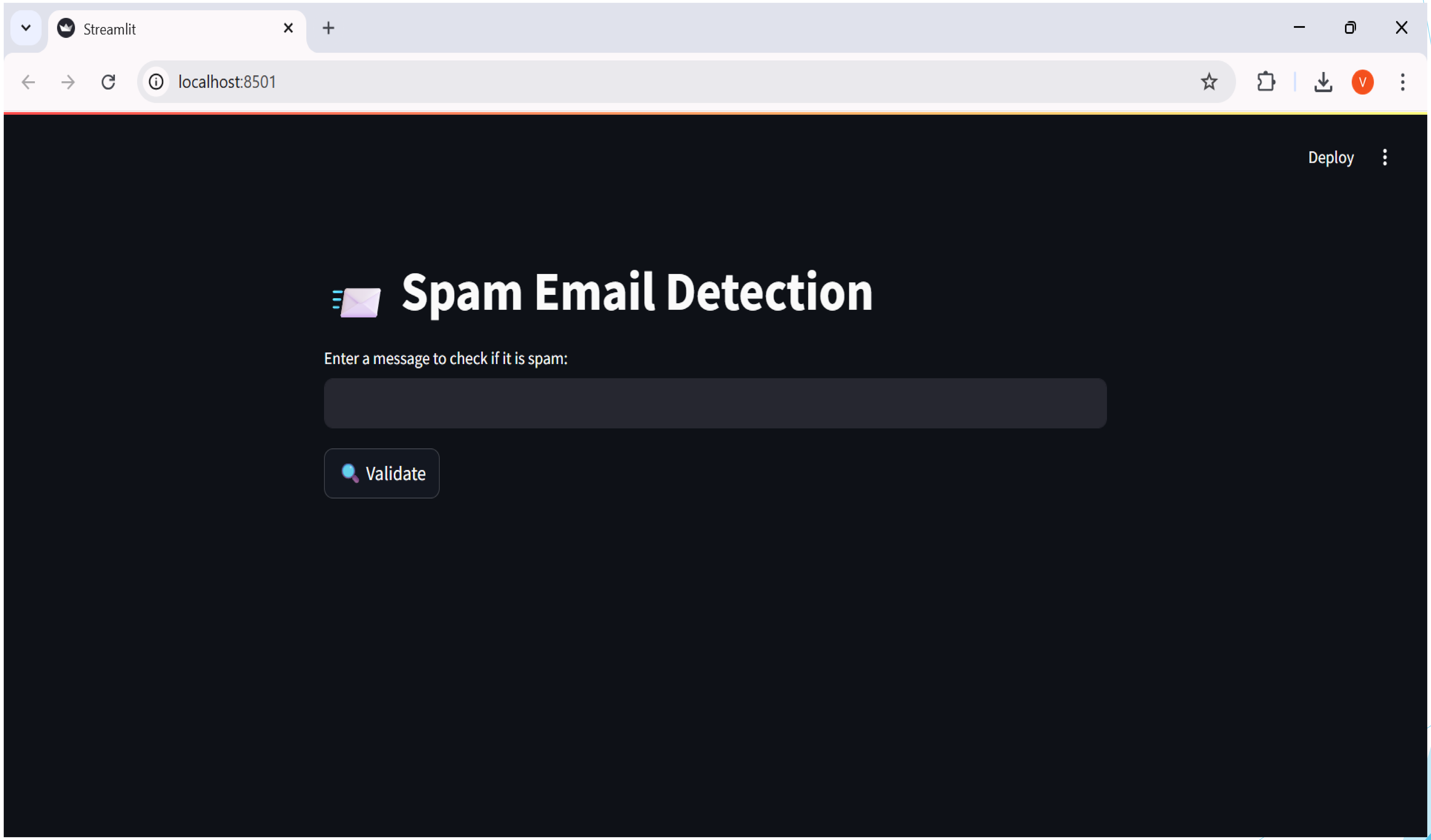


Vectorizer
stop_words (Text)
vocabulari_ (Text)
transforms
Dataset
Message (Text)
Category (Textt)
contains
Feature Vector
predicts
Model
classes_ (Text)
class_count_ (Integer)
displays
enters
Prediction Result
User Input
input_mes-sage (Text)
enters
displays
Streamlit UI
text_input (Text)
button (Text)
spinner (Text)
animation (Text)
result_display (Text)
Prediction Result

Use Case Diagram:-



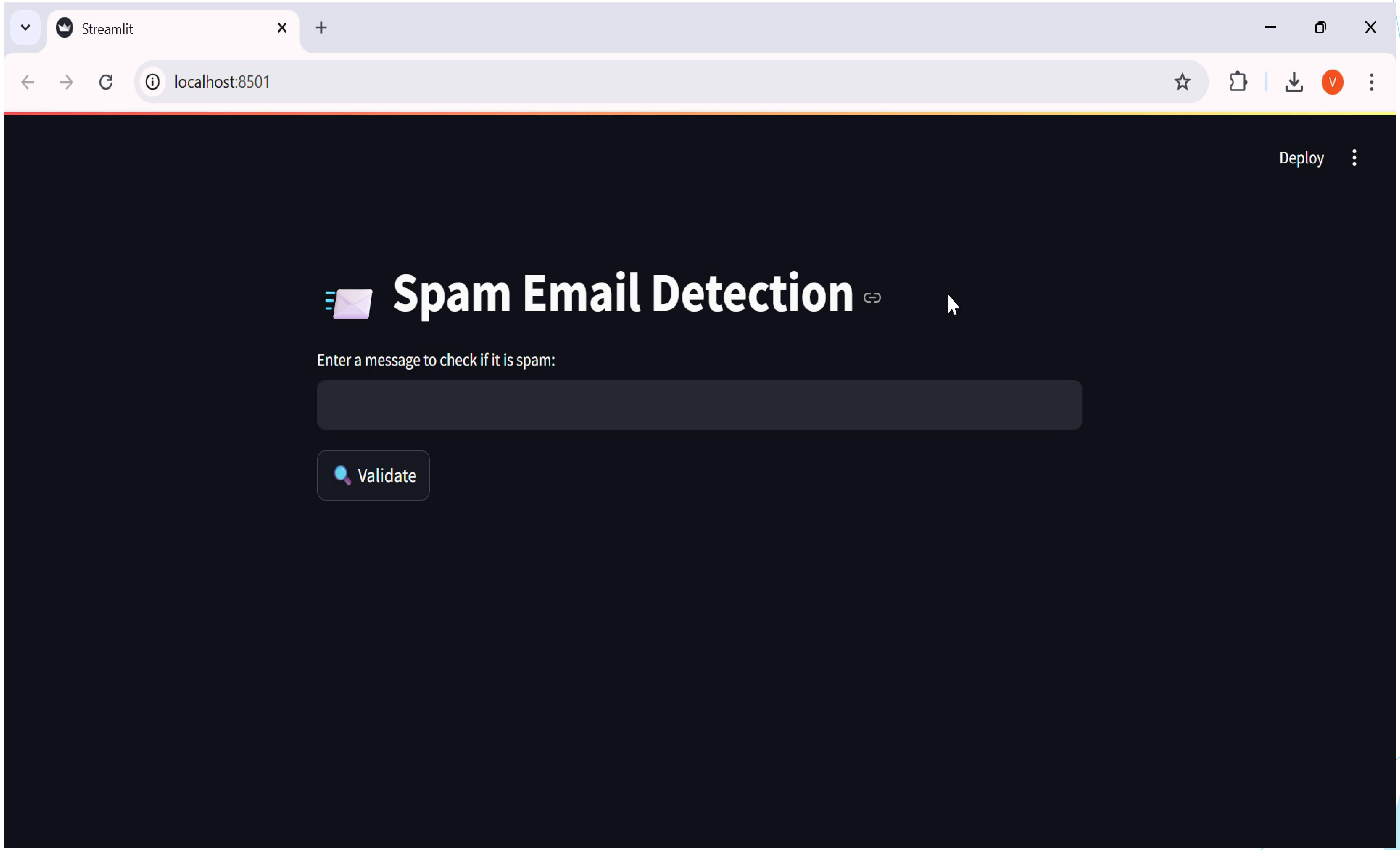
Homepage:-



Spam Email Detection

Enter a message to check if it is spam:

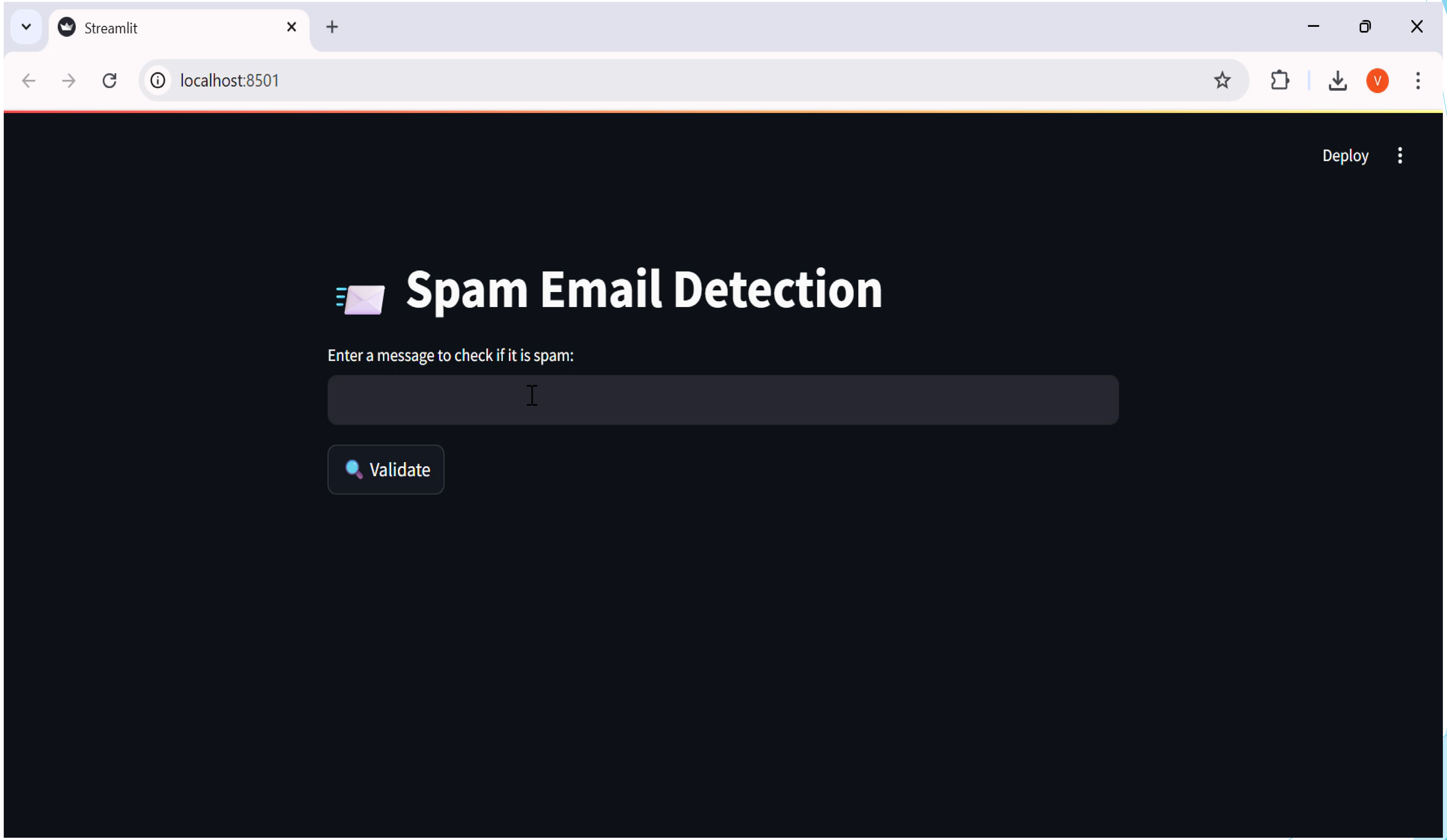
Sample Page 1:-



Spam Email Detection

Enter a message to check if it is spam:

Sample page 2:-



Spam Email Detection

Enter a message to check if it is spam:

Advantages:-

✅ **Real-World Impact**

- Detects spam in SMS/email – practical cybersecurity use

⚙️ **Efficient & Fast**

- **Naive Bayes** algorithm with **CountVectorizer**
- Fast, accurate message classification

💻 **User-Friendly Interface**

- Built with **Streamlit** – simple, interactive design
- Real-time predictions from user input

🎨 **Visual & Engaging**

- **Lottie animations** for spam, safe, and loading feedback
- Improves user experience

🧩 **Modular & Extensible**

- Clean codebase – easy to expand or modify

Conclusion:-

- This project successfully demonstrates how machine learning can be applied to real-world problems like spam detection. By leveraging the Naive Bayes algorithm and a user-friendly Streamlit interface, it offers fast, accurate predictions with an engaging user experience. The modular design and visual elements make it both functional and accessible, laying a strong foundation for future improvements such as larger datasets, additional algorithms, or email integration.

“The success of this project is shared with all who offered their guidance and support.”

THANK YOU !