📧 Email Classification with PII Masking - Project Report

1. Introduction

In today's digital era, emails are one of the most common modes of communication, often carrying sensitive personal information. As organizations grow, the volume of incoming emails increases, and manually reviewing each message for privacy and classification becomes impractical.

The core objective of this project is to develop a system that can:

- **Identify and mask Personally Identifiable Information (PII)** within email content.
- **Classify emails** into appropriate categories (e.g., Promotion, Incident, Personal, etc.) based on their content.

This solution is deployed as an interactive web application using **Gradio**, and is accessible via Hugging Face Spaces. Additionally, a **FastAPI** layer is created to support programmatic access to the classifier via API calls.

2. Approach

The project is divided into two primary components:

a. PII Masking

To maintain user privacy and comply with data regulations (like GDPR), it's essential to identify and redact any PII from the email before classification.

✔ Techniques Used:

- **Named Entity Recognition (NER)** with `spaCy` for identifying entities such as:
  - Names
  - Email addresses
  - Phone numbers
  - Credit/debit card numbers
  - CVV, expiry, Aadhar, DOB
- Custom regular expressions for detecting patterns not easily caught by standard NER models.

The identified PII elements are replaced with masked placeholders such as `[NAME]`, `[PHONE]`, `[CARD]`, etc.

b. Email Classification

After masking the PII, the sanitized content is passed to a machine learning model for classification.

✔ Pipeline:

1. **Preprocessing**: Lowercasing, stopword removal, and optional lemmatization.
2. **Vectorization**: TF-IDF (Term Frequency-Inverse Document Frequency) is used to convert text into numerical format.
3. **Model**: A traditional classifier (e.g., **Logistic Regression**) is trained on labeled email data.
4. **Prediction**: The trained model outputs the predicted category.

3. Model Selection & Training

- **Model Chosen**: Logistic Regression(Better compared to SVM and Multinomial Naive Bayes).
- **Training Dataset**: A labeled dataset of emails covering categories such as:
  - Promotional
  - Incident
  - Personal
  - Alert/Warning
- **Evaluation Metrics**: Accuracy, Precision, Recall, F1-score.

4. Deployment

- The complete solution is hosted publicly on Hugging Face Spaces:
  🔗 https://medonajugi-email-classifier.hf.space/
- FastAPI is used to expose the same logic as an API that can be consumed by external applications or integrated into enterprise systems.

5. Conclusion

This project demonstrates how NLP and ML can be effectively combined to build a secure, automated email analysis system. The masking of PII ensures user privacy, while classification supports downstream decision-making processes.

The modular design enables future enhancements such as:

- Adding more PII types (like addresses).
- Using deep learning models like BERT for classification.
- Integration into business tools like Slack or Outlook via APIs.