**Automated Email Processing System for Insurance Policyholders**

**Description:**

This document outlines the approach and implementation of an intelligent system designed to automatically process and categorize incoming email requests from insurance policyholders. The goal of the system is to streamline the handling of policyholder emails, enabling efficient categorization of requests (e.g., claim submission, policy updates, premium inquiries) and derivation of appropriate actions (e.g., creating claim records, scheduling policy reviews). The system utilizes the latest AI models and APIs, such as Gemini API, to analyze and classify email requests, providing actionable insights based on the content.

**Process**

**1. Parsing Incoming Emails**

The system begins by parsing incoming emails from insurance policyholders using the IMAP protocol. IMAP is used to retrieve and manage the emails, ensuring that all relevant emails are collected from the inbox.

**- Email Parsing:** The system accesses the mail server using IMAP and retrieves the email data.

**- Storage:** Once retrieved, the emails are structured and stored in a CSV file for easier manipulation and analysis. The CSV file contains essential fields such as subject, sender, date, and body of the email.

**2. Cleaning the Data**

Once the emails are parsed and stored, the data undergoes a cleaning process to ensure the removal of irrelevant or redundant information. This step ensures that only useful content is analyzed and categorized.

**- Noise Removal:** Unnecessary metadata, such as email signatures, headers, and disclaimers, are removed.

**- Text Preprocessing:** The email body is processed to remove punctuation, special characters, and stop words. Text is then tokenized for better understanding and classification.

**3. Analyzing the Request Using Gemini API**

Due to the lack of specific datasets for insurance email classification, we leverage the Gemini API to analyze the content of each email. Gemini API is used for its robust natural language processing capabilities and up-to-date understanding of language, which helps in identifying the type of request in each email.

- Request Identification: Gemini API analyzes the email content to identify the type of request (e.g., claim submission, policy update, premium inquiry).

**- Categorization:** Based on the analysis, the system categorizes the email into predefined categories, such as "claim," "billing," "policy update," or "premium inquiry."

**- Action Derivation:** Gemini API provides suggestions for the next steps (e.g., creating a claim record, sending premium payment details) based on the request type.

**4. Generating Recommendations and Actions**

Once the email is categorized, the system can generate actionable steps based on the identified request.

**- Automated Responses:** Depending on the category, an automated response is triggered (e.g., providing the required documents for a claim submission or offering policy review details).

**- Workflows:** Specific workflows can be suggested, such as scheduling an appointment for policy review or processing a claim.

**Reasons for Using Gemini API**

**- Lack of Dataset for Email Classification:** There is a significant challenge in finding pre-labeled datasets for classifying insurance-related email requests. By using Gemini API, which is equipped with the latest AI technology and understanding of natural language, the system can accurately categorize emails without needing a pre-trained model on insurance data.

**- Up-to-Date Language Understanding:** Gemini API continuously updates its understanding of the language, ensuring that the system can handle the latest terminologies and requests.

**- Ease of Integration:** Gemini API offers seamless integration with the email processing system, allowing for accurate content analysis and request classification in real time.

**Further Enhancements**

**1. Spam Email Classifier**

One of the planned enhancements is the introduction of a spam email classifier. This classifier will filter out unwanted or irrelevant emails at the beginning of the process. By identifying spam emails early on, the system ensures that only non-spam emails are processed, and recommendations are provided only for legitimate requests.

**2. Fine-Tuning with Large Language Models (LLMs)**

If access to relevant email data becomes available in the future, the system can be further enhanced by fine-tuning large language models (LLMs) specifically on insurance-related email datasets. This will improve the accuracy of the categorization and action recommendations and using multi language models.

**3. Live Dashboard**

A live dashboard could be created to visualize the distribution of emails based on request type. This would include:

- A real-time view of the most common email requests.

- Insights into the most frequent recommendations given by the system.

- Tracking the status of requests and their associated actions.

**4. Scalability via Database Integration**

For better scalability and performance, the system can be connected to a database to store emails, requests, actions, and other metadata. This integration would allow the system to handle large volumes of email data more efficiently and provide a better user experience for policyholders.

**Conclusion**

The intelligent system developed for automating email processing for insurance policyholders streamlines the categorization and action derivation processes, improving operational efficiency. By leveraging the Gemini API for natural language analysis and incorporating potential enhancements such as spam classification and LLM fine-tuning, the system can continue to evolve and provide even greater accuracy and functionality. Future improvements, including the addition of a live dashboard and database integration, will further enhance scalability and user experience.