**Automated Email Classification and Data Extraction for Commercial Bank Lending Service Team**

**1. Introduction**

The Commercial Bank Lending Service Team receives servicing requests via email, often containing attachments. This project aims to automate the classification and data extraction process using **Generative AI (LLMs)** to:

* **Identify the intent of emails** and classify them into predefined request types and sub-request types.
* **Extract key attributes** from email content and attachments for structured service request population.
* **Improve efficiency, accuracy, and turnaround time** by minimizing manual gatekeeping activities.
* **Enable auto-routing** of service requests to the correct processors based on skill-based routing.

This solution leverages freely available tools to automate email processing, attachment extraction, and classification, ultimately reducing human intervention and optimizing workflows.

**2. Solution Overview**

The solution consists of the following steps:

1. **Fetch unread emails** from an email inbox.
2. **Extract text** from the email body and attachments (PDF, DOCX, XLSX).
3. **Classify the email** using **LLMs (Generative AI)** based on predefined request types.
4. **Extract key attributes** from the email and attachments.
5. **Output structured data** for service request population and routing.

**3. Technology Stack**

* **Python** – Primary programming language
* **IMAP Library** – Fetching emails from an inbox
* **BeautifulSoup** – Parsing email content
* **PDFMiner** – Extracting text from PDF attachments
* **python-docx** – Extracting text from DOCX attachments
* **pandas** – Processing Excel files
* **OpenAI API** – Using LLMs for email classification and data extraction

**4. Code Breakdown**

**4.1. Configuration**

The script includes configuration settings for:

* **IMAP Server Credentials**: Email login details for fetching emails.
* **OpenAI API Key**: For calling the LLM to classify emails and extract relevant details.
* **Request Type Mapping**: A predefined list of possible request types and sub-request types.

**4.2. Fetching Emails**

* Connects to the IMAP server and retrieves unread emails.
* Extracts **email subject, sender, date, body, and attachments**.
* Parses email content using **BeautifulSoup**.
* Stores email metadata in a structured format.

**4.3. Processing Attachments**

* Identifies and extracts text from different file types:
  + **PDF:** Uses pdfminer.high\_level.extract\_text()
  + **DOCX:** Uses python-docx
  + **XLSX:** Uses pandas.read\_excel()

**4.4. Classifying Emails Using LLMs**

* Combines **email body and extracted text from attachments**.
* Constructs a structured prompt for OpenAI’s LLM to classify the request.
* Asks LLM to **extract relevant attributes**, such as:
  + Customer ID, Loan ID, Account Number, Requestor Name, Customer Name
  + Amount, Due Date, Request Timestamp, Request Summary
* Receives and parses the LLM’s structured response.

**4.5. Output & Routing**

* The final structured data is **ready for auto-routing** to processors based on skill-based handling.
* Extracted data can be stored in a database or passed to a service request system.

**5. Benefits & Impact**

* **Automates manual classification & data extraction**, reducing errors and processing time.
* **Enhances service request handling efficiency** through structured data extraction.
* **Improves turnaround time** by auto-routing service requests to the right team members.
* **Minimizes human intervention**, allowing teams to focus on higher-value activities.

**6. Future Enhancements**

* Implement a **feedback loop** for continuous model improvement.
* Integrate with **workflow automation tools** for end-to-end request processing.
* Enhance **error handling and logging** for better maintainability.
* Expand LLM training to **improve classification accuracy** based on historical data.

This solution lays the groundwork for an AI-powered workflow automation system, significantly optimizing commercial lending service operations.