Abstract: Automation of Receipt Data Entry

Team ID is Data-230436

Project Resources

- **PPT Link**: https://www.canva.com/design/DAGS6aQkviE/bO5Z2RO3K-RzjhwpSEqiVQ/edit?utm_content=DAGS6aQkviE&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton
- **GitHub Repository**: https://github.com/Jayesh-Gautam/Datamatics-Hackathon

1. Introduction

In today's fast-paced business environment, organisations are continually seeking ways to streamline processes, reduce manual tasks, and enhance operational efficiency. Automation technologies have become vital tools in achieving these goals, especially in the realm of data entry and processing. This project focuses on automating the receipt data entry process for food and petrol bill receipts, leveraging the capabilities of **Datamatics TruBot** and **TruCap+**. The proposed solution significantly minimises manual effort, improves accuracy, and speeds up financial record-keeping for the finance team.

By connecting to an email inbox, the automation solution extracts key details from receipt attachments in PDF or image formats, processes the data using Optical Character Recognition (OCR), and submits it to a **Google Form**. Additionally, the system sends an automatic email notification to the finance team, summarising the extracted data, including key details such as receipt date, receipt number, vendor name, total amount, and items purchased.

2. Project Objective

The primary goal of this project is to automate the extraction of receipt details and the subsequent data entry process. By doing so, the project seeks to:

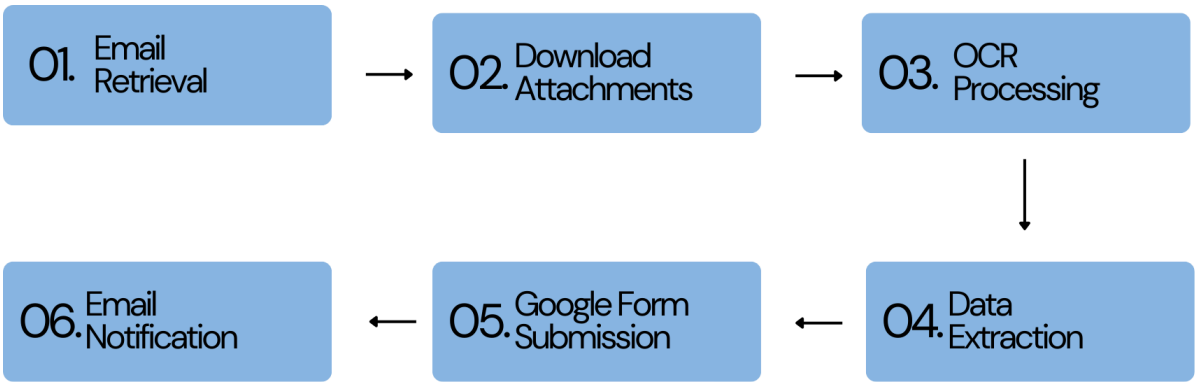
- Eliminate the need for manual data entry tasks typically performed by finance personnel.
- Improve the accuracy of the financial records by minimising human errors.
- Automate the communication process between the system and the finance team through automatic notifications.

The automation process is applied to food and petrol bill receipts, reducing the time spent processing each receipt while ensuring accurate data capture.

3. Methodology

To achieve the project's goals, a step-by-step approach was taken to design, develop, and implement the automation solution. The following workflow outlines the key phases of the process:

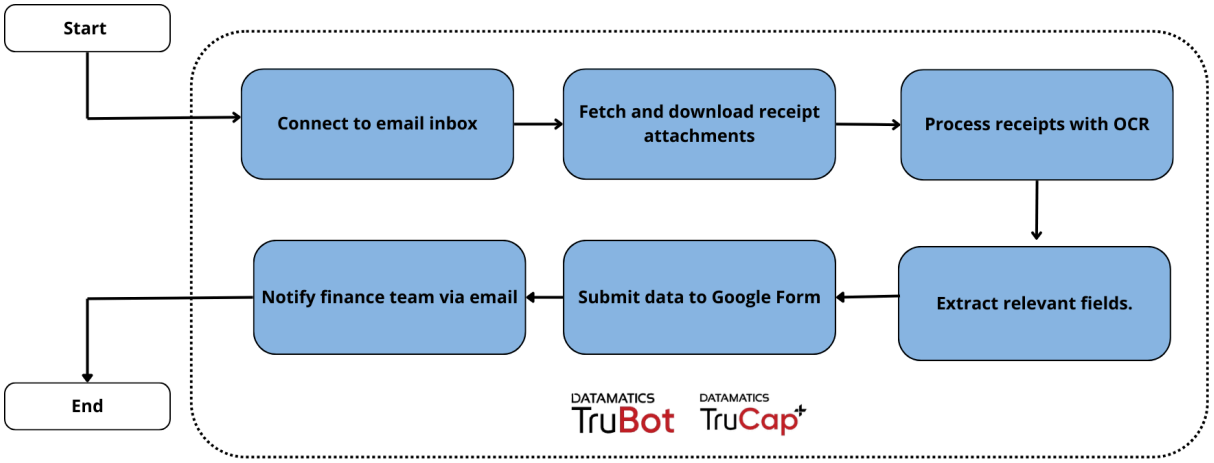
- **Phase 1: Email Retrieval** The system begins by connecting to a designated email inbox, such as Gmail, to fetch emails containing receipts. Specific filters are applied to detect receipts in the inbox. The system is capable of recognizing attachments in either **PDF** or **image** format.
- **Phase 2: Downloading Attachments** Once relevant emails are identified, the system automatically downloads the attached receipts, which are processed further. The solution ensures all receipt attachments are gathered regardless of the file format, ensuring broad compatibility.
- **Phase 3: OCR Processing** Using **TruCap+**, the system processes each receipt via **Optical Character Recognition (OCR)** technology. This step is crucial for extracting textual data from the attached images or PDFs. The OCR technology is configured to recognize key fields such as:
    - **Receipt Date:** Extracted in various common formats (e.g., MM/DD/YYYY, DD/MM/YYYY, Month Day Year).
    - **Receipt Number:** Identified through predefined patterns that include alphanumeric or numeric values.
    - **Vendor Name:** The name of the vendor is extracted by looking for common keywords such as "Vendor," "Supplier," or by analysing the first prominent text lines on the receipt.
    - **Items Purchased:** The items are usually listed under sections labelled "Items" or "Purchased," and the system is designed to extract the details following such labels.
    - **Total Amount:** The final total is detected by locating words such as "Total" or specific currency symbols (e.g., $, £).
- **Phase 4: Data Submission to Google Form** After extraction, the system prepares the data to be sent to a **Google Form** for logging. By using the Google Forms API, the system creates an HTTP request to submit the extracted fields. The submission ensures that the data is recorded in an organised manner, easily accessible for further processing or auditing.
- **Phase 5: Automated Email Notification** After each successful data extraction and form submission, an automatic email notification is sent to the finance team. This email contains a summary of the receipt details, including the total amount and vendor name. This ensures that the team is kept informed of the transactions without having to manually check each receipt.



4. Technical Implementation

The automation solution leverages several technical components and software tools to ensure seamless execution. The primary technologies used in the project include:

- **Datamatics TruBot**: A robust Robotic Process Automation (RPA) platform that automates repetitive tasks by mimicking human actions.
- **Datamatics TruCap+**: An intelligent document processing tool that uses OCR and AI to capture data from structured and unstructured documents, ensuring accurate data extraction from various formats.
- **Google Forms API**: This API is used to submit data directly into Google Forms, ensuring that the finance team has easy access to well-organised records.
- **Email Integration**: The system is designed to connect with a specified email inbox (e.g., Gmail) to fetch and download receipt attachments.



Additionally, the OCR technology has been configured with specific extraction patterns to ensure accuracy in identifying the receipt details. By using machine learning techniques and predefined patterns for common formats like dates, numbers, and item listings, the system enhances the efficiency of data capture.

5. Results and Success Rate

The implementation of this automation project has shown significant results in terms of efficiency, accuracy, and reduction of manual tasks. Based on testing, the success rate of the system in extracting and entering receipt data is estimated to be around **95%**. This high success rate underscores the reliability of the TruBot and TruCap+ tools when it comes to processing receipts.

Key performance indicators (KPIs) include:

- **Reduction in manual work**: The finance team no longer needs to spend valuable time manually entering receipt data.
- **Improved accuracy**: By automating the process, human errors in data entry are virtually eliminated.
- **Timely notifications**: The finance team receives real-time updates on processed receipts through automatic email notifications.

6. Conclusion and Impact

This automation solution has had a tangible impact on the workflow of the finance team, streamlining the process of data entry and improving communication through automatic notifications. By automating the receipt processing task, the project has enhanced both accuracy and efficiency in financial record-keeping.

The project serves as an excellent example of how emerging technologies, such as RPA and intelligent document processing, can be integrated into everyday business operations to achieve operational excellence. This solution not only demonstrates the potential for automation in the finance sector but also sets the stage for further innovation in automating other repetitive, data-driven tasks.

The success of this project paves the way for future enhancements, including the possibility of expanding the automation to handle other types of receipts or invoices and integrating it into broader enterprise resource planning (ERP) systems.

Team Information

- **Team ID**: Data-230436
- **Team Members**:
    - Jayesh Gautam
    - Rikin Parekh
    - Rudra Patel
- **University**: Pandit Deendayal Energy University, Gandhinagar, Gujarat

Contact Information

- **Jayesh Gautam**: gjayesh007@gmail.com
- **Rikin Parekh**: rikinparekh15147@gmail.com
- **Rudra Patel**: rudrapatel3504@gmail.com