

CUSTOMER RELATIONSHIP MANAGEMENT

A PROJECT REPORT

Submitted by

KEERTHIGA (220701125)

in partial fulfillment for the course

OAI1903 - INTRODUCTION TO ROBOTIC PROCESS AUTOMATION

for the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

RAJALAKSHMI NAGAR

THANDALAM

CHENNAI – 602 105

NOVEMBER 2023

RAJALAKSHMI ENGINEERING COLLEGE

CHENNAI - 602105

BONAFIDE CERTIFICATE

Certified that this project report “Generation of employee id and email” is the bonafide work of “**KEERTHIGA P(220701125)**” who carried out the project work for the subject OAI1903- Introduction to Robotic Process Automation under my supervision.

Mrs. J. Jinu Sophia, M.E., (Ph.D.),

SUPERVISOR

Assistant Professor (SG)

Department of

Computer Science and Engineering

Rajalakshmi Engineering College

Rajalakshmi Nagar

Thandalam

Chennai - 602105

Submitted to Project and Viva Voce Examination for the subject OAI1903-
Introduction to Robotic Process Automation held on_____.

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Thiru. S.Meganathan, B.E, F.I.E.**, our Vice Chairman **Mr. Abhay Shankar Meganathan, B.E., M.S.**, and our respected Chairperson **Dr. (Mrs.) Thangam Meganathan, Ph.D.**, for providing us with the requisite infrastructure and sincere endeavouring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N.Murugesan, M.E., Ph.D.**, our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P.Kumar, M.E., Ph.D.**, Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guides, **Mrs. J. Jinu Sophia, M.E., (Ph.D.)**, Assistant Professor (SG), Department of Computer Science and Engineering for their valuable guidance throughout the course of the project. We are very glad to thank our Project Coordinators, **Dr. N. Durai Murugan, M.E., Ph.D.**, Associate Professor, and **Mr. B.Bhuvaneshwaran, M.E.**, Assistant Professor (SG), Department of Computer Science and Engineering for their useful tips during our review to build our project.

KEERTHIGA(220701125)

ABSTRACT

With so many organizations moving away from purchased software and towards cloud-based business process applications, there's a need to transfer data into those new cloud applications. Data Migration is a multi-step process that begins with an analysis of the legacy data and culminates in the loading and reconciliation of data into new applications. With the rapid growth of data, organizations are in constant need of data migration. Data migration can be a complex process where testing must be conducted to ensure the quality of the data. Migration can be very expensive if the best practices are not followed and the hidden costs are not identified at the early stage. In the past, a large-scale transfer like that would require someone to write a complex database code or to task people with the mind-numbing project of manually copying and pasting information. When a content migration depends on a single hand-coded program, there's the risk that the complexity of the code will lead to unforeseen mistakes, or that the incredibly difficult trick of formatting everything correctly will fail. The other option to move large amounts of data was to have humans copy and paste everything. The drawbacks are: paying all that extra staff time, a large amount of human error possible from inattention, a painfully slow upgrade, and the most boring job in history. The goal is to achieve smooth, error-free and easy content migration, system migration, data migration – there are so many ways to refer to the same thing: moving data from one system to a new one and aim is to set up a workflow to retrieve multiple fields and input them into another database with quick speed and accuracy.

Keywords: Data Migration, Legacy Data, Database.

TABLE OF CONTENTS

| CHAPTER NO. | TITLE | PAGE NO. |
|-------------|------------------------------|------------|
| | ABSTRACT | iii |
| | LIST OF FIGURES | vi |
| | LIST OF ABBREVIATIONS | vii |
| 1. | INTRODUCTION | 8 |
| | 1.1 INTRODUCTION | 8 |
| | 1.2 OBJECTIVE | 9 |
| | 1.3 EXISTING SYSTEM | 10 |
| | 1.4 PROPOSED SYSTEM | 12 |
| 2. | LITERATURE REVIEW | 14 |
| 3. | SYSTEM DESIGN | 17 |
| | 3.1 SYSTEM FLOW DIAGRAM | 17 |
| | 3.2 ARCHITECTURE DIAGRAM | 18 |
| | 3.3 SEQUENCE DIAGRAM | 19 |
| 4. | PROJECT DESCRIPTION | 20 |
| | 4.1 MODULES | 20 |
| | 4.1.1 FILE MODULE | 20 |
| | 4.1.2 EXTRACTION MODULE | 22 |
| | 4.1.3 LOADER MODULE | 25 |
| 5. | OUTPUT SCREENSHOTS | 27 |
| 6. | CONCLUSION | 31 |
| | APPENDIX | 32 |
| | REFERENCES | 38 |

LIST OF ABBREVIATIONS

| ABBREVIATION | ACCRONYM |
|--------------|----------------------------|
| RPA | Robotic Process Automation |
| URL | Uniform Resource Locator |

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Customer Relationship Management (CRM) plays a vital role in ensuring businesses effectively manage and enhance their interactions with customers. In today's competitive landscape, providing personalized and timely service has become a cornerstone of customer satisfaction and loyalty. However, traditional CRM processes often involve repetitive, time-intensive tasks, such as entering customer data, tracking interactions, updating records, and generating reports. These manual efforts are not only resource-heavy but also susceptible to human error, which can lead to inefficiencies and reduced customer satisfaction.

To address these challenges, **Robotic Process Automation (RPA)** using UiPath offers a robust solution. RPA leverages software robots to automate mundane, rules-based tasks, enabling organizations to streamline their workflows while maintaining precision and speed. By integrating UiPath into CRM operations, businesses can optimize their processes, reduce operational costs, and focus on delivering exceptional customer experiences. This project demonstrates how RPA can revolutionize CRM by automating key processes like data management, customer support, and reporting, ultimately enhancing business efficiency and customer engagement. Through UiPath's intuitive platform, organizations can achieve seamless CRM automation, driving productivity, accuracy, and scalability.

1.2 OBJECTIVE

The primary objective of this project is to optimize and enhance the efficiency of Customer Relationship Management (CRM) processes through the implementation of Robotic Process Automation (RPA) using UiPath. By automating repetitive and manual tasks such as data entry, record updates, and report generation, the project aims to minimize human error and ensure data accuracy within the CRM system.

Another key objective is to improve the customer experience by enabling faster response times and providing consistent service. This involves streamlining workflows and integrating the CRM system with other enterprise applications, ensuring seamless data exchange and efficient process management. The project also seeks to generate actionable insights by automating the collection, processing, and analysis of customer data, empowering businesses to make data-driven decisions.

1.3 EXISTING SYSTEM

In many organizations, the existing Customer Relationship Management (CRM) systems are often heavily reliant on manual processes, which can lead to inefficiencies, human errors, and a lack of scalability. These systems typically require employees to manually input customer data, update records, and generate reports. For example, when a customer contacts the company for support or to make an inquiry, a support agent must manually enter the customer's information into the CRM system. This information may need to be updated every time the customer reaches out, which is both time-consuming and prone to errors. As businesses grow and their customer bases expand, these manual processes become even more burdensome, leading to longer response times and potential lapses in communication.

In addition to time-consuming manual data entry, existing CRM systems often struggle with integrating data from multiple platforms. For instance, the CRM might be isolated from other essential business tools such as marketing platforms, billing systems, or customer support tools. This creates a siloed environment where employees are forced to switch between multiple systems to find the information they need. When a sales representative wants to follow up with a lead, they might need to search through emails, spreadsheets, and multiple CRM databases, leading to inconsistent information and inefficiencies in tracking customer interactions. This lack of integration not only increases the chances of data duplication but also prevents the company from having a comprehensive, up-to-date view of its customers.

In many organizations, the existing Customer Relationship Management (CRM) system is heavily reliant on manual data entry, basic automation, and disjointed workflows. For example, a company might use a traditional CRM software like **Salesforce** or **Zoho CRM**, but the processes for handling customer information, communication, and reporting are still largely performed by human employees.

Reporting is another area where the existing system falls short. While CRM systems can store vast amounts of customer data, generating insights or reports often requires manual effort. Employees must extract data, perform analysis, and create reports on a periodic basis, which can be both time-consuming and error-prone. Moreover, these reports may not always be available in real time, preventing decision-makers from accessing timely data to make informed choices. This lack of real-time insights limits the organization's ability to respond quickly to changing customer needs or market trends, putting the business at a competitive disadvantage.

1.4 PROPOSED SYSTEM

The proposed system for the Customer Relationship Management (CRM) project leverages Robotic Process Automation (RPA) with UiPath to transform the way customer interactions are handled, making the entire process more efficient, accurate, and scalable. The primary goal of this system is to automate repetitive, time-consuming tasks, such as data entry, updating customer records, and generating reports, which are often prone to human error in the current manual system. By implementing automation, these tasks will be completed faster, with fewer mistakes, leading to better data accuracy and enhanced overall productivity.

One of the core strengths of the proposed system is its ability to integrate seamlessly with other enterprise tools, including marketing automation platforms, email systems, sales applications, and order management software. This integration ensures that data flows effortlessly across different systems, providing a unified view of the customer journey and eliminating the need for employees to manually switch between various applications. As a result, employees can access real-time information without delays, enabling them to serve customers more effectively and make faster, informed decisions.

The automation of customer service processes, such as managing support tickets and handling follow-ups, will directly enhance the customer experience. With RPA bots taking care of logging, assigning, and tracking service requests, customer queries will be addressed more quickly and efficiently. Moreover, automated notifications and updates will keep customers informed throughout the resolution process, leading to increased satisfaction. As a result, businesses will be able to handle a larger volume of inquiries without compromising service quality, improving customer retention rates.

The proposed system is also highly scalable, designed to accommodate the growing needs of the business. As the customer base and data volumes expand, the automated processes will ensure that the system remains efficient, without requiring additional manual intervention. Whether the business expands locally or globally, the system can scale accordingly, adapting to increased demands without sacrificing quality or performance.

Finally, the incorporation of RPA allows employees to shift their focus from repetitive tasks to more strategic, value-added activities, such as customer relationship building, sales strategy development, or personalized marketing. By reducing the manual burden on employees, the system enhances overall workforce productivity, allowing the business to focus on growing customer relationships rather than getting bogged down in administrative work.

The proposed CRM system powered by UiPath RPA will modernize the organization's approach to customer relationship management by automating manual processes, improving operational efficiency, and delivering better customer experiences. It will streamline workflows, ensure data accuracy, provide real-time insights, and scale seamlessly to meet the evolving needs of the business. This transformation will result in a more agile, responsive, and customer-centric organization.

CHAPTER 2

LITERATURE REVIEW

The study of Customer Relationship Management (CRM) has undergone significant advancements over the decades, with researchers exploring its evolution, challenges, and the impact of emerging technologies. Initially, CRM systems were introduced as tools for managing customer data and interactions, enabling businesses to improve customer retention and satisfaction. **Kotler and Keller (2006)** emphasized the strategic importance of CRM in fostering long-term relationships and enhancing customer lifetime value. However, these early systems were largely manual, focusing on basic functions such as data storage and retrieval, which limited their ability to support complex business processes effectively. **Peppers and Rogers (1993)** expanded this understanding by shifting the focus from transactional models to relationship-oriented approaches, advocating for personalized customer engagement as a means to build trust and loyalty.

While traditional CRM systems played a critical role in centralizing customer information, they were not without challenges. One major limitation identified in the literature was their reliance on manual processes, which made them time-consuming and prone to human error. Employees were tasked with repetitive activities such as data entry, updating records, and generating reports, which often led to inefficiencies and inconsistencies in customer data. Studies by **Syam and Sharma (2018)** highlighted how these inefficiencies could result in missed opportunities, reduced productivity, and fragmented customer experiences.

Furthermore, the lack of integration between CRM systems and other business tools created data silos, making it difficult to maintain a unified view of customer interactions. This fragmentation not only slowed down operations but also hindered organizations' ability to deliver consistent and personalized services.

The limitations of traditional CRM systems have driven researchers and businesses to explore the potential of emerging technologies to improve efficiency and scalability. Robotic Process Automation (RPA) has emerged as a game-changing solution for addressing the challenges associated with manual processes in CRM. RPA utilizes software bots to automate repetitive, rule-based tasks, such as entering data, updating customer records, and generating reports. Aguirre and Rodriguez (2017) demonstrated that RPA significantly reduces processing times and enhances accuracy in CRM workflows, thereby improving overall operational efficiency. By automating mundane tasks, organizations can redirect employee efforts toward more strategic activities, such as building stronger customer relationships and crafting innovative solutions.

The integration of RPA into CRM systems offers several tangible benefits, as highlighted in recent studies. Willcocks et al. (2015) emphasized that RPA enables seamless integration between CRM systems and other business tools, ensuring smooth data flow and eliminating redundancies. This integration allows businesses to maintain a comprehensive view of customer interactions across multiple platforms, facilitating better decision-making and more effective customer engagement. Furthermore, RPA-powered CRM systems can process data in real time, providing organizations with up-to-date insights and analytics. According to van der Aalst et al. (2018), these real-time capabilities empower decision-makers to respond swiftly to customer needs and market changes, enhancing the organization's agility and competitiveness.

Similarly, a retail company used RPA to automate customer feedback analysis, enabling them to identify trends and improve service delivery in real time. These examples illustrate how RPA can enhance CRM systems' ability to handle high volumes of data and interactions, making them more scalable and adaptable to growing business demands.

The literature surrounding CRM systems highlights a clear progression from manual, resource-intensive approaches to highly automated, integrated solutions. The incorporation of RPA has proven to be a pivotal development, addressing many of the inefficiencies and limitations of traditional CRM systems. By automating repetitive tasks, improving data accuracy, and enabling real-time insights, RPA has transformed CRM into a more scalable and efficient tool for managing customer relationships. As businesses continue to adopt and refine these technologies, future CRM systems are poised to become even more dynamic and customer-centric, driving long-term growth and satisfaction. This growing body of research underscores the critical role of technology in shaping the future of customer relationship management.

CHAPTER 3 SYSTEM DESIGN

3.1 SYSTEM FLOW DIAGRAM

A flowchart is a type of diagram that represents an algorithm, workflow or process. The flowchart shows the steps as boxes of various kinds, and their order by connecting the boxes with arrows. This diagrammatic representation illustrates a solution model to a given problem.

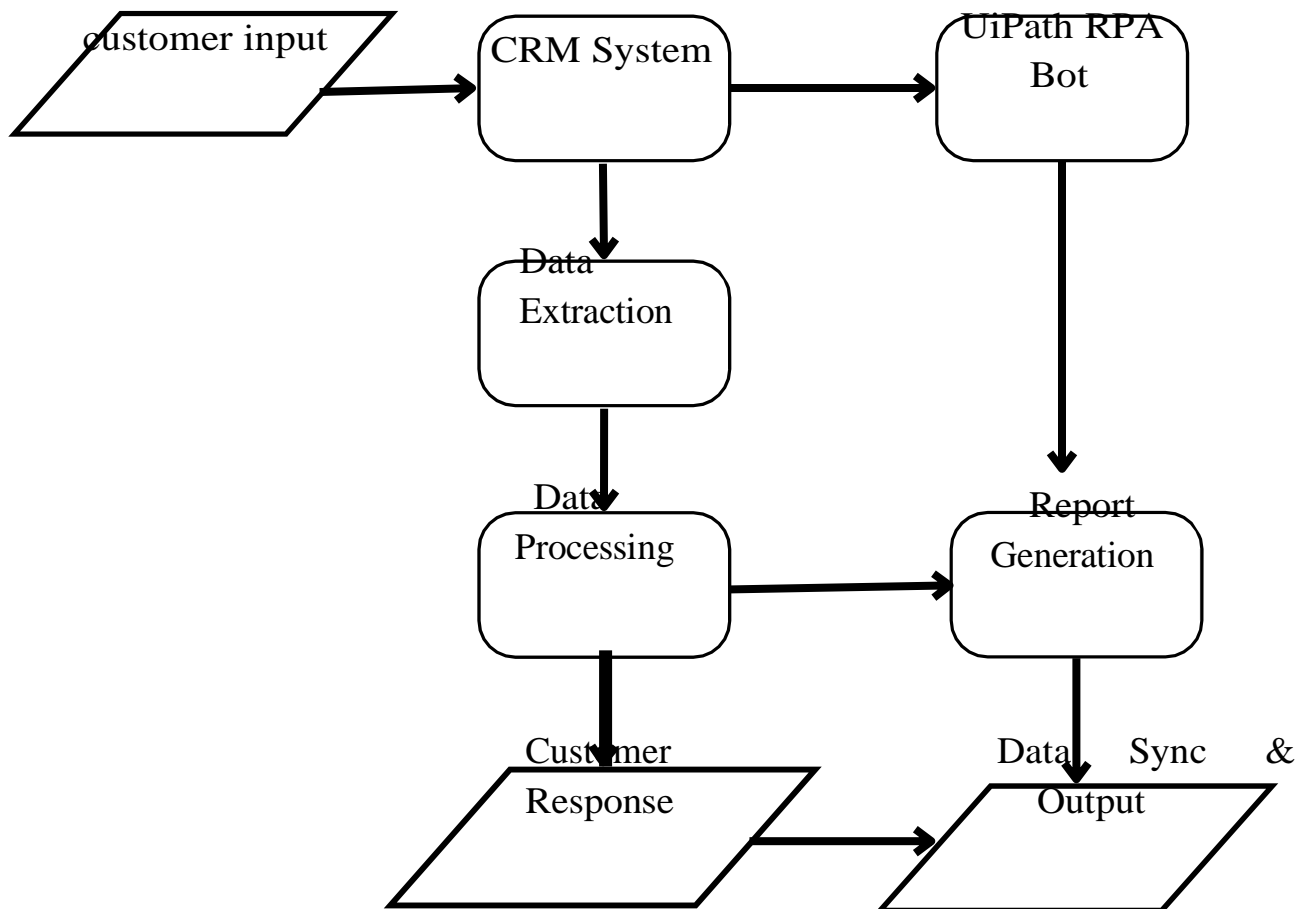


Fig 3.1 System Flow Diagram

3.2 ARCHITECTURE DIAGRAM

An architecture diagram is a graphical representation of a set of concepts, that are part of an architecture, including their principles, elements and components.

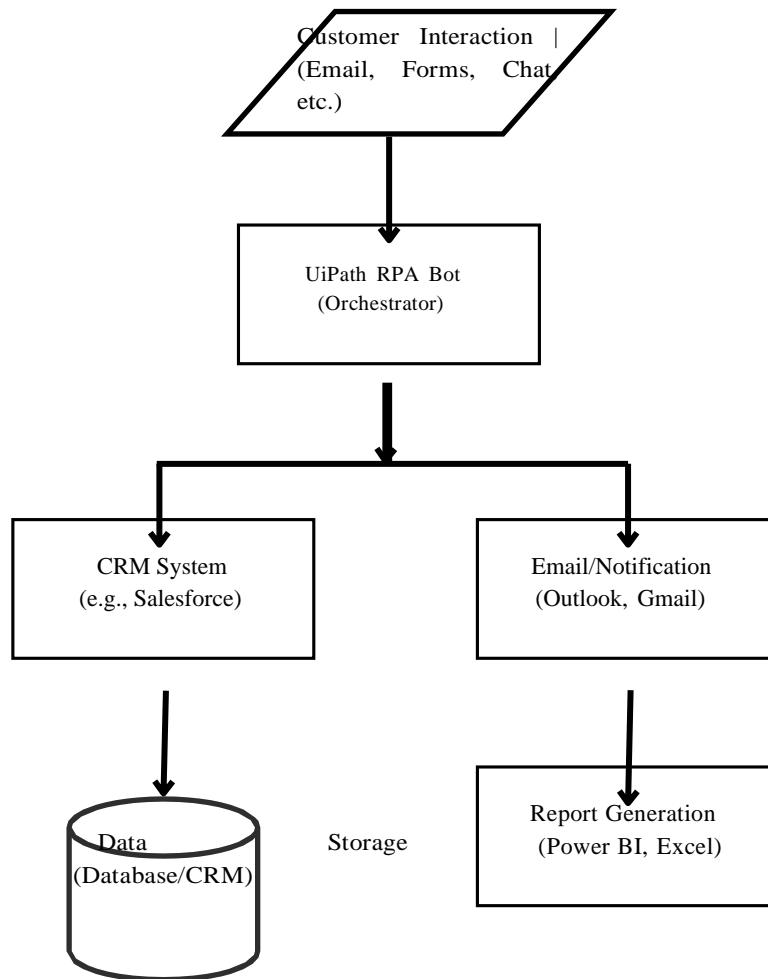


Fig 3.2 Architecture Diagram

3.3 SEQUENCE DIAGRAM

A sequence diagram is a type of interaction diagram because it describes how— and in what order—a group of objects works together.

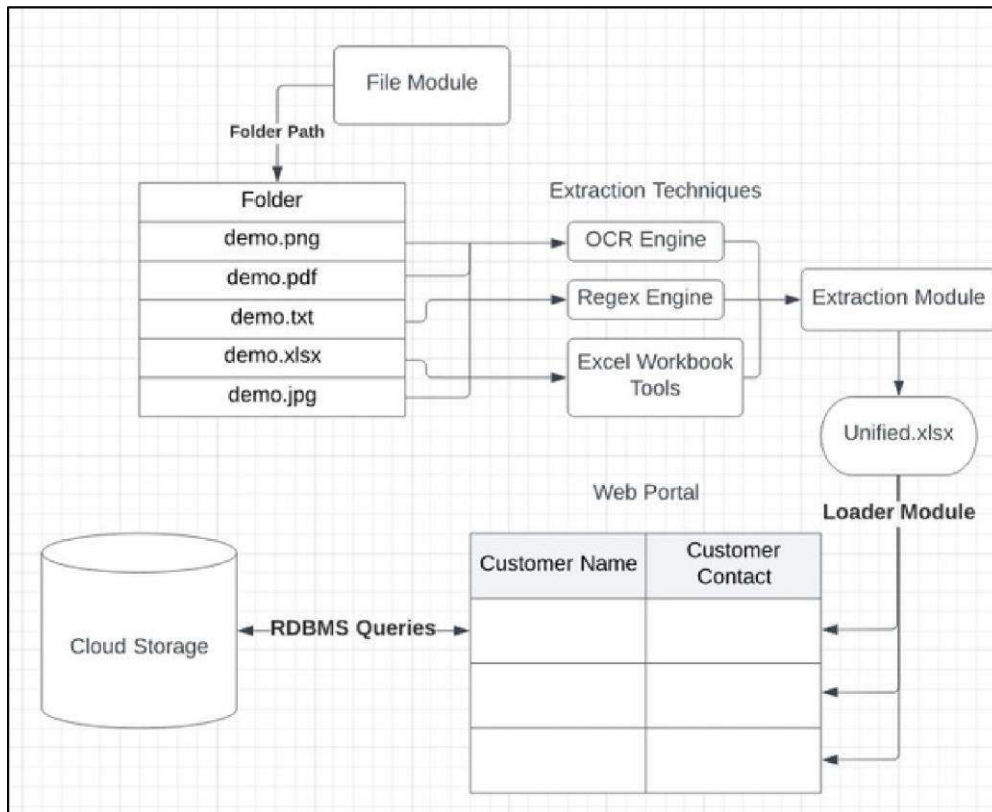


Fig 3.3 Sequence Diagram

CHAPTER 4

PROJECT DESCRIPTION

4.1 MODULES

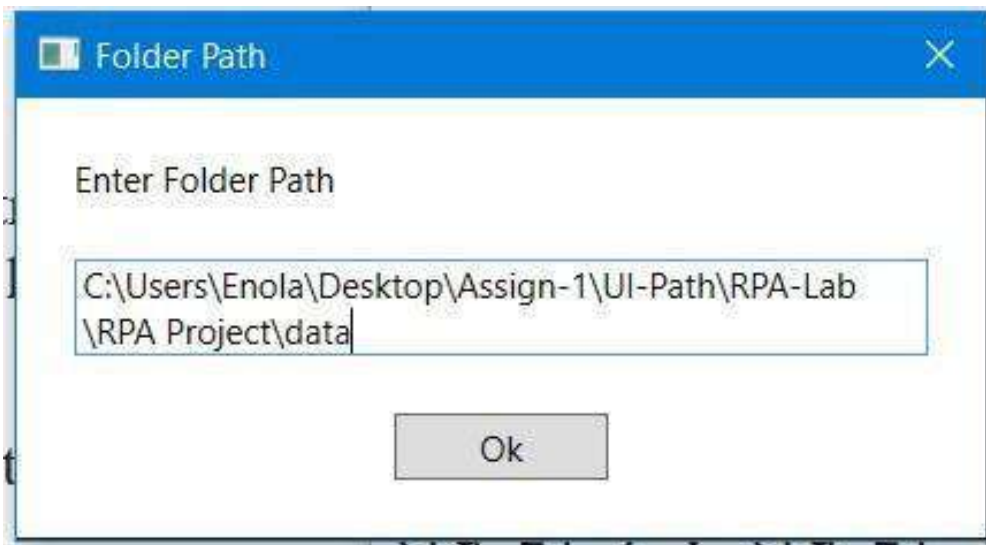
4.1.1 FILE MODULE

The File Management Module is an essential component in the CRM RPA automation project, aimed at automating the storage, retrieval, and processing of files and documents within the CRM system. This module plays a critical role in streamlining operations that involve managing customer-related files, such as contracts, support tickets, invoices, or any other documentation linked to customer interactions. Automating these tasks significantly reduces the manual labor associated with organizing, storing, and retrieving files, ensuring a more efficient workflow, especially when dealing with large volumes of customer data.

Purpose and Importance

In many CRM systems, data is not only stored in databases but also in various file formats (e.g., PDFs, Excel sheets, Word documents, etc.). These files may contain crucial customer-related information such as purchase history, service agreements, invoices, customer feedback, and other forms of communication. Manually processing these files can be time-consuming, error-prone, and inefficient. The File Management Module addresses these challenges by automating file-related tasks, ensuring that files are correctly named, stored in the appropriate directories, and processed according to business rules.

By implementing this module, companies can significantly reduce the time spent on searching for and managing customer files, allowing employees to focus on more value-added activities such as customer support and relationship building. Additionally, the automation ensures that files are organized consistently, stored securely, and retrieved efficiently, which is critical for compliance and data governance.



4.1.2 EXTRACTION MODULE

The Extraction Module is a core component of the CRM RPA project designed to automate the extraction of critical data from various unstructured and semi-structured documents, emails, forms, and other customer-related files. In traditional CRM workflows, employees are often required to manually extract key information from emails, invoices, support tickets, PDFs, and customer forms, which is not only time-consuming but also prone to errors. With the help of RPA, specifically UiPath, this process can be automated to drastically reduce manual input, enhance data accuracy, and improve overall operational efficiency.

The extraction module plays a pivotal role in processing large volumes of data, transforming it into structured, usable information that can be directly integrated into CRM systems. This automated data extraction process ensures that customer records are up-to-date and accurate, and it eliminates the need for employees to spend valuable time on mundane, repetitive tasks.

Additionally, it enables companies to respond faster to customer requests, resolve support tickets promptly, and maintain better communication with clients by always having accurate data at hand.

The module leverages a combination of technologies such as Optical Character Recognition (OCR), Data Scraping, Natural Language Processing (NLP), and AI-driven document processing to handle different types of data sources, including scanned images, web forms, emails, and structured documents like PDFs and Excel sheets. By using UiPath's native tools and integration capabilities, businesses can ensure seamless and error-free extraction of critical customer data from various channels. Let's dive deeper into the key features and capabilities of the Extraction Module.

One of the primary features of the Extraction Module is the integration of Optical Character Recognition (OCR) technology. OCR allows the RPA bot to read and extract text from scanned documents, images, and PDF files that were not originally in a machine-readable format. When a customer submits an invoice, a contract, or any other document in a non-editable format, OCR is used to scan and extract relevant data like customer names, product details, pricing information, and transaction dates.

UiPath supports several OCR engines, such as Microsoft OCR, Google OCR, and Tesseract OCR, which help improve accuracy and speed when extracting text from a variety of document types. This capability is essential when dealing with large volumes of paper-based or image-based documents, as it eliminates the need for manual data entry and ensures that data is captured accurately.

For example, when an invoice is uploaded as a PDF, the RPA bot uses OCR to extract key details such as the invoice number, customer name, invoice amount, and due date. The bot then stores this information in the CRM system and triggers any necessary workflows, such as sending a payment reminder or generating a receipt.

In many CRM systems, data is often input through web forms or email communication. The Data Scraping feature within UiPath enables the RPA bot to automatically extract specific pieces of data from web pages, HTML forms, and email bodies. This is especially useful when customers submit queries via email, contact forms on websites, or online surveys.

Data scraping within UiPath is versatile, supporting a wide range of formats, such as HTML tables, text fields, and lists. By automating the extraction of data from emails or forms, the bot reduces human involvement in copying and pasting information, ensuring that customer data is accurately captured and processed in real time.

The Text Analytics and Natural Language Processing (NLP) capabilities in the Extraction Module allow the bot to understand and extract data from unstructured text, such as customer feedback, product reviews, or support tickets. NLP techniques enable the RPA bot to analyze text, interpret sentiment, and classify the data into meaningful categories, making it easier for the CRM system to process.

For example, if a customer submits a support ticket or email that contains free-text content, the bot can analyze the text to determine the intent behind the request, categorize it (e.g., "Technical Issue," "Billing Query," or "Customer Feedback"), and then automatically route the request to the appropriate department.

This ability to extract insights from unstructured data enhances the CRM system's ability to respond quickly and effectively to customer needs. For instance, if a customer sends an email saying, "I have a problem with my last purchase," the RPA bot can identify the issue, categorize it, and update the CRM accordingly, ensuring that a customer service representative can quickly address the concern.

Another critical feature of the Extraction Module is the ability to automate the extraction of data from structured forms, including PDFs, Excel sheets, and custom form templates. Form Data Extraction involves automatically reading and processing forms to pull out specific data fields such as customer names, addresses, phone numbers, and other information typically captured in forms. For example, businesses often receive customer feedback forms, registration forms, or survey forms, all of which contain valuable data. The RPA bot can scan these forms and extract structured data based on predefined templates.

This process is particularly useful for handling documents that are filled out electronically or scanned into PDF formats, enabling the bot to process large amounts of data quickly and efficiently.

4.1.3 LOADER MODULE

The Loader Module is a key component in an RPA-based CRM system, designed to automate the process of transferring extracted data into the CRM system after it has been processed, validated, and cleaned. This module ensures that customer data is accurately entered into the CRM system in a timely and error-free manner. The Loader Module plays an essential role in enhancing the data flow and ensuring that customer-related information, such as inquiries, feedback, purchases, and support requests, are correctly updated and stored in the CRM database. By automating the loading process, businesses can eliminate manual data entry, increase efficiency, and minimize the potential for errors.

In many traditional CRM workflows, entering customer data into the system is a time-consuming and error-prone process. The Loader Module addresses these challenges by automating the task of feeding data into the CRM system after it has been extracted from emails, forms, documents, or other sources. This module also helps ensure that customer data is consistent, up-to-date, and ready for analysis or use by other departments, such as sales or customer support. The ability to automate data loading allows businesses to quickly process large volumes of customer data, facilitating faster decision-making and improved customer service.

Data Mapping and Transformation The first function of the Loader Module is to perform data mapping. Once the data is extracted, it needs to be matched with the corresponding fields in the CRM system. This process, called data mapping, ensures that data from the extracted sources, such as forms or documents, is properly aligned with the appropriate CRM fields. For example, an extracted email address from a customer inquiry must be mapped to the email field in the CRM.

Batch Processing of Data The Loader Module is designed to handle large volumes of data efficiently. Many CRM systems require batch processing when loading new customer records or updates, especially when extracting data from multiple sources at once. The batch processing feature ensures that the data is processed in bulk, reducing processing time and improving the efficiency of the system.

Error Handling and Validation Ensuring the integrity and quality of the data being loaded into the CRM system is crucial. The Loader Module performs error handling and validation at different stages of the data loading process. Before loading data into the CRM system, the module checks if all required fields are populated and that the data adheres to the expected format or business rules.

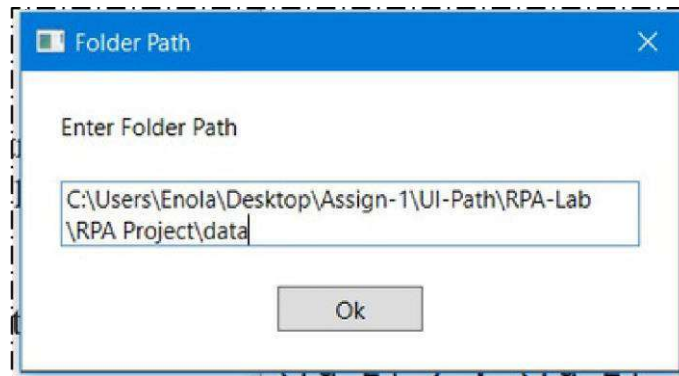
Real-Time Data Synchronization A key feature of the Loader Module is real-time data synchronization with the CRM system. As customer data is extracted and processed, it is immediately synchronized and updated in the CRM system, ensuring that customer information is always up-to-date. This is particularly important in environments where timely data updates are crucial, such as customer support, sales teams, and marketing campaigns.

The Loader Module can be set up to trigger data loads automatically whenever new data is available. For instance, if a customer submits an online inquiry or a support ticket, the bot extracts the relevant data and synchronizes it with the CRM system in real-time. This ensures that the CRM system always reflects the most current customer information, reducing the need for manual updates and enabling teams to work with accurate data at all times.

CHAPTER

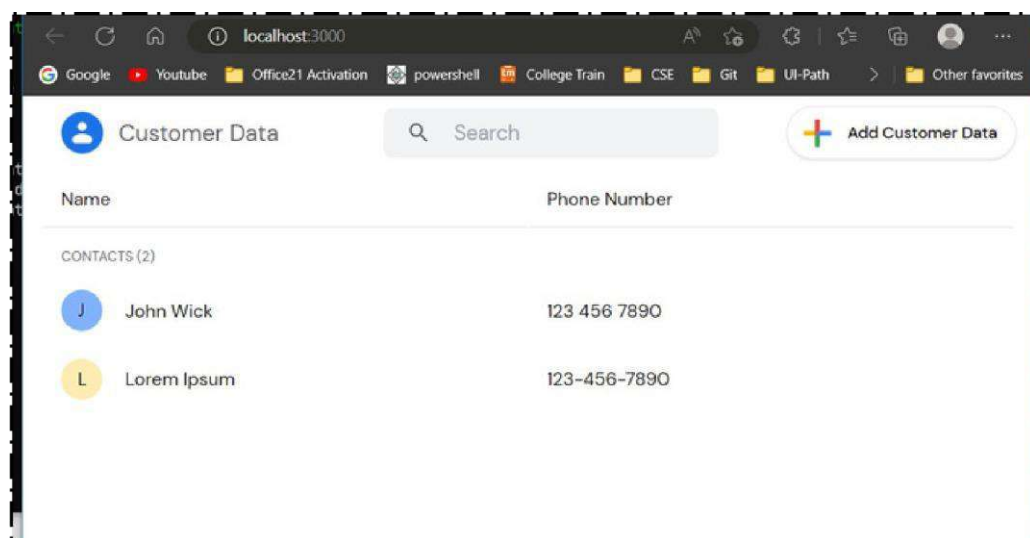
OUTPUT SCREENSHOTS

Step 1: Taking the folder path where the customer data is stored

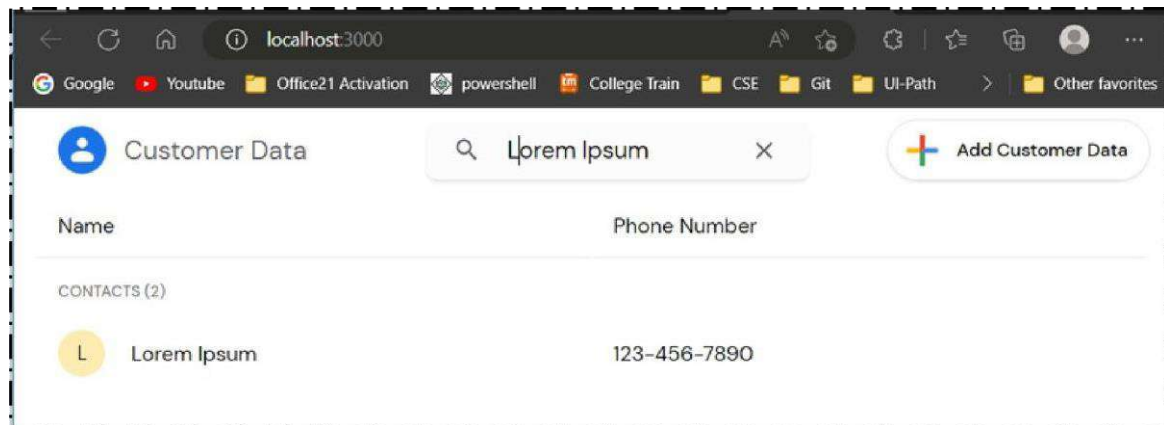


Step 2: The workflow loops through all the files present in the path and extracts data from all the files and stores it a common unified excel file

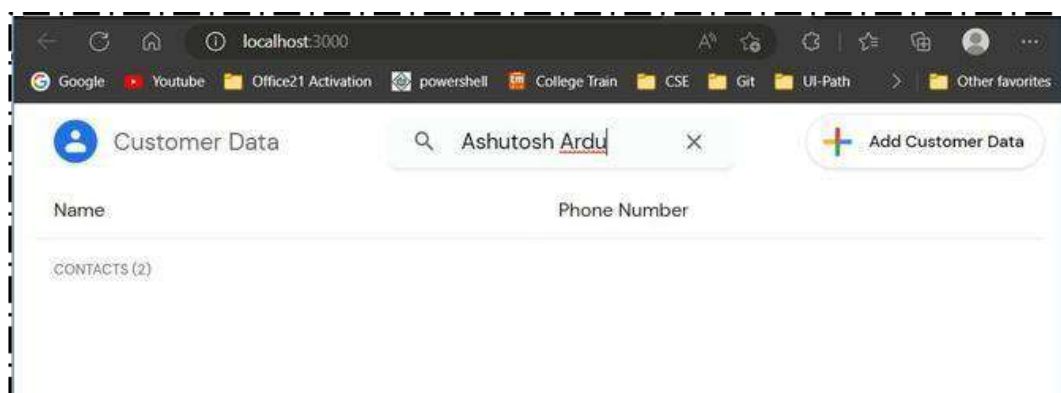
Step 3: The workflow now opens the web portal and loads the customer data onto the server/database.

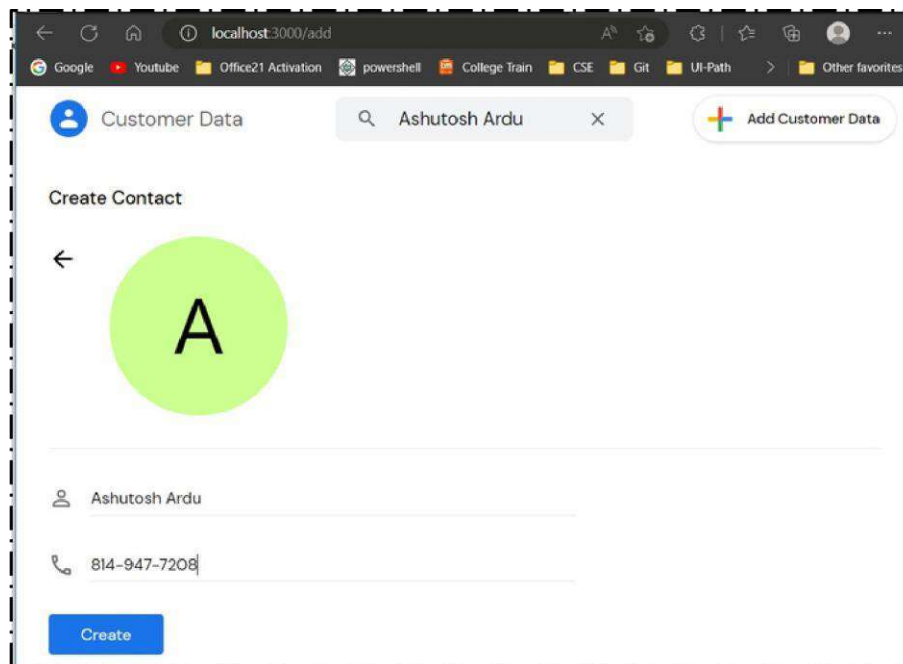


If the customer entry is already present it continues on to the next data row in the unified excel file

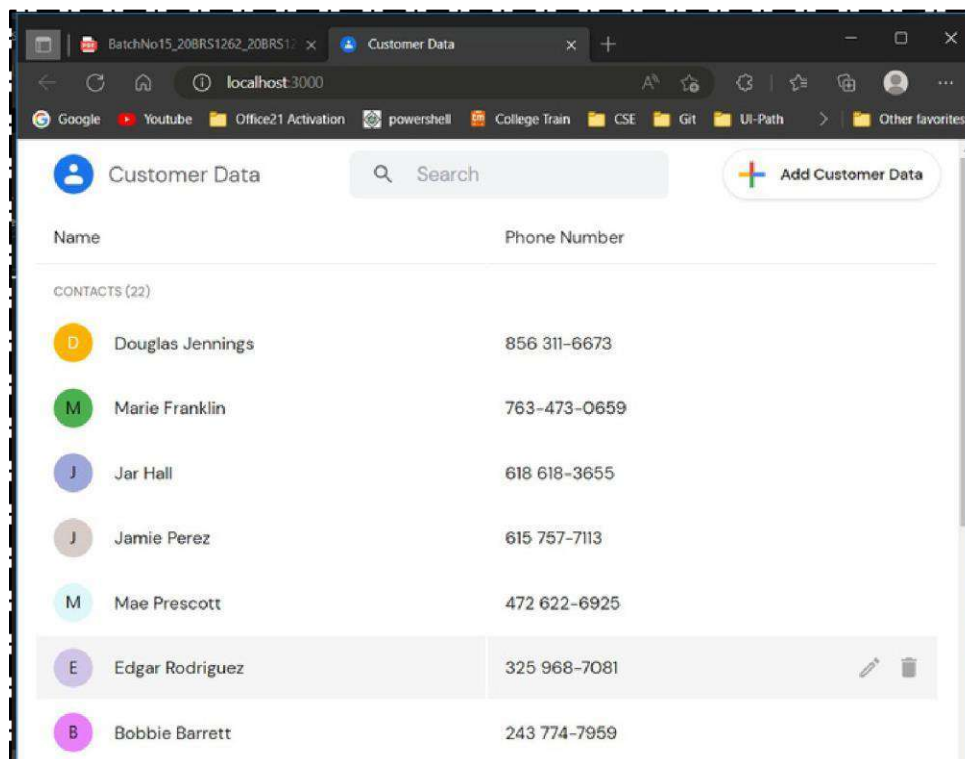


If customer data is not yet entered then it creates a new customer contact



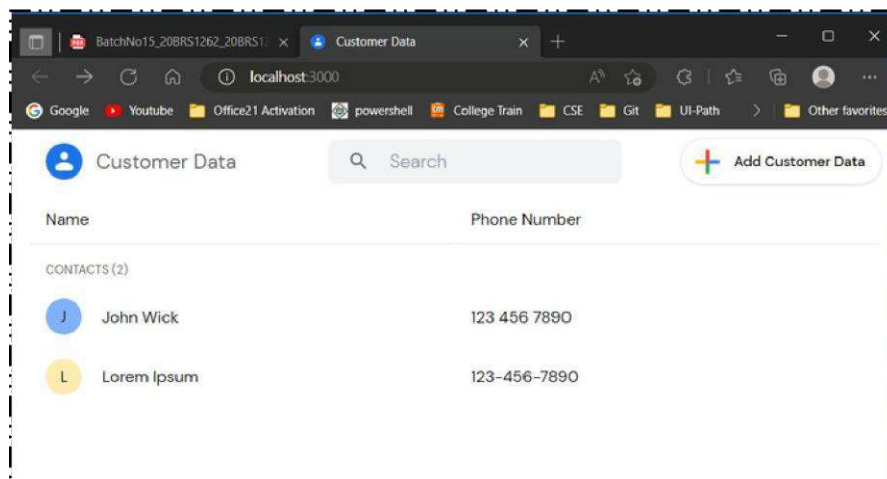


Website (Final State after running the workflow)

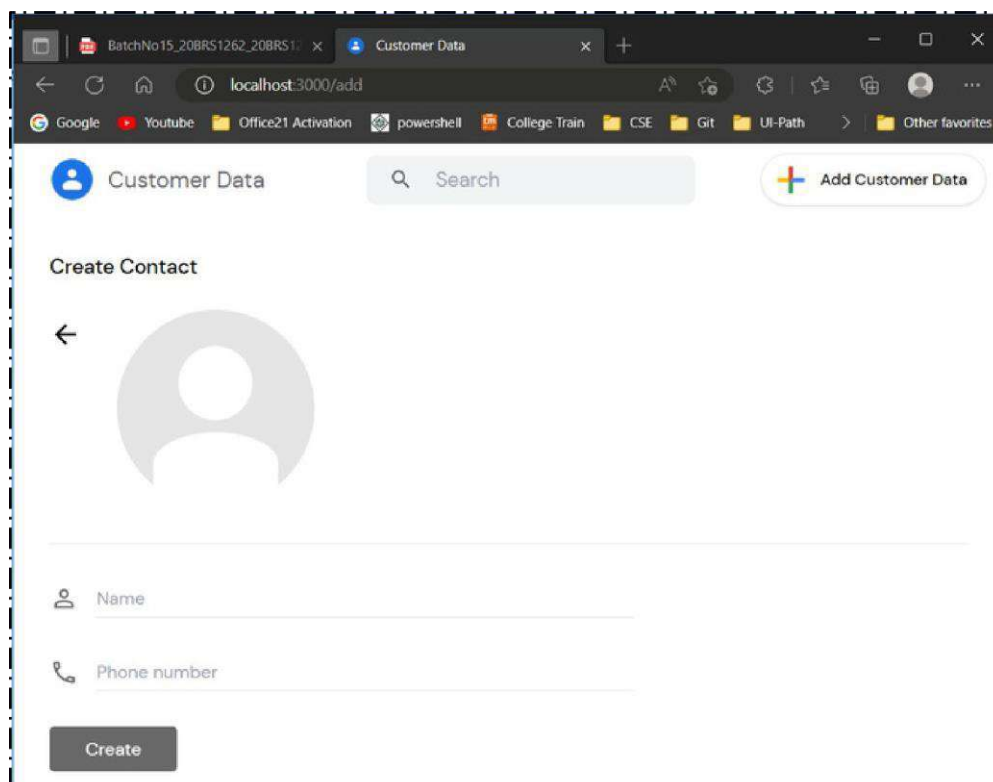


Website

Home Page (Initial State)



Create Customer Page



CHAPTER 6

CONCLUSION

The CRM RPA Project using UiPath has made a huge difference in how businesses manage their customer relationships. By automating key processes like data extraction, validation, and loading, companies can save a lot of time and reduce human errors. The Extraction Module makes it easier and faster to pull important customer information from emails, forms, and documents, while the Loader Module ensures this data gets smoothly entered into the CRM system and kept up-to-date in real time.

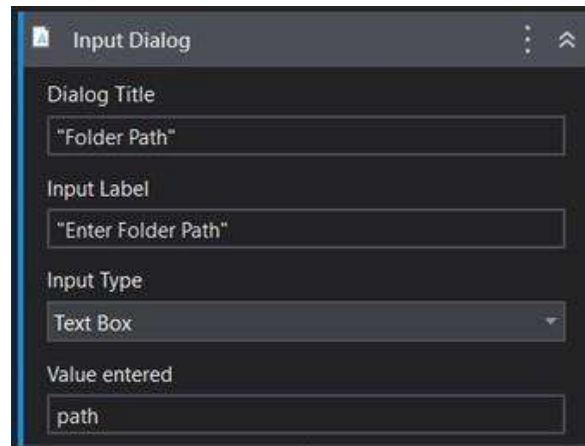
What's great is that all these automated tasks allow businesses to focus on more important things, like improving customer service or making better decisions.

With accurate and current data always available, teams can respond to customer inquiries faster, process orders more efficiently, and handle customer feedback better. This leads to better customer experiences overall.

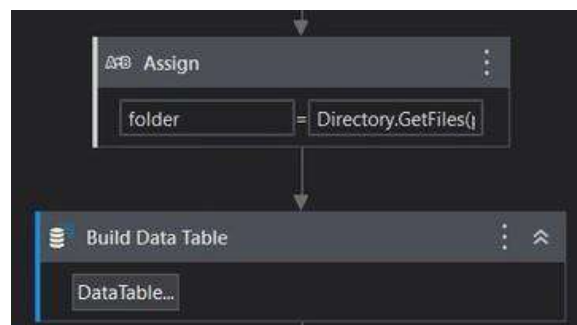
In short, using UiPath for CRM automation helps businesses save costs, increase productivity, and scale their operations without much extra effort. By reducing manual work and increasing the speed of data processing, companies can stay ahead of the competition and create stronger connections with their customers.

APPENDIX SAMPLE PROCESS

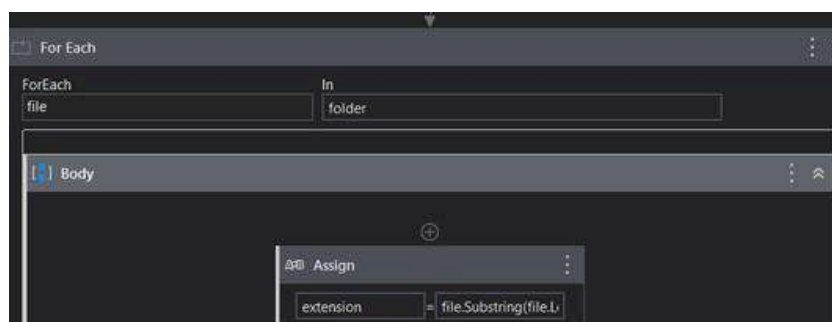
Workflow Taking the file path where all the customer data in different file format is stored



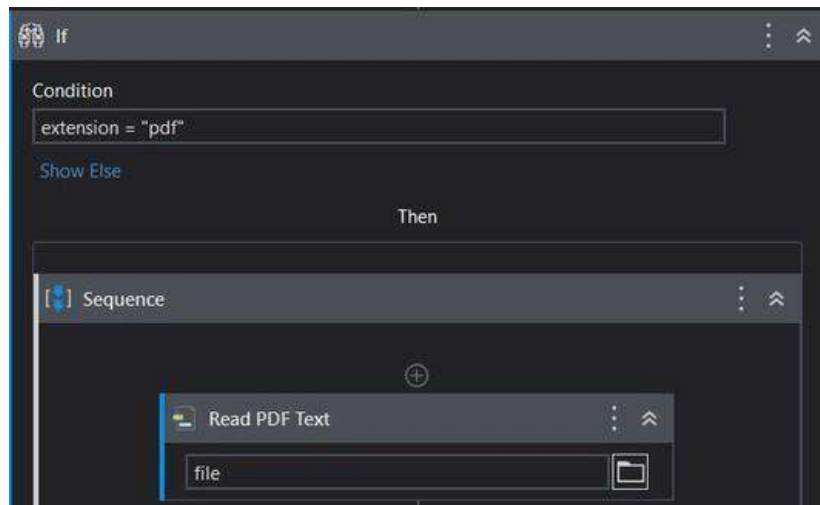
Extracting the list of files present in that directories and building a datatable to store all the customer data present in different file formats.



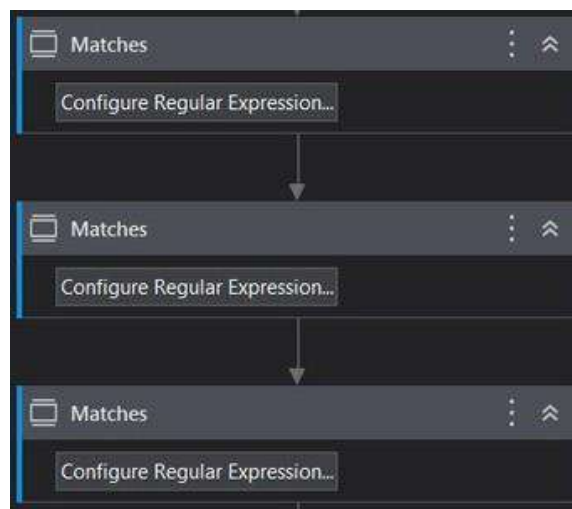
Looping through all the files present in the directory and extracting their file extension to identify the file formats



If current file type is of PDF format, extracting the text present in the pdf and applying regex modules to extract customer name, email and phone numbers



Regex Module



Regex Module for extracting name

| Value | Quantifiers |
|---|--------------------------|
| <code>(([a-z]+) : ([a-z]+(\s[a-z]+)? [a-z]+)</code> | At least one (1 or more) |

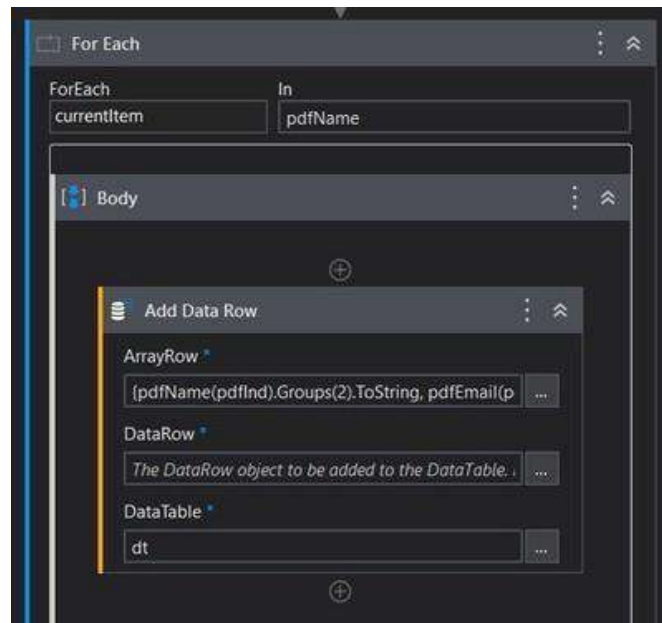
Regex Module for extracting email

| RegEx | Value | Quantifiers |
|-------|---|--------------------------|
| Email | <code>((?>[a-zA-Z\d!#\$%&*+\-\\/=?^_`{ }~]+\\x20)*"((?=[\x01-\x7f])[^"\\\])</code> | At least one (1 or more) |

Regex Module for extracting phone number

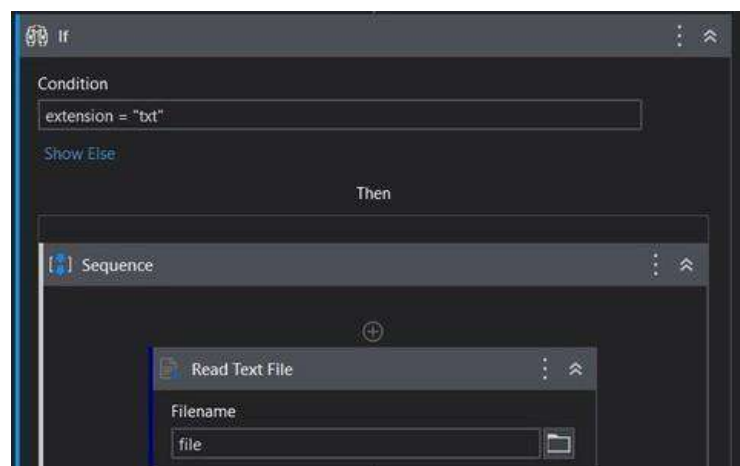
| Value | Quantifiers |
|--|--------------------------|
| <code>(91)?([0-9]{3})[-]?([0-9]{3})[-]?([0-9]{4})</code> | At least one (1 or more) |

Data Row Append Module for adding the extracted customer data using regex

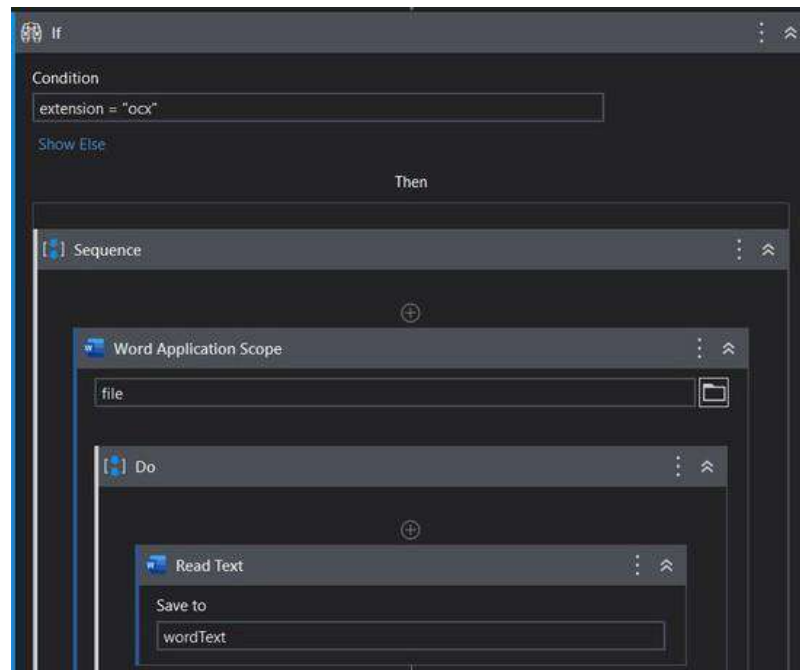


Array Object loaded into the Datatable

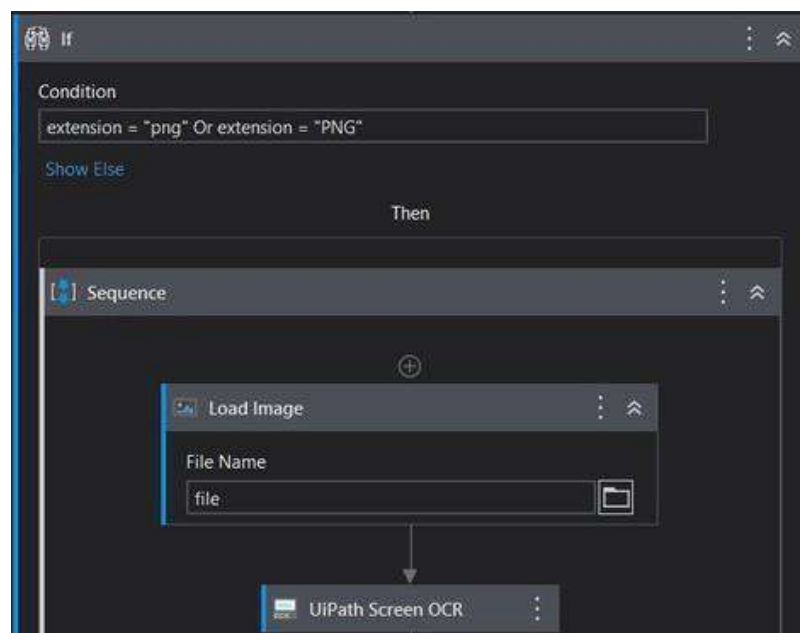
```
ArrayRow (Object[])  
1 {pdfName(pdfInd).Groups(2).ToString, pdfEmail(pdfInd).ToString, pdfNumber(pdfInd).ToString}
```



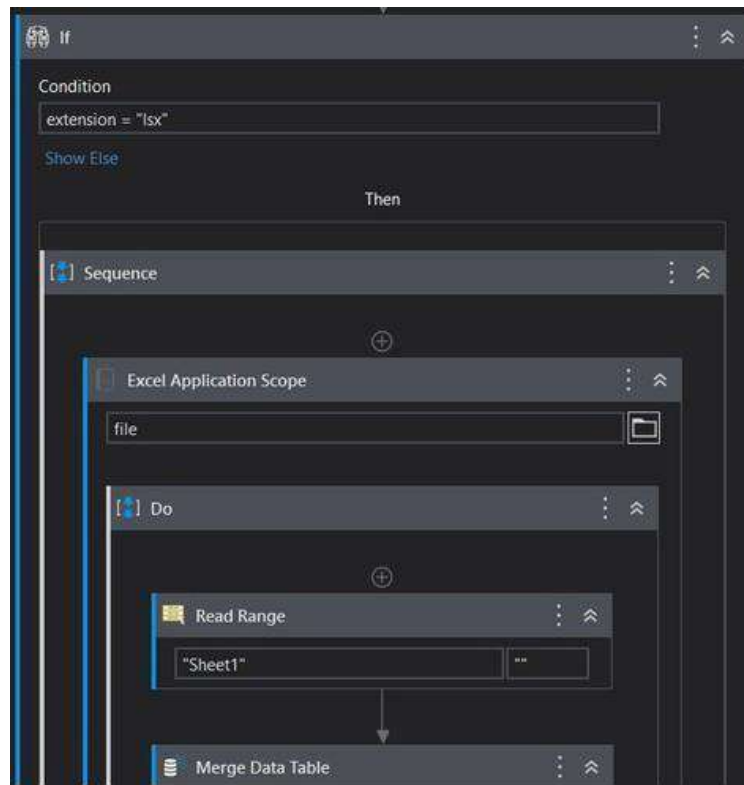
If file is a docx type file then we extract text using the following activities and later we apply the regex module to extract the customer data followed by data row append module to add the data to the common datatable



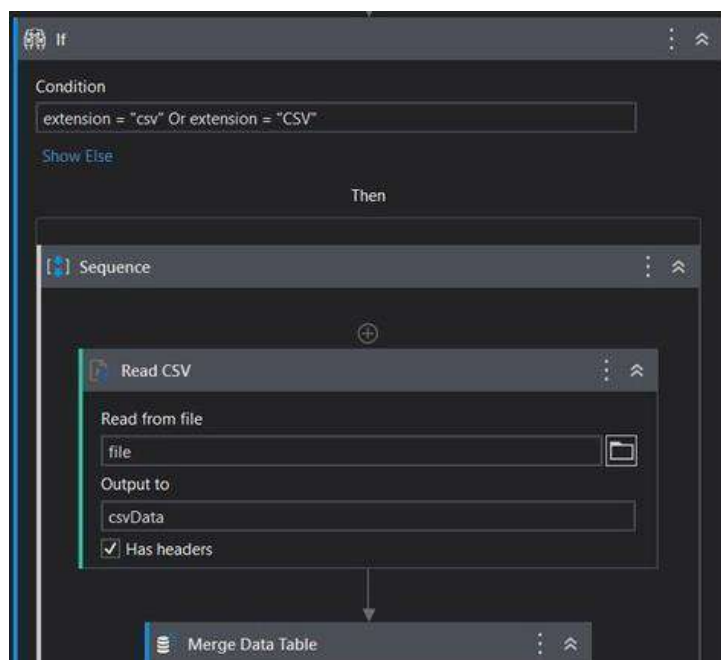
If file is an image file, then we apply OCR activities to extract the data from the image file and later apply regex module and data row append module for further data processing.



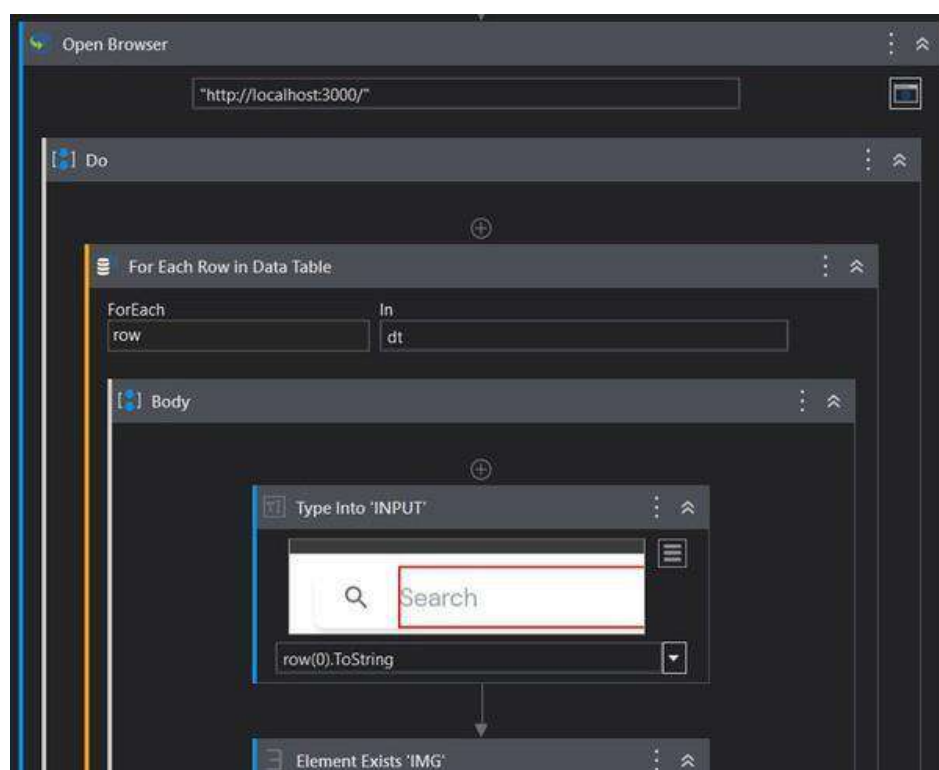
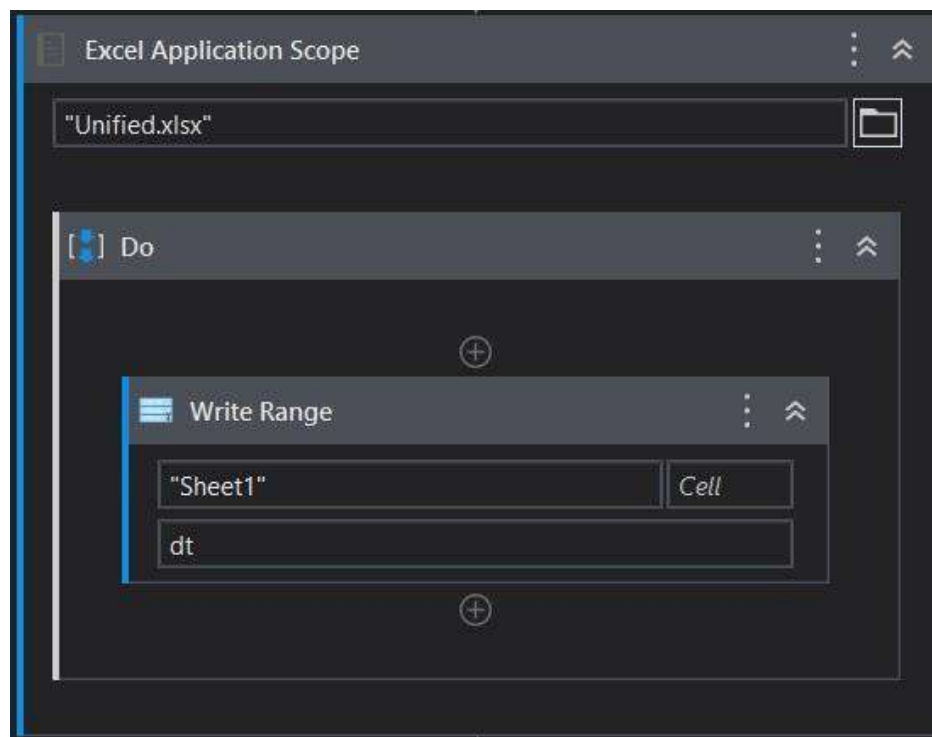
If file is of excel format, then we read the data using read range activity and later we simply merge the data table produced by read range to parent data table where we are storing the unified customer data from all file formats



If file of CSV format then we extract the data into a datatable using read CSV activity and merge this extracted data table to the parent unified datatable



After data is extracted from all the files present in the path provided, the datatable holding all the customer information is converted to an excel file



REFERENCES

1. [OCR Engine a Detailed Wiki by Wiki Grooper - Documentation](#)
2. [Tesseract OCR Documentation by UIPath - Documentation](#)
3. [Solutions to Data Migration Problems by Katie Behrens - A UIPath Blogger](#)
[A deeper dive into Data Migration by Simanta Shekhar Sarmah - Research Paper](#)
4. [Data migration: A theoretical perspective by Qing Wang - A Research Paper](#)
5. [Extraction data from JSON reference by Deepak Rai - A Video](#)
6. [Image text extraction Reference by Marcelo Cruz - A Video](#)
7. [Website using Reactjs reference by Abhisheik G - A Video](#)
8. [Data Migration in Cloud by Mehreen Ansara, Muhammad Ashraf and Mubeen Fatima - A Research Paper](#)
9. [Future Scope of Data Migration by Arif Iqbal and Ricardo Colomo Palacios - A Research Paper](#)
10. [Image Processing Using UIPath by UIPath - A Documentation](#)
11. [Extracting Data from PDF by UIPath - A Documentation](#)
12. [Backend for React JS using Express - A Video](#)
13. [React Js Does by React Js - Documentation](#)