UiPath™

Crawling Data Organization

Process Definition Document (PDD)

SEPTEMBER 30, 2020
Prepared by: ICSTAR – CoolE Team

Documents History

Date	Version	Position	Name	Organization	Function	Notes

Document Approval Flow

Condition	Position	Name	Organization	Signature & date
Created by	Developer	Kevin Dwiki Wijaya	ICSTAR	*30.09.2021*
Created by	Developer	Felix Filipi	ICSTAR	*30.09.2021*
Created by	Developer	Yohana Polin Simatupang	ICSTAR	*30.09.2021*
Approved by	Key user	Marvin Hadi	BCA	
Approved by	Key user	Vania christiana	BCA	

TABLE OF CONTENTS

I. Introduction

I.1 Purpose of the Document

The Process Definition document describes the business processes selected for automation using UiPath Robotic Process Automation (RPA) technology. This document will explain the business process of new potential customer data using data crawling and data searching processes. This document describes the sequence of steps carried out as part of the process as well as the conditions and requirements prior to the automated search process for new prospective customers. This document design serves as the basis for the RPA BCA developers, providing the necessary details to apply the automation robot to the selected business processes.

I.2 Objectives

The process that has been done by Robotic Process Automation (RPA) is part of the BCA RPA developer use case project. The goals and expected benefits of automation process are as follows:

- Obtain new data about potential users by using one robot.
- Speed up data search time up to 300% compared to searching manually.
- Obtain accurate data regarding information on potential new customers.
- Leverage automation to improve overall performance and reliability Department.
- Provide predictions to determine potential new customer from the obtained data.

I.3 Key Contacts

The specification document includes concise and complete business process requirements and is built on input provided by the Subject Matter Expert (SME) / process owner. The process owner is expected to review it and give approval for the accuracy and completion of the complete steps, context, impact and set of exceptions to the process. Their names are listed in the table below.

Role	Name	Contact details	Notes
ICStar Participants	Kevin Dwiki Wijaya	kvndwijya@gmail.com	
ICStar Participants	Felix Filipi	felixfilipi4@gmail.com	
ICStar Participants	Yohana Polin Simatupang	yohanapolinsimatupang@gmail.com	

I.4 Minimum Prerequisites for Automation

The minimum requirements for automation by robots in searching process for potential customers based on the nearest KCU location are as follows:

1. Completed KCU BCA data search data (have Bank.xlsx already)
2. Good internet connection
3. Join group Telegram https://bit.ly/Telegram_CoolE

II.   As-Is Process Description

II.1 Process Overview

No	Item	Description
1	**Process full name**	Crawling Data Organization
2	**Function**	Reporting
3	**Department**	Finance and Accounting
4	**Process short description** *(operation, activity, outcome)*	RPA will search for information from the business desired by the user through a website application called Google Maps. The information that will be retrieved such as Name of business, address, plus code, latitude, longitude, category, area, postal code, open hours, rating, total review, website, and telephone. All retrieved data will be mapped to find its nearest bank and find for its distance to predict it as a potential customer using a machine learning algorithm. Each completed or on progress process status will be sent via telegram.
5	**Role(s) required for performing the process**	End-User
6	**Process schedule and frequency**	Daily
7	**No of items processes /reference period**	1 - 5 times

8	*Average handling time per item*	15 Sec
9	*Peak period (s)*	N/A
10	*Transaction Volume During Peak period*	N/A
11	*Total # of FTEs supporting this activity*	N/A
12	*Number of persons performing the process*	*1 Person.*
13	*Expected increase of volume in the next reference period*	N/A
14	*Level of exception rate*	15 %
15	*Input data description*	Users need to input about the keyword of business they are looking for, and fill its limitation such as area, latitude-longitude, and plus code. The limitation can be filled with empty which means no limitation for the crawling process, but the user needs to fill the keyword section.
16	*Output data description*	The output for this process is data in form of an excel file containing several types of information for businesses that users are looking for such as Name of the businesses, address, plus code, latitude, longitude, category, area, postal code, open hours, rating, total review, website, telephone, the nearest bank from the business, its distance, and prediction about the potential client that the users are looking for.

II.2. Applications Used in the Process

The table includes a comprehensive list of all the applications that are used as part of the process to be automated to perform the given actions in the flow.

No	Application Name	System Language	Thin/Thick Client	Environment/ Access method	Comments
1	Chrome	EN	Thick Client	Web Browser	Task management

2	Microsoft Excel 2019	EN	Thick Client	Windows Application	Input / Output (I/O) files
3	Telegram	EN	Thick Client	Web Application, Mobile Application	Alerting process status
4	Prediction Model	EN	Thick Client	Windows Application	Determine which is the potential client

II.3 As-Is Process Map

This chapter contains a process map that contributes to a better understanding of how processes are carried out prior to automating an organization's data crawl.

High Level As-Is Process Map:

In this section we describe each general business process at the highest level to allow developers to have an understanding of each area of weakness, inefficiency, or to indicate which actions might fall within the scope of automation.
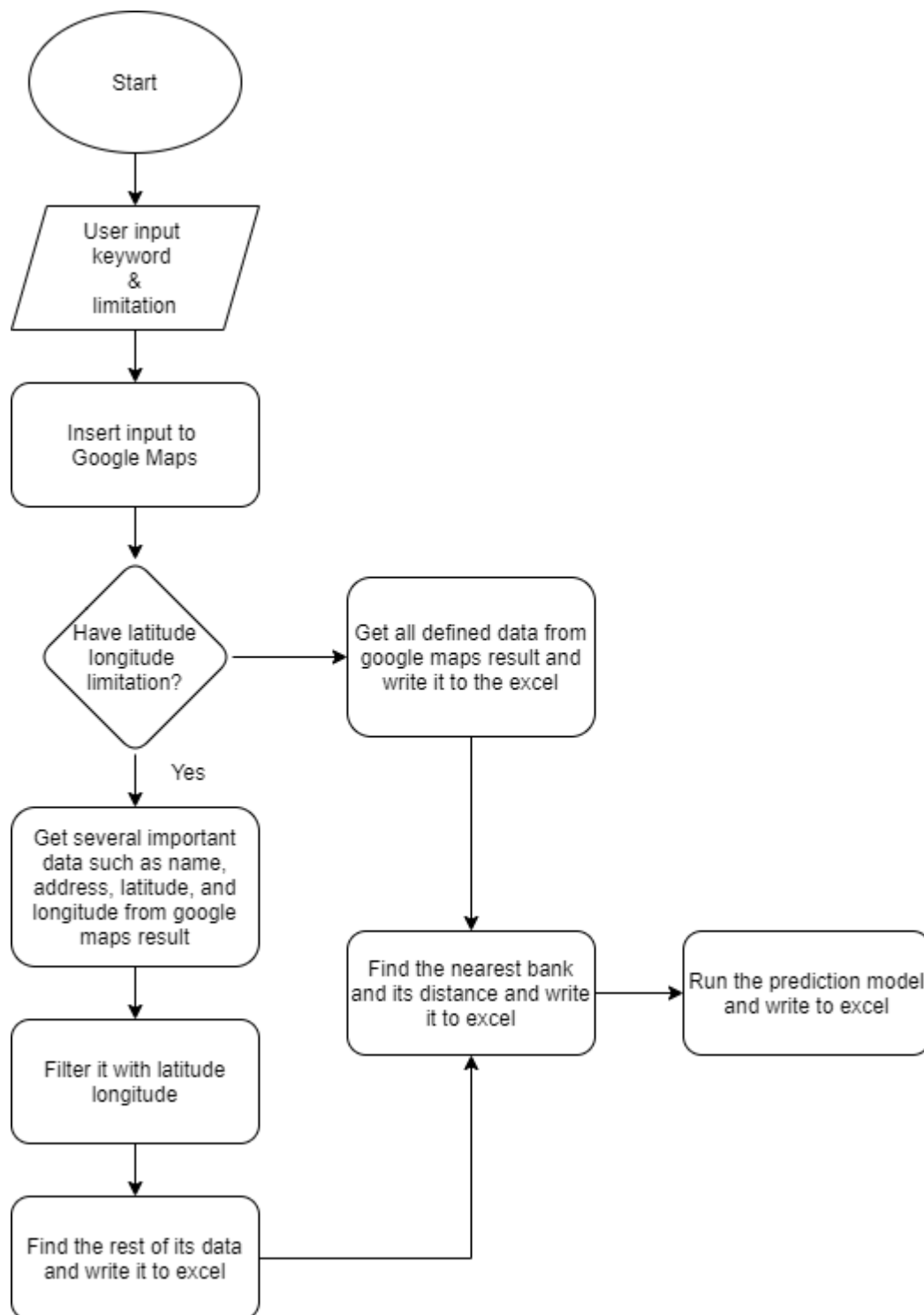
Figure 1. High-Level As Is Flowchart

II.4 Detailed As-Is Process Steps

The detailed process level map contains an explanation of the process at the key-stroke level and a section for understanding with the development team.

#Action	Input	Description	Details (Screen/Video Recording Index)	Output
1.	KCU Master Data	Master Data File Checking - Excel Window (Bank.xlsx)	Search for information about KCU that has been registered.	
2.	-	Select KCU name to search Latitude and Longitude	Filtering KCU data to be compared.	
3.	Name and address of KCU	Find Longitude and Latitude KCU	Search for Address,Pos Code, Plus Code and other detailed information from KCU BCA	
4.	-	Finding Longitude and latitude from code plus	Translating plus code to location, latitude, longitude	
5.	-	Entering KCP Latitude and Longitude (Bank.xlsx)	Save on bank.xlsx	Excel File .xlsx
6.	Name of business want to search and location.	Search location of business.	Search for Latitude, Longitude and other detailed information from business or new potential user.	
7.	-	Define Latitude and Longitude	Latitude and Longitude filtering based on data from business	
8.	-	Getting Final Output Bank.xlsx)	Save data on name Of PotensialCustomer.xlsx or business.xlsx	
9.	-	Business Screening based on the Nearest KCU.	Compare each Latitude and Longitude Bank and business data	Excel Data Crawling Final File .xlsx

See doc attached

Detailed As-Is
Process Steps.xlsx

II.5 Input Data Description

#	Step action descrition	Screenshot	Expected result	Remarks
1.1	Master Data File Checking - Excel Window (Bank.xlsx)		Checking keyword data searching	Possible exception: - Handle exception if excel file not available
1.2	Select KCU name to search Latitude and Longitude		Delete KCP,KK, BCA Excpress on column "Status Cabang"	
1.3	Find Longitude and Latitude KCU		Get Plus Code, Address, Pos code	Possible exception: - Handle exception if UI Google Maps Update
1.4	Finding Longitude and latitude from code plus		Translating plus code to get location, longitude, latitude	

Process Design Document – Crawling Data Organisasi

1.5	Getting Output KCU Latitude and Longitude (Bank.xlsx)		File Bank.xlsx created	
1.6	Search location of business.			Possible exception: - Handle exception if UI Google Maps Update
1.7	Define Latitude and Longitude from codeplus or url		Translating plus code to get location, longitude, latitude	Possible exception: - Handle exception if UI Plus Code Update
1.8	Getting Final Output		Get all of information in google maps	

1.9	Business Screening based on the Nearest, Distance, Prediction				Final additional output for nearest KCU BCA, distance, and machine learning prediction Output	

	U	V	W
	Nearest KCU BCA	Distance	Prediction
	2 BANDUNG	0.883438	Potential
	1 BANDUNG	2.023938	Potential
	2 BANDUNG	0.847935	Potential
	BANDUNG	2.406186	Potential
	BANDUNG	2.589055	Potential
	A. YANI	1.090015	Potential
	BANDUNG	1.387601	Potential
	BANDUNG	1.883	Potential
	SOEKARNO HATTA	0.697081	Potential
	BANDUNG	0.846035	Potential
	DAGO	0.362532	Potential
	A. YANI	2.601438	Tidak
	A. YANI	2.952559	Potential
	BANDUNG	1.869038	Tidak

#Action	Sample	Input Type	Location	Are inputs Natively Digital*?	Are the Inputs Structured*?
1	Bank.xlsx	Screen	n/a	YES	YES
3	KCU BCA Jakarta barat, Jakarta	Text	n/a	NO	YES
6	Martabak Jakarta barat, Jakarta.	Text	n/a	NO	YES

III.   To-Be Process Description

III.1 Detailed To-Be Process Map

This section will explain clearly the design of the automation system carried out in the organization's data crawling process. Details to be explained in the form of a flowchart then for details of the automation that has been implemented in the form of a workflow on the UiPath studio application.
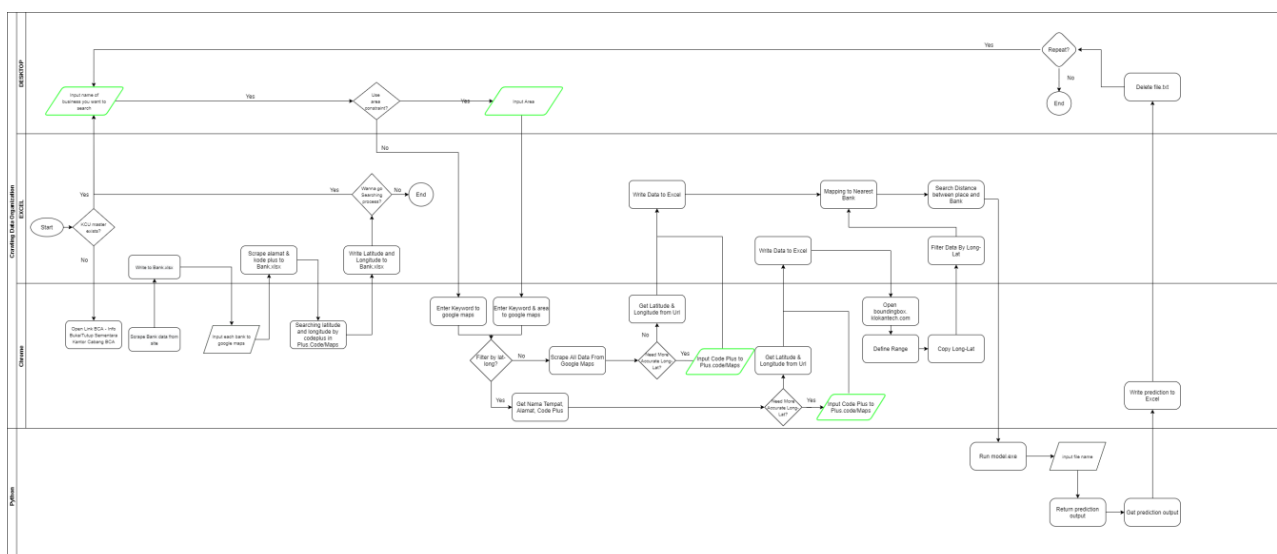

Figure 2. Detailed As-Is Flowchart

FlowChart.png

Legend	Description
Green Box	This process action remains manual (to be performed by a human agent).
Black Box	This process action is proposed for automation.

III.2 In Scope for RPA

The activities **in scope of RPA**, are listed here:

- All steps in the embedded detailed map and steps (Crawling Data Organization Process Detailed)

III.3 Out of Scope for RPA

The activities **OUT of scope of RPA**, are listed here:

- All steps in the embedded detailed map and steps (Crawling Data Organization Process Detailed)

III.4 Detail to be Process in UiPath

The initial stage of the robot is checking the availability of the master data. This checking process is carried out automatically by the robot. The checking condition depends on the KCU master data existence, when the robot found the master data, then the robot will run the business searching process. Meanwhile if not, then the robot will search for KCU data by crawling the master data on the BCA website.

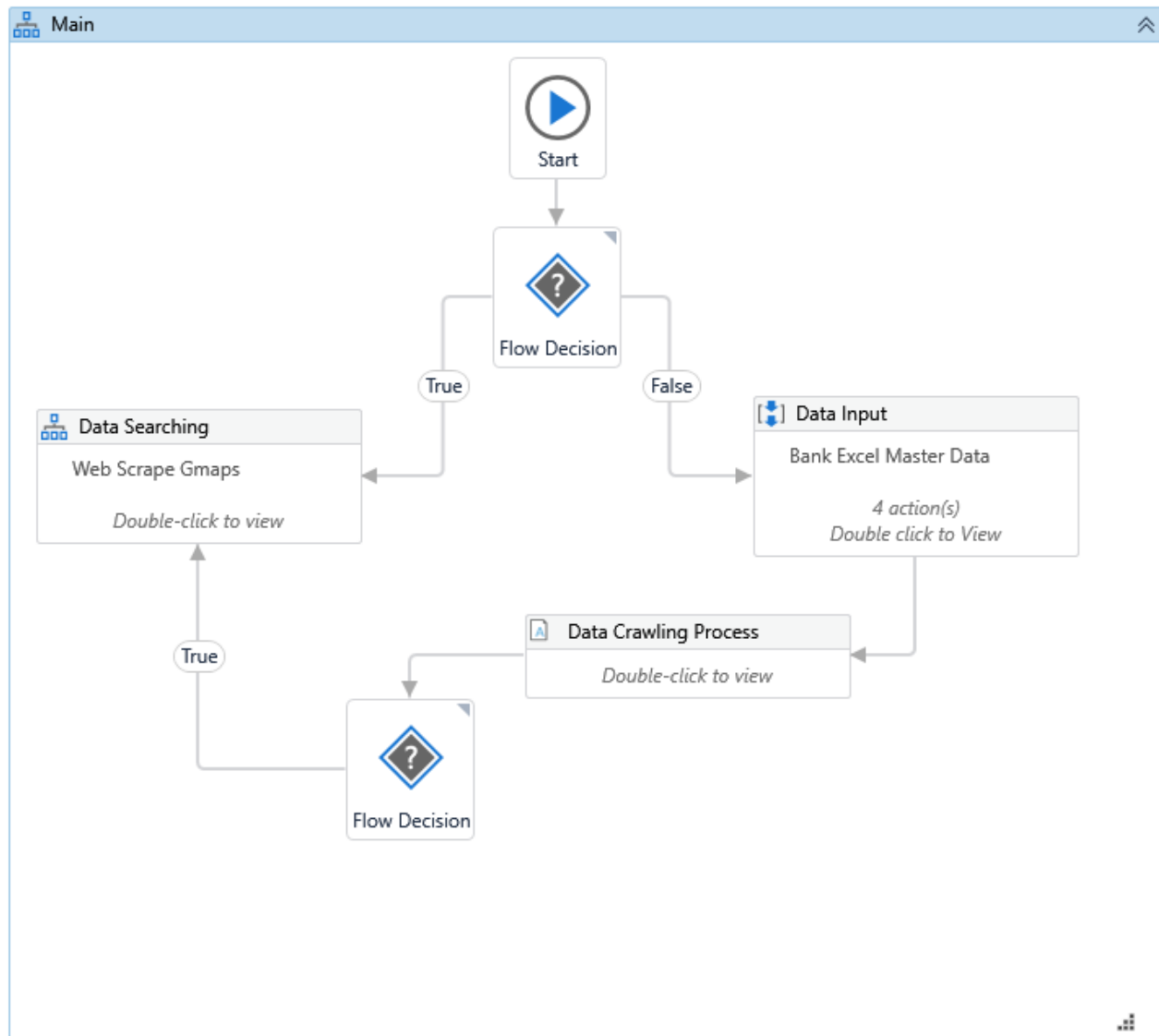## III.4 Business Exceptions Handling

Figure III. Main Workflow

Figure III is about the main workflow. Here the robot will look for the existence of master data which is Bank.xlsx file and determine which process will the robot do as the first time. If master data existence not found, then the robot will execute master data gathering process. Else, if the robot find the existence of master data, then the robot will do data searching process, to find what user are looking for.
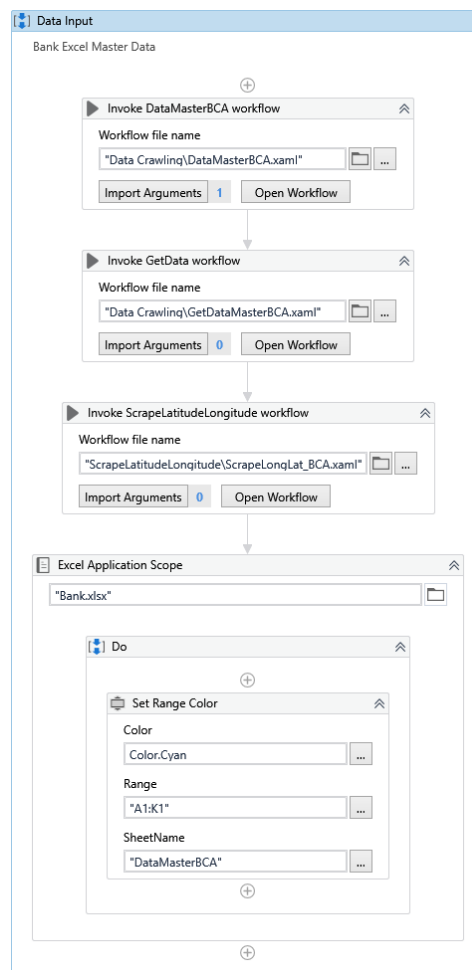
Figure IV. Data Input Workflow

Figure IV is about KCU Master data searching is performed when the master data is not available. This process starts from crawling data on the BCA website and creating an excel file that will store each KCU data. Then, every data in excel will be searched on Google Maps with input automatically by the robot. The input by the robot to Google Maps is generated by the name of the KCU and its address. Then each longitude and latitude data will be written in the excel master data file.
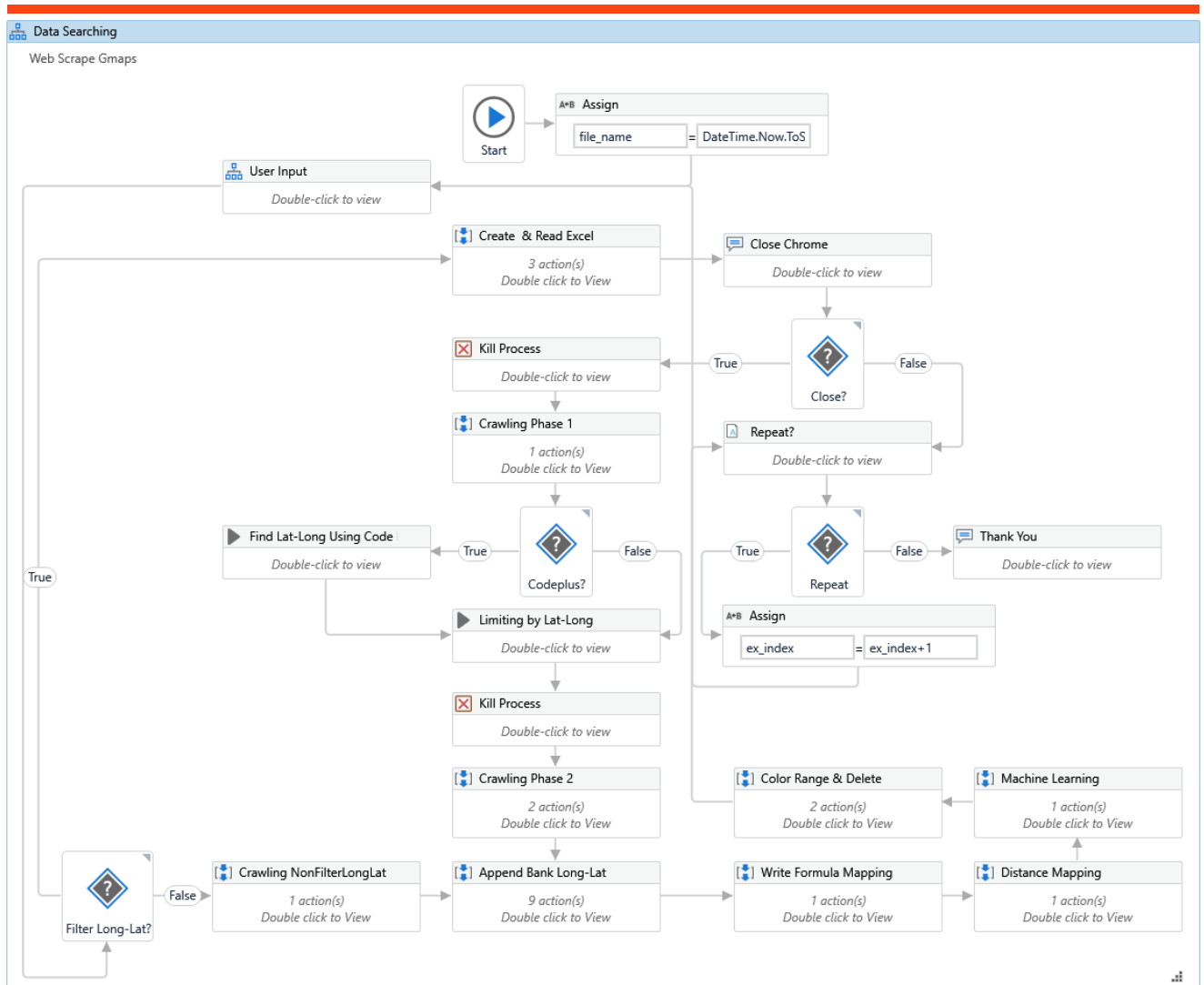
Figure V. Flow Data Crawling & Data Searching

Figure V is about Searching for information on potential customer data begins by creating a file that will store potential user data by using the current time as the name for the file. Furthermore, the user needs to input which businesses and its limitation about where the user wants to look for. Next, the robot will close all of the chrome application before executing the next stage of its process. Then the robot will ask "Do you need exact latitude-longitude?"

- Yes, then the robot will enter crawling phase 1 where the robot will search each business which users are looking for and take several types of information such as name of businesses, address, plus code, latitude, and longitude. The robot will keep it in the temporary sheet of an excel file and open the **bounding box website**, so the user can determine the distance limit to determine the latitude and longitude limitation of the crawling process they are looking for.
- If not, then the robot will automatically look all of the information from the businesses including its latitude and longitude and keep it in the excel output file.

If the user decides to user limitation on latitude and longitude, then the robot will do the crawling phase 2 where the robot will filter the information that have been gained by its latitude and longitude from the bounding box website. After that, the data from the master bank will be appended to the excel file to find its nearest bank and calculate the distance using haversine formula. Furthermore, there is an additional process, to predict which is the potential client by using machine learning algorithm to see potential client from generated data based on its total reviews, rating and the distance between businesses with its nearest bank.

The business process owners and business analysts (BAs) are expected to perform documentation to all business exceptions identified in this automation process which can be classified as:

Known	Unknown
Previously encountered. A scenario is defined with clear actions and workarounds for each case.	New situation never encountered before. It can be caused by external factors. Cannot be predicted with precision, however if it occurs, it must be communicated to an authorized person for evaluation.

Known Exceptions

The table below reflects all the business process exceptions captured during the process evaluation and documentation. These are **known exceptions,** met in practice before. For each of these exceptions, define a corresponding expected action that the robot should complete if it encounters the exception.

BE #	Exception name	Step	Parameters	Action to be taken
1	Multiple Assign (Throw) Google Maps UI Updates	N/A	Cannot assign "in_alamatLengkap" or etc Exception Type: System.InvalidOperationException	The robot will stop the process and the user should make sure have stable connection, then restart the robot processes. This exception arises because the user's internet is unstable, so elements from the website do not appear when the robot is looking for it.
2.	Wait for a Chrome page to be loaded	Any step	Excel write nothing	The robot will stop the process and the user should make sure have stable connection, then restart the robot processes. This exception arises because the user's internet is unstable, so elements from the website do not appear when the robot is looking for it.

Unknown Exceptions

For all the other **unanticipated or unknown business (process) exceptions**, the robot should:

1. Send an email notification at kvndwijya@gmail.com or felixfilipi4@gmail.com and error message screenshot attached.
2. Join group telegram https://bit.ly/Telegram_CoolE and send error message with screenshot attached.

![UiPath logo]

III.4 Application Error and Exception Handling

A comprehensive list of all errors, warnings or notifications should be consolidated here with the description and action to be taken, for each, by the Robot.

Errors identified in the automation process can be classified as:

Area	Known	Unknown
Technology/ Applications	Experienced previously, action plan or workaround available for it.	New situation never encountered before or may happened independent of the applications used in the process.

Know Errors or Exceptions

The table below reflects all the errors identifiable in the process evaluation and documentation.

For each of these errors or exceptions, define a corresponding expected action that the robot should complete if it is encountered.

No	Error name	Step	Parameters	Action to be taken
1	Application Crash / Internal Server Error	Any step	Error message	The robot will stop the process and the user should make sure have stable connection, then restart the robot processes. This exception arises because the user's internet is unstable, so elements from the website do not appear when the robot is looking for it.

Unknow Errors and Exceptions

For all the other **unanticipated or unknown application exceptions/errors**, the user should:

1. Send an email notification at kvndwijya@gmail.com or felixfilipi4@gmail.com with the error message and error screenshot attached.
2. Join group telegram https://bit.ly/Telegram_CoolE and send error message with screenshot attached.

IV. Other Observations

1. **Dataset**.

Dataset.xlsx

Using logistic regression algorithm and the gradient descent algorithm for loss function. The accuracy above 90% both in training and testing set which able to classify models without problems such as overfit or underfit.

2. **Data Master**

Bank.xlsx

Master data file, which containing about the information of BCA KCU in all regions in Indonesia.

V. Additional Sources of Process Documentation

If there is additional material created to support the process automation please mention it here, along with the supported documentation provided.

Additional Process Documentation		
Video Recording of the process demo	Video Demo	
Other documentation	https://bit.ly/ICStar_CoolE-FinalProject	All of documentation