

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	ВСА	



TABLE OF CONTENTS

I.	I. Introduction	2
	I.1 Purpose of the Document	2
	I.2 Objectives	2
	I.3 Key Contacts	2
	I.4 Minimum Prerequisites for Automation	5
11	II. As-Is Process Description	5
	II.1 Process Overview	<u>5</u>
	II.2. Applications Used in the Process	6
	II.3 As-Is Process Map	7
	II.4 Detailed As-Is Process Steps	9
	II.5 Input Data Description	10
II	III. To-Be Process Description	12
	III.1 To-Be Detailed Process Map	12
	III.2 In Scope for RPA	13
	III.3 Out of Scope for RPA	13
	III.4 Detail to be Proces in UIPath	13
	III.4 Business Exceptions Handling	13
	III.4 Application Error and Exception Handling	18
	IV. Other Observations	19
	V. Additional Sources of Process Documentation	19



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		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.



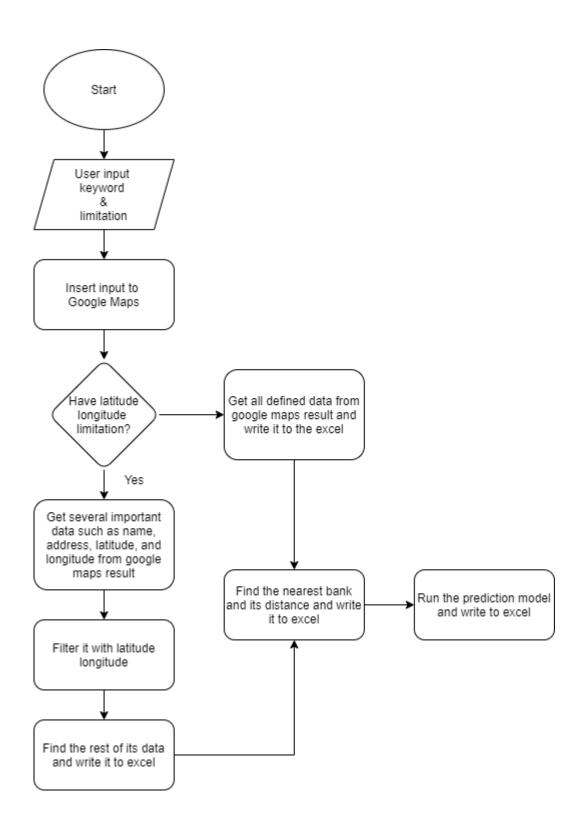


Figure 1. High-Level As Is Flowchart



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



II.5 Input Data Description

#	Step action descrition	Screenshot	Expected result	Remarks
1.1	Master Data File Checking - Excel Window (Bank.xlsx)	The second secon	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	BCA JI. Jend. Sudirman No.49-51 Q X Chika Stor Bank BCA KCU Yogyakarta 4,4 SPBU 44.552 Rute Simpan Di Sekifar Kirim ke Bagikan ponsel Anda O JI. Jend. Sudirman No.49-51, Terban, Kec. Gondokusuman, Kota Yogyakarta, Daerah	Get Plus Code, Address, Pos code	Possible exception: - Handle exception if UI Google Maps Update
1.4	Finding Longitude and latitude from code plus	No final an experience of a final information	Translating plus code to get location, longitude, latitude	



1.5	Getting Output KCU	A	В С	File Bank.xlsx created	
	Latitude and	1 Bank Latitude Bank	k Longitude Bank		
	Longitude (Bank.xlsx)	2 -6.906438	107.648438 A. YANI		
		3 -6.242812	106.652812 ALAM SUTERA		
		4 -3.695938	128.182187 AMBON		
		5 -6.138937	106.813187 ASEMKA		
		6 3.586313	98.695313 ASIA		
		7 -1.276312	116.841437 BALIKPAPAN		
		8 -5.445813	105.273062 BANDAR LAMPUNG		
		9 -6.922562	107.612687 BANDUNG		
		10 -7.026188	112.752813 BANGKALAN		
		11 -3.322812	114.592062 BANJARMASIN		
		12 -8.219188	114.368938 BANYUWANGI		
		13 1.152062	104.006062 BATAM		
		14 -6.237938	106.993813 BEKASI		
		15 2 700027	102 269427 DENIGVIIIII		
1.6	Search location of	BCAKCUTANGERANGCTTY-C= X	HERDANG + CITY/60-6.1951669.106.6399317.17z/data=I3m1Hb14m5I3m4f1s0x2e69994db3s93es9.0x871	480	Possible
	business.	■ BCA KCU TANGERANG CITY, RT. (1 Q. ×	Ille Y4 Restaurants Ma Hotels & Attractions Transit if P Parking	No.	exception:
		The same of the sa	University of High School Rudus Chicked	ear me	- Handle
			Tanghang Q B Q		exception if UI
			SMK PCRETTengerung SMAN 7 Tangerung Partier TANK	900 508	Google Maps
			NSGI SKIP PORT 1 TULL TANGEKANG. Hank BIN KC Language	ther	Update
		BCA KCU TANGERANG CITY 47 ***** 42 reviews	J. Points Knurdoust J. Tangoty Mail O B		Opuale
		Bank	Agema Tancer ang SAMSAT CIKOKOL Novobel Tangerang TANCE RANG AS (664)	394	
		○ □ ○ □ <	ta Kencana Mandriu. A. P. Ketayoo da angelang Tarran Potiet. Gaca KCU	T	
		phone save nearly serious orane	POAM Ckokel Sangrang Poam Kitman		
		 Ji Jendral Sudirman No I, Blok H0007S - H0009S Apartemen Skandinavia, RT.001/RW.005, 	UUner B JL Tesks Umer TANGERANG		
		Babakan, Kec. Tangerang, Kota Tangerang, Danten 15118			
		Located in: Tangcity Mail Closed: Opens at 08:15 V	Monier Tangerang &		
		S bca.co.id	No.11 O		
		(021) 59732000	© Kota Ayadhya		
		RJSM+W6 Babakan, Tangerong City, Banten	Avadha Azariment	4	
			Tower Sterna Google RSUD KO	TA.	
1.7	Define Latitude and	H hours approximately x ₩ his concent case there is ← → C = # skinning contribution of	9 RG/Trove-Responsy-Cocol: X → Fix-cole: X + 0 − 0 X	Translating plus code	Possible
1.7	Longitude from	Enter a plus code or an actives × mass √ s	Custor Alaska Clinore Mognetis Warung Bapan Abas Clinore Mognetis Warung Bapan Abas Clinore Mognetis Conditional Conditions (Conditional Conditions)	to get location,	exception:
		10 Seption:	Majordia Pritace Outling Screen Address Control Agent Balks Emak Tarih Agent Balks Emak Tarih		
	codeplus or url	Clays Save C	Story Tourist State Control of Story Con	longitude, latitude	- Handle
			Only served		exception if UI
		Makening Countries Countri	Miss and B		Plus Code
		PUWC+6J Curing Sangereng, Tangerang, Banton, Bartodr Aver Indonesia	Sociality Sociality See Social		Update
		S	Goting Being (
		O APPRIMENTAL O	All Remarkshore Office Control Office Office Office Office Control Office Offic		
		© RESERVORAL ®	Toleran		
		. Jid See Kaling Sey	have an a Fame load		
		iliana.	Prumouri 850		
		0	O Out Class Forumeuss		
			-		
1.8	Getting Final Output	1 G H I J K L 1 Langitude Kategori Wilayah Kode Fox Secon Selasa Rabu	M N O F Q E S T U V W sareis serus Sabbi Mengge Rating Total Resi Website Tolepon Research Christian	Get all of information	
1.0	Coung i mai Odiput	2 1077-6791 Murtabak Kota Band 40282 11AM-131 11AM-131 11AM- 3 1077-6254 Murtabak Kota Band 40284 4-10.3054, 4-10.3054, 4-305 4 1077-6381 Health Fox Kota Band 4012 4-10594 4-10594 4-10594	13/13AM-13/13A	in google maps	
		5 007.6341 Murtabak Kota Band 40273 6PM-1AM69M-1AM69M-1 0 1077.6353 Murtabak Kota Band 4032 3-109M 3-11PM 3-12P 7 107.6353 Murtabak Kota Band 4032 3-109M 3-11PM 3-12PM 0 07 5-123 Murtabak Kota Band 4031 31M-134 20M-134	## 4-1009 #-1094 #-1094 #-1094 ## 2 3 0877-882 BAXCOUX 0.85735 Peternial AMB 91-1049 -1049 #-1049 #-1049 ## 2 32 0877-882 BAXCOUX 0.85735 Peternial ## 1299 91-1049 -1149 1-1049 #-1049 ## 3 3,378 (022)08540-04000 Peternial 125 1234-125 1234-125 1234-125 1234-125 ## 2 1000000000000000000000000000000000	in google maps	
		8 507.6381 Mustabok Kota Band 4013 5946-124-5944-124-5944-1 5 107.599 Mustabok Kota Band 4027 3-11940 3-11940 3-1194	284594-1245994-124594-124594-124594-1246 0.7 19 0857.0546 BANDUNG 1.87098 Percental of 9-1109 9-11996 9-11996 9-11996 0.5 1109 0.		
		11 107 687 Perturamina cultum Sanda 100	15 15 15 15 15 15 15 15		
		27 107 6167 November State Const. 46054 S 11084 S 11084 4 1108	B B BANCINE 2,8998 Tidak		
		17 UT. 6155 Murtabak Kota Band 40254 0-11PM 0-11PM 0-11PM 18 107.6129 Rosta krankota Band 40222 18 107.6129 Murtabak Kota Band 40201 0-11PM 0-11PM 0-11PM	## + 11PM + 11PM + 11PM + 11PM 42 5 0076-2084 MAXCHER 2509 Fotemal ### + 11PM + 11PM + 11PM + 12PM 44 30 0076-2229 MAXCHER 25085 TOOM ### + 11PM + 11PM + 11PM + 12PM 44 30 0076-2229 MAXCHER 250857 Fotemal ### + 11PM + 11PM + 12PM 4 30 0076-2229 MAXCHER 250857 Fotemal ### + 11PM + 11PM + 12PM 4 30 0076-2229 MAXCHER 250857 Fotemal		
		30 07.5459 Sectavis-costs Band 45022	M 5-11PM 5-11PM 5-11PM 5-11PM 6.7 3 0866-0887 0MAGO 1.38003 Potential 1 0866-0887 0MAGO 2.58073 0MAGO 1.38003 Potential 1 0866-0887 0MAGO 2.58073 0MAGO 2.58		
		24 107-602 Murtabak Kota Band 40561 1-109M 1-109M 1-2094	## 1-100M 1-100M 1-100M ## 2 022 00230465 1,3993 Petersul ## 1-100M 1-100M 1-100M ## 2 022 00230465 1,3993 Petersul ## 2154M-1154M-1154M-1154M-1 ## 45 8 0009-738 84XCSM 1,3987 Petersul ## 2154M-1154M-1154M-1154M-1 ## 45		
		28 307.5823 Martistati Bankhang 40223 3-009M 3-109M 3-009 27 307.6955 Restauraminota Band 40261 2215-4PR 1215-4PR 1215- 28 307.5957 Restauraminota Band 4024 3-009M 3-109M 4-1097 29 307.5957 Restauraminota Band 4032	M 3-100M 2-100M 2-100M 2-100M 5 1 SOCKARW 2-2298G Petersial PRE215-498 125-598 1215-698 1215-698 4 6 389 BANDUM (0.73589 Petersial M 3-99M 4-109M 4-209M 3-109M 41 24 BANDUM (1.73589 Petersial D 2002MB M M 2-100M 2-10		
		25 107.5572 Bestauran Kota Band 40032 107 K688 Mustabak Kota Band 40523 SPAH-GARASPM-	0 0 003-1346 DAGG 1.30003 Triels AM 8PM - AAM 8PM - EAR 8PM - EAR 8PM - EAR 6 48 0057-2129 A, YANG 4.471254 PC contail 2 2 A YANG 4.471254 PC contail		



1.9	Business	U	V	W	Final additional
	Screening based	Nearest KCU BCA	Distance	Prediction	output for nearest
	on the Nearest,	2 BANDUNG	0.883438	Potential	
	Distance,	1 BANDUNG	2.023938	Potential	
	Prediction	2BANDUNG	0.847935	Potential	machine learning
	1 Todiotion	BANDUNG	2.406186	Potential	prediction Output
		BANDUNG	2.589055	Potential	prediction Output
		5A. YANI	1.090015	Potential	
		1BANDUNG	1.387601	Potential	
		BANDUNG	1.883	Potential	
		SOEKARNO HATTA	0.697081	Potential	
		1BANDUNG	0.846035	Potential	al entre
		DAGO	0.362532	Potential	
		A. YANI	2.601438	Tidak	
		A. YANI	2.952559	Potential	al Control of the Con
		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.

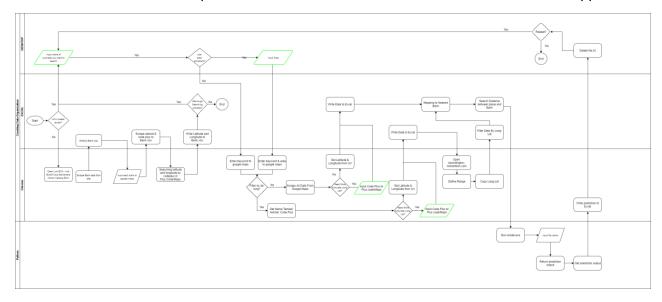


Figure 2. Detailed As-Is Flowchart





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

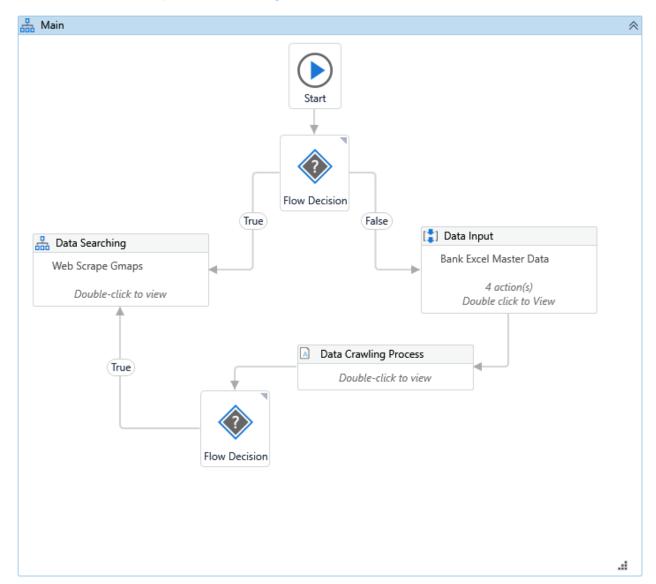


Figure III. Main Workflow

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.



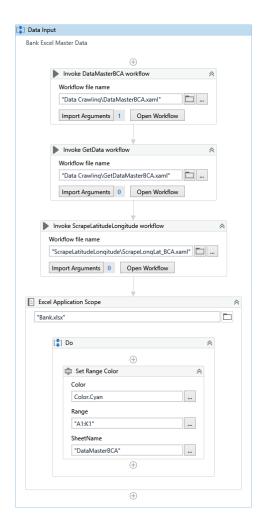


Figure IV. Data Input Workflow

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.



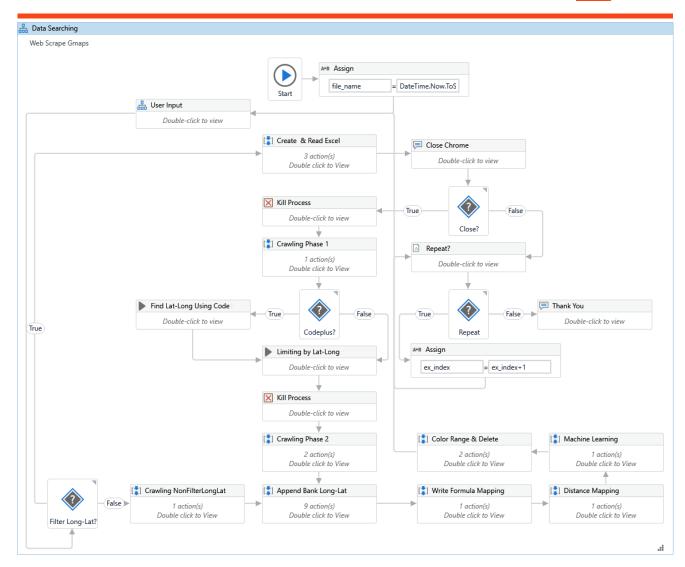


Figure V. Flow Data Crawling & Data Searching

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	New situation never encountered before. It can be caused by	
defined with clear actions and	external factors. Cannot be predicted with precision, however if	
workarounds for each case.	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.



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.



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



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	Video Demo		
demo			
Other documentation	https://bit.ly/ICStar CoolE-FinalProject	All of documentation	