|  |
| --- |
| Case study for document understanding |
| Analysis of documents using data wrangling and RPA methodologies. |

|  |
| --- |
| Vignesh Ganesan  2-8-2021 |

[1. Introduction 2](#_Toc63718942)

[2. Case Study 2](#_Toc63718943)

[3. Requirements 2](#_Toc63718944)

[4. Solution Model 2](#_Toc63718945)

[5. Application of solution 2](#_Toc63718946)

[6. Analysis of methodologies 2](#_Toc63718947)

[7. Hurdles 2](#_Toc63718948)

[8. Conclusion 2](#_Toc63718949)

[9. Appendix 2](#_Toc63718950)

# Introduction

# Case Study

# Requirements

* Extract data from LSP invoices
* Filter data to extract only relevant fields
* Save data in Excel workbook
* Mail data to customer

# Solution Model

## RPA model in UiPath

### Software Used:

UiPath Studio Pro Community Edition 2020.10.4

### Packages Used:

documentUnderstanding v1.5.2 (Requires 3 additional packages)

### Process

Using the documentUnderstanding package provided in UiPath Studio

## Model based on Python 3.9.1

### Libraries Required:

pdfminer.six

nltk

pandas

os

slate

pdf2txt

# Viable Solution

UiPath:

load packages

create taxonomy

insert field names

save taxonomy

digitize document>select text and path>output variable

create variable

load “omniPage OCR reader” (must have the package installed)

load document scope>keyword based classifier>load empty json and select keyword?????

load machine learning extractor?????

Python 3.9.1:

load libraries

convert pdf to text using pdf2txt.py -o path/to/output.txt path/to/input.pdf

# Analysis & Meeting targets

# Hurdles

# Conclusion

# Appendix