Table of Contents

[Abstract 2](#_Toc163566328)

[Introduction 2](#_Toc163566329)

[Problem statement 2](#_Toc163566330)

[Background 2](#_Toc163566331)

[Aims 3](#_Toc163566332)

[Objectives 3](#_Toc163566333)

[Literature review 4](#_Toc163566334)

[Receipt Extractor 4](#_Toc163566335)

[Paper 1 - Utilize OCR text to extract receipt data and classify receipts with common Machine Learning algorithms. 4](#_Toc163566336)

[Paper 2 - Information Extraction from Scanned Invoices using Machine Learning, OCR, and Spatial Feature Mapping Techniques 5](#_Toc163566337)

[Paper 3 - Computer Vision for Document Image Analysis and Text Extraction 7](#_Toc163566338)

[Mobile Application 8](#_Toc163566339)

[Paper 4 - React Native vs Flutter, cross-platform mobile application frameworks. 8](#_Toc163566340)

[Conclusion 10](#_Toc163566341)

[Requirements specification 10](#_Toc163566342)

[Receipt Section Detection 10](#_Toc163566343)

[Text Extraction 11](#_Toc163566344)

[Format and Classify text extraction. 11](#_Toc163566345)

[User Interface 11](#_Toc163566346)

[Design 11](#_Toc163566347)

[Server 12](#_Toc163566348)

[Receipt extraction feature. 12](#_Toc163566349)

[User Correction Feature. 13](#_Toc163566350)

[Server Communication 13](#_Toc163566351)

[Database 13](#_Toc163566352)

[Application 14](#_Toc163566353)

[Receipt upload for receipt extraction. 14](#_Toc163566354)

[Implementation 14](#_Toc163566355)

[Testing 14](#_Toc163566356)

[Result & Analyzing 14](#_Toc163566357)

[Conclusion 14](#_Toc163566358)

# Abstract

# Introduction

## Problem statement

This project of ‘smart receipt management and Extraction’ aim to enhance the efficiency of company accounting process, which “plays a significant role in the effective management process” (Alabdullah, 2019). A big part of this accounting process consists of the logging of varied expenses and for numerous raisons has been a manual process. Such a repetitive has always require a significant amount of time and energy. Therefore, this project has for purpose to develop a solution to the problematic: How to facilitate the receipt logging process?

## Background

Has said while elaborating the process statement, the purpose of this project is to facilitate the logging of receipt. This problematic come from the different impacts of this manual receipt logging task, has said before it requires humans’ action to log the different receipt one by one using a lot of time and energy, automating this process could considerably assist the users in this task, leading to a digitalization of this task and logging. Some solution already exists in the market such has “Recipator AI” (AI, no date) or “Veryfi” (*Transform Documents into Actionable Data in Seconds using Veryfi OCR API*, no date) but are limited in functionalities not adapted to the Mauritians market, lacking easy to use system. More other, digitalize the process of receipt handling is also providing a more ecologic move by reducing the ecologic impact from this task which produce a high amount of paper waste and business carbon footprint. In summary, this project aims to tackle the specific task of handling and managing receipt through a user-friendly system using the latest state-of-the-art technologies.

## Aims

This project of “Smart Receipt Management and Extraction” system is designed to streamline and simplify the repetitive and time consuming of handling, organizing and informatize receipt, whether physical or digital. Therefore, the primary purpose of this project is to serve a comprehensive receipt extraction and management system, offering the user a set of feature leading to a user friendly user-interface.

## Objectives

Literature Review

Gather the similar project and research paper to retrieve the possible technoilogies that can used with their advantage and disadvantages to use appropriate state-of-the-art technologies for the different elements of the system.

Receipt extraction model

Develop an receipt extraction model which will be use to efficiantly extract the data out of the receipt, offering to the user the wanted receipt extraction automation to facilitate the receipt logging.

Receipt dataset

Gather the biggest receipt dataset in order to have the most precise and efficiant model to extract as efficiently possibly the user’s receipts, this dataset will be use for the training and testing process of this receipt prediction model.

Mobile Application

Devellop a Mobile application has front-enfd of the project, it will be use by the users has a gateway for the system which will be partially in a server.

Local Server

Develop a local server which will be used to host the extraction model and the database. It will contain all the “heavy” process need by the system.

System Test

Create tests for the system and his different component such as the mobile application, the server, the communication between then and the different feature which will be provided to the user.

# Literature review

## Receipt Extractor

### Paper 1 - Utilize OCR text to extract receipt data and classify receipts with common Machine Learning algorithms.

This paper from Joel Odd and Emil Theologou is a study to “investigated if it was feasible to use machine learning tools on OCR extracted text data to classify receipts and extract specific data points.”(Odd and Theologou, 2018). This process firstly extracts the receipt data through different OCR, then classify through a model into different category. They have test different technology for the OCR extraction and the model prediction, all list in *Table 1*

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Technologies** | **Pros** | **Cons** |
| **Optical Character Recognition (OCR)** | Azure Computer Vision API | * Good accuracy in text recognition. | * The security is uncontrollable due to third party dependance. * May require significant data preprocessing. |
| Google Cloud Vision API | * Good performance and time response. | * The security is uncontrollable due to third party dependance. * Managing variations in receipt formats can be complex |
| Tesseract OCR | * Deployable locally * Not dependent of third party. * Strong community | * Varying receipt formats may affect accuracy. * Accuracy depend on image quality. * Slow process for large amount of data. |
| **Machine Learning Models** | Linear Support Vector Classification (LinearSVC) | * 94% accuracy achieved. * Efficient with large text dataset | * Not has efficient for non-linear data relationship. * Require optimal parameters tuning, otherwise reduction in performance. |
| Multilayer Perceptron Classifier (MLPClassifier) | * Can capture complex relationship (non-linear). * Flexibility with a large number of parameters. | * May be computationally intensive as the model become complex. * Risk of overfitting if the parameters are not well set. |
| Naive Bayes Classifier | * Easy implementation to handle large dataset. * May be use has baseline for classification problem. | * Can make naïve assumption which will lead to a reduction in accuracy for complex data. * Training dataset balance strongly influence the accuracy. |

Table 1 - Advantage and Limitation of technology used in "Utilize OCR text to extract receipt data and classify receipts with common Machine Learning algorithms" written by Odd and Theologou

For the OCR, three principal technologies are tested “Azure Computer Vision API” provided by Microsoft, “Google Drive REST API” provide by Google and “Tesseract OCR”.

The “Azure Computer Vision API” such has “Google Drive REST API” are efficient and powerful OCR tool but are third party dependent and all their processing power are deported in their own servers. This deported strategy allows powerful OCR but therefore create delay using API to upload and download the data and create a dependance to their services and could be costly. On the other hand, “Tesseract OCR” is an open-source OCR which can be deployed locally and therefore have a quicker response time compared to “API” OCR. In the context our project, the user will have the ability to correct any error from the extraction, therefore a quick response time is essential for a nice user experience and is preferable other a slice reduction of extraction accuracy.

After extracted the text from the receipt, the output is categorized through models. Different models were used such has Linear Support Vector Classification, Multi-layer Perceptron classifier and Naive Bayes Classifier all having their advantage and disadvantage has showed in *Table 1*. Their model is based on a text extract before using the model strategy, and therefore is not used for image recognition.

### Paper 2 - Information Extraction from Scanned Invoices using Machine Learning, OCR, and Spatial Feature Mapping Techniques

This second research paper “Information Extraction From Scanned Invoices using Machine Learning, OCR and Spatial Feature Mapping Techniques” (Darsha, 2023) is a project focusing one extracting information from scanned invoice using different technologies for different step. In the context of our receipt extractor project, we will focus on the receipt detection/classification and the text extraction step.

For the Optical Character Recognition, like in the paper 1 in the section *Paper 1 - Utilize OCR text to extract receipt data and classify receipts with common Machine Learning algorithms.*, similar technology were used such as Tesseract OCR and Google Cloud Vision API. Since we have already discussed the advantage and disadvantage of these technology, there is no need to reanalyze them. There it shows that these technologies are the state of the art in their domain and prove their efficiency.

As for the text detection and classification, it uses the YOLO (You Only Look Once) model which is a state-of-the-art object detection provided by Ultralitics in python. Has showed in *Table 2*, the YOLO model is well known for is high speed in the prediction, while having a low background mistake which is an important characteristic for a receipt extraction application where the picture would be taken quickly from a mobile phone. Therefore, it also has is limitation such has the trade of between speed and accuracy which could cause problem in case of too low accuracy, also these processing could be resource intense for training and prediction.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Technologies** | **Pros** | **Cons** |
| **Object detection model** | YOLO version 5 | * Good accuracy in detecting and classifying multiple objects. * Fast inference time and global speed. * Reduction of background error by processing all picture | * Require extensive computation power for many classes. * Potential trade-off between speed and accuracy. * Less densely pack data may reduce accuracy. |
| **Optical Character Recognition (OCR)** | Tesseract OCR | See *Table 1* | |
| Google Cloud Vision API | See *Table 1* | |
| Convolutional Neural Networks (CNNs) | * Good at extracting hierarcchical feature from images. * Versatile for a wide range of image recognition. | * High computational resources for training. * Risk of overfitting training data. * Need a lot o tuning and optimization for optimal performance. |

Table 2 - Advantage and Limitation of technologies used in "Information Extraction from Scanned Invoices using Machine Learning, OCR and Spatial Feature Mapping Techniques" written by Darsha.

### Paper 3 - Computer Vision for Document Image Analysis and Text Extraction

This third paper “Computer Vision for Document Image Analysis and Text Extraction” (Benchekroun, 2022) is a research paper aiming to improve Optical Character Recognition (OCR) systems, particularly for image processing. It explores multiple technologies such has Convolutional Neural Network (CNN) + Long Short-Term Memory (LSTM) Network for feature extraction from images, Deep CNN + Transformer/seq2seq Network to handle sequential data, Generative Adversarial Networks (GANs) to generate synthetic training data and Morphological Operations for preprocessing training data with different operation.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Technologies** | **Pros** | **Cons** |
| **Optical Character Recognition (OCR)** | Convolutional Neural Networks (CNN) + Long Short-Term Memory (LSTM) | * Effective for non-handwritten text. * Shows significant accuracy improvement with synthetic data. | * Are computationally intensive and may require large dataset for high accuracy. * Struggle with comp[lex patterns in data |
| Deep CNN + Transformer/seq2seq Network | * Capable of parallel processing, offering faster execution. * Can extract complex features from images. | * Require significan computation resources. * Difficulty to optimize due to his complexity. |
| **Synthetic Data Generation** | Generative Adversarial Networks (GANs) | * Increase training data. * Help model learn feature that may not be present in real-world cenario. | * Generated data might not always represent real-world cenarios. * Can reduce accuracy in case of generated data not accurate to the cenario. |

Table 3- Advantage and Limitation of technologies used in "Computer Vision for Document Image Analysis and Text Extraction" written by Benchekrou.

The advantage and disadvantage of these technologies can be found in *Table 3*. Furthermore, the usage of GANs to create synthetic data is an interesting feature to increase the training dataset and therefor increase the accuracy if an OCR algorithm is developed from scratch such has the different one used in this third research paper. Developing an OCR from scratch without using an already made system such has Tesseract or Google Cloud Vision API could allow a more precise extraction specially for Mauritian receipt, therefore it’ll increase the development time and will be limited by his training data which in out context is limited and therefore would not be recommended compared to powerful pretrained system.

## Mobile Application

### Paper 4 - React Native vs Flutter, cross-platform mobile application frameworks.

This last research paper “React Native vs Flutter, cross-platform mobile application frameworks” (Wu, 2018) focus on comparing two frameworks to develop cross-platform mobile application: React Native and Flutter. These two frameworks being prominent actors in mobile development with their advantage and disadvantage available in *Table 4*. React Native developed by Facebook is using JavaScript and React which are famous programing language but for complex development will require specific programming language per platform. One the other hand, Flutter is developed by google using Dart which is not as widely adopted as JavaScript but use a single code base for both iOS and android, however this single code base create a larger app size compared to React Native app.

In the context of the receipt extractor application, the emphasis of the mobile app will be on the execution speed and stability of the app. Therefore, Flutter has shown in *Table 4* is fast with different feature for development such has the hot reload with the portability from android to iOS without code modification.

|  |  |  |
| --- | --- | --- |
| **Programming Language** | **pros** | **Cons** |
| **React Native** | * Strong Community Support with big tech contribution. * Bring modern web techniques to mobile support. * Use JavaScript syntax extension for designing UI. * Access native hardware feature like camera and storage. * Encourage modularity and reusable of component. | * Have performance limitation compared to native app in complexes scenario using JavaScript thread or memory optimization. * Highly dependable on third-party libraries for navigation and file system operations which may affect consistency and the reliability. |
| **Flutter** | * Hight-Performance using his own rendering engine for view component offering a close performance to native application. * Using Dart Programming language which is efficient for memory management and garbage collection offering fast performance. * Providing customizable widget for development. * Hot-Reload feature for development. * Assuring a consistency across the different platform. | * Produce larger application size due to the different widget renderer that may affect the app size. * Relatively new community compared to react native which may affect the resources available. |

Table 4 - Advantage and limitation of Flutter and React Native from "React Native vs Flutter, cross-platform mobile application frameworks" written by Wu.

## Conclusion

# Requirements specification

[ tell the feature, why these features related with literature review].

* Detect Receipt Part
  + Inspired by paper 2.
  + Help for classification.
* Extract receipt data.
  + Key feature in all the papers
* Classify Receipt
  + Important for analytics
  + Use of an LLM
  + Refer to paper 1 and 2.
* User interface
  + Give access system to users.
  + No viable solution out of the paper.
* Database management
  + No referring in current paper.
  + Storing of historic for user -> may provide analytics.

## Receipt Section Detection

Has referred while analyzing *Paper 2 - Information Extraction from Scanned Invoices using Machine Learning, OCR, and Spatial Feature Mapping Techniques*  referring to Darsha work (Darsha, 2023), it showed technologies of detection of the different section of the receipt before applying the text extraction. These strategies will be kept by detecting the key element of the receipt to increase the formatting and classification of the receipt data by developing an image recognition model to identify the precises area representing some element of the receipt such has the total, items list, time and shop information.

## Text Extraction

A key feature of the receipt extraction will be the character recognition which will allow the digitalization of the data from the paper to the receipt management system. As per the inside from the work from Odd and Theologou (Odd and Theologou, 2018)(*Paper 1*), Benchekrou (Benchekroun, 2022)(*Paper 3*) and Darsha (Darsha, 2023)(*Paper 2*) showing Optical Character Recognition (OCR) technologies such has Tesseract OCR and Google Cloud Vision API, along with advance machine learning models. Has present in all three research papers criticizes, show the importance of the text extraction has a key feature impacting the precision of the global system.

## Format and Classify text extraction.

After extracting the raw data from the receipt, a formatting and classification showed in *Paper 1 - Utilize OCR text to extract receipt data and classify receipts with common Machine Learning algorithms* and *Paper 2 - Information Extraction from Scanned Invoices using Machine Learning, OCR, and Spatial Feature Mapping Techniques* underscore the effeteness of machine learning algorithm for text classification using model such has LinearSVC and MLPClassifier. This approach facilitates the classification of the receipt and the formatting of the output to facilitate the uniformities of the data structure for a possible database storage and analytics.

## User Interface

Even if the papers criticize for the receipt extraction show multiple technologies for the processing, none of them provide user interface for common user to use these technologies, which without will restrict the usage of the software to specialized set of users. Therefore, a User Interface (UI) is required increase the spectrum of possible user. As showed in *Paper 4 - React Native vs Flutter, cross-platform mobile application frameworks* base on the research paper from Wu (Wu, 2018) comparing Flutter and React Native mobile developing framework. These frameworks demonstrating the portability of the UI development through multiple devise common use of the system.

# Design

[ design of these feature (globally) with the technologies I think I’ll use (note at this point I haven’t start working in the artefact, so purely theorical]

After discussing the existing work in *Literature review* and discussing the different requirements in *Requirements specification*, we can now discuss of the design of the system of application. At that stage of development, it can be separated into two distinct part which will be working together: the server side which will manage all the computationally exhausting processing such has the “receipt extraction feature” to reduce the execution time and provide a better user experience while communicating with the database, all build in a docker container to offer an easy deployment on any device. And on the other side the mobile application which will be use has gateway for the user to use the system and communicate with the server. In the next sections we can have a list of the feature which will be integrated.

## Server

* + Receipt section detection model.
  + Receipt extraction feature (section detection + OCR + Classification + user).
  + Prediction user correction feature.
  + Server Communication.
  + Database.
  + Database management.

### Receipt extraction feature.

As previously mentioned, the receipt extraction is the key feature of the system which has for global purpose to extract from a receipt image into a specific format while predicting the type of the receipt. The field which will be extracted are the shop information, the list of items, the total and the date of the transaction to be formatted into a json format to be stored into a database. The structure of this feature is composed of three major parts, the receipt section image recognition model, followed by the text extraction of these section and finally the formatting and classification of the receipt extracted data.

The receipt section detection model is the first step of the receipt extraction feature. The model will compose of a Yolo v8 model (You Only Look Once) to predict the different receipt sections explain in the precedent paragraph trained by Mauritian receipt to better predict the section of the receipt. To enter more in detail, the prediction will take has input a picture of 640 per 640 pixel and return for each class (chop information, item list, …) their coordinate on the picture.

The second step is this feature is the extraction the text display of these class extracts through Optical Character Recognition (OCR), for each detected class, a sub-image compose of coordinate of the class prediction and the text from it are extracted. As for the choice of the OCR, based on the different OCR technologies showed in the *Literature review*, the OCR used would be Tesseract OCR due to his high efficiency and capacity to be deployed locally without using a third party services.

Finally, will come the classification and formatting of the of the output of the previous function output. This task will be performed by a Custom Multi-Layer Perceptron (MLP) to firstly classify the receipt in categories (groceries, restaurant, cosmetics, electronics, etc.), then format all the information into a specific format (see *Figure 1 - Receipt Extraction Feature output format*) in json for an easy manipulation afterward.



Figure 1 - Receipt Extraction Feature output format.

### User Correction Feature.

While the data will be automatically extracted, the user will have the opportunity to “review” the extracted data, through the flutter application interface, to ensure the correctness of the data which will be store into the database. The corrected data will be set has “reviewed” and will be accessible by the analytics feature to ensure the accuracy of the data. The structure of the feature can be seen in the sequence diagram [ sequence diagram to be input]

### Server Communication

Ad for the communication between the server and the application, the HTTP (Hypertext Transfer Protocol) which will be integrally handle by a FastAPI application host into the server. The primary function of the FastAPI app is to execute the requested server feature and relay back the result to the mobile app. Each endpoint will be related to a feature of the server and therefore forming the server’s operation core.

### Database

The server architecture will include a database system which will maintain user and receipt’s related data. Using Docker technology, the server will use a PostgreSQL database initialize alongside the server’s image. During the server’s image building, the PostgreSQL image will be initialized and if in absence of pre-existing data, the database schema will be initialized through the execution of an SQL scripts. The usage of docker will ensure a seamless integration of the system.

The database schema is composed of three primary tables, outline as follows:

* Users table: A table for user related information.
* Raw Receipt Table: A table to store the unprocessed data before any data manipulation.
* Receipts Tables: A table which will contain all data related to the receipt extraction.

Each table use a specific purpose and all tables’ keys can be found in the database schema showed in [insert reference to database schema]

## Application

* + Upload receipt picture for receipt extraction.
  + List receipt historical.
  + Analytics generation
  + Application communication with server.

### Receipt upload for receipt extraction.

On the user side, the key feature of the application is the upload of the receipt of the extraction of his data to be store in the database. The first step is the loading oh the receipt either through the loading from the phone gallery, or a live capture through the camera. It is then followed by the upload to the server through http post request where the receipt extraction will be done, the result will be store as “pending” status in the database before being send back to the mobile application. With the result of the extraction, the user will have the option to review and modify the result to correct potential error through before being send back to the server to update the receipt data stored into the database and set is status as “reviewed”.

### List Receipt Historical

The application will allow the user to review his historical entries in a form of a list where key information about the receipt will be display. It’ll query the “reviewed” status receipt from the user before formatting the data for display.

### Analytics

The application will offer to the user, the possibilities to show analytics through their data stored into the database. When this page will be load, the current user’s data will be requested to the server which will query it from the database back to the mobile application, the data will be process directly format by the application during the building of the widget two create different analytics to the user. The first graph provided will be a line graph showing the sum of purchase per day within a range of a month with the capacity to display for older month. It’ll allow the user to keep track of the money spend during the along the month. A second graph generated would be a pie graph showing the sum of money spend per receipt categories (groceries, restaurant, etc.) per month also with the abilities to see older month data, it’ll allow the user to the amount used for each categories and allow it to manage a potential busget.

# Implementation

[ implementation that I have done, how I have done, reason if change from design section (e.g.: using chatgpt instead of homemade llm)]

# Testing

[ In waterfall model: explain the different test case I have done, then testing of model (accuracy…)

# Result & Analyzing

[Result of the accuracy , then explaining the reason of these result]

# Conclusion