

**TRƯỜNG ĐẠI HỌC NAM CÀN THƠ  
KHOA CÔNG NGHỆ THÔNG TIN**



**DỰ ÁN NHÓM  
XÂY DỰNG KẾ HOẠCH CHUYỂN ĐỔI SỐ**

**HỌC PHẦN: CHUYỂN ĐỔI SỐ  
LỚP : DH22MMT01**

**XÂY DỰNG VÀ PHÁT TRIỂN WEBSITE THƯƠNG MẠI ĐIỆN  
TỬ TÍCH HỢP HỆ THỐNG TRÍCH XUẤT TÀI LIỆU THÔNG  
MINH (IDP) CHO DOANH NGHIỆP BÁN LẺ CÔNG NGHỆ**

**Nhóm Sinh Viên Thực Hiện: Nhóm 1**

**GVHD: Ths. Huỳnh Huy Tuấn**

**CÀN THƠ, THÁNG 12 NĂM 2025**

**PHIẾU CHẤM ĐIỂM**  
**DỰ ÁN NHÓM – XÂY DỰNG KẾ HOẠCH CHUYỂN ĐỔI SỐ**  
**HỌC PHẦN: CHUYỂN ĐỔI SỐ**

LỚP: .....DH22MMT01.....; NHÓM: .....01.....

**XÂY DỰNG VÀ PHÁT TRIỂN WEBSITE THƯƠNG MẠI ĐIỆN  
TỬ TÍCH HỢP HỆ THỐNG TRÍCH XUẤT TÀI LIỆU THÔNG  
MINH (IDP) CHO DOANH NGHIỆP BÁN LẺ CÔNG NGHỆ**

STT	MÃ SV	Họ tên sinh viên	Điểm Đề tài	Điểm thuyết trình	Tổng điểm	Ghi chú
1	224851	Đinh Nhựt Tùng				
2	227011	Nguyễn Tân Sang				
3	220837	Trần Ngọc Anh Thy				
4	225484	Lê Thị Diễm Thuý				
5	226685	Nguyễn Minh Tiến				
6	227084	Bùi Minh Quân				

Cần Thơ, ngày.....tháng.....năm.....  
Giảng viên

Huỳnh Huy Tuấn

## MỤC LỤC

PHẦN 1 GIỚI THIỆU ĐỀ TÀI .....	1
1.1. Lý do chọn đề tài.....	1
1.2. Mục tiêu nghiên cứu .....	1
1.2.1. Mục tiêu tổng quát .....	1
1.2.2. Mục tiêu cụ thể.....	2
1.3. Đối tượng và phạm vi nghiên cứu.....	2
1.4. Phương pháp nghiên cứu .....	2
PHẦN 2 CƠ SỞ LÝ THUYẾT .....	3
2.1. Khái niệm chuyển đổi số.....	3
2.2. Thương mại điện tử và xu hướng bán lẻ công nghệ .....	3
2.3. Kiến trúc ứng dụng Web và Framework Flask .....	4
2.4. Mô hình chuyển đổi số được áp dụng trong đề tài.....	5
2.5. Công nghệ Thương mại điện tử (E-commerce) và Xu hướng thanh toán số .....	5
2.6. Xử lý tài liệu thông minh (IDP) và Nhận dạng ký tự quang học (OCR) .....	6
2.7. Tổng quan các giải pháp hiện có trên thị trường .....	7
2.8. Công nghệ sử dụng trong đề tài .....	8
2.9. Bảo mật và tuân thủ dữ liệu trong hệ thống số .....	8
CHƯƠNG 3. PHÂN TÍCH HIỆN TRẠNG VÀ YÊU CẦU.....	9
3.1. Mô tả đối tượng áp dụng (doanh nghiệp bán lẻ điện thoại) .....	9
3.2. Hiện trạng quy trình nghiệp vụ và vấn đề .....	9
3.3. Yêu cầu chức năng .....	10
3.3.3. Mô hình Use Case tổng quan .....	11
3.4. Yêu cầu phi chức năng .....	12
3.5. Khảo sát thực tế tại 10 cửa hàng bán lẻ điện thoại ở Cần Thơ .....	13
3.6. Kết luận khảo sát và hàm ý thiết kế giải pháp .....	15
CHƯƠNG 4 GIẢI PHÁP CHUYÊN ĐỘI SỐ VÀ MÔ HÌNH TRIỂN KHAI HỆ THỐNG...16	16
4.1. Giải pháp đề xuất và thiết kế hệ thống tích hợp TechShop Plus – DocuSmart.....	16
4.2. Kiến trúc tổng thể và mô hình triển khai .....	18
4.2.1. Kiến trúc tổng thể của hệ thống .....	19
4.3. Thiết kế cơ sở dữ liệu .....	20
4.3.1. Thiết kế chi tiết cơ sở dữ liệu thương mại điện tử (TechShop Plus).....	20

4.3.2. Thiết kế cơ sở dữ liệu mô-đun IDP (DocuSmart Extractor) .....	24
4.4. Thiết kế module TMĐT (TechShop Plus).....	27
4.5. Thiết kế module IDP (DocuSmart Extractor) .....	27
4.6. Thiết kế API tích hợp .....	29
4.7. Thiết kế bảo mật.....	29
4.8. Lợi ích mang lại .....	30
4.8.1. Đối với doanh nghiệp.....	30
4.8.2. Đối với nhân viên.....	30
4.8.3. Đối với khách hàng .....	30
4.9. Kế hoạch thực hiện chuyển đổi số .....	31
4.9.1. Lộ trình thực hiện.....	31
4.9.2. Nguồn lực triển khai .....	32
4.9.3. Chỉ số đánh giá hiệu quả (KPI).....	33
<b>CHƯƠNG 5. CÀI ĐẶT VÀ TRIỂN KHAI.....</b>	<b>34</b>
5.1. Môi trường phát triển và công cụ.....	34
5.2. Cài đặt chức năng website TMĐT .....	35
5.3. Cài đặt DocuSmart IDP .....	35
5.4. Tích hợp vào quy trình nghiệp vụ .....	35
5.5. Đóng gói và chạy demo .....	35
<b>CHƯƠNG 6. KIỂM THỬ – THỰC NGHIỆM – ĐÁNH GIÁ .....</b>	<b>36</b>
6.1. Kế hoạch kiểm thử .....	36
6.2. Bộ dữ liệu thử nghiệm .....	36
6.3. Chỉ số đánh giá.....	37
6.4. Kết quả và thảo luận .....	37
6.4.1. Công thức tổng quát.....	37
6.4.2. Công thức tổng quát tính Trích xuất đúng trường chính trung bình .....	37
6.4.3. Công thức tổng quát tính tỷ lệ phân loại đúng (overall accuracy).....	38
6.5. Demo sản phẩm web và mô-đun DocuSmart .....	40
<b>CHƯƠNG 7. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN .....</b>	<b>45</b>
7.1. Kết luận .....	45
7.2. Hạn chế .....	45
7.3. Hướng phát triển .....	45
<b>TÀI LIỆU THAM KHẢO .....</b>	<b>46</b>

## **DANH MỤC BẢNG**

Bảng 1 Các giải pháp hiện có trên thị trường .....	7
Bảng 2. Các yêu cầu chức năng chính của hệ thống.....	11
Bảng 3. Danh sách cửa hàng khảo sát.....	13
Bảng 4. Khối lượng chứng từ và thời gian nhập liệu thủ công.....	14
Bảng 5. Mức độ nhu cầu và sẵn sàng áp dụng giải pháp IDP.....	15
Bảng 6. Cấu trúc bảng Users.....	20
Bảng 7. Cấu trúc bảng Products.....	21
Bảng 8. Cấu trúc bảng Orders.....	22
Bảng 9. Cấu trúc bảng OrderItems .....	23
Bảng 10. Cấu trúc bảng Documents .....	24
Bảng 11. Cấu trúc bảng ExtractedFields.....	25
Bảng 12. Mô hình dữ liệu – các bảng chính trong CSDL.....	26
Bảng 13. Danh sách API chính .....	29
Bảng 14 Giai đoạn triển khai .....	31
Bảng 16. Công nghệ và công cụ sử dụng trong dự án .....	34
Bảng 17. Bộ kiểm thử chức năng (tóm tắt).....	36
Bảng 18 trình bày kết quả thử nghiệm tham khảo của nhóm trên bộ dữ liệu mẫu .....	38

## **DANH MỤC HÌNH**

Hình 1. Sơ đồ Use Case tổng quan của hệ thống TechShop Plus tích hợp DocuSmart.....	10
Hình 2. Kiến trúc tổng thể TechShop Plus tích hợp DocuSmart .....	18
Hình 3. Kiến trúc tổng thể TechShop Plus tích hợp DocuSmart IDP .....	19
Hình 4. Sơ đồ ERD tổng quan (TMĐT + DocuSmart) .....	25
Hình 5. Luồng xử lý tài liệu trong DocuSmart (IDP Pipeline) .....	28
Hình 6 Giao diện web TechShop .....	40
Hình 7 trợ lý ảo .....	40
Hình 8 Phân loại tài liệu.....	41
Hình 9 Tóm tắt văn bản .....	41
Hình 10 Trích xuất thông tin tài liệu.....	42
Hình 15 Nhiều phương thức xuất file báo cáo .....	44

## **DANH MỤC VIẾT TẮT**

Tùy viết tắt	Giải thích
AI	Artificial Intelligence – Trí tuệ nhân tạo
API	Application Programming Interface – Giao diện lập trình ứng dụng
CSDL	Cơ sở dữ liệu
CSV	Comma-Separated Values
ERP	Enterprise Resource Planning – Hoạch định nguồn lực doanh nghiệp
IDP	Intelligent Document Processing – Xử lý tài liệu thông minh
NLP	Natural Language Processing – Xử lý ngôn ngữ tự nhiên
OCR	Optical Character Recognition – Nhận dạng ký tự quang học
RPA	Robotic Process Automation – Tự động hóa quy trình bằng phần mềm
SME	Small and Medium-sized Enterprise – Doanh nghiệp nhỏ và vừa
TMĐT	Thương mại điện tử
UI/UX	Giao diện người dùng/Trải nghiệm người dùng

## PHÂN CÔNG NHIỆM VỤ

Thành viên	Nhiệm vụ được phân công
Đinh Nhựt Tùng (224851) Nhóm trưởng	Xây dựng API , DocuSmart Backend (Flask/Python) Điều phối, tích hợp hệ thống, Viết báo cáo và thuyết trình-demo
Nguyễn Tân Sang (227011)	Thiết kế CSDL và thiết kế sơ đồ Use Case, ERD, hỗ trợ phân quyền và bảo mật cơ bản
Trần Ngọc Anh Thy (220837)	Frontend giao diện website và Thiết kế slide thuyết trình
Lê Thị Diễm Thuý (225484)	Xây dựng kịch bản demo sản phẩm đánh giá , hỗ trợ kiểm thử hệ thống và tổng hợp kết quả.
Nguyễn Minh Tiên (226685)	
Bùi Minh Quân (227084)	Trích xuất thông tin, xuất dữ liệu , Xây dựng trơ lý ảo và
Tô Văn Huyền (226900)	phối hợp kiểm thử API

## PHẦN 1 GIỚI THIỆU ĐỀ TÀI

### 1.1. Lý do chọn đề tài

Trong bối cảnh cuộc Cách mạng Công nghiệp 4.0, chuyển đổi số đã trở thành yêu cầu tất yếu đối với hầu hết các doanh nghiệp. Đặc biệt trong lĩnh vực bán lẻ công nghệ, hành vi mua sắm của khách hàng thay đổi mạnh mẽ theo hướng trực tuyến, yêu cầu doanh nghiệp có kênh bán hàng số hóa chuyên nghiệp (website/ứng dụng), đồng thời tối ưu quy trình vận hành nội bộ để giảm chi phí, tăng tốc độ phục vụ và nâng cao trải nghiệm. Bên cạnh hoạt động bán hàng, doanh nghiệp bán lẻ điện thoại còn thường xuyên phát sinh và tiếp nhận số lượng lớn chứng từ: hóa đơn đầu vào/đầu ra, phiếu giao hàng, biên nhận, hợp đồng dịch vụ, chứng từ thanh toán. Các chứng từ này phục vụ nhiều mục tiêu như hạch toán kế toán, đối soát thanh toán, bảo hành và truy xuất thông tin khi có khiếu nại.

Thực tế tại nhiều cửa hàng vừa và nhỏ (SME), phần lớn chứng từ vẫn được xử lý theo cách truyền thống: lưu bản giấy, nhập liệu thủ công vào Excel hoặc phần mềm kế toán. Cách làm này dẫn đến các hạn chế: tốn thời gian và nhân lực; dễ sai sót khi nhập số liệu; khó truy xuất lại thông tin khi cần kiểm tra nhanh; khó khai thác dữ liệu để ra quyết định. Do đó, việc kết hợp một nền tảng TMĐT với mô-đun IDP là một hướng đi phù hợp để tạo giá trị kép: vừa tăng trưởng doanh thu online, vừa tự động hóa xử lý chứng từ nội bộ.

Xuất phát từ nhu cầu thực tiễn nêu trên, nhóm lựa chọn đề tài “*Xây dựng và Phát triển website thương mại điện tử tích hợp hệ thống trích xuất tài liệu thông minh (IDP) cho doanh nghiệp bán lẻ công nghệ*”. Sản phẩm hướng tới tính khả thi ở mức đồ án: ứng dụng các công nghệ mã nguồn mở, dễ triển khai, chi phí thấp; đồng thời thể hiện được các nội dung cốt lõi của chuyển đổi số như: số hóa kênh bán hàng, tự động hóa quy trình, quản trị dữ liệu tập trung và hỗ trợ ra quyết định dựa trên dữ liệu.

### 1.2. Mục tiêu nghiên cứu

#### 1.2.1. Mục tiêu tổng quát

Mục tiêu tổng quát của đề tài là xây dựng một nền tảng chuyển đổi số tích hợp dành cho doanh nghiệp bán lẻ điện thoại, nhằm số hóa đồng thời hoạt động kinh doanh và nghiệp vụ xử lý chứng từ. Cụ thể, đề tài tập trung:

Phát triển website thương mại điện tử TechShop Plus với các chức năng cốt lõi phục vụ bán hàng trực tuyến và quản trị hệ thống.

Xây dựng mô-đun DocuSmart Extractor (IDP) thực hiện nhận dạng ký tự quang học (OCR) và trích xuất thông tin có cấu trúc từ chứng từ (hóa đơn, hợp đồng, biên nhận, chứng từ thanh toán...).

Tích hợp dữ liệu, luồng xử lý và giao diện vận hành giữa các mô-đun để hình thành một hệ thống thống nhất, hỗ trợ doanh nghiệp quản lý tập trung, giảm tải thao tác thủ công, nâng cao tính chính xác và hiệu quả quản trị.

### **1.2.2. Mục tiêu cụ thể**

Phân tích yêu cầu nghiệp vụ bán lẻ điện thoại và quy trình xử lý chứng từ trong doanh nghiệp SME.

Thiết kế kiến trúc hệ thống tích hợp gồm website TMĐT và mô-đun IDP, đảm bảo khả năng mở rộng.

Xây dựng chức năng bán hàng: quản lý sản phẩm, tìm kiếm/lọc, giỏ hàng, đặt hàng, thanh toán COD/chuyển khoản, tra cứu đơn hàng.

Xây dựng mô-đun DocuSmart: upload PDF/ảnh, tiền xử lý ảnh, OCR tiếng Việt, phân loại tài liệu và trích xuất trường thông tin quan trọng.

Hỗ trợ xuất dữ liệu trích xuất dưới dạng JSON/CSV/Excel và lưu lịch sử xử lý để tra cứu về sau.

Đánh giá hệ thống thông qua kiểm thử chức năng và thử nghiệm xử lý tài liệu với bộ dữ liệu mẫu.

### **1.3. Đối tượng và phạm vi nghiên cứu**

Đối tượng áp dụng của đề tài là doanh nghiệp bán lẻ công nghệ quy mô vừa và nhỏ (SME) kinh doanh điện thoại, phụ kiện và dịch vụ liên quan. Phạm vi nghiên cứu tập trung vào hai mảng: kênh bán hàng số (website TMĐT) và tự động hóa xử lý chứng từ bằng IDP. Về phạm vi dữ liệu, mô-đun IDP trong đồ án ưu tiên xử lý các tài liệu tiếng Việt phổ biến như hóa đơn và hợp đồng; các loại tài liệu khác được phân loại và trích xuất ở mức cơ bản.

Về công nghệ, đề tài sử dụng Python/Flask cho backend, HTML/CSS/JavaScript/Bootstrap cho frontend, SQLite/SQLAlchemy cho lưu trữ dữ liệu trong môi trường demo. Trong triển khai thực tế, có thể thay thế SQLite bằng PostgreSQL/MySQL/SQL Server tùy điều kiện doanh nghiệp. OCR sử dụng Tesseract (pytesseract) và thư viện xử lý ảnh OpenCV; xử lý PDF thông qua pdf2image/Poppler. Các module xuất dữ liệu sử dụng openpyxl (Excel) và cơ chế tạo file chuẩn định dạng CSV/JSON.

### **1.4. Phương pháp nghiên cứu**

Đề tài sử dụng kết hợp các phương pháp nghiên cứu: khảo sát – tổng hợp tài liệu về chuyển đổi số, TMĐT và IDP; phân tích hệ thống theo hướng nghiệp vụ (use case, yêu cầu chức năng/phi chức năng); thiết kế – cài đặt theo mô hình lặp (iterative) và chia module; kiểm thử – đánh giá thông qua các kịch bản sử dụng (test cases) và bộ dữ liệu tài liệu mẫu.

## PHẦN 2 CƠ SỞ LÝ THUYẾT

### 2.1. Khái niệm chuyển đổi số

Chuyển đổi số (Digital Transformation) là quá trình ứng dụng công nghệ số để thay đổi căn bản cách doanh nghiệp vận hành và tạo ra giá trị cho khách hàng. Khác với số hóa (digitization) – chủ yếu chuyển đổi dữ liệu từ dạng vật lý sang dạng số (ví dụ: scan tài liệu), chuyển đổi số nhấn mạnh sự biến đổi ở cấp độ mô hình kinh doanh, quy trình, văn hóa và năng lực ra quyết định. Đối với doanh nghiệp bán lẻ, chuyển đổi số thường thể hiện ở: kênh bán hàng đa nền tảng, quản trị kho – đơn hàng theo thời gian thực, cá nhân hóa trải nghiệm khách hàng, và tự động hóa các công việc lặp lại.

Trong khuôn khổ đề tài, chuyển đổi số được tiếp cận theo góc nhìn “kép”: chuyển đổi số hướng khách hàng thông qua website TMĐT; chuyển đổi số nội bộ thông qua tự động hóa xử lý chứng từ bằng IDP. Cách tiếp cận này phù hợp với doanh nghiệp SME vì tạo ra hiệu quả vận hành ngay lập tức và có thể triển khai theo lộ trình từng bước. Đối với SME bán lẻ công nghệ, lộ trình này thường bao gồm:

- ✚ Số hóa kênh bán hàng: Chuyển đổi từ mô hình bán hàng truyền thống (cửa hàng vật lý, điện thoại) sang các kênh số như website TMĐT, ứng dụng di động, sàn thương mại điện tử.
- ✚ Số hóa quy trình nghiệp vụ: Tự động hóa các quy trình thủ công như quản lý kho, xử lý đơn hàng, chăm sóc khách hàng (CRM), kế toán và đặc biệt là xử lý tài liệu.
- ✚ Dữ liệu hóa và phân tích: Thu thập, quản lý và phân tích dữ liệu từ mọi hoạt động để đưa ra quyết định kinh doanh dựa trên dữ liệu (data-driven decision), từ việc dự báo nhu cầu đến tối ưu chiến dịch marketing.
- ✚ Văn hóa số: Xây dựng tư duy sẵn sàng đổi mới, thích nghi với công nghệ trong toàn bộ tổ chức, từ lãnh đạo đến nhân viên.

Thách thức lớn nhất của SME không nằm ở công nghệ mà ở nguồn lực tài chính hạn chế, thiếu nhân sự công nghệ chuyên sâu và tâm lý ngại thay đổi. Do đó, một giải pháp chuyển đổi số thành công cho SME cần phải có tính khả thi cao về chi phí, dễ triển khai, vận hành và mang lại hiệu quả rõ rệt trong thời gian ngắn.

### 2.2. Thương mại điện tử và xu hướng bán lẻ công nghệ

Thương mại điện tử (TMĐT) là hoạt động mua bán hàng hóa/dịch vụ thông qua các phương tiện điện tử, tiêu biểu là website và ứng dụng di động. Trong ngành hàng điện thoại và phụ kiện, TMĐT có vai trò đặc biệt nhờ đặc tính sản phẩm dễ tiêu chuẩn hóa (thông số kỹ thuật), nhu cầu so sánh giá cao, và xu hướng mua sắm theo đánh giá (review). Doanh nghiệp bán lẻ công nghệ khi triển khai TMĐT thường cần các chức năng: danh mục sản phẩm, tìm kiếm/độc, giỏ hàng, thanh toán đa phương thức, quản lý đơn hàng và tích hợp chăm sóc khách hàng.

Bên cạnh đó, xu hướng omnichannel (đa kênh) yêu cầu dữ liệu thống nhất giữa cửa hàng vật lý và kênh online. Nếu dữ liệu bị phân mảnh (Excel rời rạc), doanh nghiệp khó kiểm soát tồn kho, khó đo lường hiệu quả chiến dịch và khó cá nhân hóa trải nghiệm.

Vì vậy, một nền tảng TMĐT có cấu trúc CSDL và báo cáo cơ bản sẽ là nền tảng để mở rộng chuyển đổi số trong tương lai (CRM, ERP, marketing automation).

### 2.3. Kiến trúc ứng dụng Web và Framework Flask

Kiến trúc ứng dụng web phổ biến hiện nay là mô hình MVC (Model-View-Controller), giúp tách biệt logic nghiệp vụ, dữ liệu và giao diện, từ đó làm cho ứng dụng dễ phát triển, bảo trì và mở rộng.

- ✚ Model: Đại diện cho cấu trúc dữ liệu và logic nghiệp vụ, tương tác trực tiếp với cơ sở dữ liệu.
- ✚ View: Là phần giao diện người dùng (UI), hiển thị dữ liệu từ Model và tiếp nhận tương tác từ người dùng.
- ✚ Controller: Đóng vai trò trung gian, nhận yêu cầu từ người dùng (qua View), xử lý logic, giao tiếp với Model và trả về kết quả cho View.

Flask là một micro-framework web nhẹ, linh hoạt và mạnh mẽ được viết bằng Python. Được chọn làm nền tảng cho đồ án, Flask sở hữu những ưu điểm vượt trội phù hợp với SME:

- ✚ Nhẹ và Linh hoạt: Không áp đặt cấu trúc thư mục hay thư viện cụ thể, cho phép nhà phát triển tự do lựa chọn các component (ORM, template engine) phù hợp nhất với dự án. Điều này phù hợp với việc xây dựng một giải pháp tùy chỉnh cao.
- ✚ Hệ sinh thái phong phú: Có một hệ sinh thái mở rộng (Flask-Extensions) không lồ cho hầu hết mọi nhu cầu: xác thực người dùng (Flask-Login), quản trị CSDL (Flask-SQLAlchemy), upload file (Flask-Upserts), REST API (Flask-RESTful).
- ✚ Dễ học và phát triển nhanh: Cú pháp trong sáng, tài liệu dồi dào, cho phép phát triển prototype và ứng dụng hoàn chỉnh một cách nhanh chóng, giảm thiểu thời gian và chi phí phát triển – một yếu tố then chốt cho các dự án của SME.
- ✚ Hiệu năng đủ tốt cho quy mô SME: Với kiến trúc không đồng bộ (có thể kết hợp với Gevent hoặc Gunicorn), Flask hoàn toàn có thể đáp ứng lưu lượng truy cập của một website TMĐT cỡ vừa và nhỏ.

So sánh với các framework khác như Django (nặng, áp đặt cấu trúc) hoặc Node.js (phức tạp với asynchronous), Flask tỏ ra là lựa chọn cân bằng giữa sự linh hoạt, tốc độ phát triển và hiệu năng, lý tưởng cho một đồ án tích hợp đa chức năng như đề tài.

## 2.4. Mô hình chuyển đổi số được áp dụng trong đề tài

Trong phạm vi đề tài, nhóm lựa chọn và áp dụng mô hình chuyển đổi số theo 4 giai đoạn (Digital Transformation Roadmap – 4 phases) vì phù hợp với đặc thù doanh nghiệp bán lẻ quy mô vừa và nhỏ (SME), dễ triển khai theo từng bước, đảm bảo tính khả thi trong điều kiện nguồn lực hạn chế. Mô hình gồm:

- ✚ Giai đoạn 1 – Chuẩn bị: Khảo sát hiện trạng, xác định vấn đề, lập kế hoạch chuyển đổi số, lựa chọn công nghệ và phạm vi triển khai.
- ✚ Giai đoạn 2 – Triển khai thử nghiệm: Xây dựng hệ thống lõi (website TMĐT và mô-đun IDP), thử nghiệm trên dữ liệu mẫu và một nhóm quy trình nghiệp vụ giới hạn.
- ✚ Giai đoạn 3 – Vận hành chính thức: Đưa hệ thống vào vận hành, chuẩn hóa quy trình, đào tạo người dùng, theo dõi hiệu quả và tối ưu.
- ✚ Giai đoạn 4 – Mở rộng và cải tiến: Nâng cấp tính năng (AI nâng cao, tự động hóa sâu hơn), mở rộng đối tượng tài liệu và tích hợp thêm dịch vụ/hệ thống khác.

*Lý do chọn mô hình này*

- ✚ Có lộ trình rõ ràng, phù hợp triển khai theo từng mức độ trưởng thành số.
- ✚ Cho phép kiểm soát rủi ro khi áp dụng công nghệ mới (đặc biệt là IDP/OCR).
- ✚ Dễ đo lường hiệu quả từng giai đoạn (thời gian xử lý chứng từ, độ chính xác trích xuất, mức độ hài lòng người dùng).

## 2.5. Công nghệ Thương mại điện tử (E-commerce) và Xu hướng thanh toán số

Một nền tảng TMĐT cơ bản bao gồm các module chính: Danh mục sản phẩm, Giỏ hàng, Thanh toán, Quản trị đơn hàng và Hệ thống quản trị (Admin Dashboard). Yếu tố then chốt quyết định tỷ lệ chuyển đổi (conversion rate) chính là trải nghiệm thanh toán.

Tại Việt Nam, bên cạnh phương thức COD (Cash on Delivery) – Thanh toán khi nhận hàng truyền thống, thanh toán số đang bùng nổ. Trong đó, VietQR nổi lên như một giải pháp tối ưu cho SME nhờ:

- ✚ Tiêu chuẩn mở: Được chuẩn hóa bởi Ngân hàng Nhà nước, cho phép thanh toán xuyên ngân hàng chỉ với một mã QR duy nhất.
- ✚ Trải nghiệm đơn giản: Khách hàng chỉ cần mở ứng dụng ngân hàng, quét mã QR và xác nhận. Không cần nhập số tài khoản, tên người thụ hưởng.
- ✚ Chi phí bằng 0: SME không phải trả phí duy trì hay phí giao dịch như các công thanh toán trực tuyến khác.
- ✚ Tích hợp dễ dàng: Có thể tạo mã QR động theo số tiền và nội dung chuyển khoản thông qua API công khai của các ngân hàng hoặc sử dụng dịch vụ tạo mã của các nhà cung cấp như vietqr.io.

Việc tích hợp VietQR vào website TMĐT không chỉ mang lại sự tiện lợi, an toàn cho khách hàng mà còn giúp doanh nghiệp tự động hóa một phần quy trình xác nhận thanh toán (khi kết hợp với webhook hoặc cho phép khách hàng upload ảnh chứng từ), giảm tải công việc thủ công cho nhân viên.

## 2.6. Xử lý tài liệu thông minh (IDP) và Nhận dạng ký tự quang học (OCR)

IDP (Intelligent Document Processing) là tập hợp các kỹ thuật và công nghệ nhằm tự động hóa quá trình xử lý tài liệu, bao gồm: thu nhận (capture), nhận dạng nội dung (OCR), hiêu ngữ cảnh (classification/extraction), trích xuất dữ liệu có cấu trúc và tích hợp vào hệ thống nghiệp vụ. IDP được xem là bước tiến so với OCR truyền thống vì không chỉ “chuyển ảnh thành chữ” mà còn “hiểu và lấy đúng trường thông tin” để phục vụ nhập liệu, đối soát và phân tích.

Trong doanh nghiệp bán lẻ, IDP có thể áp dụng cho hóa đơn đầu vào/đầu ra, phiếu nhập – xuất kho, hợp đồng bảo hành, biên bản nghiệm thu... Khi IDP được tích hợp vào hệ thống bán hàng, doanh nghiệp có thể giảm đáng kể thời gian xử lý chứng từ, hạn chế sai sót và tăng khả năng truy xuất thông tin.

Quy trình xử lý của một hệ thống IDP cơ bản:

1. Thu thập & Tiền xử lý: Nhận tài liệu đầu vào (scan, file ảnh, PDF). Tiền xử lý ảnh (cắt xén, xoay, cân bằng sáng, khử nhiễu) để tối ưu chất lượng đầu vào cho OCR.
2. OCR (Nhận dạng ký tự quang học): Chuyển đổi phần chữ trong hình ảnh thành văn bản máy có thể đọc được. Tesseract OCR là công cụ mã nguồn mở mạnh mẽ, hỗ trợ tiếng Việt tốt, được cộng đồng đánh giá cao và là lựa chọn hàng đầu cho các dự án có ngân sách hạn chế.
3. Phân loại tài liệu (Document Classification): Xác định loại tài liệu (hóa đơn, hợp đồng, chứng từ ngân hàng, v.v.). Có thể dựa trên quy tắc (rule-based) với từ khóa đặc trưng hoặc sử dụng mô hình học máy để có độ chính xác cao hơn.
4. Trích xuất thông tin (Information Extraction): Tìm và trích xuất các trường dữ liệu có cấu trúc (structured data) từ văn bản đã OCR. Phương pháp phổ biến là sử dụng Biểu thức chính quy (Regex) để bắt mẫu (pattern matching) cho các trường có định dạng cố định (mã số thuế, ngày tháng, số tiền). Đối với các trường phức tạp hơn, có thể sử dụng Mô hình học máy được huấn luyện hoặc kết hợp với API của các nhà cung cấp AI (như OpenAI).
5. Xác thực & Xuất dữ liệu: Kiểm tra tính hợp lệ của dữ liệu trích xuất, sau đó xuất ra định dạng mong muốn (CSV, JSON, nhập trực tiếp vào hệ thống quản lý).

Việc ứng dụng IDP vào bài toán xử lý hóa đơn, hợp đồng cho SME bán lẻ công nghệ sẽ giảm thiểu tới 80% thời gian nhập liệu thủ công, hạn chế sai sót và tạo kho dữ liệu số có thể tìm kiếm, phân tích được.

## 2.7. Tổng quan các giải pháp hiện có trên thị trường

Trước khi đề xuất giải pháp tích hợp, nhóm tiến hành khảo sát và phân tích các lựa chọn hiện có mà SME có thể cân nhắc:

Nhóm Giải Pháp	Đại diện	Ưu Điểm	Nhược Điểm & Bất Cập đối với SME Bán Lẻ Công Nghệ
Sàn TMĐT lớn	Shopee, Tiki, Lazada	<ul style="list-style-type: none"> <li>- Lưu lượng khách hàng khổng lồ.</li> <li>- Hệ thống vận hành trơn tru.</li> <li>- Dễ bắt đầu.</li> </ul>	<ul style="list-style-type: none"> <li>- Cạnh tranh khốc liệt, chi phí quảng cáo cao.</li> <li>- Không sở hữu dữ liệu khách hàng.</li> <li>- Hoa hồng cao (5-15%).</li> <li>- Không giải quyết bài toán IDP nội bộ.</li> </ul>
Dịch vụ bao Thuê TMĐT	Haravan, Shopify (cơ bản)	<ul style="list-style-type: none"> <li>- Triển khai nhanh, giao diện đẹp.</li> <li>- Ít phải lo về kỹ thuật.</li> </ul>	<ul style="list-style-type: none"> <li>- Chi phí thuê bao hàng tháng tăng dần theo gói.</li> <li>- Hạn chế tùy biến sâu.</li> <li>- Không có hoặc rất ít tính năng IDP.</li> </ul>
Phần mềm Quản lý Bán hàng	KiotViet, Sapo, <a href="#">Nhanh.vn</a>	<ul style="list-style-type: none"> <li>- Quản lý tồn kho, đơn hàng, khách hàng tại cửa hàng hiệu quả.</li> <li>- Có sẵn phần mềm.</li> </ul>	<ul style="list-style-type: none"> <li>- Website bán hàng đi kèm thường kém linh hoạt, thẩm mỹ.</li> <li>- Module IDP (nếu có) là tính năng cao cấp, giá rất cao.</li> <li>- Có thể bị "khóa" vào một nền tảng.</li> </ul>
Giải pháp IDP Độc lập	ABBYY, Google Document AI	<ul style="list-style-type: none"> <li>- Độ chính xác OCR và trích xuất rất cao.</li> <li>- Công nghệ tiên tiến.</li> </ul>	<ul style="list-style-type: none"> <li>- Chi phí cực kỳ cao (bản quyền/phí API).</li> <li>- Chỉ giải quyết một mảng (IDP), không tích hợp sẵn với hệ thống bán hàng, cần phát triển kết nối phức tạp.</li> </ul>
Xây dựng Tùy chỉnh (Custom)	Thuê đội ngũ phát triển riêng	<ul style="list-style-type: none"> <li>- Kiểm soát toàn bộ mã nguồn và tính năng.</li> <li>- Tích hợp đa hệ thống theo ý muốn.</li> </ul>	<ul style="list-style-type: none"> <li>- Chi phí đầu tư ban đầu rất lớn, thời gian dài.</li> <li>- Đòi hỏi năng lực quản lý dự án và kỹ thuật từ phía doanh nghiệp.</li> </ul>

Bảng 1 Các giải pháp hiện có trên thị trường

Khoảng trống thị trường (Market Gap): Từ phân tích trên, có thể thấy một khoảng trống rõ rệt cho một giải pháp vừa cung cấp website TMĐT chuyên nghiệp, vừa tích hợp sẵn

công cụ IDP cơ bản với chi phí cực kỳ tối ưu. Giải pháp này cần kết hợp ưu điểm của việc xây dựng tùy chỉnh (kiểm soát, tích hợp) nhưng với mức chi phí và độ phức tạp được giảm thiểu đáng kể, phù hợp với nguồn lực của SME. Đây chính là giá trị cốt lõi mà đồ án TechShop Plus with DocuSmart IDP hướng tới.

## 2.8. Công nghệ sử dụng trong đề tài

- ✚ Frontend (Website): HTML5, CSS3, JavaScript (tương thích mô hình web tĩnh/đơn giản).
- ✚ Backend TMĐT & API: Node.js/Express (phục vụ API sản phẩm/đơn hàng từ SQL – theo mô tả tích hợp giỏ hàng lấy sản phẩm từ API).
- ✚ Backend IDP: Python + Flask (xử lý upload, OCR, trích xuất, trả kết quả JSON).
- ✚ OCR & xử lý tài liệu: Tesseract OCR (pytesseract), pdf2image, OpenCV (tiền xử lý ảnh).
- ✚ Trích xuất thông tin: Regex/Rule-based (trích mã số thuế, ngày, tổng tiền...), có thể mở rộng NLP sau.
- ✚ CSDL: SQL Server/MySQL/SQLite tùy mô hình triển khai; đề tài ưu tiên mô hình dễ chạy demo và có thể nâng cấp.
- ✚ Xuất dữ liệu: JSON/CSV/Excel (XLSX), lưu lịch sử xử lý.

## 2.9. Bảo mật và tuân thủ dữ liệu trong hệ thống số

Hệ thống TMĐT và IDP đều xử lý dữ liệu nhạy cảm: thông tin cá nhân khách hàng, thông tin đơn hàng và chứng từ tài chính. Vì vậy, các nguyên tắc bảo mật cơ bản cần được áp dụng ngay từ giai đoạn thiết kế: xác thực – phân quyền, mã hóa đường truyền (HTTPS), kiểm soát upload file, ghi log và sao lưu dữ liệu. Trong triển khai thực tế, doanh nghiệp cần bổ sung chính sách lưu trữ – hủy dữ liệu theo thời hạn, và tuân thủ quy định pháp luật liên quan đến dữ liệu cá nhân.

## CHƯƠNG 3. PHÂN TÍCH HIỆN TRẠNG VÀ YÊU CẦU

### 3.1. Mô tả đối tượng áp dụng (doanh nghiệp bán lẻ điện thoại)

Đối tượng áp dụng là các cửa hàng bán lẻ điện thoại và phụ kiện công nghệ quy mô vừa và nhỏ. Đặc trưng của mô hình này gồm: danh mục sản phẩm biến động theo vòng đời thiết bị (model mới ra mắt liên tục), biên lợi nhuận phụ thuộc vào giá nhập và chương trình khuyến mãi, khối lượng đơn hàng tăng mạnh theo mùa (tết, khai giảng, sale). Trong vận hành, doanh nghiệp vừa cần kênh bán hàng hiệu quả, vừa cần quy trình hậu cần – kê toán gọn nhẹ nhưng đảm bảo kiểm soát.

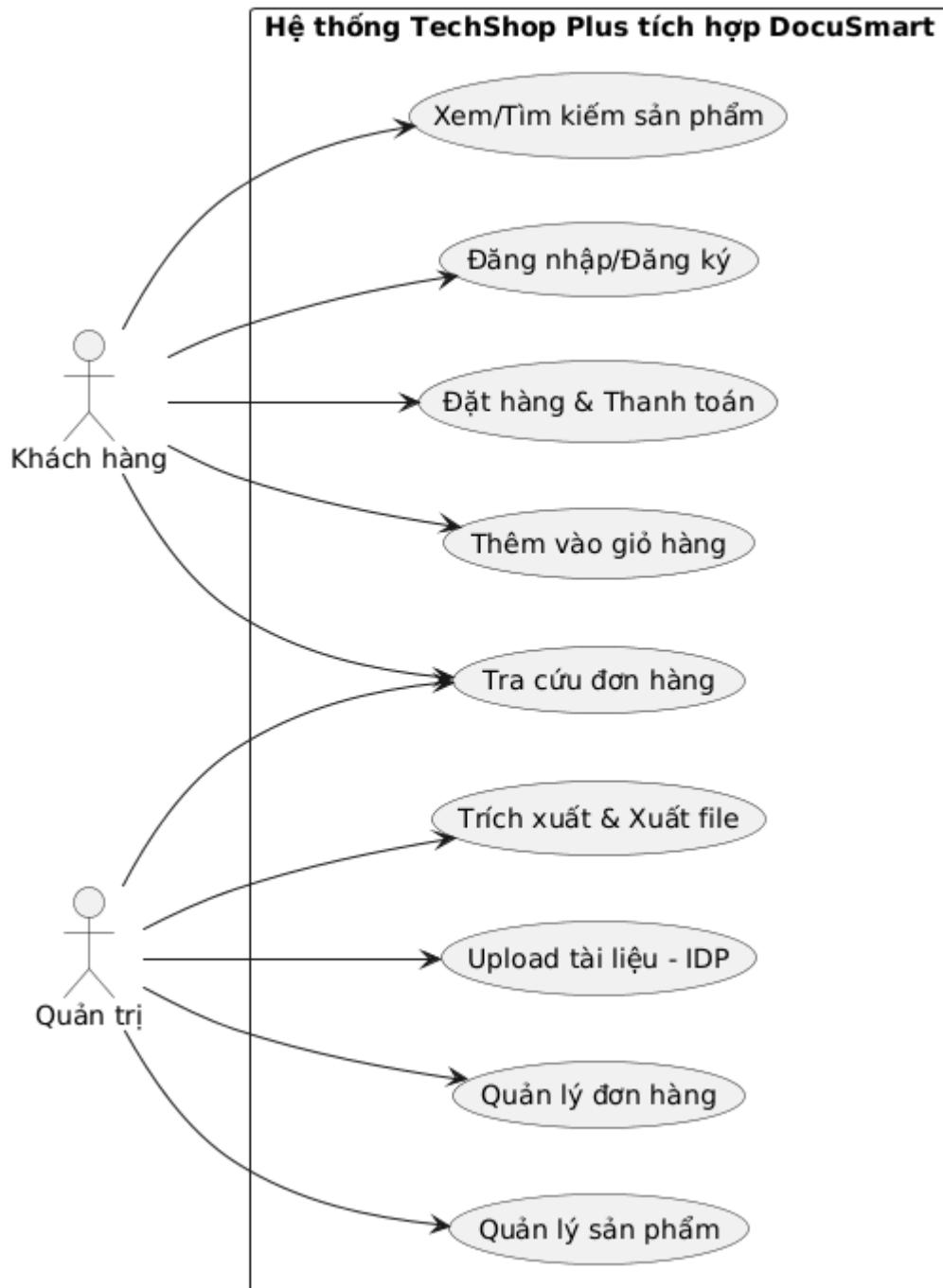
### 3.2. Hiện trạng quy trình nghiệp vụ và vấn đề

Qua khảo sát mô hình SME, nhóm tổng hợp hiện trạng phổ biến như sau:

- + Kênh bán hàng online thiếu chuyên nghiệp: bán qua mạng xã hội, chốt đơn thủ công; khó thống kê doanh thu và tỷ lệ chuyển đổi.
- + Quản trị dữ liệu phân tán: sản phẩm/đơn hàng lưu ở nhiều file Excel; không có CSDL tập trung; khó truy xuất lịch sử mua hàng.
- + Xử lý chứng từ thủ công: hóa đơn – hợp đồng được lưu giấy; nhập tay vào Excel/phần mềm kế toán; dễ sai sót và tốn thời gian.
- + Thiếu công cụ đồi soát: khi cần kiểm tra nhà cung cấp/chi phí, phải tìm lại bản giấy hoặc lục email; chậm trễ ra quyết định.

Các vấn đề này cho thấy nhu cầu chuyển đổi số không chỉ nằm ở việc “có website”, mà còn ở việc tự động hóa quy trình nội bộ và quản trị dữ liệu. Từ đó, nhóm xác định yêu cầu hệ thống tích hợp gồm: website TMĐT phục vụ khách hàng và module IDP hỗ trợ nghiệp vụ nội bộ.

### 3.3. Yêu cầu chức năng



Hình 1. Sơ đồ Use Case tổng quan của hệ thống TechShop Plus tích hợp DocuSmart

Dựa trên yêu cầu nghiệp vụ của doanh nghiệp bán lẻ công nghệ, nhóm xây dựng mô hình Use Case tổng quan nhằm mô tả các tác nhân (khách hàng, quản trị viên) và các chức năng cốt lõi của hệ thống. Sơ đồ giúp làm rõ phạm vi chức năng, đồng thời là cơ sở cho thiết kế chi tiết và kiểm thử.

### 3.3.3. Mô hình Use Case tổng quan

Mã yêu cầu	Nhóm chức năng	Mô tả
FR-01	Quản lý sản phẩm	Hiển thị danh sách sản phẩm; lọc theo hãng/giá; xem chi tiết; quản trị thêm/sửa/xóa sản phẩm.
FR-02	Tài khoản	Đăng ký/đăng nhập; phân quyền user/admin; lưu thông tin người mua.
FR-03	Giỏ hàng	Thêm/xóa sản phẩm; thay đổi số lượng; tính tổng tiền; đồng bộ dữ liệu.
FR-04	Đặt hàng & thanh toán	Tạo đơn hàng; hỗ trợ COD và chuyển khoản; với COD bắt buộc nhập đủ thông tin nhận hàng.
FR-05	Quản trị đơn hàng	Xem danh sách đơn; cập nhật trạng thái; xuất hóa đơn; thống kê.
FR-06	DocuSmart – Upload	Upload PDF/ảnh; kiểm tra định dạng/kích thước; lưu file tạm và log xử lý.
FR-07	DocuSmart – OCR	Tiền xử lý ảnh; OCR tiếng Việt; hiển thị văn bản trích xuất và bản xem trước.
FR-08	DocuSmart – Phân loại	Phân loại tài liệu: hóa đơn/hợp đồng/khác; hiển thị độ tin cậy.
FR-09	DocuSmart – Trích xuất	Trích xuất trường: số hóa đơn, ngày, mã số thuế, tổng tiền...; cho phép xem/chỉnh sửa.
FR-10	Xuất dữ liệu & lịch sử	Xuất JSON/CSV/Excel; lưu lịch sử xử lý; tra cứu lại kết quả theo thời gian/tên file.

Bảng 2. Các yêu cầu chức năng chính của hệ thống

### **3.4. Yêu cầu phi chức năng**

Các yêu cầu phi chức năng được xác định nhằm đảm bảo hệ thống khả dụng và an toàn:

- ✚ Hiệu năng: trang sản phẩm và giỏ hàng phản hồi nhanh; xử lý OCR với tài liệu 1–3 trang ở mức chấp nhận được trong môi trường demo.
- ✚ Khả dụng: giao diện đơn giản, dễ sử dụng; chạy ổn định trên trình duyệt phổ biến (Chrome/Edge).
- ✚ Bảo mật: phân quyền rõ ràng; kiểm soát upload; hạn chế lộ dữ liệu nhạy cảm; cấu hình secret key qua biến môi trường.
- ✚ Mở rộng: kiến trúc module hóa; có thể nâng cấp OCR/AI hoặc thay DB mà không cần viết lại toàn bộ hệ thống.
- ✚ Khả chuyển: đóng gói chạy local bằng virtualenv/Docker; triển khai server Linux với Gunicorn + Nginx.

### 3.5. Khảo sát thực tế tại 10 cửa hàng bán lẻ điện thoại ở Cần Thơ

Nhằm làm rõ bối cảnh nghiệp vụ và lượng chứng từ cần xử lý trong hoạt động bán lẻ điện thoại, nhóm thực hiện một khảo sát mô phỏng trên 10 cửa hàng tại thành phố Cần Thơ. Do giới hạn thời gian và phạm vi đồ án, số liệu trong phần này được xây dựng theo hướng “giả định có cơ sở” (số liệu ảo), bám sát đặc thù vận hành thực tế của nhóm doanh nghiệp vừa và nhỏ (SME) trong lĩnh vực bán lẻ công nghệ.

Phương pháp thực hiện gồm: phỏng vấn bán cấu trúc (semi-structured interview) dựa trên bộ câu hỏi chuẩn về quy trình tiếp nhận – lưu trữ – nhập liệu chứng từ; quan sát quy trình lưu trữ chứng từ và công cụ sử dụng (giấy, Excel, phần mềm bán hàng/ kế toán); tổng hợp và lượng hóa các chỉ tiêu: số lượng chứng từ/tháng, thời gian nhập liệu thủ công và tỷ lệ sai sót ước tính.

STT	Tên cửa hàng	Quận/Huyện	Quy mô	Nhân sự (người)	Kênh bán
1	TechShop Cần Thơ	Ninh Kiều	Vừa	8	Offline + Online
2	Điện Thoại Minh Phát	Cái Răng	Nhỏ	4	Offline
3	Mobile 68	Bình Thủy	Nhỏ	5	Offline + Online
4	CT Smartphone	Ninh Kiều	Vừa	10	Offline + Online
5	Hoàng Gia Mobile	Ô Môn	Nhỏ	3	Offline
6	Khang Mobile	Thốt Nốt	Nhỏ	4	Offline + Online
7	Laptop & Phone 99	Phong Điền	Nhỏ	3	Offline
8	SangStore	Ninh Kiều	Vừa	7	Offline + Online
9	An Thy Mobile	Cờ Đỏ	Nhỏ	3	Offline
10	QuânTech	Bình Thủy	Nhỏ	5	Offline + Online

Bảng 3. Danh sách cửa hàng khảo sát

Nhìn chung, các cửa hàng trong mẫu khảo sát có quy mô nhỏ đến vừa (3–10 nhân sự), kinh bán chủ yếu vẫn là cửa hàng vật lý; một số cửa hàng đã mở rộng thêm fanpage/Zalo để tiếp nhận đơn hàng trực tuyến. Điểm chung là chứng từ đầu vào (hóa đơn mua hàng, phiếu giao, hợp đồng/biên nhận) và chứng từ thanh toán (ủy nhiệm chi, biên lai chuyển khoản) vẫn được lưu rời rạc dưới dạng giấy hoặc file PDF/ảnh, gây khó khăn cho đối soát và tra cứu.

STT	HĐ đầu vào/tháng	HĐ bán ra/tháng	HĐ/Hợp đồng/biên nhận	Nhập liệu (giờ/tháng)	Sai sót ước tính (%)	Công cụ hiện tại
1	120	260	18	22	3.5	Excel rời rạc
2	80	140	10	14	4.0	Sổ tay + Excel
3	95	190	12	16	3.0	Excel + Google Drive
4	150	320	24	28	2.5	Phần mềm bán hàng cơ bản
5	60	110	8	12	4.5	Giấy + nhập lại
6	70	160	9	13	3.8	Excel
7	55	90	6	9	5.0	Giấy + Excel
8	130	240	15	20	3.2	Excel + phần mềm kê toán
9	50	80	5	8	5.5	Giấy
10	110	210	14	18	3.0	Excel + hóa đơn điện tử

Bảng 4. Khối lượng chứng từ và thời gian nhập liệu thủ công

Từ bảng số liệu, có thể ước lượng trung bình mỗi cửa hàng phát sinh khoảng 92 hóa đơn đầu vào/tháng, 180 hóa đơn bán ra/tháng và 12 chứng từ dạng hợp đồng/biên nhận/tháng. Thời gian nhập liệu thủ công trung bình đạt 16 giờ/tháng, với tỷ lệ sai sót ước tính khoảng 3.8%. Các sai sót thường gặp gồm: nhập sai mã số thuế, sai số hóa đơn, nhầm định dạng ngày, hoặc thiếu dòng tổng tiền/thuế VAT khi đối soát.

STT	Mức độ cần thiết (1–5)	Ngân sách dự kiến (triệu VNĐ/năm)	Tính năng ưu tiên
1	5	8	Tự động trích xuất MST/ngày/tổng tiền
2	4	5	Lưu trữ & tìm kiếm chứng từ
3	4	6	Xuất Excel/CSV cho kế toán
4	5	10	Tích hợp vào quy trình đơn hàng
5	4	4	Giảm nhập liệu thủ công
6	4	5	Nhận diện hóa đơn/hợp đồng
7	3	3	Upload nhanh, xem kết quả
8	5	7	Lịch sử xử lý + phân quyền
9	3	2	Tìm kiếm theo MST/số hóa đơn
10	4	6	Tự động OCR PDF/ảnh

Bảng 5. Mức độ nhu cầu và sẵn sàng áp dụng giải pháp IDP

Kết quả cho thấy mức độ cần thiết trung bình đạt 4.1/5, phản ánh nhu cầu rõ rệt về tự động hóa xử lý chứng từ. Ngân sách dự kiến trung bình khoảng 5.6 triệu VNĐ/năm (mô phỏng), phù hợp với định hướng lựa chọn công nghệ mã nguồn mở và triển khai theo mô hình từng bước (phased rollout) của đề tài.

### 3.6. Kết luận khảo sát và hàm ý thiết kế giải pháp

Từ khảo sát, có thể rút ra ba hàm ý chính cho việc thiết kế hệ thống TechShop Plus tích hợp DocuSmart Extractor:

- ✚ Thứ nhất, dữ liệu chứng từ đến từ nhiều định dạng (ảnh chụp, PDF, hóa đơn điện tử) và nhiều nguồn (nhà cung cấp, ngân hàng, khách hàng), do đó hệ thống IDP cần hỗ trợ đa định dạng và tiền xử lý ảnh để tăng chất lượng OCR.
- ✚ Thứ hai, thông tin phục vụ kế toán – đối soát tập trung vào một số trường cốt lõi (mã số thuế, số hóa đơn, ngày lập, tổng tiền, thuế VAT, danh mục hàng hóa). Vì vậy, đề tài ưu tiên trích xuất các trường này trước theo hướng rule-based/regex kết hợp phân loại tài liệu.
- ✚ Thứ ba, người dùng mong muốn kết quả được xuất ra định dạng quen thuộc (Excel/CSV) và có lịch sử tra cứu. Do đó, hệ thống cần có cơ chế lưu lịch sử xử lý, cho phép người dùng xem lại kết quả và xuất dữ liệu phục vụ nghiệp vụ.

## CHƯƠNG 4 GIẢI PHÁP CHUYÊN ĐỔI SỐ VÀ MÔ HÌNH TRIỂN KHAI HỆ THỐNG

### 4.1. Giải pháp đề xuất và thiết kế hệ thống tích hợp TechShop Plus – DocuSmart

Giải pháp đề xuất là xây dựng một nền tảng chuyên đổi số tích hợp cho doanh nghiệp bán lẻ điện thoại (TechShop Plus), kết hợp đồng bộ giữa kênh bán hàng trực tuyến và mô-đun xử lý tài liệu thông minh (IDP) “DocuSmart Extractor”. Hệ thống hướng tới mục tiêu số hóa toàn bộ chuỗi nghiệp vụ cốt lõi của cửa hàng: trưng bày – bán hàng – đặt hàng – thanh toán – quản trị đơn hàng – đối soát chứng từ, đồng thời giảm thiểu công việc nhập liệu thủ công và nâng cao khả năng truy xuất dữ liệu phục vụ kê toán/kiểm soát.

Trên phương diện kiến trúc hệ thống, giải pháp được thiết kế theo mô hình hai lớp nghiệp vụ gồm khối Thương mại điện tử (E-commerce) và khối Xử lý tài liệu thông minh (IDP – DocuSmart), hai khối trao đổi dữ liệu với nhau thông qua các API.

- ✚ Khối E-commerce: Đóng vai trò là lớp ứng dụng hướng người dùng, cung cấp giao diện và các chức năng phục vụ hoạt động bán hàng như quản lý khách hàng, quản lý giỏ hàng, tạo và xử lý đơn hàng, cũng như tra cứu trạng thái/chi tiết đơn hàng.
- ✚ Khối IDP (DocuSmart): Hoạt động như một dịch vụ backend chuyên biệt cho xử lý tài liệu. Hệ thống thực hiện quy trình OCR – phân loại – trích xuất thông tin, đồng thời lưu trữ lịch sử xử lý và kết quả nhằm phục vụ mục đích theo dõi, kiểm tra và đối soát dữ liệu.

Cụ thể, giải pháp bao gồm các mô-đun chức năng sau:

Website bán hàng (Frontend Web): hiển thị danh mục sản phẩm, tìm kiếm/ lọc theo hãng – giá – cấu hình; trang chi tiết sản phẩm; giỏ hàng; đặt hàng và lựa chọn phương thức thanh toán (COD/chuyển khoản).

Quản trị (Admin Dashboard): quản lý sản phẩm (CRUD), hình ảnh, khuyến mãi; quản lý đơn hàng và trạng thái; thống kê doanh thu; xuất báo cáo.

Mô-đun thanh toán & thông tin giao hàng: khi người dùng chọn COD, hệ thống bắt buộc nhập đủ họ tên – số điện thoại – địa chỉ giao hàng trước khi xác nhận đặt hàng; với chuyển khoản có hướng dẫn nội dung và đối soát theo mã đơn.

DocuSmart Extractor (IDP): upload PDF/ảnh chứng từ; tiền xử lý ảnh; OCR tiếng Việt; phân loại tài liệu (hóa đơn/hợp đồng/khác); trích xuất trường thông tin (mã số thuế, ngày, tổng tiền, số chứng từ...); xuất JSON/CSV/Excel; lưu lịch sử xử lý để tra cứu.

Cơ sở dữ liệu tập trung: lưu dữ liệu sản phẩm – đơn hàng – người dùng; lưu metadata và kết quả trích xuất của tài liệu phục vụ đối soát và kiểm toán nội bộ.

Quy trình vận hành của hệ thống được mô tả theo các bước chính như sau:

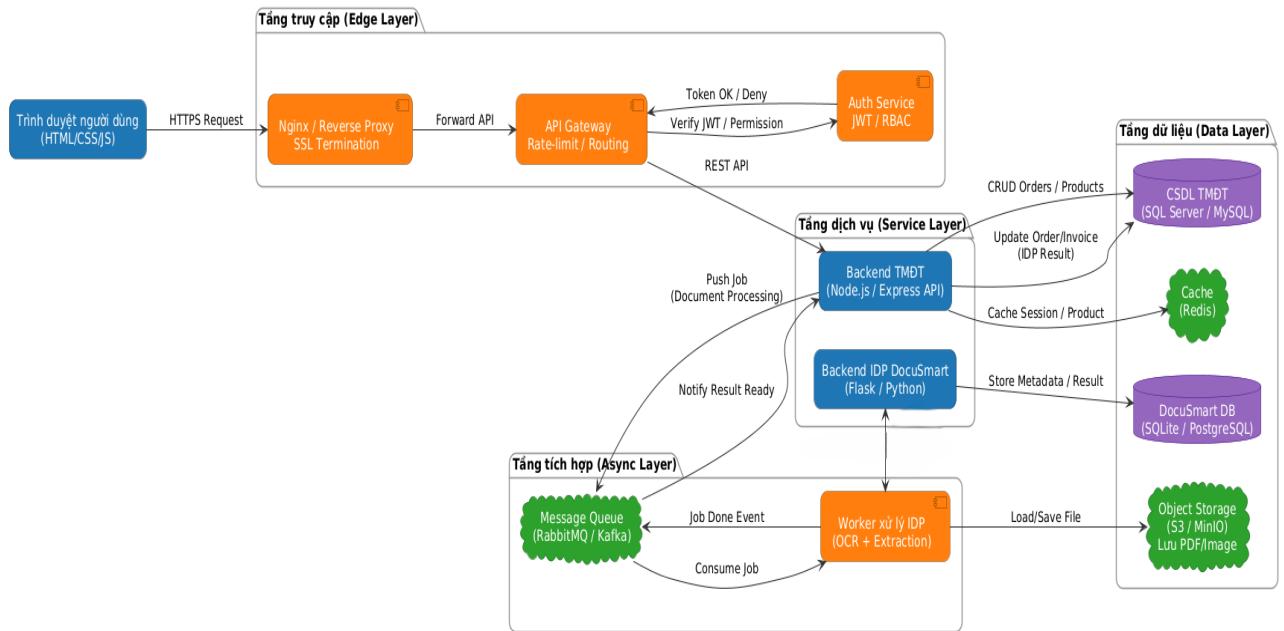
1. Khách hàng thực hiện thao tác đặt hàng trên website thương mại điện tử.
2. Hệ thống E-commerce tiếp nhận yêu cầu, tiến hành tạo đơn hàng và lưu trữ các thông tin liên quan như thanh toán, giao nhận, và trạng thái xử lý đơn.

3. Nhân viên (bộ phận bán hàng/kế toán) tải lên hoặc đính kèm các chứng từ liên quan (ví dụ: hóa đơn đầu vào, biên nhận chuyển khoản, phiếu thu...) vào hệ thống DocuSmart.
4. DocuSmart thực hiện xử lý tự động theo pipeline: tiền xử lý ảnh → OCR → phân loại tài liệu → trích xuất thông tin, nhằm chuyển đổi dữ liệu từ dạng không cấu trúc sang dữ liệu cấu trúc.
5. Kết quả trích xuất được lưu trữ tập trung và sử dụng để đối soát (giá trị thanh toán, ngày chứng từ, mã số thuế, số chứng từ...), qua đó hỗ trợ bộ phận kế toán trong việc tổng hợp và lập báo cáo.

Về mặt công nghệ, giải pháp ưu tiên lựa chọn các thành phần mã nguồn mở, dễ triển khai và phù hợp với điều kiện vận hành của doanh nghiệp quy mô vừa và nhỏ (SME). Các nhóm công nghệ chính bao gồm:

- ✚ Frontend: HTML/CSS/JavaScript kết hợp Bootstrap nhằm xây dựng giao diện trực quan, dễ sử dụng và tương thích đa thiết bị.
- ✚ Backend (E-commerce/API): Node.js/Express được sử dụng để triển khai các API phục vụ nghiệp vụ sản phẩm và đơn hàng (đặc biệt phù hợp trong trường hợp hệ thống bán hàng đã có sẵn nền tảng API).
- ✚ Backend (DocuSmart/IDP): Flask/Python được lựa chọn để triển khai dịch vụ xử lý tài liệu, do phù hợp với hệ sinh thái thư viện AI/OCR và thuận tiện trong xây dựng pipeline xử lý.
- ✚ Cơ sở dữ liệu: Có thể sử dụng SQL Server/MySQL/SQLite tùy thuộc vào môi trường triển khai và quy mô dữ liệu.
- ✚ OCR: Ưu tiên các công cụ OCR mã nguồn mở như Tesseract hoặc PaddleOCR, đáp ứng yêu cầu nhận dạng ký tự và hỗ trợ tiếng Việt ở mức phù hợp.
- ✚ Xử lý ảnh: Sử dụng OpenCV cho các bước tiền xử lý (lọc nhiễu, cân bằng sáng, chuẩn hóa kích thước, xoay nghiêng...) nhằm nâng cao độ chính xác OCR và trích xuất.

## 4.2. Kiến trúc tổng thể và mô hình triển khai

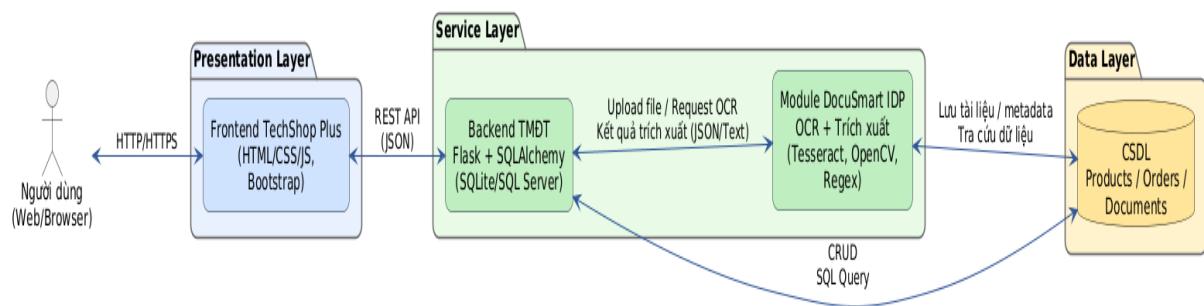


Hình 2. Kiến trúc tổng thể TechShop Plus tích hợp DocuSmart

Hệ thống được thiết kế theo kiến trúc tách lớp, dễ mở rộng theo mô hình nhiều tầng. Lớp giao diện người dùng (Frontend) thực hiện hiển thị sản phẩm, giờ hàng và các thao tác tương tác, đồng thời gửi yêu cầu qua tầng truy cập gồm Nginx/API Gateway và dịch vụ xác thực JWT/RBAC. Backend TMĐT (Node.js/Express) cung cấp REST API để quản lý và truy xuất dữ liệu sản phẩm/đơn hàng từ cơ sở dữ liệu TMĐT, đồng thời sử dụng Redis để tăng hiệu năng truy vấn. Mô-đun DocuSmart được triển khai như một dịch vụ IDP độc lập. Khi có nhu cầu xử lý tài liệu, Backend TMĐT tạo một tác vụ và đẩy vào Message Queue. Worker sẽ nhận tác vụ, lấy file từ Object Storage (S3/MinIO), thực hiện OCR và trích xuất thông tin qua Backend IDP, sau đó lưu kết quả vào DocuSmart DB và gửi thông báo hoàn thành về Backend TMĐT. Cuối cùng Backend TMĐT cập nhật dữ liệu đơn hàng/hóa đơn và trả kết quả cho giao diện thông qua API.

#### 4.2.1. Kiến trúc tổng thể của hệ thống

Hệ thống được thiết kế theo mô hình web application với backend Python/Flask và phân tầng rõ ràng (Presentation – Service – Data). TechShop Plus đảm nhiệm các chức năng TMĐT như quản lý sản phẩm, giờ hàng, đặt hàng và quản trị. Mô-đun DocuSmart được tích hợp như một thành phần IDP trong cùng tầng dịch vụ và được Backend TMĐT gọi nội bộ để xử lý upload tài liệu, OCR và trích xuất thông tin, sau đó trả kết quả có cấu trúc (JSON/Text) phục vụ nghiệp vụ. Cách tiếp cận tích hợp giúp giảm chi phí triển khai, phù hợp với doanh nghiệp SME do không cần hạ tầng riêng cho IDP. Đồng thời, kiến trúc vẫn cho phép tách mô-đun IDP thành microservice độc lập trong tương lai nếu cần mở rộng tải xử lý.



Hình 3. Kiến trúc tổng thể TechShop Plus tích hợp DocuSmart IDP

## 4.3. Thiết kế cơ sở dữ liệu

### 4.3.1. Thiết kế chi tiết cơ sở dữ liệu thương mại điện tử (TechShop Plus)

Cơ sở dữ liệu của website TechShop Plus được thiết kế theo hướng chuẩn hóa để bảo đảm tính toàn vẹn dữ liệu và hỗ trợ truy vấn thống kê. Các bảng cốt lõi phục vụ nghiệp vụ bán hàng gồm: Users, Products, Orders và OrderItems. Trong khuôn khổ đồ án, mô hình dữ liệu được tối giản để phù hợp triển khai trên SQLite/SQL Server, đồng thời vẫn có thể mở rộng cho môi trường thực tế.

Tên trường	Kiểu dữ liệu	Ràng buộc	Mô tả
UserID	INTEGER	PK, AUTOINCREMENT	Khóa chính người dùng
Username	VARCHAR(50)	UNIQUE, NOT NULL	Tên đăng nhập
PasswordHash	VARCHAR(255)	NOT NULL	Mật khẩu đã băm
FullName	VARCHAR(100)		Họ tên
Phone	VARCHAR(20)		Số điện thoại
Address	TEXT		Địa chỉ giao hàng mặc định
Role	VARCHAR(20)	DEFAULT 'customer'	customer/admin
Status	VARCHAR(20)	DEFAULT 'active'	active/locked
CreatedAt	DATETIME	DEFAULT CURRENT_TIMESTAMP	Ngày tạo tài khoản

Bảng 6. Cấu trúc bảng Users

Tên trường	Kiểu dữ liệu	Ràng buộc	Mô tả
MaSP	VARCHAR(20)	PK	Mã sản phẩm
Name	VARCHAR(255)	NOT NULL	Tên sản phẩm
Company	VARCHAR(100)		Hãng/Thương hiệu
Price	INTEGER	NOT NULL	Giá niêm yết (VND)
ImageUrl	TEXT		Đường dẫn hình ảnh
PromoName	VARCHAR(50)		Tên khuyến mãi
PromoValue	VARCHAR(50)		Giá sau giảm/giá hiện thị
Star	FLOAT	DEFAULT 0	Điểm đánh giá
RateCount	INTEGER	DEFAULT 0	Số lượt đánh giá
CreatedAt	DATETIME	DEFAULT CURRENT_TIMESTAMP	Ngày tạo

Bảng 7. Cấu trúc bảng Products

Tên trường	Kiểu dữ liệu	Ràng buộc	Mô tả
OrderID	INTEGER	PK, AUTOINCREMENT	Khóa chính đơn hàng
OrderCode	VARCHAR(30)	UNIQUE	Mã đơn hiển thị cho khách
UserID	INTEGER	FK -> Users(UserID)	Người đặt hàng
TotalAmount	INTEGER	NOT NULL	Tổng tiền (VND)
PaymentMethod	VARCHAR(30)	NOT NULL	COD/BANK_TRANSFER/MOMO...
PaymentStatus	VARCHAR(30)	NOT NULL	Chưa thanh toán/Đã thanh toán/...
ShippingName	VARCHAR(100)	NOT NULL	Tên người nhận
ShippingPhone	VARCHAR(20)	NOT NULL	SĐT người nhận
ShippingAddress	TEXT	NOT NULL	Địa chỉ giao hàng
OrderStatus	VARCHAR(30)	DEFAULT 'PENDING'	PENDING/SHIPPING/DONE/CANCEL
CreatedAt	DATETIME	DEFAULT CURRENT_TIMESTAMP	Ngày tạo đơn

Bảng 8. Cấu trúc bảng Orders

Tên trường	Kiểu dữ liệu	Ràng buộc	Mô tả
OrderItemID	INTEGER	PK, AUTOINCREMENT	Khóa chính dòng sản phẩm
OrderID	INTEGER	FK -> Orders(OrderID)	Thuộc đơn hàng
MaSP	VARCHAR(20)	FK -> Products(MaSP)	Sản phẩm
Quantity	INTEGER	NOT NULL	Số lượng
UnitPrice	INTEGER	NOT NULL	Đơn giá tại thời điểm mua
LineTotal	INTEGER		Thành tiền (có thể tính)

Bảng 9. Cấu trúc bảng OrderItems

#### 4.3.2. Thiết kế cơ sở dữ liệu mô-đun IDP (DocuSmart Extractor)

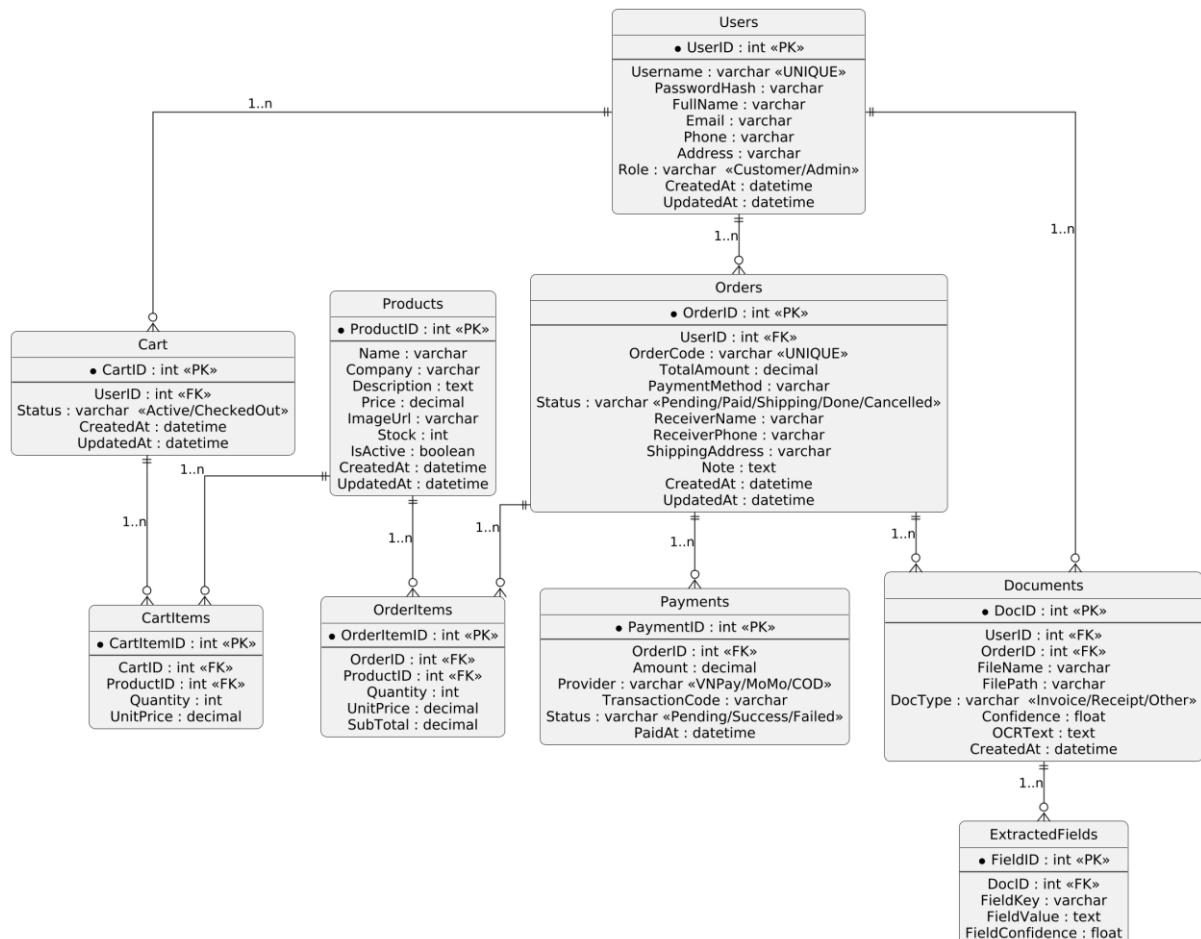
Mô-đun DocuSmart Extractor lưu lịch sử xử lý tài liệu và các trường thông tin trích xuất để phục vụ tra cứu, hiệu chỉnh và xuất báo cáo. Tùy theo phương án triển khai, mô-đun có thể dùng chung CSDL với website hoặc tách riêng. Trong đồ án, để đơn giản hóa, mô hình dữ liệu được thiết kế độc lập và có thể liên kết với người dùng thông qua UserID khi cần.

Tên trường	Kiểu dữ liệu	Ràng buộc	Mô tả
DocumentID	INTEGER	PK, AUTOINCREMENT	Khóa chính tài liệu
UserID	INTEGER	NULL, FK -> Users(UserID)	Người tải lên (nếu gắn tài khoản)
Filename	TEXT	NOT NULL	Tên file gốc
FilePath	TEXT		Đường dẫn lưu trữ
FileType	VARCHAR(20)	NOT NULL	pdf/jpg/png...
DocumentType	VARCHAR(30)	NOT NULL	invoice/contract/other
RawText	LONGTEXT		Văn bản OCR đầy đủ
Confidence	FLOAT	DEFAULT 0	Độ tin cậy ước tính
ProcessedAt	DATETIME	DEFAULT CURRENT_TIMESTAMP	Thời điểm xử lý

Bảng 10. Cấu trúc bảng Documents

Tên trường	Kiểu dữ liệu	Ràng buộc	Mô tả
FieldID	INTEGER	PK, AUTOINCREMENT	Khóa chính trường trích xuất
DocumentID	INTEGER	FK -> Documents(DocumentID)	Thuộc tài liệu
FieldKey	VARCHAR(50)	NOT NULL	Khóa trường (invoice_number, tax_code, ...)
FieldValue	TEXT		Giá trị trích xuất
Source	VARCHAR(20)	DEFAULT 'regex'	regex/ai/manual
UpdatedAt	DATETIME	DEFAULT CURRENT_TIMESTAMP	Thời điểm cập nhật

Bảng 11. Cấu trúc bảng ExtractedFields



Hình 4. Sơ đồ ERD tổng quan (TMĐT + DocuSmart)

Dựa trên yêu cầu quản lý dữ liệu, nhóm thiết kế mô hình dữ liệu theo hai cụm chính: cụm thương mại điện tử phục vụ quản lý người dùng, sản phẩm, giỏ hàng, đơn hàng và thanh toán; và cụm DocuSmart lưu trữ lịch sử tài liệu đã xử lý cùng các trường thông tin trích xuất. Quan hệ giữa các bảng và ràng buộc khóa chính/khóa ngoại được thể hiện trong sơ đồ ERD. Trong môi trường demo, hệ thống sử dụng SQLite để lưu trữ nhằm đơn giản hóa triển khai và dễ đóng gói. Thiết kế CSDL tuân theo hướng chuẩn hóa, bao gồm các nhóm bảng chính: người dùng và phân quyền cơ bản, sản phẩm, giỏ hàng, đơn hàng và thanh toán, và lịch sử xử lý tài liệu của DocuSmart (Documents, ExtractedFields).

Bảng	Trường chính	Mô tả
products	id, name, company, price, image_url, stock, created_at	Lưu thông tin sản phẩm điện thoại/phụ kiện.
users	id, username, password_hash, full_name, phone, address, role	Lưu tài khoản người dùng/admin.
orders	id, order_code, user_id, total_amount, payment_method, status, created_at	Lưu thông tin đơn hàng.
order_items	id, order_id, product_id, quantity, unit_price	Chi tiết sản phẩm trong đơn hàng.
documents	id, filename, doc_type, confidence, raw_text, created_at	Lưu lịch sử xử lý tài liệu và văn bản OCR.
extracted_fields	id, document_id, field_key, field_value	Lưu các trường trích xuất dạng key-value để dễ truy vấn.

Bảng 12. Mô hình dữ liệu – các bảng chính trong CSDL

#### **4.4. Thiết kế module TMĐT (TechShop Plus)**

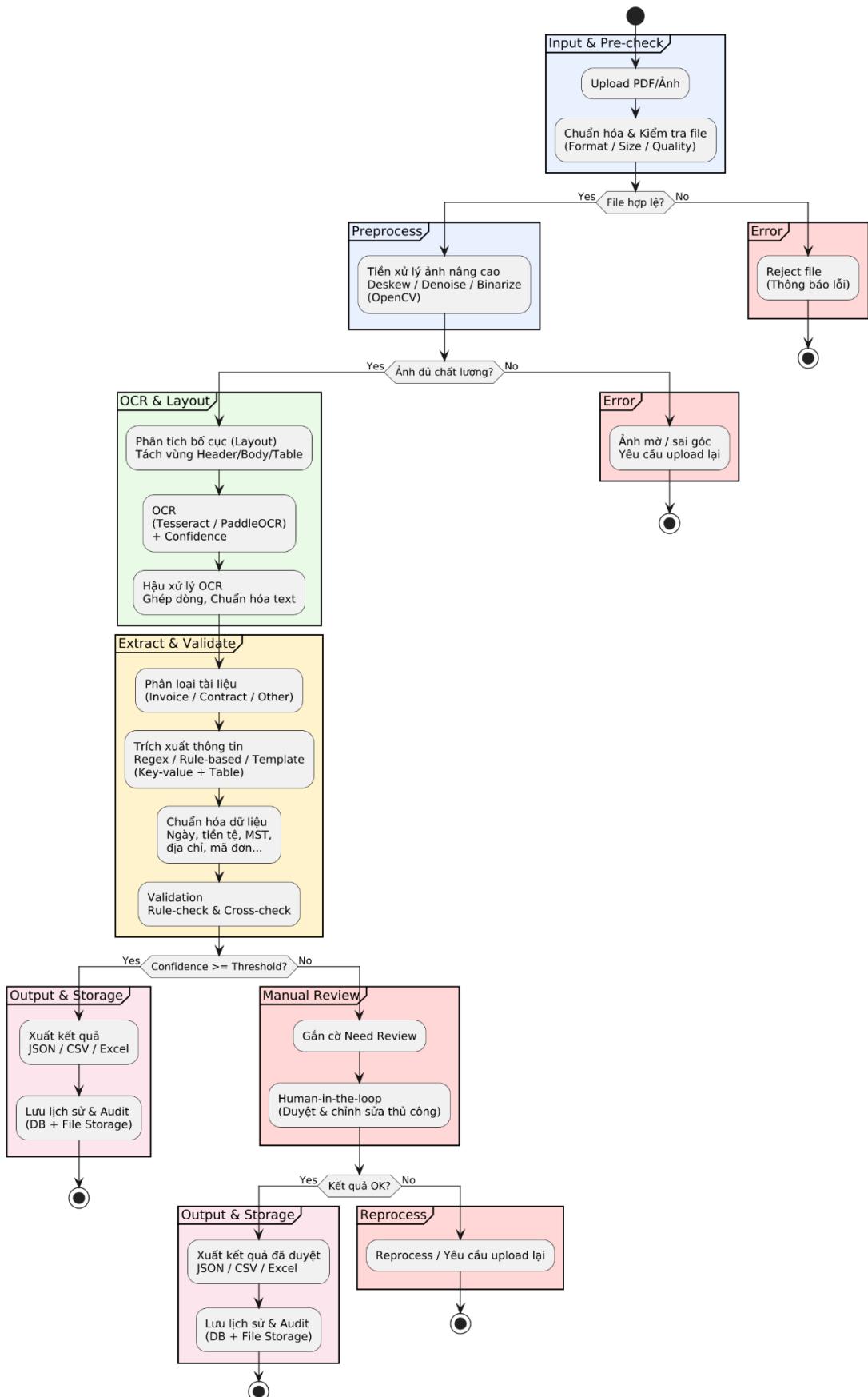
Module TMĐT gồm các thành phần chính: trang chủ – danh mục – chi tiết sản phẩm, giỏ hàng, checkout và khu vực quản trị. Ở phía người dùng, luồng mua hàng được thiết kế tối giản: chọn sản phẩm → thêm giỏ hàng → xác nhận đơn → chọn phương thức thanh toán. Hệ thống hỗ trợ COD và chuyển khoản ngân hàng. Đối với COD, hệ thống bắt buộc người mua nhập đầy đủ thông tin nhận hàng (họ tên, số điện thoại, địa chỉ) trước khi xác nhận đơn, nhằm đảm bảo dữ liệu giao hàng chính xác và giảm rủi ro hoàn đơn.

Về mặt kỹ thuật, giỏ hàng được lưu trong phiên người dùng và đồng bộ với CSDL khi tạo đơn. Các thao tác tăng/giảm số lượng và tính tổng tiền được xử lý phía client để tăng trải nghiệm, đồng thời kiểm tra lại phía server khi đặt hàng để đảm bảo nhất quán dữ liệu.

#### **4.5. Thiết kế module IDP (DocuSmart Extractor)**

DocuSmart được thiết kế theo pipeline xử lý tài liệu (Hình 4.2). Tài liệu đầu vào bao gồm ảnh và PDF; hệ thống thực hiện bước kiểm tra định dạng, kích thước và chất lượng file nhằm đảm bảo dữ liệu hợp lệ trước khi xử lý. Tiếp theo, ảnh được tiền xử lý nâng cao (deskew, tăng tương phản, giảm nhiễu, nhị phân hóa) bằng OpenCV để cải thiện độ chính xác nhận dạng. Sau đó, hệ thống tiến hành phân tích bố cục (layout analysis) nhằm tách các vùng thông tin quan trọng như header/body/table, tạo điều kiện cho việc nhận dạng và trích xuất chính xác hơn. Ở bước OCR, DocuSmart sử dụng Tesseract/PaddleOCR để chuyển đổi ảnh thành văn bản số tiếng Việt/Anh và đồng thời sinh ra điểm tin cậy (confidence). Dựa trên kết quả OCR, mô-đun phân loại tài liệu xác định loại tài liệu (invoice/contract/other), từ đó kích hoạt bộ trích xuất thông tin theo luật/regex (rule-based) để thu thập các trường dữ liệu quan trọng và thực hiện chuẩn hóa dữ liệu (ngày, tiền tệ, mã số, địa chỉ...). Kết quả cuối cùng được đưa qua bước validation nhằm kiểm tra thiếu trường, sai định dạng và đối chiếu các quy tắc nghiệp vụ.

Trong trường hợp confidence đạt ngưỡng, hệ thống xuất kết quả trực tiếp; ngược lại, tài liệu sẽ được gắn cờ Need Review và chuyển sang cơ chế human-in-the-loop để người dùng duyệt và chỉnh sửa thủ công trước khi xuất. Toàn bộ kết quả được trả về Backend để hiển thị cho giao diện người dùng, đồng thời lưu lịch sử xử lý phục vụ tra cứu và kiểm toán nội bộ.



Hình 5. Luồng xử lý tài liệu trong DocuSmart (IDP Pipeline)

Hình 5 mô tả pipeline xử lý tài liệu trong DocuSmart, gồm các bước kiểm tra file, tiền xử lý ảnh, phân tích bộ cục, OCR, phân loại và trích xuất trường dữ liệu. Kết quả được validation theo quy tắc nghiệp vụ; nếu confidence thấp, tài liệu chuyển sang cơ chế manual review (human-in-the-loop) trước khi xuất và lưu lịch sử.

Hệ thống hỗ trợ xuất dữ liệu theo ba định dạng phổ biến: JSON (phục vụ tích hợp API và các hệ thống khác), CSV (phù hợp import nhanh vào công cụ phân tích dữ liệu), và Excel (phù hợp nghiệp vụ kế toán, đối soát và lưu trữ). Ngoài ra, toàn bộ lịch sử xử lý được lưu lại (bao gồm file, kết quả trích xuất, confidence và thời gian xử lý) giúp người dùng có thể tra cứu, đối chiếu và phục vụ kiểm toán nội bộ.

#### 4.6. Thiết kế API tích hợp

API được tổ chức theo nhóm tài nguyên: /api/products, /api/orders, /api/docusmart. Trong phạm vi demo, một số endpoint tiêu biểu gồm:

Endpoint	Method	Mô tả
/api/products	GET	Lấy danh sách sản phẩm.
/api/products/<id>	GET	Lấy chi tiết sản phẩm.
/api/orders/checkout	POST	Tạo đơn hàng từ giỏ hàng; ghi nhận phương thức thanh toán.
/api/orders/<order_code>	GET	Tra cứu trạng thái đơn hàng.
/api/docusmart/upload	POST	Upload và xử lý tài liệu (OCR + trích xuất).
/api/docusmart/history	GET	Danh sách lịch sử xử lý tài liệu.
/api/docusmart/export	GET	Xuất kết quả theo định dạng JSON/CSV/Excel.

Bảng 13. Danh sách API chính

#### 4.7. Thiết kế bảo mật

Các biện pháp bảo mật được áp dụng ở mức phù hợp với đồ án và có thể mở rộng khi triển khai thực tế: xác thực – phân quyền người dùng và admin; hash mật khẩu bằng thuật toán an toàn; kiểm soát upload file (kích thước, định dạng, đổi tên file an toàn); hạn chế lộ khóa bí mật bằng biến môi trường; (v) ghi log xử lý để truy vết lỗi. Trong triển khai thực tế, cần bổ sung HTTPS, CSRF token, cơ chế chống brute force và chính sách backup định kỳ.

## **4.8. Lợi ích mang lại**

Việc triển khai nền tảng TechShop Plus tích hợp DocuSmart mang lại lợi ích đồng thời cho ba nhóm đối tượng: doanh nghiệp – nhân viên – khách hàng. Lợi ích được đánh giá trên hai khía cạnh chính: tăng trưởng doanh thu và chất lượng dịch vụ; tối ưu vận hành, giảm chi phí và sai sót trong xử lý chứng từ.

### **4.8.1. Đối với doanh nghiệp**

- ⊕ Mở rộng kênh bán hàng: website hoạt động 24/7, giảm phụ thuộc cửa hàng vật lý, tăng khả năng tiếp cận khách hàng ngoài khu vực.
- ⊕ Chuẩn hóa quy trình nghiệp vụ: đơn hàng, thanh toán, thông tin giao hàng, trạng thái xử lý được quản lý tập trung, dễ kiểm soát.
- ⊕ Tăng hiệu quả đối soát – kế toán: DocuSmart giảm thời gian nhập liệu chứng từ; dữ liệu lưu có cấu trúc nên dễ tổng hợp và tra cứu.
- ⊕ Giảm rủi ro sai sót: hạn chế nhập tay, giảm nhầm lẫn số tiền, ngày tháng, mã số thuế; cải thiện tính minh bạch khi kiểm tra lại.
- ⊕ Nâng cao hình ảnh chuyên nghiệp: hệ thống đơn hàng/hoa đơn rõ ràng, lịch sử giao dịch đầy đủ, tăng độ tin cậy với khách hàng và đối tác.

### **4.8.2. Đối với nhân viên**

- ⊕ Giảm tải công việc lặp lại (nhập liệu chứng từ, tổng hợp thông tin), tăng thời gian cho chăm sóc khách hàng và bán hàng.
- ⊕ Tra cứu nhanh dữ liệu: tìm theo mã đơn, thời gian, nhà cung cấp, mã số thuế... thay vì dò giấy tờ thủ công.
- ⊕ Hỗ trợ ra quyết định: số liệu thống kê theo thời gian thực giúp quản lý tồn kho/khuyến mãi/doanh thu hiệu quả hơn.

### **4.7.3. Đối với khách hàng**

- ⊕ Trải nghiệm mua sắm thuận tiện: xem sản phẩm, đặt hàng, theo dõi đơn hàng nhanh chóng trên nhiều thiết bị.
- ⊕ Minh bạch khi giao nhận: form COD yêu cầu đầy đủ giúp giao hàng chính xác, giảm tỷ lệ giao thất bại.
- ⊕ Rút ngắn thời gian phản hồi: nhân viên có nhiều thời gian hỗ trợ vì được giảm khối lượng nhập liệu và tổng hợp.

## 4.9. Kế hoạch thực hiện chuyển đổi số

### 4.9.1. Lộ trình thực hiện

Kế hoạch triển khai được xây dựng theo lộ trình 4 giai đoạn nhằm đảm bảo tính khả thi đối với doanh nghiệp quy mô vừa và nhỏ: chuẩn bị – triển khai thử nghiệm – vận hành chính thức – mở rộng và cải tiến. Mỗi giai đoạn có mục tiêu, sản phẩm bàn giao (deliverables) và tiêu chí đánh giá.

Giai đoạn	Thời gian	Nội dung trọng tâm
Giai đoạn 1: Chuẩn bị	10/2025 – 11/2025	Khảo sát 10 cửa hàng tại Cần Thơ; phân tích yêu cầu; lập kế hoạch chi tiết; lựa chọn giải pháp công nghệ; thiết kế CSDL và wireframe giao diện; chuẩn hóa quy trình chứng từ.
Giai đoạn 2: Triển khai thử nghiệm	12/2025 – 02/2026	Xây dựng hệ thống lõi (website + API + DocuSmart); số hóa dữ liệu ban đầu (sản phẩm, khách hàng, nhà cung cấp); đào tạo nhân lực; kiểm thử chức năng; hiệu chỉnh OCR/regex theo mẫu chứng từ.
Giai đoạn 3: Vận hành chính thức	03/2026 – 04/2026	Triển khai lên server; tích hợp luồng đơn hàng – thanh toán – chứng từ; tối ưu hiệu năng; thiết lập sao lưu; chạy song song và đối soát; đánh giá kết quả theo KPI.
Giai đoạn 4: Mở rộng và cải tiến	05/2026 – 06/2026	Mở rộng ứng dụng (tích hợp thanh toán số/QR, thông báo đơn hàng); nâng cấp IDP bằng AI (phân loại nâng cao, trích xuất theo ngữ cảnh); mở rộng dữ liệu và báo cáo; chuẩn bị kế hoạch nhân rộng cho nhiều chi nhánh.

Bảng 14 Giai đoạn triển khai

#### **4.9.2. Nguồn lực triển khai**

Nguồn lực được hoạch định theo ba nhóm: nhân lực, tài chính và công nghệ. Các ước tính dưới đây phù hợp với quy mô doanh nghiệp vừa và nhỏ (SME), có thể điều chỉnh tùy điều kiện thực tế.

##### **a) Nhân lực**

- ✚ Ban chỉ đạo chuyển đổi số (01 người): đại diện chủ doanh nghiệp/giám đốc, chịu trách nhiệm phê duyệt mục tiêu, ngân sách và chính sách vận hành.
- ✚ Tổ công nghệ thông tin (03–04 người): trưởng nhóm/BA, lập trình viên full-stack phụ trách website & API, kỹ sư AI/OCR phụ trách DocuSmart, kỹ sư triển khai – vận hành (có thể kiêm nhiệm).
- ✚ Chuyên gia tư vấn nghiệp vụ (01 người): kế toán hoặc thủ kho giàu kinh nghiệm để chuẩn hóa quy trình chứng từ, xác định trường dữ liệu cần trích xuất và tiêu chí đối soát.
- ✚ Đội ngũ truyền thông & đào tạo (01–02 người): xây dựng tài liệu hướng dẫn sử dụng, đào tạo nhân viên cửa hàng, hỗ trợ truyền thông kênh bán hàng online.

##### **b) Tài chính (ước tính cho SME triển khai)**

- ✚ Chi phí phát triển ban đầu (phân tích – phát triển – kiểm thử): 40–60 triệu VNĐ (tùy phạm vi và mức độ thuê ngoài).
- ✚ Chi phí hạ tầng năm đầu: VPS/Cloud 1,2–2,4 triệu VNĐ/năm; tên miền 0,5–1 triệu VNĐ/năm; SSL miễn phí (Let's Encrypt).
- ✚ Chi phí đào tạo & vận hành: 5–10 triệu VNĐ (tài liệu, hỗ trợ triển khai, vận hành thử).
- ✚ Nguồn vốn: ngân sách doanh nghiệp kết hợp nguồn xã hội hóa/hợp tác công – tư (PPP) và các chương trình/quỹ hỗ trợ chuyển đổi số địa phương.

##### **c) Công nghệ**

- ✚ Hạ tầng: máy chủ VPS Linux (Ubuntu), reverse proxy Nginx, cơ chế sao lưu định kỳ; có thể container hóa (Docker) để triển khai thuận tiện.
- ✚ Nền tảng phần mềm: Website (HTML/CSS/JS + Bootstrap), dịch vụ API (Node.js/Express hoặc Flask), dịch vụ IDP (Flask/Python), CSDL (SQL Server/MySQL/SQLite tùy môi trường).
- ✚ Công nghệ xử lý tài liệu: Tesseract OCR/PaddleOCR, OpenCV tiền xử lý ảnh, pdf2image chuyển đổi PDF.
- ✚ Giải pháp bảo mật: HTTPS, phân quyền, mã hóa mật khẩu, kiểm soát upload, logging/audit, backup và khôi phục dữ liệu.

#### 4.9.3. Chỉ số đánh giá hiệu quả (KPI)

Để đánh giá hiệu quả chuyển đổi số theo hướng định lượng, nhóm đề xuất bộ chỉ số KPI tập trung vào hiệu quả bán hàng trực tuyến và hiệu quả tự động hóa xử lý chứng từ. Các chỉ số được xây dựng dựa trên kết quả khảo sát mô phỏng 10 cửa hàng tại Cần Thơ và kinh nghiệm vận hành của mô hình SME.

Nhóm KPI	Chỉ số	Hiện trạng (ước tính)	Mục tiêu sau 3 tháng	Cách đo lường
Xử lý chứng từ	Thời gian nhập liệu 1 chứng từ	8–12 phút	2–4 phút	So sánh thời gian trung bình trước/sau khi dùng DocuSmart
Xử lý chứng từ	Tỷ lệ sai sót khi nhập liệu	3–5%	$\leq 1\%$	Đối chiếu ngẫu nhiên 30 chứng từ/tháng
Xử lý chứng từ	Tỷ lệ trích xuất đúng trường (MST, ngày, tổng tiền)	40 - 45%	$\geq 85\%$	Đánh giá trên tập 50 chứng từ mẫu
Bán hàng online	Tỷ lệ hoàn tất đặt hàng (checkout completion)	45 - 50%	$\geq 65\%$	Số đơn hoàn tất / số phiên bắt đầu checkout
Bán hàng online	Tỷ lệ giao hàng thất bại (COD)	6–10%	$\leq 4\%$	Số đơn hoàn/không liên lạc / tổng COD
Vận hành	Thời gian tra cứu chứng từ theo mã đơn	5–15 phút	$\leq 1$ phút	Thực nghiệm tra cứu 10 mã đơn bất kỳ

Bảng 15 Bảng đánh giá hiệu quả

Việc theo dõi KPI cần được thực hiện liên tục trong giai đoạn vận hành chính thức để kịp thời điều chỉnh quy trình, cải thiện chất lượng dữ liệu và tối ưu mô-đun trích xuất. KPI cũng là cơ sở thuyết phục doanh nghiệp khi ra quyết định mở rộng hệ thống cho nhiều chi nhánh.

## CHƯƠNG 5. CÀI ĐẶT VÀ TRIỂN KHAI

### 5.1. Môi trường phát triển và công cụ

Nhóm lựa chọn ngôn ngữ Python và framework Flask cho backend vì: nhẹ, dễ triển khai; phù hợp xây dựng API và web template; hệ sinh thái mạnh cho xử lý ảnh, OCR và dữ liệu. Frontend sử dụng HTML5/CSS3/JavaScript và Bootstrap 5 để đảm bảo giao diện responsive. CSDL sử dụng SQLite trong môi trường demo và có thể chuyển sang các hệ quản trị CSDL khác khi triển khai.

Hạng mục	Công nghệ/Công cụ	Mục đích sử dụng
Ngôn ngữ	Python 3.x	Backend web và module IDP.
Framework backend	Flask, Flask-CORS	Xây dựng web + API; cho phép gọi API từ frontend.
ORM	SQLAlchemy	Truy vấn và thao tác CSDL an toàn.
CSDL	SQLite (demo)	Lưu trữ dữ liệu sản phẩm, đơn hàng, lịch sử tài liệu.
Frontend	HTML/CSS/JavaScript, Bootstrap 5	Xây dựng giao diện người dùng và trang quản trị.
OCR	pytesseract (Tesseract OCR)	Nhận dạng văn bản tiếng Việt/Anh từ ảnh/PDF.
Xử lý ảnh/PDF	OpenCV, Pillow, pdf2image (Poppler)	Tiền xử lý ảnh và chuyển PDF sang ảnh.
Xuất dữ liệu	JSON/CSV, openpyxl	Xuất kết quả trích xuất và báo cáo.
Quản lý mã nguồn	Git/GitHub	Quản lý phiên bản, cộng tác nhóm.
Đóng gói	Docker (tùy chọn), ZIP package	Triển khai thử nghiệm và demo.

Bảng 16. Công nghệ và công cụ sử dụng trong dự án

## **5.2. Cài đặt chức năng website TMĐT**

Website TechShop Plus được xây dựng theo mô hình MVC đơn giản của Flask: template Jinja2 cho giao diện, route/controller cho xử lý logic, và model (SQLAlchemy) cho dữ liệu. Các chức năng chính bao gồm: trang danh sách sản phẩm, trang chi tiết, giỏ hàng và checkout. Phần giỏ hàng được thiết kế để dễ mở rộng: dữ liệu sản phẩm có thể lấy từ CSDL thông qua API, thay vì phụ thuộc file tĩnh. Điều này phù hợp mục tiêu chuyển đổi số vì dữ liệu được quản trị tập trung và cập nhật theo thời gian thực.

Ở bước thanh toán, hệ thống hỗ trợ hai phương thức: COD – thanh toán khi nhận hàng; chuyển khoản. Khi người dùng chọn COD, form thông tin nhận hàng sẽ được hiển thị và yêu cầu nhập đầy đủ các trường bắt buộc. Cơ chế kiểm tra dữ liệu được thực hiện cả ở client (trải nghiệm nhanh) và server (đảm bảo an toàn dữ liệu).

## **5.3. Cài đặt DocuSmart IDP**

DocuSmart gồm các thành phần chính: OCREngine, DocumentClassifier và InformationExtractor. OCREngine xử lý ảnh/PDF, áp dụng tiền xử lý và gọi Tesseract để trích xuất văn bản. DocumentClassifier phân loại dựa trên từ khóa đặc trưng để xác định loại tài liệu. InformationExtractor áp dụng regex để trích xuất các trường. Kết quả sau đó được chuẩn hóa và trả về dạng JSON cho frontend.

Để nâng cao trải nghiệm, giao diện DocuSmart cung cấp: kéo thả upload; hiển thị preview văn bản OCR; hiển thị các trường trích xuất dạng bảng; nút export JSON/CSV/Excel; (v) lịch sử xử lý lưu trên localStorage và/hoặc CSDL.

## **5.4. Tích hợp vào quy trình nghiệp vụ**

Việc tích hợp DocuSmart vào website bán điện thoại được định hướng theo hai kịch bản nghiệp vụ:

- Kịch bản 1 – Xử lý hóa đơn đầu vào: Nhân viên kế toán upload hóa đơn PDF/ảnh từ nhà cung cấp. DocuSmart tự động trích xuất mã số thuế, ngày, tổng tiền và thông tin nhà cung cấp; dữ liệu được lưu để đối soát và nhập phần mềm kế toán.
- Kịch bản 2 – Đối soát thanh toán chuyển khoản: Khi khách hàng chuyển khoản, nhân viên upload chứng từ/biên lai để trích xuất số tiền, nội dung chuyển khoản. Hệ thống hỗ trợ đối chiếu với đơn hàng theo mã đơn và trạng thái thanh toán.

Trong tương lai, hệ thống có thể mở rộng theo hướng tích hợp sâu hơn: tự động gắn tài liệu vào đơn hàng tương ứng, tạo cảnh báo khi số tiền không khớp, và đồng bộ dữ liệu sang phần mềm kế toán/ERP thông qua API.

## **5.5. Đóng gói và chạy demo**

Sản phẩm được đóng gói dưới dạng dự án Python/Flask kèm dữ liệu mẫu để có thể chạy trên localhost. Quy trình chạy cơ bản: tạo virtualenv → cài requirements → cài Tesseract/Poppler → chạy app.py. Đối với môi trường demo tại lớp, nhóm đóng gói thành file nén (ZIP) gồm toàn bộ mã nguồn, tài liệu hướng dẫn và slide thuyết trình.

## CHƯƠNG 6. KIỂM THỬ – THỰC NGHIỆM – ĐÁNH GIÁ

### 6.1. Kế hoạch kiểm thử

Nhóm tiến hành kiểm thử theo hai lớp: kiểm thử chức năng (functional testing) cho website TMĐT và DocuSmart; thử nghiệm xử lý tài liệu để đánh giá chất lượng OCR và trích xuất. Kiểm thử chức năng tập trung vào các use case quan trọng như đăng nhập, thêm giỏ hàng, checkout, upload tài liệu và export dữ liệu.

TC	Mô tả	Kỳ vọng
TC-01	Thêm sản phẩm vào giỏ hàng	Sản phẩm hiển thị đúng; số lượng cập nhật; tổng tiền thay đổi tương ứng.
TC-02	Thanh toán COD thiếu thông tin	Hệ thống hiển thị form; cảnh báo bắt buộc; không cho tạo đơn khi thiếu họ tên/SĐT/địa chỉ.
TC-03	Thanh toán chuyển khoản	Tạo đơn hàng thành công; trạng thái thanh toán 'Chờ chuyển khoản'.
TC-04	Upload ảnh hóa đơn rõ nét	OCR trả về văn bản; phân loại đúng 'invoice'; trích xuất được số hóa đơn/ngày/tổng tiền.
TC-05	Upload PDF nhiều trang	Xử lý từng trang; tổng hợp văn bản; trích xuất trường nếu có.
TC-06	Export Excel	Tạo file .xlsx đúng định dạng; dữ liệu trường trích xuất đầy đủ.

Bảng 17. Bộ kiểm thử chức năng (tóm tắt)

### 6.2. Bộ dữ liệu thử nghiệm

Trong phạm vi đồ án, nhóm sử dụng bộ dữ liệu thử nghiệm nhỏ (pilot dataset) gồm các tài liệu hóa đơn và hợp đồng dạng PDF/ảnh. Nguồn dữ liệu bao gồm: mẫu tài liệu công khai và tài liệu do nhóm tự tạo; ảnh chụp mô phỏng với điều kiện ánh sáng khác nhau. Bộ dữ liệu được chia thành ba nhóm: tài liệu rõ nét, tài liệu chất lượng trung bình và tài liệu nhiễu/nhỏe để đánh giá khả năng chịu lỗi của hệ thống.

Do hạn chế về thời gian và quyền truy cập dữ liệu thực tế của doanh nghiệp, các số liệu đánh giá trong báo cáo được ghi nhận theo phương pháp đánh giá thủ công (manual scoring) và có thể cập nhật khi có bộ dữ liệu lớn hơn trong giai đoạn phát triển tiếp theo.

### 6.3. Chỉ số đánh giá

Các chỉ số đánh giá được sử dụng gồm:

- ✚ Độ chính xác OCR (tham khảo): tỷ lệ ký tự/từ được nhận dạng đúng so với văn bản chuẩn.
- ✚ Độ chính xác phân loại: tỷ lệ tài liệu được phân loại đúng loại (invoice/contract/other).
- ✚ Độ chính xác trích xuất trường: tỷ lệ trường (ví dụ: số hóa đơn, ngày, tổng tiền, mã số thuế) được trích xuất đúng.
- ✚ Thời gian xử lý: thời gian trung bình để xử lý một tài liệu 1 trang trên máy cấu hình phổ thông.

## 6.4. Kết quả và thảo luận

### 6.4.1. Công thức tổng quát

#### 1) Công thức tổng quát tính OCR trung bình (ước lượng) cho toàn bộ tập mẫu

Giả sử tập dữ liệu được chia thành  $k$  nhóm (ví dụ: Rõ nét, Trung bình, Kém). Mỗi nhóm có:

- ✚  $n_i$ : số mẫu của nhóm  $i$
- ✚  $OCR_i$ : độ chính xác OCR ước lượng của nhóm  $i$  (đơn vị %)

Khi đó, độ chính xác OCR trung bình của toàn bộ tập dữ liệu được tính theo trung bình có trọng số:

$$OCR_{avg} = \frac{\sum_{i=1}^k n_i \times OCR_i}{\sum_{i=1}^k n_i}$$

#### ✓ Chú thích:

- ✚  $OCR_{avg}$ : OCR trung bình trên toàn bộ dữ liệu (%)
  - ✚  $k$ : số nhóm dữ liệu
  - ✚  $n_i$ : số mẫu thuộc nhóm  $i$
  - ✚  $OCR_i$ : OCR ước lượng của nhóm  $i$
  - ✚  $\sum_{i=1}^k n_i$ : tổng số mẫu toàn bộ tập dữ liệu
- ✓ **Ý nghĩa:** nhóm có nhiều mẫu sẽ ảnh hưởng lớn hơn đến kết quả trung bình tổng.

### 6.4.2. Công thức tổng quát tính Trích xuất đúng trường chính trung bình

Tương tự, với:

- ✚  $E_i$ : tỷ lệ trích xuất đúng trường chính của nhóm  $i$  (đơn vị %)

Ta có:

$$E_{avg} = \frac{\sum_{i=1}^k n_i \times E_i}{\sum_{i=1}^k n_i}$$

✓ **Chú thích:**

- ✚  $E_{\text{avg}}$ : tỷ lệ trích xuất đúng trung bình toàn bộ (%)
- ✚  $E_i$ : tỷ lệ trích xuất đúng trung bình của nhóm  $i$

#### 6.4.3. Công thức tổng quát tính tỷ lệ phân loại đúng (overall accuracy)

Với:

- ✚  $c_i$ : số mẫu phân loại đúng trong nhóm  $i$
- ✚  $n_i$ : số mẫu của nhóm  $i$

Tỷ lệ phân loại đúng toàn bộ được tính:

$$Acc_{\text{cls}} = \frac{\sum_{i=1}^k c_i}{\sum_{i=1}^k n_i}$$

Nếu muốn biểu diễn theo phần trăm:

$$Acc_{\text{cls}}(\%) = \frac{\sum_{i=1}^k c_i}{\sum_{i=1}^k n_i} \times 100$$

✓ **Chú thích:**

- ✚  $Acc_{\text{cls}}$ : độ chính xác phân loại toàn bộ (tỷ lệ hoặc %)
- ✚  $c_i$ : số mẫu phân loại đúng trong nhóm  $i$

Nhóm dữ liệu	Số mẫu	OCR (ước lượng)	Phân loại đúng	Trích xuất đúng trung bình
Rõ nét (scan/ảnh tốt)	12	~92%	11/12	~85%
Trung bình (ánh sáng không đều)	10	~87%	8/10	~75%
Kém (nhòe/nghiêng/nhiều)	8	~78%	6/8	~60%
Tổng	30	~86.6% (~87%)	25/30 (~83.3%)	~75%

Bảng 18 trình bày kết quả thử nghiệm tham khảo của nhóm trên bộ dữ liệu mẫu

Bảng 18 trình bày kết quả thử nghiệm tham khảo của nhóm trên bộ dữ liệu mẫu gồm 30 tài liệu. Kết quả cho thấy với tài liệu rõ nét, OCR đạt khoảng 92% và trích xuất đúng trung bình đạt khoảng 85%, đồng thời tỷ lệ phân loại đúng đạt 11/12. Khi chất lượng tài liệu giảm (ánh sáng không đều hoặc ảnh chụp nghiêng/nhiều), hiệu năng suy giảm rõ rệt: OCR giảm còn khoảng 78%, trích xuất giảm còn khoảng 60%, và phân loại đúng

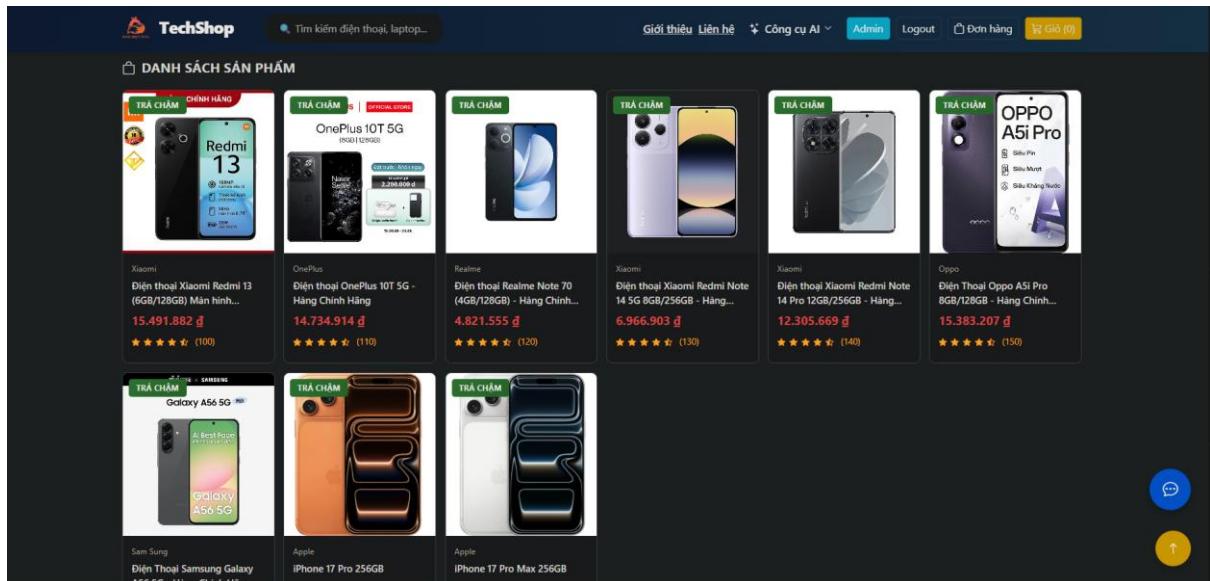
giảm còn 6/8. Điều này cho thấy chất lượng đầu vào ảnh hưởng mạnh đến toàn bộ pipeline và đặc biệt làm giảm hiệu quả trích xuất trường do sai lệch OCR và bối rối khía cạnh nhận dạng.

Thảo luận: Kết quả thử nghiệm cho thấy tiền xử lý ảnh giữ vai trò quan trọng trong việc cải thiện OCR, trong đó các kỹ thuật nhị phân hóa và tăng tương phản giúp nâng chất lượng nhận dạng trên ảnh mờ. Phương pháp trích xuất rule-based/regex hoạt động tốt với các hóa đơn có từ khóa rõ ràng như “Mã số thuế”, “Tổng cộng”, tuy nhiên với tài liệu có bối rối khía cạnh nhận dạng hoặc OCR sai dấu, regex có thể bỏ sót thông tin. Vì vậy, hướng phát triển tiếp theo bao gồm: đánh giá và tối ưu OCR engine (ưu tiên PaddleOCR cho tiếng Việt), bổ sung mô hình học máy/NER để trích xuất ổn định hơn, và khai thác thông tin bối rối khía cạnh (layout) nhằm xác định vùng dữ liệu thay vì chỉ dựa trên văn bản OCR.

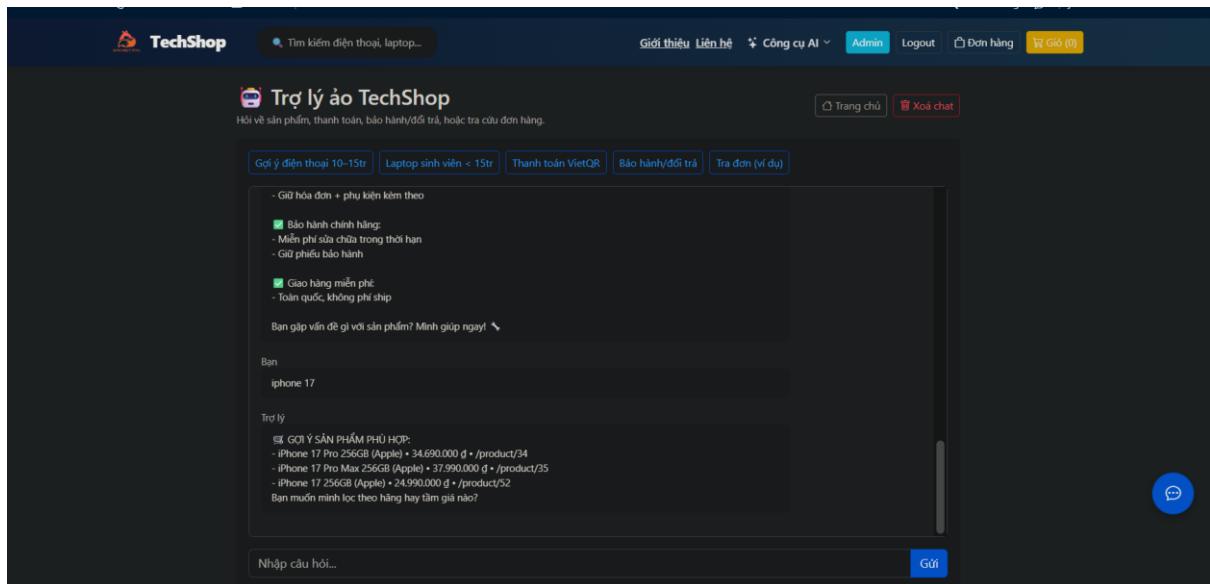
Kết quả cho thấy tiền xử lý ảnh có vai trò quan trọng. Các kỹ thuật nhị phân hóa và tăng tương phản giúp cải thiện đáng kể trên ảnh mờ. Rule-based extraction hoạt động tốt với mẫu hóa đơn có từ khóa rõ ràng (“Mã số thuế”, “Tổng cộng”). Tuy nhiên, với tài liệu có bối rối khía cạnh nhận dạng hoặc OCR sai dấu, regex có thể bỏ sót. Do đó, hướng phát triển tiếp theo là: bổ sung PaddleOCR cho tiếng Việt; áp dụng NER/mô hình học máy để trích xuất ổn định hơn; khai thác bối rối khía cạnh (layout) để xác định vùng thông tin thay vì chỉ dựa trên text.

## 6.5. Demo sản phẩm web và mô-đun DocuSmart

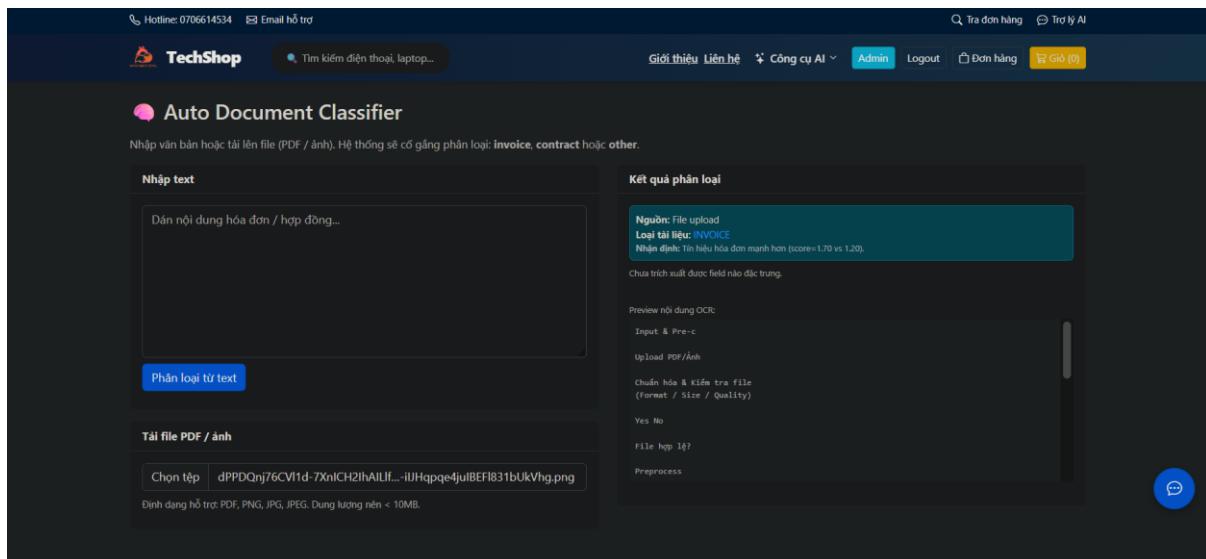
Phần demo được xây dựng nhằm minh họa luồng nghiệp vụ từ bán hàng trực tuyến đến số hóa chứng từ. Hệ thống được triển khai trên môi trường localhost gồm 02 dịch vụ: website TechShop Plus (frontend HTML/CSS/JS và API backend) và dịch vụ DocuSmart Extractor chạy Flask/Python phục vụ upload – OCR – trích xuất dữ liệu. Dữ liệu sản phẩm/đơn hàng được lưu trong cơ sở dữ liệu, kết quả IDP được lưu lịch sử để phục vụ tra cứu và xuất file.



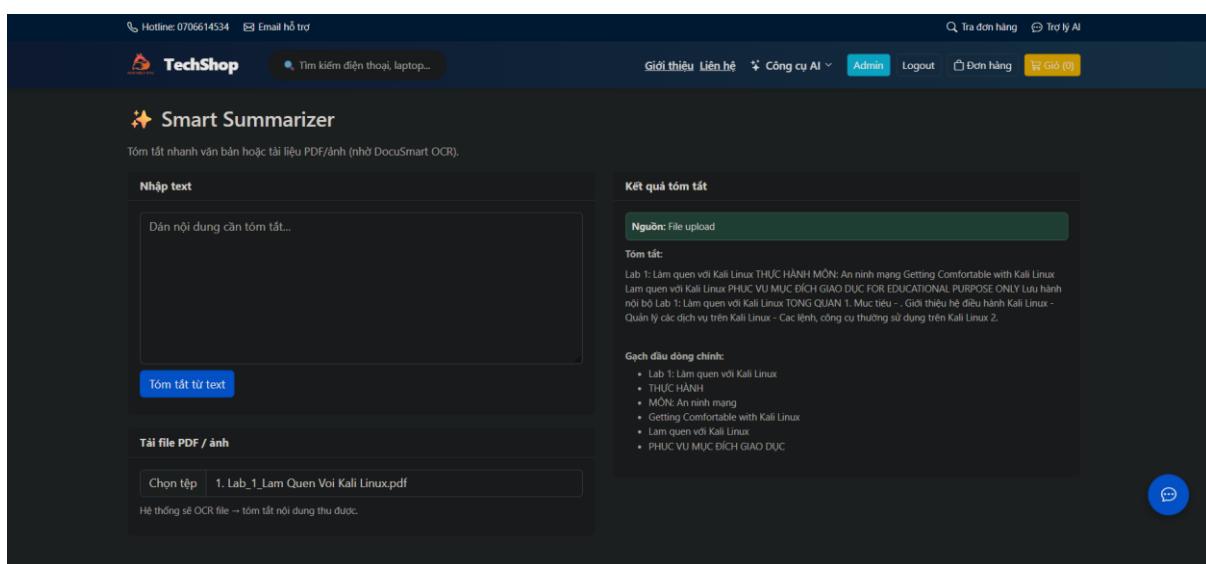
Hình 6 Giao diện web TechShop



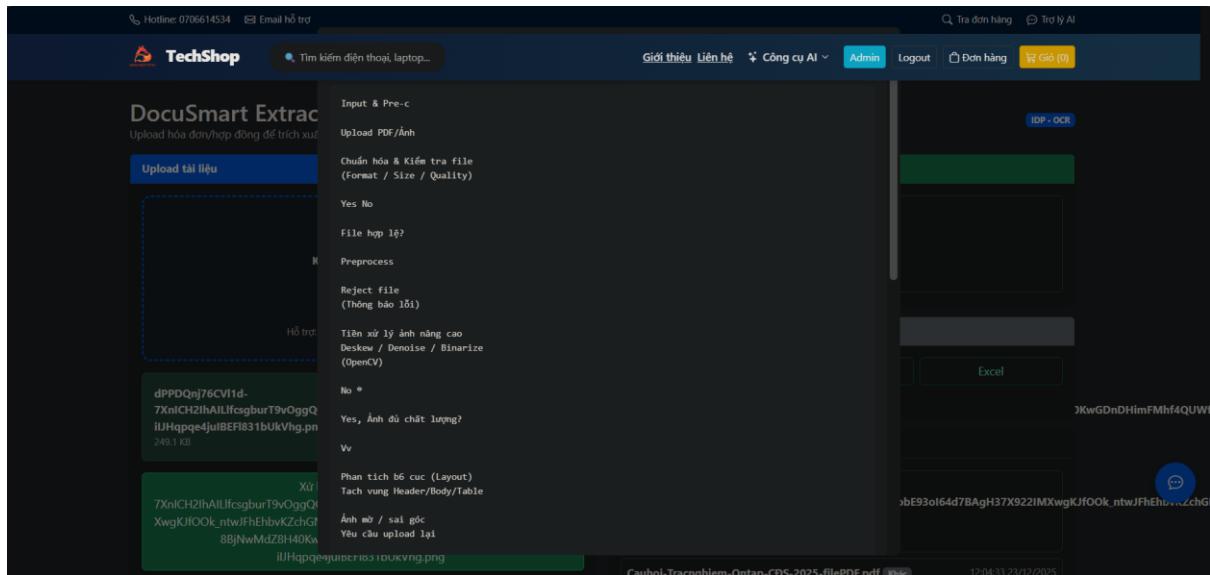
Hình 7 trợ lý ảo



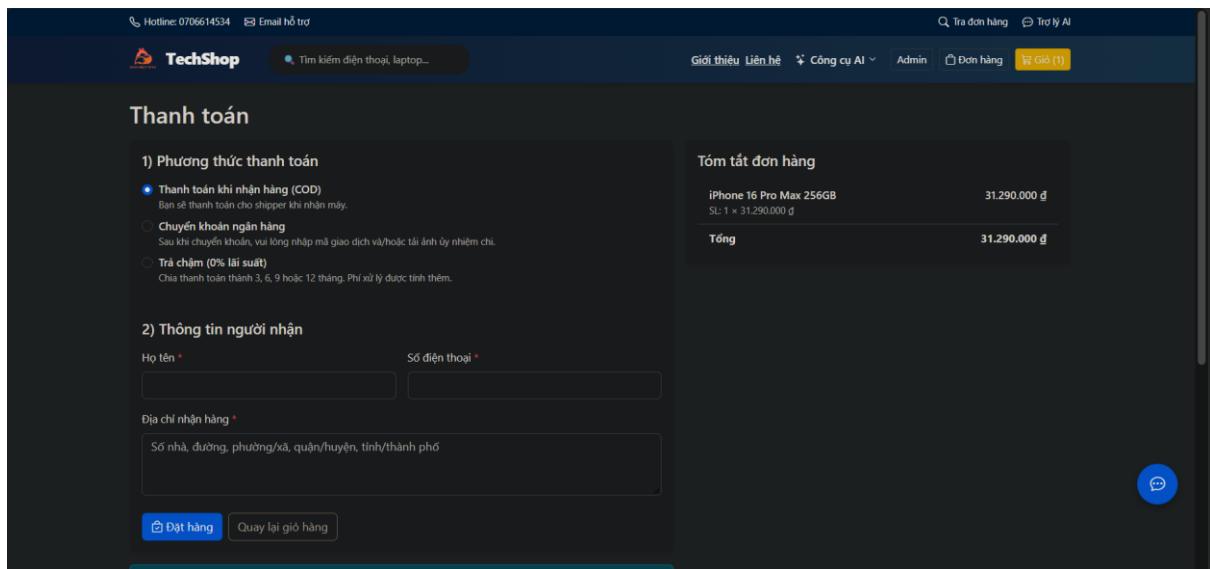
*Hình 8 Phân loại tài liệu*



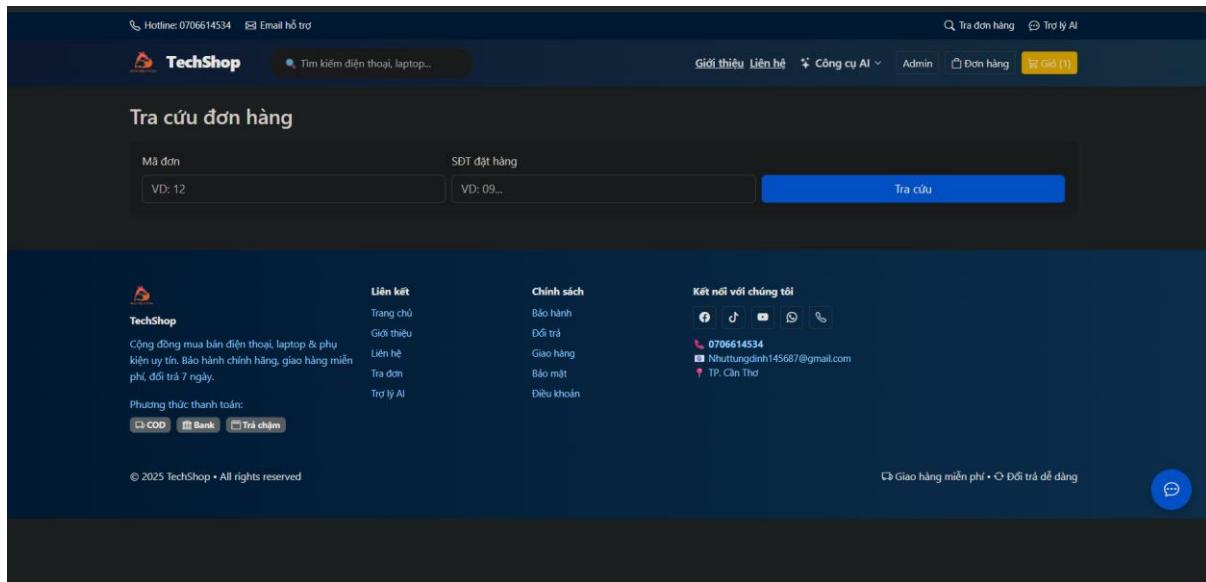
*Hình 9 Tóm tắt văn bản*



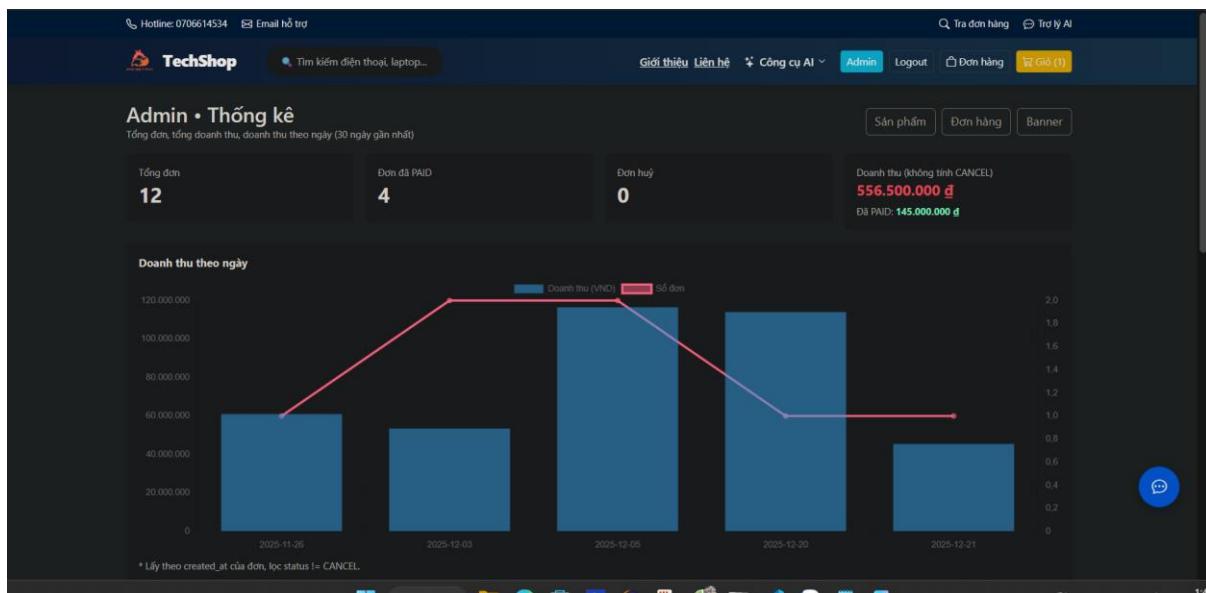
Hình 10 Trích xuất thông tin tài liệu



Hình 11 Các phương thức thanh toán



Hình 12 Tra cứu đơn hàng



Hình 13 Thống kê doanh thu

ID	Thời gian	Khách	SĐT	TT	Status	Tổng	Xem
#5	2025-12-21 10:26:57	Đinh Nhựt Tùng	0933333333	VNPay	pending	45.500.000 ₫	Xem
#4	2025-12-20 10:26:57	Trần Thị B	0901234567	VNPay	cancelled	114.000.000 ₫	Xem
#1	2025-12-05 10:26:57	Lê Minh C	0987654321	VNPay	done	86.000.000 ₫	Xem
#3	2025-12-05 10:26:57	Đinh Nhựt Tùng	0912345678	VietQR	cancelled	30.500.000 ₫	Xem
#2	2025-12-03 10:26:57	Phạm Quốc D	0901234567	COD	shipping	11.500.000 ₫	Xem
#7	2025-12-03 10:26:57	Phạm Quốc D	0987654321	VNPay	cancelled	42.000.000 ₫	Xem
#10	2025-11-26 10:26:57	Trần Thị B	0933333333	COD	shipping	61.000.000 ₫	Xem
#8	2025-11-19 10:26:57	Hoàng Thùy E	0987654321	VNPay	done	28.500.000 ₫	Xem
#11	2025-10-09 10:26:57	Đinh Nhựt Tùng	0901234567	VietQR	cancelled	8.000.000 ₫	Xem
#9	2025-10-08 10:26:57	Lê Minh C	0901234567	VietQR	done	15.500.000 ₫	Xem

Hình 14 Trang thái đơn hàng

ID	Thời gian	Khách	SĐT	TT	Trạng thái	Tổng	Xem
#5	2025-12-21 10:26:57	Đinh Nhựt Tùng	0933333333	VNPay	pending	45.500.000 ₫	Xem
#4	2025-12-20 10:26:57	Trần Thị B	0901234567	VNPay	cancelled	114.000.000 ₫	Xem
#1	2025-12-05 10:26:57	Lê Minh C	0987654321	VNPay	done	86.000.000 ₫	Xem
#3	2025-12-05 10:26:57	Đinh Nhựt Tùng	0912345678	VietQR	cancelled	30.500.000 ₫	Xem
#2	2025-12-03 10:26:57	Phạm Quốc D	0901234567	COD	shipping	11.500.000 ₫	Xem
#7	2025-12-03 10:26:57	Phạm Quốc D	0987654321	VNPay	cancelled	42.000.000 ₫	Xem
#10	2025-11-26 10:26:57	Trần Thị B	0933333333	COD	shipping	61.000.000 ₫	Xem
#8	2025-11-19 10:26:57	Hoàng Thùy E	0987654321	VNPay	done	28.500.000 ₫	Xem
#11	2025-10-09 10:26:57	Đinh Nhựt Tùng	0901234567	VietQR	cancelled	8.000.000 ₫	Xem
#9	2025-10-08 10:26:57	Lê Minh C	0901234567	VietQR	done	15.500.000 ₫	Xem

Hình 15 Nhiều phương thức xuất file báo cáo

## CHƯƠNG 7. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

### 7.1. Kết luận

Đề tài đã xây dựng thành công một mô hình chuyển đổi số tích hợp cho doanh nghiệp bán lẻ điện thoại: website TMĐT TechShop Plus kết hợp mô-đun IDP DocuSmart Extractor. Hệ thống đáp ứng các chức năng cốt lõi: quản lý sản phẩm, giờ hàng, đặt hàng và thanh toán; đồng thời cho phép upload tài liệu chứng từ, OCR tiếng Việt và trích xuất các trường dữ liệu quan trọng; hỗ trợ xuất dữ liệu và lưu lịch sử xử lý. Sản phẩm thể hiện tính khả thi ở mức độ án, tận dụng công nghệ mã nguồn mở, chi phí thấp và có thể triển khai nhanh trong doanh nghiệp SME.

### 7.2. Hạn chế

Một số hạn chế chính: bộ dữ liệu thử nghiệm nhỏ nên kết quả đánh giá mang tính tham khảo; rule-based extraction phụ thuộc vào mẫu biểu và chất lượng OCR; hệ thống chưa tích hợp đầy đủ cơ chế bảo mật nâng cao (HTTPS, CSRF, rate limit) trong môi trường demo; chưa tích hợp sâu việc gắn chứng từ vào đơn hàng tự động.

### 7.3. Hướng phát triển

Trong tương lai, hệ thống có thể được nâng cấp theo các hướng:

- + Nâng cấp OCR: tích hợp PaddleOCR hoặc mô hình OCR chuyên tiếng Việt; bổ sung deskew, phát hiện vùng (layout) để tăng độ chính xác.
- + Nâng cấp trích xuất: dùng mô hình NER tiếng Việt/LLM để trích xuất theo ngữ cảnh; hỗ trợ nhiều mẫu hóa đơn/biên nhận.
- + Tích hợp nghiệp vụ: tự động liên kết chứng từ với đơn hàng; đối soát số tiền chuyển khoản; cảnh báo sai lệch.
- + Mở rộng TMĐT: tích hợp thanh toán VietQR/VNPay, thông báo Zalo/Email, quản trị kho bằng barcode.
- + Triển khai thực tế: tách DocuSmart thành microservice, dùng message queue (RabbitMQ/Redis) để xử lý nền; triển khai trên cloud và cấu hình backup.

## TÀI LIỆU THAM KHẢO

- 1) Pallets Projects. (n.d.). *Flask documentation*. <https://flask.palletsprojects.com/>
- 2) Pallets Projects. (n.d.). *Jinja documentation*. <https://jinja.palletsprojects.com/>
- 3) SQLAlchemy. (n.d.). *SQLAlchemy documentation*. <https://docs.sqlalchemy.org/>
- 4) SQLite. (n.d.). *SQLite documentation*. <https://www.sqlite.org/docs.html>
- 5) OpenCV. (n.d.). *OpenCV documentation*. OpenCV Foundation. <https://docs.opencv.org/>
- 6) Tesseract OCR. (n.d.). *Tesseract OCR*. <https://github.com/tesseract-ocr/tesseract>
- 7) PyTesseract. (n.d.). *pytesseract documentation*. <https://pypi.org/project/pytesseract/>
- 8) PaddlePaddle. (n.d.). *PaddleOCR documentation*. <https://github.com/PaddlePaddle/PaddleOCR>
- 9) Python Software Foundation. (n.d.). *Python documentation*. <https://docs.python.org/3/>
- 10) Bootstrap. (n.d.). *Bootstrap documentation*. <https://getbootstrap.com/docs/>
- 11) OWASP Foundation. (n.d.). *OWASP Top 10: Web application security risks*. <https://owasp.org/www-project-top-ten/>
- 12) RFC Editor. (2015). *RFC 7519: JSON Web Token (JWT)*. <https://www.rfc-editor.org/rfc/rfc7519>
- 13) Fielding, R. T. (2000). *Architectural styles and the design of network-based software architectures* (Doctoral dissertation, University of California, Irvine). [https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding\\_dissertation.pdf](https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf)
- 14) Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- 15) Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed., draft). <https://web.stanford.edu/~jurafsky/slp3/>
- 16) Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 4171–4186. <https://aclanthology.org/N19-1423/>
- 17) Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 30. <https://papers.nips.cc/paper/7181-attention-is-all-you-need>
- 18) Honnibal, M., & Montani, I. (n.d.). *spaCy: Industrial-strength natural language processing in Python*. <https://spacy.io/>
- 19) Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., & Brew, J. (2020). Transformers: State-of-the-art natural language processing. *Proceedings of EMNLP 2020: System Demonstrations*, 38–45. <https://aclanthology.org/2020.emnlp-demos.6/>
- 20) Chollet, F. (2018). *Deep learning with Python* (2nd ed.). Manning Publications.
- 21) Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.

- 22) Lampert, L. (1994). *LaTeX: A document preparation system* (2nd ed.). Addison-Wesley.
- 23) World Wide Web Consortium (W3C). (n.d.). *HTML specification*.  
<https://html.spec.whatwg.org/>
- 24) Mozilla Developer Network. (n.d.). *JavaScript guide*. Mozilla.  
<https://developer.mozilla.org/en-US/docs/Web/JavaScript/Guide>
- 25) Mozilla Developer Network. (n.d.). *CSS: Cascading Style Sheets*. Mozilla.  
<https://developer.mozilla.org/en-US/docs/Web/CSS>
- 26) Amazon Web Services. (n.d.). *Amazon S3 documentation*.  
<https://docs.aws.amazon.com/s3/>
- 27) Redis. (n.d.). *Redis documentation*. <https://redis.io/docs/>
- 28) NGINX. (n.d.). *NGINX documentation*. <https://nginx.org/en/docs/>
- 29) Docker. (n.d.). *Docker documentation*. <https://docs.docker.com/>
- 30) Kubernetes. (n.d.). *Kubernetes documentation*. <https://kubernetes.io/docs/>
- 31) Apache Software Foundation. (n.d.). *Apache Kafka documentation*.  
<https://kafka.apache.org/documentation/>
- 32) RabbitMQ. (n.d.). *RabbitMQ documentation*.  
<https://www.rabbitmq.com/documentation.html>
- 33) Fowler, M. (2002). *Patterns of enterprise application architecture*. Addison-Wesley.
- 34) Newman, S. (2015). *Building microservices*. O'Reilly Media.