



AI POWERED DOCUMENT CLASSIFICATION AND INTELLIGENT INDEXING

BY CodeNova





TABLE OF CONTENTS

• PROBLEM STATEMENT	01
• SYSTEM ARCHITECTURE	02
• KEY FEATURE	03
• TECH STACK	04
• DEMO WORKFLOW	05
• FUTURE SCOPE	06
• Conclusion	07



PROBLEM STATEMENT

Managing, analyzing, and retrieving insights from large volumes of documents is time-consuming and inefficient.

Artificial Intelligence

- Extract and summarize text from documents
- Classify documents into categories
- Provide secure access with authentication
- Enable quick search and retrieval



SYSTEM ARCHITECTURE



- User Interface
- Text Extraction Layer
- AI Processing Layer
- Database Layer (MySQL - XAMPP)
- Storage Layer
- Access Control



KEY FEATURES

- Upload Documents (PDF, DOCX, TXT)
- Automatic Text Extraction
- AI-Powered Classification & Summarization
- Store Metadata in MySQL Database
- Search Keyword
- Show Results Ranked by Meaning
- Secured Role-based Access

TECH STACK

01

STREAMLIT (FRONTEND + DASHBOARD)

02

HUGGING FACE TRANSFORMERS (CLASSIFICATION+SUMMARIZATION)
PYTORCH (MODEL BACKEND)

03

MYSQL (XAMPP) (USERS, DOCUMENTS STORAGE)

04

PYPDF2 / PYTHON-DOCX (TEXT EXTRACTION)

DEMO WORKFLOW

- 1. USER LOGS IN
- 2. UPLOADS A DOCUMENT
- 3. SYSTEM EXTRACTS TEXT
- 4. AI CLASSIFIES AND SUMMARIZES
- 5. DOCUMENT SAVED WITH METADATA
- 6. USER SEARCHES OR VIEWS DOCUMENTS
- 7. LOG OUT

FUTURE SCOPES

- - ADD SUPPORT FOR MORE FILE FORMATS (EXCEL, IMAGES WITH OCR)
- - ADVANCED SEMANTIC SEARCH USING EMBEDDINGS
- - INTEGRATION WITH CLOUD STORAGE (AWS, GCP, AZURE)
- - REAL-TIME COLLABORATION FEATURES



CONCLUSION



- AUTOMATED TEXT EXTRACTION: EFFICIENTLY EXTRACTS TEXT FROM PDF, DOCX, AND TXT FILES, REDUCING MANUAL EFFORT.
- AI-POWERED INSIGHTS: USES ZERO-SHOT CLASSIFICATION TO CATEGORIZE DOCUMENTS AND SUMMARIZATION TO PROVIDE QUICK OVERVIEWS.
- SECURE & ROLE-BASED ACCESS: STORES DOCUMENTS AND METADATA IN MYSQL WITH ACCESS CONTROL BASED ON USER ROLES.
- QUICK SEARCH & RETRIEVAL: ALLOWS SEARCHING BY TITLE, SUMMARY, OR CATEGORY FOR FAST DOCUMENT ACCESS.
- ENHANCED PRODUCTIVITY: STREAMLINES DOCUMENT MANAGEMENT, SAVES TIME, AND PROVIDES INTELLIGENT INSIGHTS FOR BETTER DECISION-MAKING.



THANK YOU!

