**Smart Expense Tracker with Receipt OCR and Auto-Categorization**

**1. Project Topic:**

Development of a Python-based **Smart Expense Tracker** that allows users to log expenses manually or automatically via **receipt image upload**. The project integrates **EasyOCR** for text extraction and **auto-categorization** of expenses to improve user convenience and tracking accuracy.

**2. Objective:**

* Automate expense logging from receipts
* Extract key details: amount, date, vendor
* Suggest expense categories automatically
* Provide interactive visualization of spending by category and over time

**3. Dataset:**

Sample receipts for testing OCR functionality are sourced from:

* 1. [High-Quality Invoice Images for OCR](https://www.kaggle.com/datasets/osamahosamabdellatif/high-quality-invoice-images-for-ocr?utm_source=chatgpt.com)
  2. [Invoice OCR Dataset](https://www.kaggle.com/datasets/senju14/invoice-ocr?utm_source=chatgpt.com)

These datasets are used **only for testing and validating OCR**, not for ML training. EasyOCR is pre-trained.

**4. Tools & Technologies:**

* **Python 3.x**
* **EasyOCR** for OCR
* **Streamlit** for interactive UI
* **Pandas** for data management
* **Matplotlib / Streamlit charts** for visualization

**5. Features:**

* **Manual expense entry**
* **Receipt upload with OCR**
* **Auto-categorization** (keyword-based or similarity-based)
* **Confirmation interface** for user to adjust category
* **Data storage** in CSV/JSON
* **Charts & summaries** of expenses

**6. Workflow / Pipeline:**

1. User adds expense manually or uploads receipt image
2. OCR (EasyOCR) extracts text from receipt
3. Parsing logic extracts **amount, date, vendor**
4. Auto-categorization suggests category based on vendor
5. User confirms or edits category
6. Expense data saved to CSV/JSON
7. Streamlit UI shows charts and recent transactions

A diagram of a software flow

AI-generated content may be incorrect.

**7. Deliverables:**

* **Source code** (src/) including tracker logic, OCR module, and UI
* **Documentation** (Project\_Description.docx, screenshots)
* **WBS** (tasks, responsibilities, progress tracking)
* **Sample dataset** for testing

**8. Team Members & Roles:**

* **Monika:** Core expense tracker logic
* **Sovann:** Receipt OCR module (EasyOCR + parsing)
* **Rithy:** Streamlit UI, charts, documentation

**9. Timeline (2 Weeks):**

* Week 1: Core tracker + OCR module development + testing
* Week 2: Streamlit UI, integration, debugging, documentation, ZIP submission