project\_root/

│

├── src/

│ ├── api/ ← FastAPI routes

│ │ ├── endpoints/

│ │ │ ├── upload.py ← POST /upload endpoint

│ │ │ ├── classify.py ← Calls Google Doc AI classifier

│ │ │ ├── extract.py ← Routes to OCR pipeline

│ │ │ └── validate.py ← Field-level validation

│ │ └── \_\_init\_\_.py

│ │

│ ├── services/ ← Logic layer (modular + swappable)

│ │ ├── classification/

│ │ │ ├── google\_docai.py ← Google Doc AI classifier handler

│ │ ├── extraction/

│ │ │ ├── azure\_bank\_ocr.py ← Azure OCR for bank statement

│ │ │ ├── google\_bank\_ocr.py ← Google OCR alternative

│ │ ├── ml/

│ │ │ └── transaction\_classifier.py ← ML model logic

│ │ ├── normalization/

│ │ │ └── bank\_statement.py ← Field-level normalization rules

│ │ ├── mapping/

│ │ │ └── bank\_mapper.py ← Map raw → schema format

│ │ ├── db/

│ │ │ └── insert.py ← Populate DB

│ │ └── \_\_init\_\_.py

│ │

│ ├── models/ ← Pydantic schemas

│ │ ├── doc\_input.py

│ │ ├── doc\_output.py

│ │ └── enums.py

│ │

│ ├── utils/

│ │ └── logging.py ← Logging helper, retries, etc.

│ │

│ ├── config/

│ │ ├── settings.py ← API keys, endpoints, secrets

│ │ └── .env

│ │

│ └── main.py ← FastAPI app entry point

│

├── streamlit\_ui/ ← Optional: Streamlit visualization

│

├── poetry.lock

├── pyproject.toml

└── README.md