MIS 382N Advanced Machine Learning Project Proposal

Smart Doc Approver - Adaptive ML Agent for Receipt and Invoice Processing

Emily Caraher, Luke Hartfield, John MacDonald, Michael Ovassapian, Raghu Subramanian

I. Problem Description

Manual processing of receipts and invoices is a very time consuming process, but is necessary for many business workflows. The varied and unstructured nature of these documents makes the review and approval a significant challenge. Traditional automation systems are commonly used to lessen the human intervention required in this process, but these systems are typically non-adaptive and cannot learn from historical approval data or adapt to document variations across vendors and formats.

Our project aims to deliver an end-to-end ML based agentic solution that will ingest the uploaded document (invoice / receipt), extract and structure the fields, classify the document and finally approve / reject the same, while also accounting for anomaly detection. This will ultimately reduce human workload and error rates, while having a continuous adaptation with HITL feedback.

We will utilize freely available, academically recognized datasets to ensure reproducibility and practical relevance. The methods applied also align with both the classification and neural network concepts taught in class, and highlight the agentic AI principles as emphasized in today's ML applications.

II. Data Sources

- RVL-CDIP: Large, labeled archive of 400,000 scanned business documents (receipts, invoices, letters, memos).
- SROIE & CORD: Annotated receipts with field-level labels (e.g., amount, vendor, date). FUNSD: Annotated forms dataset for layout and key-value extraction.
- Synthetic Approval Logs: Custom CSV data simulating approvals, rejections, flags, and anomalies, aligned with rules.
- Human Feedback Logs: Record of manual corrections and overrides for continuous model retraining.

III. Proposed Methodology

Our project implements an agentic ML workflow for automated document approval, as illustrated in Figure 1. Upon user upload, the system applies open-source OCR (EasyOCR) to extract text from the document image. Key fields such as vendor, amount, and date are identified through LayoutLM, trained and evaluated with examples from the SROIE dataset. Next, documents are classified using a CNN/Transformer-based model (RVL-CDIP), and structured features are processed by a tabular classifier (XGboost) to predict approval, rejection, or the need for manual review. In parallel, anomaly detection (Iso forests) monitors for duplicates and suspicious submissions.

As depicted, high-confidence documents are auto-approved, with users notified. Documents flagged as low-confidence / anomalous are routed for HITL review, and all feedback is recorded for retraining—ensuring continuous improvement of the agent's decision accuracy.

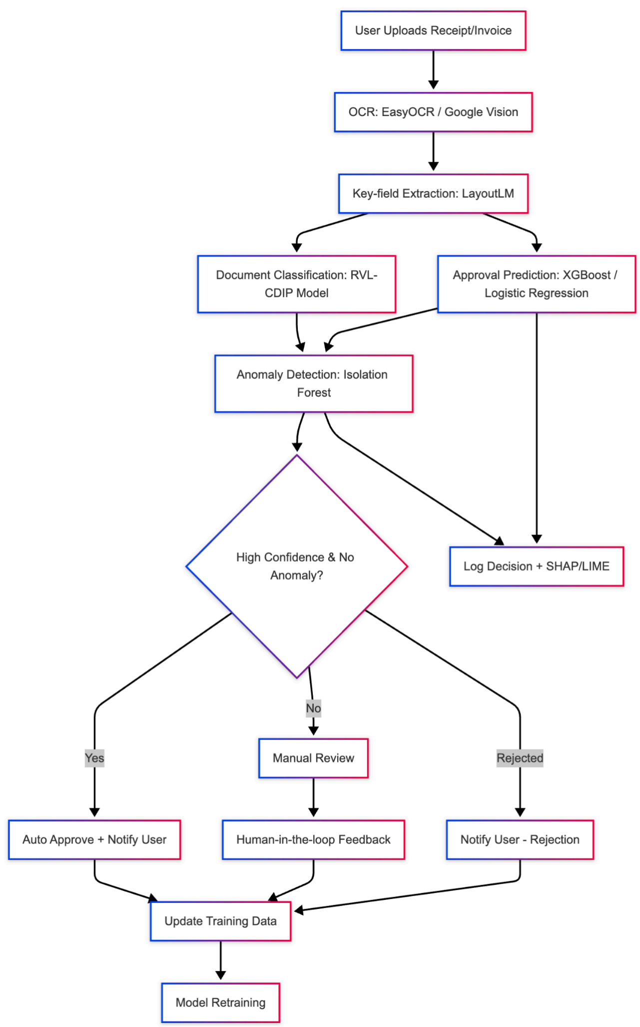

Figure 1: End-to-end agentic ML workflow for automated document approval.

IV. Evaluation Metrics

- OCR Accuracy (EasyOCR + SROIE): Quality of extracted text will be assessed using character- and word-level accuracy of extracted text against labeled examples in the SROIE dataset.
- Field Extraction Performance (SROIE): The key fields like vendor, date, etc will be extracted and evaluated on LayoutLM using precision and recall against the labeled examples.
- Document Classification Accuracy (CNN / Transformers + RVL-CDIP): The classification CNN model will be measured using overall accuracy and confusion matrices to assess performance across document types.
- Approval Prediction (Boosting): Automated decisions will be based on the structured features and evaluated using accuracy, precision, recall and AUC-ROC for approval and rejection predictions.
- Anomaly and Duplicate Detection: Measured using true positive and false positive rates for identifying unusual or repeated submissions, potentially via some Keras libraries.
- Finally, the entire workflow will also be monitored to identify the ratio of auto-approval and manual reviews, and the improvements in accuracy following the HITL reasoning.

V. References

- Harley, A. W., Ufkes, A., & Derpanis, K. G. (2015). Evaluation of deep convolutional nets for document image classification and retrieval. In 2015 13th International Conference on Document Analysis and Recognition (ICDAR). https://adamharley.com/icdar15/harley_convnet_icdar15.pdf
- Xu, Y., Li, M., Cui, L., Huang, S., Wei, F., Zhou, M., & Hon, H.-W. (2020). LayoutLM: Pre-training of text and layout for document image understanding. arXiv preprint arXiv:1912.13318.
- RVL-CDIP Dataset. (n.d.). Kaggle. https://www.kaggle.com/datasets/abdellatifsassioui/rvlcdip
- SROIE Dataset. (n.d.). Kaggle. https://www.kaggle.com/datasets/urbikn/sroie-datasetv2
- Park, S., Shin, S., Lee, B., Lee, S., Surh, J., Seo, M., & Lee, H. (2019). CORD: A consolidated receipt dataset for post-OCR parsing. GitHub. https://github.com/clovaai/cord
- Jaume, G., Ekenel, H. K., & Thiran, J.-P. (2019). FUNSD: Form understanding in noisy scanned documents. GitHub. https://github.com/crcresearch/FUNSD