Certified SDG Classifier – Informatory Document

Project Title:

Certified SDG Classifier: AI-Powered Document Categorization Tool Based on UN Sustainable Development Goals

Objective:

To build an AI-enabled web application that automatically classifies uploaded documents (PDFs or DOCX files) into the most relevant UN Sustainable Development Goal (SDG) category and generates a formal certificate indicating its alignment.

Key Features:

- • Multi-format Document Support: Accepts PDF and DOCX files.
- • Text Extraction: Utilizes PyMuPDF and python-docx to extract content.
- • NLP Preprocessing: Tokenizes and cleans the extracted text using nltk.
- • Keyword-Based SDG Mapping: Uses predefined keywords for each of the 17 SDGs to guide classification.
- • LLM-based Classification: Employs the llama3.1:8b model via Ollama and LangChain to provide contextual classification.
- • Result Presentation: Displays SDG prediction along with the reasoning.
- • Certificate Generation: Automatically generates an HTML-based certificate with dynamic SDG classification and logos.
- • Download Option: Users can download the certificate as an HTML file.

Technical Stack:

Component	Technology/Library
Frontend	Streamlit
Backend AI Model	LLaMA 3.1 (via Ollama)
NLP Preprocessing	NLTK
Document Handling	PyMuPDF, python-docx
LLM Integration	LangChain, Langchain-Ollama
Certificate Output	HTML + Base64 Images
File Types Supported	PDF, DOCX

Workflow:

- • Upload Documents: Users upload one or more PDF or DOCX files.
- • Text Extraction: PDF: Extracted via fitz (PyMuPDF); DOCX: Extracted via python-docx.
- • Preprocessing: Sentences are tokenized and cleaned.
- • Prompt Creation: The cleaned content is structured into a LangChain prompt containing relevant SDG keywords.
- • Model Prediction: The LLaMA model predicts the most relevant SDG goal with an explanation.
- • Result Display: Displays the filename and predicted SDG with supporting explanation.
- • Certificate Generation: An HTML certificate is generated using the predicted SDG category and embedded logos.
- • Download Certificate: Option to download the certificate as an HTML file.

Example Use Cases:

- • Universities/Researchers: Validate academic work related to SDGs.
- • NGOs: Showcase project alignment with global goals.
- • Corporates: Certify policy or report documents for sustainability compliance.
- • Policy Makers: Organize and classify large volumes of public documents.

Data Privacy:

All processing is performed locally on the user's machine. No files or content are sent to any server, ensuring complete confidentiality.