AI-Powered Legal Contract Analyzer: Final Project Submission

CothonSolutions Internship Team

May 29, 2025

1 Project Overview

The AI-Powered Legal Contract Analyzer, developed by CothonSolutions, is a web-based tool designed to streamline the analysis of legal contracts for legal and business professionals. The tool automates the extraction of key clauses, identifies risky or unusual clauses using machine learning, generates plain-English summaries, and provides an interactive dashboard. This project leverages open-source NLP and ML tools to enhance efficiency in contract review.

2 Objectives

- Extract key clauses (e.g., indemnity, confidentiality, governing law) from legal contracts.
- Identify risky or unusual clauses using machine learning models.
- Summarize contracts in plain English using Legal-BERT.
- Provide a visual and interactive dashboard for easy navigation.

3 Technical Implementation

3.1 Technology Stack

- **NLP & Machine Learning**: SpaCy for Named Entity Recognition (NER), Scikit-learn for clause risk classification, HuggingFace's Legal-BERT for summarization, and the CUAD dataset for training.
- **Document Handling**: `pdfplumber` for PDF text extraction and `docx2txt` for DOCX files.
- **User Interface**: Streamlit for the web-based interface and Plotly/Matplotlib for visualizations.
- **Storage**: SQLite for data storage and Pandas/JSON for data manipulation.

3.2 System Architecture

1. **Document Processing**: Extracts text from PDF and DOCX files.
2. **Clause Extraction**: Uses SpaCy NER to identify key clauses.
3. **Risk Classification**: Trains Scikit-learn models to classify clause risk levels.
4. **Summarization**: Generates plain-English summaries with Legal-BERT.
5. **Storage**: Stores data in SQLite and manipulates it with Pandas.
6. **Interface**: Displays results via a Streamlit dashboard with Plotly visualizations.

3.3 Code Quality

The project adheres to best practices:

- Followed PEP 8 guidelines using `black` and `flake8`.
- Wrote modular code with separate modules for extraction, models, and the main app.
- Implemented robust error handling with try-except blocks.
- Wrote unit tests using `pytest`.
- Maintained a clear project structure (see README.md).

4 Challenges and Solutions

- **Complex PDF Formatting**: Used `pdfplumber`'s advanced parsing and added error handling.
- **Model Accuracy**: Fine-tuned SpaCy and Legal-BERT on the CUAD dataset.
- **Performance**: Optimized text processing with batching and efficient SQLite storage.
- **Dependency Management**: Used virtual environments and pinned package versions.

5 Results and Impact

The AI-Powered Legal Contract Analyzer successfully:

- Extracts key clauses with high accuracy using SpaCy.
- Classifies risky clauses with Scikit-learn models.
- Generates concise summaries with Legal-BERT.
- Provides an intuitive Streamlit dashboard for users.

The tool reduces manual contract review time and enhances decision-making for legal professionals.

6 Future Improvements

- Support for scanned PDFs using OCR.
- Enhanced risk classification with ensemble ML models.
- Cloud deployment on AWS or Heroku.
- Real-time collaboration features in the dashboard.

7 Conclusion

Developed under CothonSolutions, the AI-Powered Legal Contract Analyzer is a robust tool that leverages cutting-edge NLP and ML to automate contract analysis. It demonstrates the power of open-source technologies in solving real-world legal challenges, with potential for further enhancements.