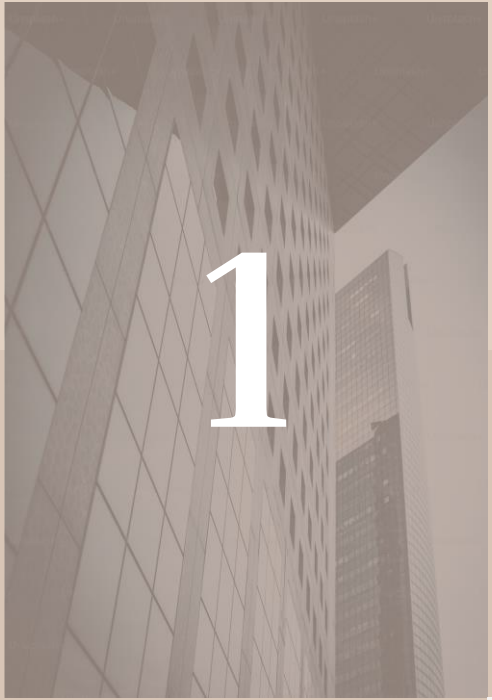
Intel® Unnati Industrial Training Program 2024

BUSINESS CONTRACT VALIDATION

ClauseX

Business Contract Validation - To Classify Content within the Contract Clauses and Determine Deviations from Templates and Highlight Them



Problem Statement

In the modern business environment, managing a vast number of complex contracts poses significant challenges in terms of accuracy and compliance. Manual validation of these contracts is not only time-consuming but also prone to errors, which can lead to legal and financial risks. There is a pressing need for an automated solution that can efficiently parse, analyze, and validate contract contents to ensure they meet predefined standards and highlight any deviations.

Unique Idea Brief

Our project uses advanced machine learning techniques to automate the verification process of the employment contract. Using PDF parsers, text classifiers, OCR, and Named Entity Recognition (NER), our system can analyze and segment contract features, identify deviations from predefined templates, and confirm for additional analysis This innovative solution increases the accuracy and efficiency of contract integrity Significantly reduces time and effort.

Features Offered

- **Automated Parsing and Classification**: Efficiently parses contract documents and classifies contents using machine learning models.
- **Deviation Detection**: Identifies and highlights deviations from predefined templates for easy review.
- **OCR Integration**: Converts scanned documents into machine-readable text for processing.
- **NER Implementation**: Extracts key entities and clauses from contracts to ensure comprehensive validation.
- **User-Friendly Interface**: Provides an intuitive interface for uploading contracts and viewing validation results.
- **Text Comparison and Summarization**: Uses advanced text comparison techniques to match contract sections with templates and summarize findings.

Process Flow

1. Contract Upload

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

2. Data Loading and Pre-processing

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

3. Disclosure Time

Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.

4. Template Comparison

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

5. Validation Result

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

ClauseX

Technology Used

Programming Language: Python

Libraries:

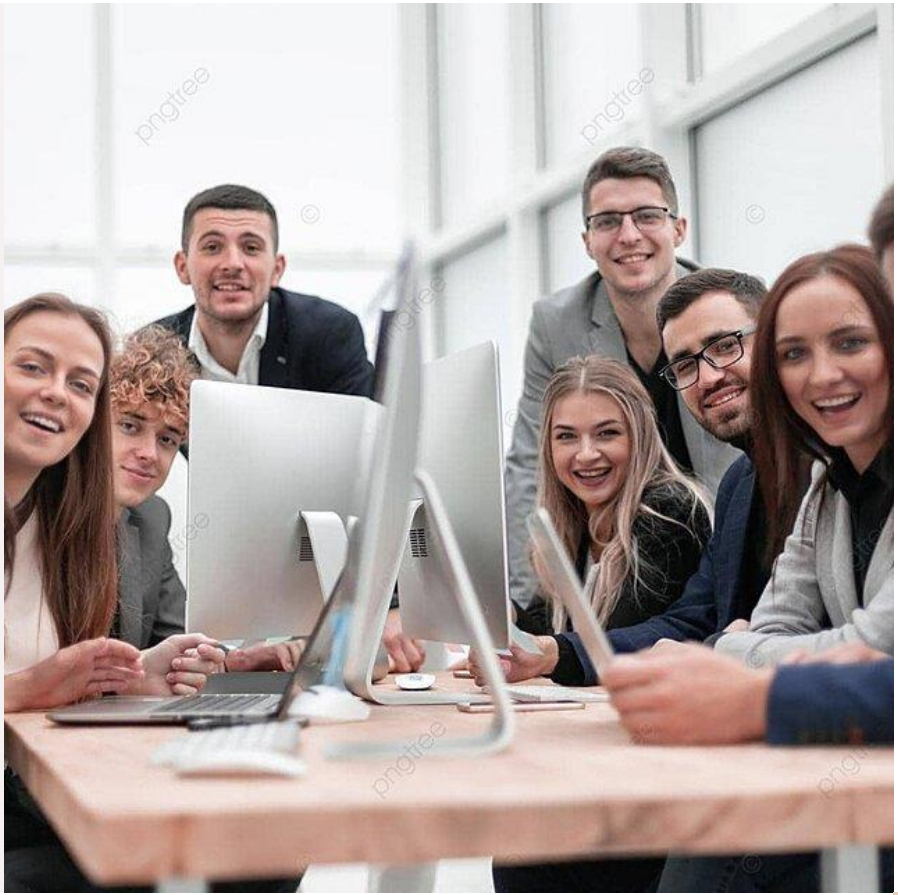
- PyMuPDF
- Pandas
- PyPDF2
- Torch
- EasyOCR
- Transformers (Hugging Face)
- Sklearn
- Flask
- Numpy

Frontend: HTML, JavaScript

Machine Learning Models: BERT for text classification and feature extraction

Team Members

- **Shane Sam:** UI/UX Design, Frontend Development
- **Megha Rajeev:** Model Training, Backend Development
- **Aiswarya Lakshmi:** Model Training, Backend Development
- **Neethu Benny:** Report and Documentation
- **Bestin Biju:** Report and Documentation



Conclusion

Our contract automation solution uses state-of-the-art machine learning to streamline the contract review process. By automating discrepancy detection and ensuring compliance, our solution significantly reduces the risk of manual errors and increases operational efficiency. This project addresses current challenges in contract management by increasing accuracy and saving time. An intuitive user interface simplifies contract uploads and provides transparent verification results. Ultimately, our program sets a new standard for future developments in law and practice, and improves accuracy, efficiency and compliance in contract management