

Business Contract Validation: A Streamlined Approach

Problem Statement :

Business contracts are crucial agreements outlining legal rights and obligations. However, manually reviewing and analyzing contracts for deviations from standard templates is a time-consuming and error-prone process. This project addresses the challenge of Business Contract Validation, aiming to automate the classification of content within contracts and identify deviations from pre-defined templates using innovative Machine Learning (ML) and Natural Language Processing (NLP) techniques.

Unique Idea

Our web application integrates advanced models with an intuitive interface for enhanced document processing and analysis.

- **Document Processing Model:**

Efficiently manages PDF and image formats through modular design. Tasks include parsing, text extraction, and document structure analysis, ensuring adaptability across various document types.

- **Intelligent Deviation Detection Model:**

Utilizes NLP techniques to identify document deviations accurately. It goes beyond keyword matching, providing insights into variations from predefined templates and standards.

- **Platform Benefits:**

Combines advanced technologies to improve efficiency, flexibility, and user experience. Offers precise document processing and analysis capabilities that streamline workflows and enhance data accuracy.

Features Offered

1. **Multi-format Document Handling:**

- Processes contracts in PDF or image formats for comprehensive document accessibility.

2. **Intelligent Clause Identification:**

- Automatically identifies and highlights clauses using predefined rules and templates.
- Facilitates efficient deviation detection during document review.

3. **Advanced Comparison Techniques:**

- Utilizes the Cosine Similarity function (from scikit-learn) to ensure accurate comparison of extracted clauses with templates.
- Enhances reliability in detecting document deviations.

4. **User-Centric Interface:**

- Intuitive frontend developed with HTML, CSS, and JavaScript.
- Enables easy navigation and interpretation of analysis results for improved user experience.

Process Flow

1. Contract Upload

- Users upload contracts in PDF or image format through the web interface.

2. Parsing and Text Extraction:

- The system extracts text from uploaded documents.

3. Clause Segmentation:

- Text is segmented into clauses based on predefined rules or natural language processing techniques.

4. Highlighted Text Extraction:

- Each clause is automatically highlighted to emphasize key content and potential deviations

.

5. Pre-processing:

- Extracted text undergoes cleaning to remove common words and noise, enhancing clarity.

6. Side-by-Side Comparison:

- Original contract and template are displayed side by side, with deviations highlighted for comparison.

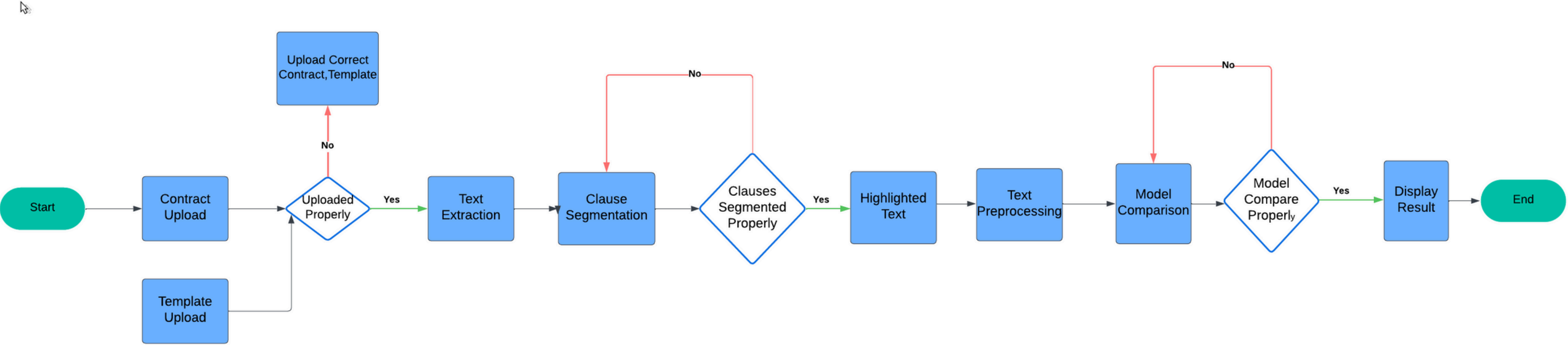
7. Deviation Detection:

- Advanced NLP techniques compare clauses, identifying and highlighting discrepancies.

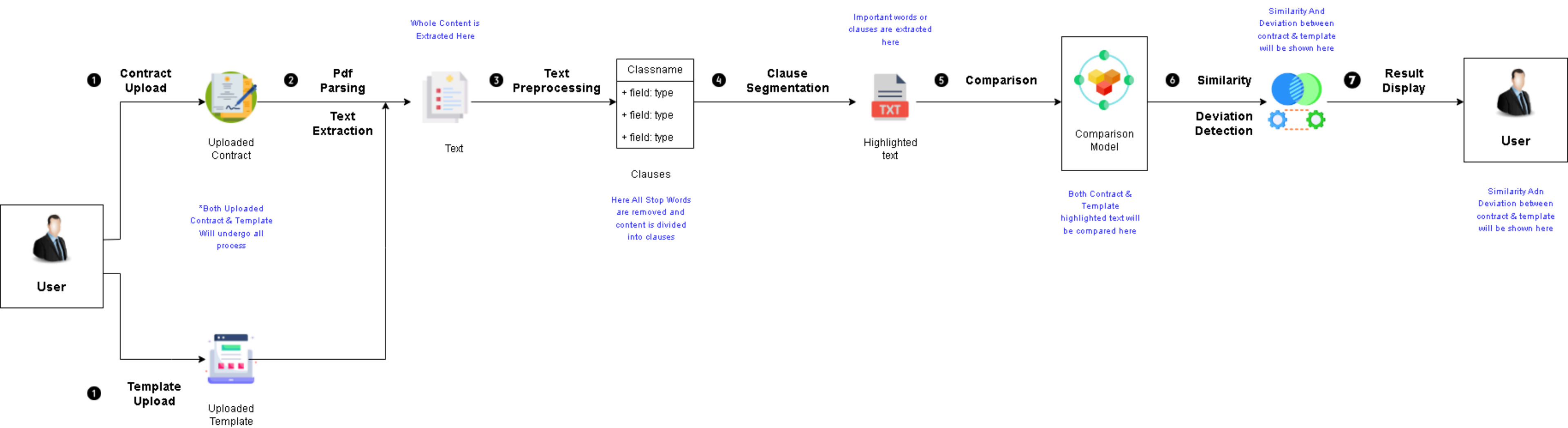
8. Display Results and Accuracy Assessment:

- Results of deviation detection and model accuracy are displayed, providing insights into document reliability and performance.

FlowChart



Architecture



Technologies Used

- Flask: A lightweight web framework for building the backend application.
- HTML, CSS, and JavaScript: The foundation for building the user interface.
- SpaCy: A powerful NLP library for text processing, analysis, and highlighting.
- PyMuPDF: A Python library for efficiently handling PDF documents.
- scikit-learn: A machine learning library used for the Cosine Similarity function to compare text and measure similarity.
- Streamlit: A Python library used during the testing phase for individually testing separate module functionalities.

Team Members and Contributions

Sr	Member Name	Role
1	Sagar Ahire	backend development, requirements analysis, functionality, testing, integration
2	Darshan Deshmukh	backend development, testing, logic, alternatives, optimization
3	Mrunmai Bhade	led front-end development, ensured user-friendly interface, aligned with requirements
4	Raj Nandale	assisted with AI prompting, researched Python library integration, provided feedback, documented project

Conclusion

This project introduces an innovative approach to Business Contract Validation via a web application. Its modular architecture offers flexibility and scalability, integrating advanced NLP and ML techniques for automated contract analysis and accurate deviation identification. The user-friendly interface includes side-by-side comparison and real-time model accuracy display, providing valuable insights into contracts. This has the potential to streamline contract review processes, reduce errors, and enhance operational efficiency in managing legal documents.