



International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AI-Powered Legal Clause Risk Analyzer using MERN Stack and Advanced Natural Language Processing Techniques

Miss. K.R. Nishmitha & P. Brindhaa & M. Harini & G. Abinayasri

Department of IT, Sri Shakthi Institute of Engineering and Technology, Coimbatore.

¹nishmithait@siet.ac.in, ²brindhapaooja@gmail.com, ³harinimanogaran05@gmail.com, ⁴abinayasrig2@gmail.com

ABSTRACT:

The AI-Powered Legal Clause Risk Analyzer is an intelligent web application designed to automatically identify and evaluate risky clauses in legal documents. Using Natural Language Processing (NLP) and Machine Learning, the system analyses legal texts such as contracts, agreements, and policies to detect potential risks, ambiguities, or unfair terms. The project is developed using the MERN stack (MongoDB, Express.js, React.js, Node.js) to ensure full-stack functionality — from user interaction on the frontend to data storage and analysis on the backend. The application enables users to upload documents (PDF or text), preprocess them, and receive a detailed risk report with clause-level recommendations. The system also maintains a record of analyses in MongoDB, providing users with a comprehensive report history for review and export.

Keywords: AI-Powered Legal Clause Analyzer, Natural Language Processing, Risk Assessment, Legal Document Analysis, Contract Review Automation, Clause Classification, MERN Stack, MongoDB, React.js, Node.js, Express.js, Tailwind CSS, PDF/Text Document Processing,

Risk Report Generation, Legal Tech, Document Preprocessing, Machine Learning in Legal Domain, Contract Risk Detection, Automated Legal Analysis, Legal Compliance

Introduction:

In today's digital era, legal documents such as contracts, agreements, and policies are often lengthy and written in complex legal language, making them difficult to interpret. Manual review of such documents is time-consuming and prone to human error, potentially leading to financial losses, legal disputes, or compliance issues. Identifying risky or ambiguous clauses requires expert knowledge, which is not always easily accessible.

The AI-Powered Legal Clause Risk Analyzer addresses this challenge by leveraging Artificial Intelligence (AI) and Natural Language Processing (NLP) to automatically detect and evaluate risky clauses. The system analyses legal texts to identify vague, high-liability, or unfavorable terms, reducing the dependency on manual review and improving accuracy. Users can upload PDF or text documents, which are preprocessed for clause extraction and analyzed by the AI engine to provide actionable recommendations.

Built using the MERN stack (MongoDB, Express.js, React.js, Node.js), the platform offers a seamless full-stack experience. All reports and analyses are securely stored in MongoDB, allowing users to track historical data and export detailed risk reports. This system is particularly valuable for law firms, corporate legal teams, and individuals, as it streamlines contract review, improves transparency, and empowers informed decision-making through AI-driven legal insights.

LEGALTECH AUTOMATION:

The AI-Powered Legal Clause Risk Analyzer is a Legal Tech automation system designed to streamline the review and analysis of legal documents. By leveraging Artificial Intelligence (AI) and Natural Language Processing (NLP), the system automatically identifies, classifies, and evaluates potentially risky clauses in contracts, agreements, and policies. Users can upload PDF or text documents, which are preprocessed and analyzed by AI models to highlight ambiguous, high-liability, or unfavorable terms.

Built on the MERN stack (MongoDB, Express.js, React.js, Node.js), the platform provides a fully automated workflow—from document upload and clause extraction to risk assessment and report generation. All analyses are securely stored in MongoDB, enabling historical tracking and easy retrieval of past reports. This automation reduces manual effort, ensures consistent and accurate evaluation, and empowers law firms, corporate legal teams, and individuals to make informed legal decisions efficiently.

By combining advanced AI techniques with modern web technologies, this Legal Tech solution transforms traditional contract review into an automated, scalable, and reliable process, improving transparency, minimizing legal risks, and enhancing operational efficiency.

NLP IN CONTRACT ANALYSIS:

Natural Language Processing (NLP) enables the AI-Powered Legal Clause Risk Analyzer to automatically read and understand legal documents. It extracts clauses, detects risks, classifies terms, and provides actionable recommendations, identifying vague, ambiguous, or high-liability clauses. By automating these tasks, NLP ensures accurate, consistent, and efficient contract analysis, reducing manual effort and helping users make informed legal decisions.

REQUIREMENTS ANALYSIS:

The system allows users to upload legal documents (PDF or text), automatically extract clauses, and analyse them for risks using AI and NLP. It generates detailed reports, stores them securely in MongoDB, and enables users to view, track, and export reports. Key requirements include scalability, security, fast processing, and an intuitive user interface.

MERN ARCHITECTURE:

The platform is built on the MERN stack: MongoDB stores documents, clauses, and reports securely. Express.js manages backend APIs and integrates AI/NLP processing. React.js provides a responsive frontend for uploading documents and viewing results. Node.js runs the server and handles communication between frontend, backend, AI modules, and database. This architecture ensures a seamless, scalable, and maintainable system for automated legal document analysis.

DATA FLOW DIAGRAM (DFD):

The DFD shows the flow of data: 1) Users upload documents via the frontend. 2) Backend preprocesses documents and extracts clauses. 3) NLP models analyse clauses and detect risks. 4) Reports are generated and stored in MongoDB. 5) Users can view, track, or export reports. This flow ensures efficient, accurate, and automated contract analysis from upload to report generation.

DOCUMENT PREPROCESSING:

Document preprocessing is the first crucial step in the AI-Powered Legal Clause Risk Analyzer. Users upload legal documents in PDF or text format, which the system then converts into a structured form suitable for analysis. Preprocessing involves extracting individual clauses, removing unnecessary formatting, handling line breaks, and normalizing text. This ensures that the AI engine can accurately interpret the content of the legal document. Techniques such as tokenization, stop-word removal, and sentence segmentation are applied to prepare the text for risk classification. By automating this step, the system significantly reduces manual effort and ensures consistent and clean input for the AI model.

RISK CLASSIFICATION MODEL:

The core of the system is the risk classification model, which uses NLP techniques to analyze each extracted clause. The model identifies clauses that may contain ambiguous language, high liability, or unfair terms and classifies them as safe, risky, or critical. Additionally, the system generates actionable recommendations, helping users understand the potential impact of each clause. The AI model is trained on legal data and learns common patterns of risky clauses, making the analysis accurate and efficient. This automation transforms traditional contract review, which is time-consuming and error-prone, into a fast and reliable process.

BACKEND DEVELOPMENT (NODE.JS):

The backend of the system is developed entirely in Node.js, handling all server-side operations. It manages file uploads, document preprocessing, and communication with the AI/NLP model for clause analysis. The backend also integrates with MongoDB to securely store uploaded documents, extracted clauses, analysis reports, and historical data. Node.js ensures efficient handling of multiple user requests, fast processing of documents, and smooth interaction between the document input and AI engine. Additional backend functions include API endpoints for report retrieval, export functionality, and real-time status updates.

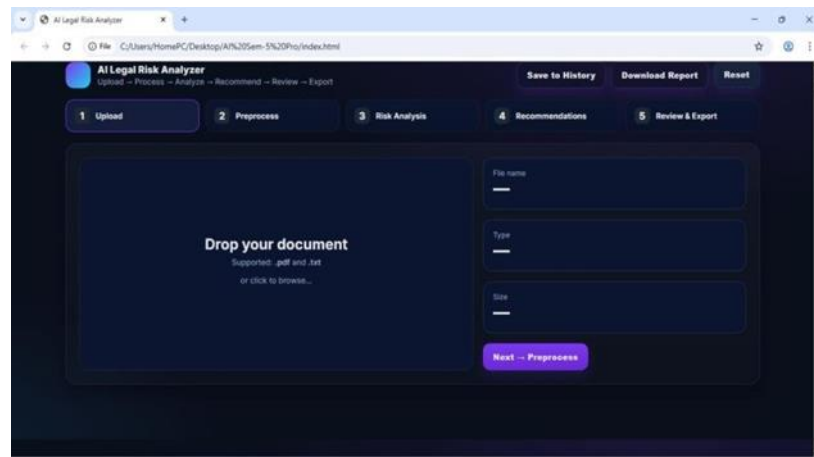
FRONTEND DEVELOPMENT:

The frontend provides a simple and intuitive interface for users to interact with the system. Users can upload documents, view clause-level risk analysis, and track historical reports stored in MongoDB. They can also download or export reports in PDF or text format for legal or business use. The frontend communicates with the Node.js backend through API calls, ensuring real-time updates and a seamless workflow. The design focuses on ease of use, clarity, and responsiveness, allowing users with no technical background to efficiently perform legal document analysis.

RESULTS AND DISCUSSION

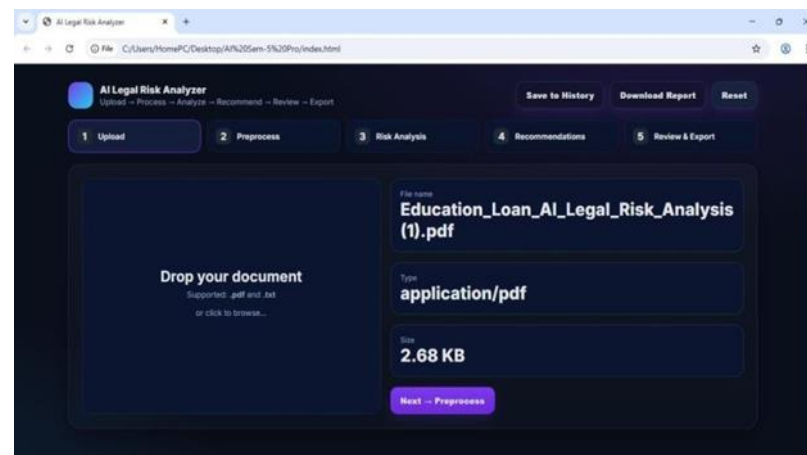
INITIAL DOCUMENT UPLOAD SCREEN:

The entry point of the AI Legal Risk Analyzer where users upload PDF/TXT legal documents through drag-and-drop functionality. The interface displays a five-stage workflow (Upload → Preprocess → Analyze → Recommend → Review → Export) with metadata fields for file information.



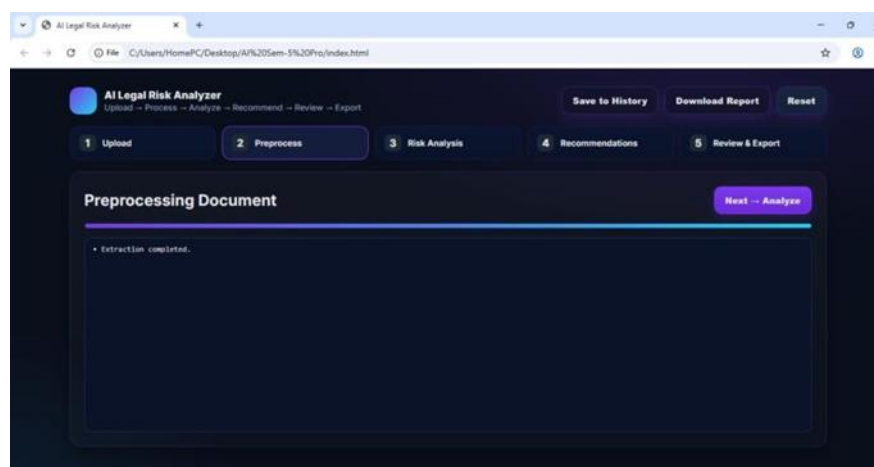
DOCUMENT UPLOAD VALIDATION SCREEN:

This interface enables users to upload legal documents (PDF or TXT) into the analyzer. It automatically displays file metadata such as name, type, and size. The dark-themed design ensures a clean, modern user experience. It serves as the entry point of the AI Legal Risk Analyzer workflow. This stage ensures seamless and secure document ingestion for analysis.



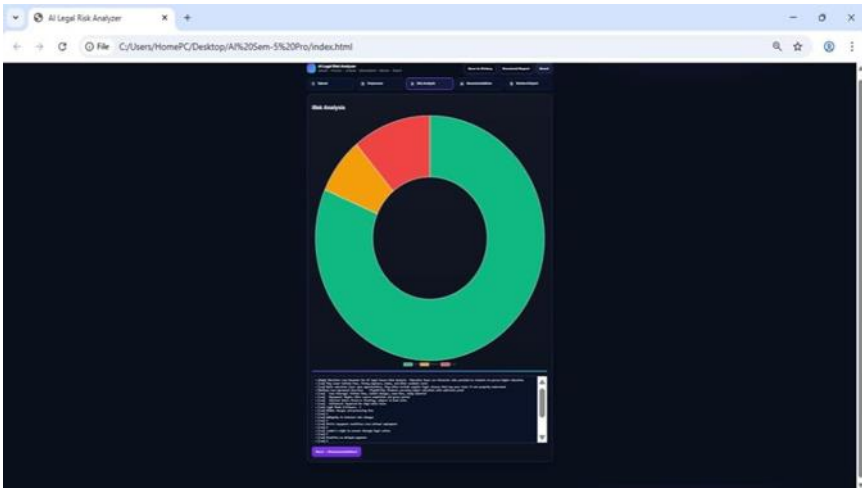
TEXT EXTRACTION AND PREPROCESSING PHASE:

After uploading, the system confirms success and displays key document details. The “Next → Preprocess” button becomes active for the next step. This ensures accuracy and transparency before further processing. It bridges the gap between frontend validation and backend file handling. This step reflects smooth interaction between user interface and MERN stack APIs.



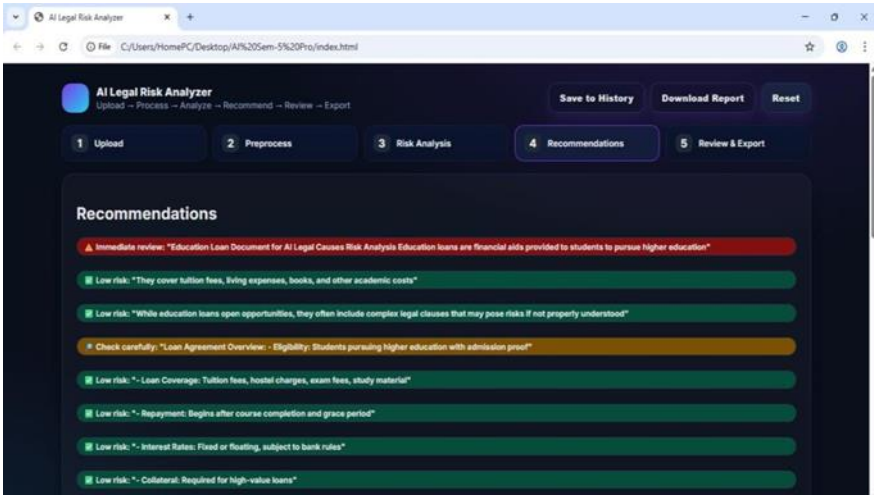
RISK ANALYSIS DASHBOARD WITH VISUAL DISTRIBUTION:

Interactive dashboard displaying risk severity distribution through a donut chart. The visualization categorizes identified legal risks into high (red), medium (orange), and low (green) severity levels, enabling quick assessment of contract risk profiles through AI-powered NLP analysis.



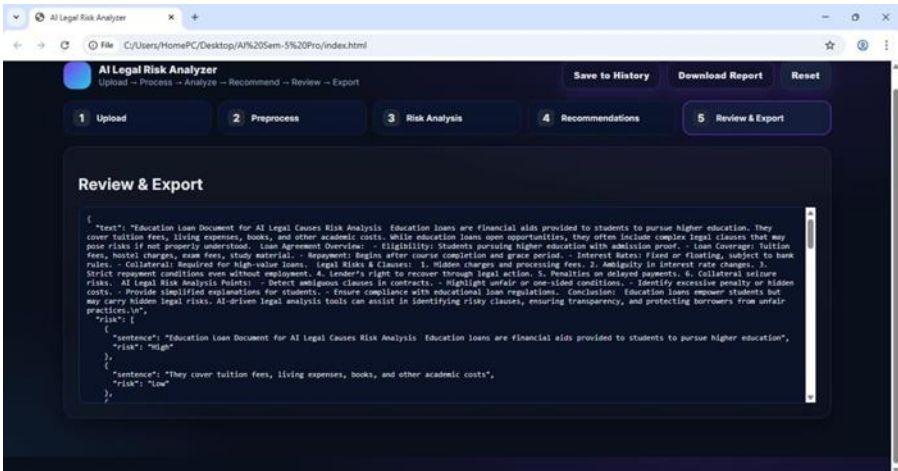
AI-DRIVEN RECOMMENDATIONS PANEL:

Intelligent recommendations interface presenting color-coded risk assessments. The system automatically categorizes clauses by urgency—immediate review (red), caution required (orange), and low concern (green)—helping legal professionals prioritize document review efficiently.



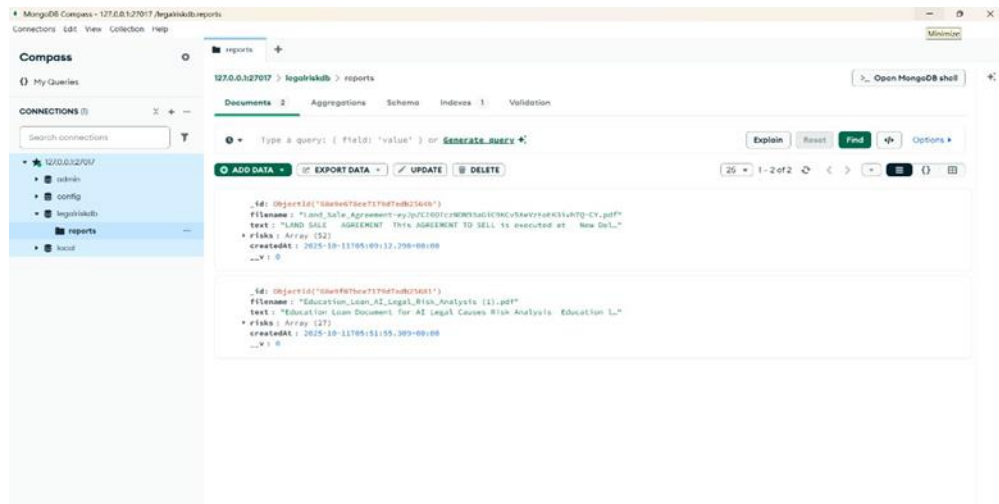
STRUCTURED JSON OUTPUT FOR EXPORT:

Review and export module displaying analysis results in JSON format. Shows comprehensive risk data including extracted text, severity classifications, and clause-level assessments, enabling seamless integration with external legal management systems.



MONGODB DATABASE SCHEMA ARCHITECTURE:

Backend database structure in MongoDB Compass showing the "reports" collection. Stores document metadata, extracted text, risk arrays, and timestamps using NoSQL architecture for scalable legal document management and historical analysis tracking.



CONCLUSION:

The *AI-Powered Legal Clause Risk Analyzer* successfully demonstrates how Artificial Intelligence and Natural Language Processing (NLP) can automate the complex process of legal document review. By combining AI-based risk detection with the MERN stack architecture, the system efficiently identifies risky, ambiguous, or unfair clauses in contracts and agreements. This reduces manual effort, enhances accuracy, and helps users make more informed legal decisions. With its ability to analyze documents, generate detailed risk reports, and store data securely in MongoDB, the project represents a step toward modernizing legal technology and promoting transparency in contract analysis.

FUTURE SCOPE:

1. Integration with Advanced AI Models:

Incorporate deep learning models like BERT or GPT-based legal analyser's for more precise and context-aware clause interpretation.

2. Multi-Language Support:

Enable the system to analyse legal documents written in multiple languages, making it accessible to global users.

3. Real Time Collaboration:

Introduce features for lawyers or teams to review and annotate documents collaboratively within the platform.

4. Automated Legal Drafting:

Extend the system to automatically suggest improved versions of risky clauses or even generate new contract drafts based on best practices.

5. Cloud Deployment and Accessibility:

Host the platform on cloud services (like AWS or Azure) to allow secure, scalable, and remote access for users and legal organizations.