



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 5, May 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Legal Document Generation using Artificial Intelligence

Mr. Pradeep K¹, Ganesh N Hegde², Gurukiran B³, Harsha M S⁴, Kishan Navali⁵

Assistant Professor, Department of CSE, PES Institute of Technology and Management, Shivamogga,
Karnataka, India¹

U.G Student, Department of CSE, PES Institute of Technology and Management, Shivamogga, Karnataka, India²⁻⁵

ABSTRACT: Legal assistance is an automated approach to providing accurate references, predictions and judgements to lawyers and legal professionals for case preparation using Machine Learning and Artificial intelligence. The legal document generator project aims to automate the process of creating various types of agreements, including rent agreements, business agreements, and loan agreements, using artificial intelligence techniques. The system takes basic details from the user, such as party names and addresses, and generates a comprehensive agreement document. By automating the document generation process, the project aims to improve efficiency, reduce errors, and provide users with legally sound agreements tailored to their specific needs.

KEYWORDS: LSTM.

I. INTRODUCTION

The legal landscape is constantly evolving, and technology is playing an increasingly important role in streamlining processes and making legal services more accessible. This project explores the development of a legal document generator specifically designed to create agreements through user input and user-defined conditions. This project on legal document generation offers a unique combination of user input, conditional logic, and LSTM-based summarization to create customized and efficient agreements. By automating basic document creation, the generator saves time and reduces errors. User input and conditional logic allow for tailored agreements, while LSTM summaries improve clarity and consistency. This technology empowers users with greater access to legal services, while still recognizing the importance of lawyers for complex situations. Overall, the project contributes to advancements in legal technology that can streamline processes, improve efficiency, and empower users. This project prioritizes user experience with a well-designed interface that simplifies the legal document generation process. Upon starting, users encounter a series of input fields tailored to capture essential details specific to their legal requirement. These fields might typically include information about parties involved. Following the input section, users are presented with a clear and concise menu for selecting the type of document they require. The system offers three primary options: rent agreements, business agreements, and loan agreements. To ensure the agreements are clear, concise, and easy to understand, the project integrates an LSTM (Long Short-Term Memory) model. This advanced machine learning technique comes into play after users define their conditions. The system reads the conditions from a comma-separated values (CSV) file, a common format for storing tabular data. This file is then fed into the LSTM model. The LSTM model plays a critical role by analyzing the sequence of words within each user-provided condition. By understanding the context and relationships between words, the model can generate a concise and clear summary of each condition. These LSTM-generated summaries are then seamlessly integrated into the appropriate sections of the chosen agreement template.

II. LITERATURE SURVEY

[1] Xiaolin Li, et al. [1] The experiments conducted have demonstrated the efficiency of TST-GAN in automating the conversion of case descriptions into legal texts, thereby alleviating the burden on legal practitioners and improving overall workflow. Comparative analyses with existing models, including RNN (LSTM), SeqGAN, and ED-GAN.

[2] V.Vaissnave et al. [2] The deep learning methods are in various tasks such as, legal data search, information retrieval, document classification, text translation, extract the relevant data, and conversational agents. Legal data search, retrieval information applications are implemented by building domain specific word embeddings.

internet content. The research addresses a critical problem in computer vision, proposing a CNN-based model for accurate and robust detection of explicit images

[3] Daniel Martin Katz et al. [3] The analysis conducted on the corpus reveals noteworthy trends, including a surge in the sophistication of methods employed in applied NLP in the legal context. The gradual alignment of Legal NLP with the methodological standards of general NLP and the broader scientific community, encompassing data availability and code reproducibility.

[4] Parth Shah [4] discussed by using sequential classification approach we can extract clause from legal contract by detecting paragraph boundaries from raw text data and we can also configure paragraph boundary according to application and domain by train model with an additional dataset.

III. PROBLEM STATEMENT

A person needs Legal Document to carry out certain legal procedures such as buying a property or an agreement. It can be difficult to prepare a document on their own or should consult a legal document preparer for this purpose and it is a time consuming task. Preparing a document on their own may lead to incorrect documentation and it may lose its purpose.

IV. PROBLEM DESCRIPTION

The current process of creating basic legal document can be time-consuming, error-prone, and require legal expertise even for straightforward situations. This can limit accessibility to legal services for individuals and small businesses who may not have the resources to hire a lawyer for every document.

V. OBJECTIVES

- Dataset Collection
- Preprocess the dataset
- Apply LSTM to the pre-processed dataset
- Generate Document

VI. METHODOLOGY

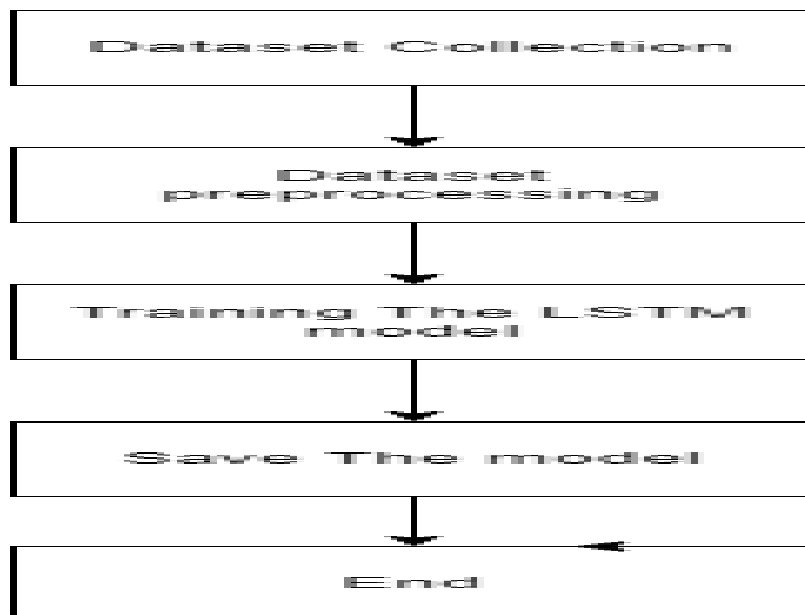


Fig 1 Methodology

1. **Data Collection:** The dataset consists of csv file that is used to train the model. The text dataset is used to train.

2. **Data Cleaning and Preprocessing:** Clean the extracted clauses by removing punctuation, stop words (common words with little meaning), and performing spelling normalization. Tokenization, splitting them into sequences of individual words.
3. **Data Splitting:** Dividing prepared data (tokens and corresponding summaries) into training, validation, and testing sets. The training set is used to train the model, the validation set helps fine-tune hyperparameters, and the testing set evaluates the final model performance.

LSTM Model Training

1. **Model Architecture:** Design an LSTM model architecture suitable for text summarization. This typically involves embedding layers, LSTM layers, and a dense output layer.
2. **Embedding Layer:** Convert each token in the training data into a numerical vector representation using an embedding layer. This allows the model to understand the relationships between words.
3. **LSTM Training:** Train the LSTM model on the prepared training data. The model learns to identify the most important parts of the clause and generate a shorter summary that retains the key points. Use the validation set to monitor the training process and adjust hyperparameters (learning rate, optimizer) for optimal performance.

Document Generation Process

1. **User Input:** The user selects the desired document type (Rent Agreement, Loan Agreement, Business Agreement). The system collects necessary user input like party names, dates, and other relevant details.
2. **Clause Selection:** Based on the chosen document type and user input, identify the relevant clauses that need to be included in the document.
3. **Clause Summarization:** For each identified clause, retrieve the full text from the CSV file. Feed the full clause text into the trained LSTM model. The model generates a concise summary of the clause.
4. **Document Assembly:** Populate a legal document template specific to the chosen agreement type with the user-provided information. Insert the LSTM-generated summaries of the relevant clauses into the appropriate sections of the template.
5. **User Review and Output:** Allow the user to review the generated document. Provide the user with the option to download the completed document as a PDF or printable format.

VII. SYSTEM DESIGN

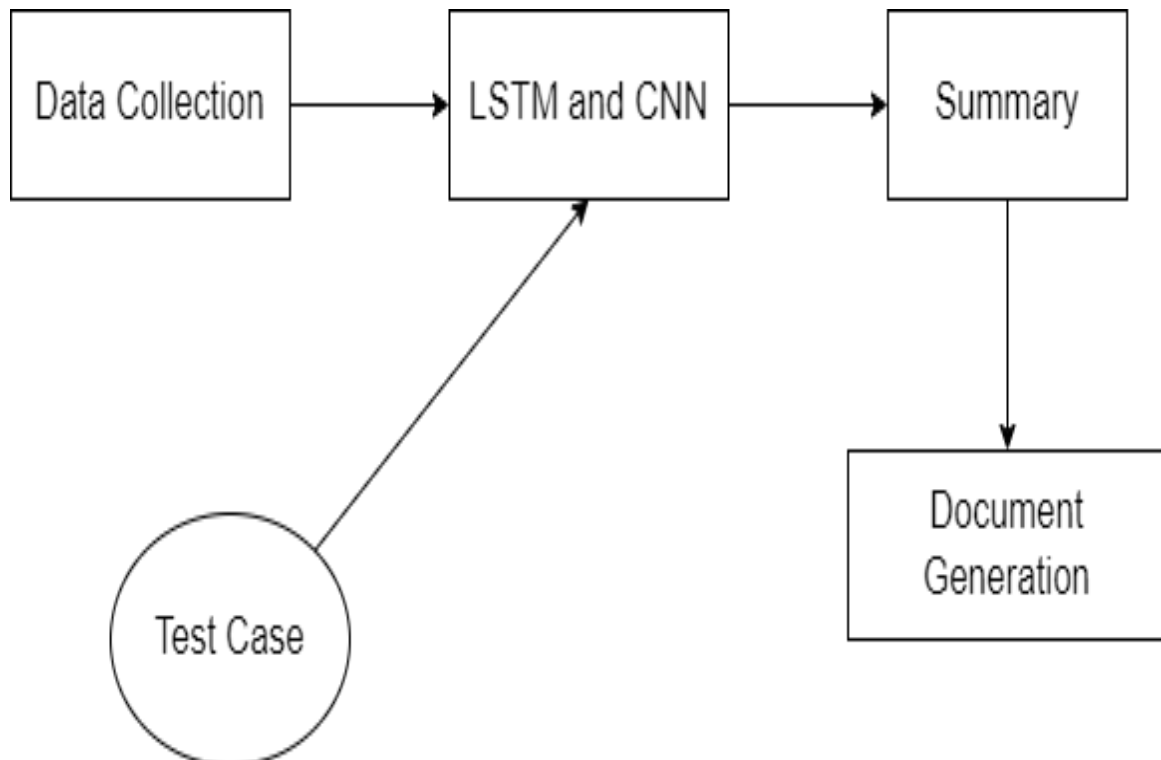


Fig 2: System Design

1. User Interface (UI):

- This is the front-end that interacts with the user.
- It allows users to:
- Enter information like party names and document type.
- Select the desired document type (rent, loan, business agreement).

2. User Input Processing:

- This module receives user input from the UI.
- It performs basic data validation to check for missing information or inconsistencies.
- The validated user input is then passed on to the Document Assembly module..

3. Condition Data Store:

- This is the database to store conditions (CSV) that holds the pre-defined legal conditions for each document type.
- Each row in the data store represents a specific condition.
- The data might include:
- Condition description (text)
- Applicable scenarios (flags or text)
- Legal citations (optional)

4. LSTM Model:

- This is the core AI component responsible for summarizing conditions.
- It's a pre-trained sequential LSTM model specifically designed for legal text processing (if available).
- The retrieved conditions are fed into the LSTM model one by one.

5. Summary Generation:

- The LSTM model generates a concise summary for each fed condition.
- This summary condenses the legal jargon into clear and understandable text.

6. Document Generation Engine:

- This module integrates the user input, document template, and LSTM-generated summaries.
- It populates the template with the relevant information, ensuring proper formatting and placement of summaries.
- The final legal document is generated.

7. Document Output:

- The final reviewed document is delivered to the user in a secure format (e.g., PDF)

VIII. SYSTEM IMPLEMENTATION AND RESULTS



The screenshot shows a web interface titled "Legal Documentor Generator". It features several input fields on the left and right sides. The fields are: "First Party Name" (text: test), "First Aadhar No" (text: 123456789010), "First Party Address" (text: test addr), "Second Party Name" (text: second), "Second Aadhar No" (text: 012345678900), "Second Party Address" (text: sse), and "Type of Agreement" (dropdown menu: Rent Agreement). Below the input fields is a "Predict" button.

Fig3 Giving input to the system

The above fig 5.2 shows the interface for user to input data and list of option to select the type of document to generate. The user is given three types of document to select. After selecting the type of document and clicking the predict button the system will generate a document

Rent Agreement

First Party Name : test
First Aadhar No : 123456789010
Second Party Name : second
Second Aadhar No : 012345678900

Agreement Details

('This rental agreement is entered into between test, residing at test addr , hereinafter referred to as the Landlord', 'and second , residing at sse , hereinafter referred to as the Tenant')

[this Section Provides Detailed Information About The Rented Property, Including Its Complete Address, Dimensions, And Any Specific Features Or Amenities Provided. Specifying The Details Of The Property Ensures Clarity Regarding The Premises Being Rented.with Section 108(l) Of The Transfer Of Property Act, 1882

This Clause Outlines The Duration Of The Tenancy, Including The Start Date And End Date Of The Lease Period. It May Also Include Provisions For Renewal Or Termination Of The Agreement.with Indian Easements Act, 1882

This Clause Specifies The Monthly Rent Amount Payable By The Tenant, The Due Date For Rent Payment, Accepted Modes Of Payment, And Any Penalties For Late Payments. Clearly Outlining The Rent Amount And

Fig 4: Document to be downloaded

IX. CONCLUSION

Promising tool for generating common legal documents like rent agreements, business agreements, and loan agreements. It leverages user input and a condition CSV file to customize documents with the help of an LSTM model. While this offers advantages in customization, scalability, and efficiency, remember that the tool excels as an assistant. For legal soundness, users should be aware of the importance of comprehensive conditions, robust model training, and seeking legal counsel for complex situations. This project lays a great foundation for future advancements in AI-powered legal tools.

REFERENCES

1. Xiaolin Li, Lei Huang , Yifan Zhou and Changcheng Shao "TST-GAN: A Legal Document Generation Model Based on Text Style Transfer "
2. Masha Medvedeva, Michel Vols and Martijn Wieling." Using machine learning to predict decisions of the European Court of Human Rights". Published on Springer on 26 June 2019 open access, Creative Commons Attribution 4.0 International License. link.springer.com/article/10.1007/s10506-019-09255-y.
3. Sandeep Bhupatiraju, Daniel L, Chen and Shareen Joshi." The Promise of Machine Learning for the Courts of India". The Promise of Machine Learning for the Courts of India. National Law School of India Review, 2021, 33 (2), pp.462-474. (hal-03629734)
4. Jane Mitchell, Simon Mitchell and Cliff Mitchell. "Machine learning for determining accurate outcomes in criminal trials". Law, Probability and Risk, Volume 19, Issue 1, March 2020, Pages 43–65, <https://doi.org/10.1093/lpr/mgaa003>
5. Riya Sil and Abhishek Roy. "A Novel Approach on Argument based Legal Prediction Model using Machine Learning". Proceedings of the International Conference on Smart Electronics and Communication (ICOSEC 2020). 10.1109/ICOSEC49089.2020

6. Huseyin Umutcan Ay, Alime Aysu Oner and Tolga Kaya. "A Machine Learning- Based Decision Support System Design for Restraining Orders in Turkey".2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC).10.1109/COMPSAC51774.2021.00226
7. V Gokul Pillai and Lekshmi R Chandran. "Verdict Prediction for Indian Courts Using Bag of Words and Convolutional Neural Network". Proceedings of the Third International Conference on Smart Systems and Inventive Technology (ICSSIT2020). 10.1109/ICSSIT48917.2020.9214278
8. Hui Wang, Tieke He, Zhipeng Zou, Siyuan Shen and Yu Li". Using Case Facts to Predict Accusation Based on Deep Learning".2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C). 10.1109/QRSC.2019.00038
9. J Haitham Hmoud Alshibly and Mohammad atwah Al- ma'aitah. "Artificial Intelligence in Law Enforcement". International Journal of Advanced Information Technology (IJAIT) Vol. 4, No. 4, August 2014. 10.1109/5254.653229
10. Kevin D, Ashley." Artificial Intelligence and Legal Analytics (New Tools for Law Practice in the Digital Age) Machine Learning with Legal Texts". University of Florida, Book - Artificial Intelligence and Legal Analytics (New Tools for Law Practice in the Digital Age) Machine Learning with Legal Texts
11. Nishant Jain and Gaurav Goel. "An Approach to Get Legal Assistance Using Artificial Intelligence". 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) Amity University, Noida, India. June 4-5, 2020 .10.1109/ICRITO48877.2020.9198029
12. Mariana Y. Noguti, Eduardo Vellasques, Luiz S. Oliveira. "Legal Document Classification: An Application to Law Area Prediction of Petitions to Public Prosecution Service".2020 International Joint Conference on Neural Networks (IJCNN). 10.1109/IJCNN48605.2020.9207211
13. Hemlata Sharma, Aakanksha. "Artificial Intelligence and Law: An Effective and Efficient Instrument".2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO).10.1109/ICRITO51393.2021.9596503
14. Vishu Madaan, Subratho Kumar Das, Prateek Agrawal, Charu Gupta and Dhruv Goel. "Fusion of ML models to Identify Sexual Harassment Cases".2021 International Conference on Computing Sciences (ICCS). 10.1109/ICCS54944.2021.00058'
15. Flora Amato, Antonino Mazzeo, Antonio Penta and Antonio Picariello. "Using NLP and Ontologies for Notary Document Management Systems".2008 19th International Workshop on Database and Expert Systems Applications. 10.1109/DEXA.2008.86
16. Suncana Roksandic, Nikola Protrka and Marc Engelhart. "Trustworthy Artificial Intelligence and its use by Law Enforcement Authorities: where do we stand?" 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO). 10.23919/MIPRO55190.2022.9803606
17. Akash Kumar, Aniket Verma, Gandhali Shinde, Yash Sukhdeve and Nidhi Lal. "Crime Prediction Using K-Nearest Neighboring Algorithm".2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE). 10.1109/ic-ETITE47903.2020.155
18. Saqeeb Abdullah, Farah Idid Nibir, Suraiya Salam, Akash Dey, Md Ashraful Alam and Md Tanzim Reza. "Intelligent Crime Investigation Assistance Using Machine Learning Classifiers on Crime and Victim Information".2020 23rd International Conference on Computer and Information Technology (ICCIT). 10.1109/ICCIT51783.2020.9392668
19. Marijn Schraagen, Floris Bex. "Extraction of Semantic Relations in Noisy User-Generated Law Enforcement Data".2019 IEEE 13th International Conference on Semantic Computing(ICSC).10.1109/ICOSC.2019.8665497
20. [17] Jishitha Kuppala, K. Kalyana Srinivas, P. Anudeep, R. Sravanth Kumar and P. A Harsha Vardhini. "Benefits of Artificial Intelligence in the Legal System and Law Enforcement".2022 International Mobile and Embedded Technology Conference (MECON). 10.1109/MECON53876.2022.9752352



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details