Natural Language Processing for Auditing Financial Contracts and Agreements

K Naresh[1], Singiri Tejasree[2]

[1]Assistant Professor, Department of MCA, Annamacharya Institute of Technology and Sciences, Tirupati, Andhra Pradesh, India.

[2]Postgraduate, Department of MCA, Annamacharya Institute of Technology and Sciences, Tirupati, Andhra Pradesh, India.

How to Cite this Article:

Tejasree, S. (2026). Natural Language Processing for Auditing Financial Contracts and Agreements. International Journal of Creative and Open Research in Engineering and Management, 02(04). https://doi.org/10.55041/ijcope.v2i4.075

https://doi.org/10.55041/ijcope.v2i4.075

Abstract

Complex legal language used in financial contracts and agreements necessitates close examination to guarantee correctness, compliance, and risk reduction. Conventional auditing of such papers is a labor-intensive, time-consuming procedure that frequently relies on subject matter specialists. Natural language processing (NLP) methods have become effective tools for automating document analysis due to the quick development of artificial intelligence.

This study describes an AI-powered system that uses machine learning and natural language processing to audit financial contracts and agreements. The technology analyzes textual data from contracts, extracts important information, and finds any irregularities, hazards, or discrepancies. To allow users to upload documents and get automated audit results, a Django-based web application is created.

The algorithm successfully scans contractual language and accurately emphasizes key terms, according to experimental data. The suggested method makes financial auditing procedures more dependable, more efficient, and requires less manual labor.

Keywords

Financial contracts, document analysis, machine learning, text mining, information extraction, Django, natural language processing, and artificial intelligence in finance.

I.Introduction

Bank agreements and deals, which regulate transactions, obligations, and legal responsibilities between parties, are crucial parts of corporate operations. Manual auditing is difficult because these documents are frequently long and complicated. Significant financial and legal dangers can result from mistakes or ignored clauses.

Traditionally, it takes a lot of effort and specialized knowledge to audit financial paperwork. However, there is a growing need for automated solutions that can help with document analysis due to the growing number of contracts in modern enterprises.

The ability to comprehend and analyze unstructured text input has been demonstrated by it (NLP). Financial documents can have significant information extracted, important clauses identified, and discrepancies found by using NLP approaches.

The goal of this research is to create an AI-powered system that will automate financial contract auditing. The system seeks to enhance the effectiveness, precision, and accessibility.

II. Dataset

Financial contracts and agreements, such as loan paperwork, legal agreements, and policy documents, make up the dataset. Important clauses like obligations, hazards, and financial terms are identified by preprocessing and labeling these texts.

III. Statement of the Problem

Financial contract auditing by hand takes a lot of time, is prone to mistakes, and necessitates specialized knowledge. An automated system that can effectively evaluate contracts, retrieve important data, and spot possible hazards or discrepancies is required.

IV. Objectives

This study's main goal is to create an intelligent system that can use natural language processing methods to analyze and audit financial contracts. The system is built to handle massive amounts of textual data and glean insightful information from intricate legal papers.

Accurately identifying crucial elements like obligations, payment terms, and risk-related declarations is another vital goal. The approach improves transparency and facilitates improved decision-making by identifying these components.

Additionally, the examination seeks to guarantee the effectiveness and dependability of the manner of auditing. The system's rapid analytical results cut down on the amount of time needed for manual document inspection.

To guarantee simple accessibility, the system is also included into a web-based interface. Without technological knowledge, users can upload contracts and obtain automated audit findings. Lastly, the system seeks to improve the general quality of financial auditing procedures by helping financial analysts, auditors, and businesses identify risks and guarantee compliance.

V. Methodology

Financial contract documentation are first gathered and prepared. To guarantee uniformity, this entails cleaning the text, eliminating superfluous symbols, and normalizing the material. The material is prepared for analysis by tokenization and stopword removal.

Feature extraction is carried out using techniques such as TF-IDF and word embeddings. These techniques transform textual data into numerical representations so that machine learning models can process it.

Key items including dates, monetary values, companies, and contractual phrases, are identified using Named Entity Recognition (NER). To find crucial passages in the papers, clause extraction techniques are also used.

After that, a machine learning model is built to categorize and evaluate the extracted features. The model recognizes possible dangers or inconsistencies by learning patterns connected to various clause types.

A web application built on Django incorporates the trained model, enabling users to upload documents and obtain automated audit reports.

VI. Implementation

Python and the Django framework for backend development are used in the system's implementation. Financial contracts in text or PDF format can be uploaded via the web interface. Libraries like NLTK, spaCy, and Transformers are used for natural language processing applications. Tokenization, entity recognition, and text preprocessing are used to extract pertinent data from the documents.

Text is transformed into numerical features using TF-IDF vectorization and embedding algorithms. For classification and analysis, machine learning models like Random Forest, Logistic Regression, and transformer-based models are employed.

VII. Result and Discussion

A wide range of financial deals and contracts, including legal documents with different terms pertaining to obligations, standard payments, and compliance requirements, were used to assess the proposed system. The evaluation's findings show that the system does a good job of recognizing important sentences and extracting important details from complicated textual input. The model exhibits a great capacity to analyze unstructured contract text and transform it into insightful information that is crucial for auditing. The system does very well at identifying important elements like dates, monetary amounts, obligations, and financial words, according to a thorough investigation. The model is able to precisely capture key elements of the contracts by utilizing natural language processing techniques including tokenization, named entity recognition, and feature extraction.

The system exhibits consistency in extracting and classifying clauses across various contract forms, in addition to entity recognition. Sections pertaining to payment terms, obligations, penalties, and legal conditions might be highlighted. This structured extraction increases productivity and efficiency by enabling users to rapidly comprehend the content of long texts without having to manually go over every detail. Nevertheless, a number of restrictions were noted throughout the assessment procedure. The intricacy of the legal terminology employed in financial contracts is one of the primary obstacles. Performance can also be impacted by differences in writing styles, formatting standards, and document structure amongst contracts. Clauses may occasionally be challenging to correctly categorize because of ambiguous wording or inadequate context.

Despite these difficulties, the approach drastically cuts down on the time and effort needed for manual audits. It reduces the possibility of human error and guarantees increased consistency by automating the extraction and analysis process. The findings unequivocally show that NLP-based methods can be vital to the modernization of financial document audits. These systems have the potential to develop into extremely dependable instruments for financial analysts, auditors, and businesses managing substantial amounts of contractual data with additional enhancements.

VIII. Conclusion

This study describes an AI-powered system that uses cutting-edge natural language processing methods to audit financial contracts and agreements. The suggested method focuses on automating the study of complicated legal papers, which typically calls for a large amount of manual labor, time, and experience. The system can process massive amounts of unstructured textual material and turn it into useful, organized information by utilizing NLP techniques. This lowers the possibility of human error in the auditing process while simultaneously increasing efficiency.

The study's findings show how well the system extracts important data, including dates, financial terms, obligations, and crucial phrases. Furthermore, the model demonstrates a great ability to detect possible dangers, irregularities, and contradictions in contracts.

Additionally, the system can be expanded by adding sophisticated features like compliance checking, which would confirm that the document complies with legal requirements, and risk scoring, which would measure the degree of risk connected with a contract. The system's utility in practical applications could be further increased by adding capabilities like automated summarization, clause comparison, and real-time notifications.

All things considered, the suggested AI-based method shows great promise for revolutionizing conventional financial auditing procedures. The technology promotes better decision-making, increased openness, and increased operational efficiency by making contract analysis quicker, more accurate, and scalable. It is an important step toward the financial and legal industries' adoption of intelligent technologies.

References

[1] BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, NAACL, 2019.

[2] "Efficient Estimation of Word Representations in Vector Space," T. Mikolov et al., ICLR, 2013.
[3] Natural Language Processing with Python, S. Bird, E. Klein, and E. Loper, O'Reilly Media, 2009.
[4] Neural Network Approaches for Natural Language Processing, Y. Goldberg, 2017.
[5] Speech and Language Processing, D. Jurafsky and J. H. Martin, 2021.