# Enhancing Legal Guidance by Utilizing Natural Language Processing-Based Document Embeddings

MSc Research Project
MSc Artificial Intelligence


Tanmay Laxmikant Mukim
Student ID: x22127933

School of Computing
National College of Ireland


Supervisor: Dr. Muslim Jamal Syed

**National College of Ireland**
**Project Submission Sheet**
**School of Computing**

| | |
|---|---|
| **Student Name:** | Tanmay Laxmikant Mukim |
| **Student ID:** | x22127933 |
| **Programme:** | MSc Artificial Intelligence |
| **Year:** | 2023 |
| **Module:** | MSc Research Project |
| **Supervisor:** | Dr. Muslim Jamal Syed |
| **Submission Due Date:** | 14/12/2023 |
| **Project Title:** | Enhancing Legal Guidance by Utilizing Natural Language Processing-Based Document Embeddings |
| **Word Count:** | 9764 |
| **Page Count:** | 25 |

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

**ALL** internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

| | |
|---|---|
| **Signature:** | |
| **Date:** | 29th January 2024 |

## PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

| | |
|---|---|
| Attach a completed copy of this sheet to each project (including multiple copies). | ☐ |
| **Attach a Moodle submission receipt of the online project submission**, to each project (including multiple copies). | ☐ |
| **You must ensure that you retain a HARD COPY of the project**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | ☐ |

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

| **Office Use Only** | |
|---|---|
| Signature: | |
| Date: | |
| Penalty Applied (if applicable): | |

# Enhancing Legal Guidance by Utilizing Natural Language Processing-Based Document Embeddings

Tanmay Laxmikant Mukim
x22127933

**Abstract**

The creation of a state-of-the-art recommendation framework for legal case study recommendations is examined in this corresponding work, which tends to utilize the four various types of Natural Language Processing (NLP) methods. Two different datasets, which mainly represent the legal citation data, are utilized in the implementation as well as the assessment of each of the models, which specifically comprise Word2Vec in combination with TF-IDF with Bidirectional Encoder Representations from Transformers (BERT), ALBERT together with DEBERTa. The specified main goal is to put forth a complex framework that can make recommendations for specific legal citations by utilizing the special advantages of these Natural Language Processing (NLP) methods accordingly. This specific research explores the effectiveness and comparative analysis of various Natural Language Processing (NLP) architectures, driven by the necessity for cutting-edge innovation in legal research. The implemented models are taught to execute the extraction of connotational significance from legal text together with understanding its subtleties by authorizing precisely with pertinent legal citation recommendations. A thorough evaluation of each specified Natural Language Processing (NLP) method's capability to provide precise legal citations is one of the main conclusions this research draws. Additionally, to provide clarity on the evenness of the variations in the recommendations that are made because of the implementation of these corresponding four models, the research compares how comparable the implemented Natural Language Processing (NLP) methods are. In conclusion, by offering a solid foundation for the improvisation of a legal recommendation framework via the integration of several Natural Language Processing (NLP) procedures, this study advances different informative aspects in the legal sector.
***Keywords***— Legal Recommendation Framework, Natural Language Processing, Comparative Survey, and Textual Similarity

## 1 Introduction

It is observed that finding pertinent citations in complicated literature can be especially challenging when conducting legal research, which frequently requires sifting through enormous amounts of data. More advanced techniques are desperately needed because traditional procedures are not up to the task of precisely together with quickly finding specified relevant legal citations. The goal of this particular study is to revolutionize the efficiency as well as the correctness of legal citation recommendations by addressing this demand by exploring the section on Natural Language Processing (NLP), which mainly includes Word2Vec in combination with TF-IDF with Bidirectional Encoder Representations from Transformers (BERT), ALBERT together with DEBERTa, respectively.

This corresponding study was specifically motivated by the ongoing limitations of concurrent works, as noted Zhong et al. (2020). These particular drawbacks result from older approaches' incapacity to understand the complex phrasing as well as contextual nuances that are observed as common in legal documentation. This factor makes it necessary to investigate sophisticated computational models specifically designed for the comprehension of the legal sector.

What primary effects does Natural Language Processing (NLP)-based document embedding have on a legal recommendation framework, together with how does the improvisation of the accuracy as well as the usefulness of legal advice This is the related research question that serves as the foundation for the corresponding relevant study. To address this specified question, the correspondent research entailed gathering sizable legal datasets together, laboriously preparing the text, creating document embeddings utilizing Natural Language Processing (NLP) techniques, and crafting an expert recommendation regarding the machine learning framework accordingly.

This specified study makes a significant contribution to the scientific literature by examining the effectiveness of several Natural Language Processing (NLP) architectures on legal case datasets. Additionally, key insights into the suitability as well as optimization of Natural Language Processing (NLP) approaches for improvising legal information retrieval frameworks are provided by the empirical data together with the comparative analysis offered in this research.

At last, this research study is structured as follows for the remainder: A specified summary of the research on spelling as well as grammar error detection is provided in the corresponding Section 2 as Related Work. Moreover, the research methodology is described in Section 3. Similarly, sections 5 as well as 6 deal with implementation together with evaluation, respectively, whereas Section 4 offers the design specification. Finally, the Conclusion with Future Work is covered in Section 7, as mentioned accordingly.

# 2 Literature Review

The construction, as well as the adjustment of legal suggestion frameworks, are the main topics of discussion in this corresponding section's exploration of recent research. Prior research has explored the development with modification of legal advice frameworks through the application of machine learning as well as natural language processing (NLP) procedures. These corresponding methodologies address problems such as topic modeling together with sentiment analysis, and the classification of legal citations, respectively. Enhancing precision with the relevance of legal advice is the primary aim, taking into account the section's potential as well as its constraints.

## 2.1 Prior research on frameworks for legal recommendations

Recognizing the corresponding growing problem of information overload in the digital age, the respective study Gan and Jiang (2013) focuses on collaborative filtering procedures as well as recommendation frameworks. Popularity bias together with accuracy-diversity trade-offs are two different queries that these methodologies struggle with, despite their superior capability to supply tailored recommendations that are based on customer preferences. Prior work significantly offers remedies for these current problems, including new protocols as well as item weighting techniques and modifications to similarity scoring. None, nevertheless, succeed in striking the ideal harmony between diversity with accuracy. The results of this investigation highlight the urgent necessity for novel strategies that balance precision as well as diversity in a way that effectively addresses popularity bias. To provide more objective as well as worthwhile customized recommendations in contemporary recommender frameworks, it emphasizes the significance of novel procedures like the Power Law modifications of User Similarities (PLUS) suggested in this study accordingly. Moreover, the corresponding study Cui et al. (2023) utilizes a variety of languages with court case datasets to demonstrate advancements in natural language processing procedures for legal judgment prediction (LJP). Although these assessments show progress, they frequently lack a unified strategy, which appropriately makes it difficult to do attribute analysis as well as evaluate methodologies and fully address overarching queries accordingly. To develop LJP approaches in the legal sector and get a deeper understanding of them, a specific, cohesive strategy is the main requirement. To further enhance legal judgment prediction, this corresponding paper highlights the necessity for a comprehensive analysis that synthesizes several qualities as well as addresses unresolved questions. Furthermore, pre-trained language models (PLMs) are increasingly being utilized in the sector of legal artificial intelligence (LegalAI) to do proper management of lengthy legal texts efficiently.

Even though these corresponding studies Xiao et al. (2021) show impressive types of achievement in modifying pre-trained language models for different legal procedures, for instance, retrieving cases, predicting judgments, understanding legal texts, and responding to different types of questions from clients, the limitations of traditional pre-trained language models prevent these aspects from processing large amounts of legal data efficiently, particularly when processing documents in languages such as Chinese. These corresponding analyses demonstrate how inadequate concurrent solutions are to fully address the complexities of legal texts as well as emphasize the urgent necessity for new kinds of approaches, such as the Lawformer architecture that is presented in this respective study, to do the management of the processing of long legal data in the sector of LegalAI in a specific efficient manner. Additionally, the impact of artificial intelligence on different disciplines, including law, is examined in the study Sil et al. (2019), which properly highlights the potential of expert frameworks to improve productivity as well as simplify procedures. These efforts demonstrate how artificial intelligence may be utilized to understand semantic legal information with predict legal outcomes, which could lead to increased judicial efficiency even in the face of budgetary constraints. The subtleties as well as complexity present in particular legal circumstances, however, are frequently difficult for them to acknowledge together with anticipate, especially when it comes to legal semantics. As a result, a significant gap in the efficacy of concurrent frameworks to provide consistently accurate legal forecasts still exists. To fulfill this gap, new approaches that make utilization of AI capabilities as well as navigate the complexities of legal context and language are observed as needed.

Bidirectional encoder representations from transformers (BERT) are optimized for domain-specific corpora such as scientific as well as biomedical literature Elwany et al. (2019), demonstrating the efficacy of this corresponding approach. But despite its respective potential benefits in legal natural language processing, it is observed that there isn't much research on optimizing BERT for legal data. Secret rules make it difficult to obtain permission for large legal data collections, which puts limitations on research on contract analysis together with a thorough investigation of BERT's benefits when reviewing commercial contracts. The full potential as well as appropriateness of optimizing language frameworks for legal NLP applications are therefore still poorly acknowledged. This emphasizes how vital it is to conduct a thorough study to ascertain the advantages in addition to the effectiveness of this strategy, especially when it comes to analyzing commercial contracts in the legal environment accordingly. In addition, deep neural networks are increasingly being utilized in the legal industry for text classification as well as information extraction, according to the literature Chalkidis and Kampas (2019). These experiments show how well deep learning generates semantic elements that are observed to be essential for deciphering legal data. While some research focuses on legal word embeddings trained on a variety of legal corpora utilizing frameworks such as Word2Vec, it does not provide thorough coverage across different types of legal situations with jurisdictions. One significant restriction is the lack of a comprehensive strategy that takes into account the subtleties of foreign legal frameworks. Broad, flexible methodologies covering many legal topics, as well as languages and jurisdictions, are essential for the effective application of deep learning in legal analytics.

It is observed that there is an increasing interest in employing several types of algorithms, such as GRUs together with RFs as well as SVMs and BiLSTMs, to predict legal outcomes from text descriptions, according to the literature Mumcuoğlu et al. (2021) in legal natural language processing. Predicting legal verdicts in well-established legal frameworks has been a promising area of research in the past, but certain legal frameworks, most notably the Republic of Turkey, have not had enough investigation concerning observations. This corresponding research gap demonstrates how inadequate the predictive legal analytics available now are for wider utilization, such as in Turkey. To emphasize the significance as well as the applicability of predicting legal court decisions in that particular legal environment, it is imperative to close this gap by implementing prediction frameworks that are based on natural language processing in less-studied legal frameworks, such as Turkey's. Moreover, in order to facilitate the effective identification of relevant decisions for legal argumentation, the literature Dhanani et al. (2021) highlights the appropriate necessity of a legal document recommendation framework in the Indian legal framework. Previous approaches, which generally utilize Doc2Vec on processed judgment corpora, have corresponding queries with noise as well as scalability that impact the precision of recommendations. This work suggests a novel methodology for processing as well as corpus refinement utilizing the Generalized English with Indian Legal Dictionary, which greatly increases Doc2Vec's space as well as time efficiency, respectively. The effectiveness of the suggested LDRS is demonstrated by the experimental findings, which significantly indicate improved accuracy in the F1-Score and MCC-Score measures accordingly. The uniqueness is in applying the new legal dictionary as well as domain-specific processing to improve the precision of judgment recommendations, which could benefit online legal search engines.

A corresponding framework for successful legal document recommendations is the prime necessity, according to the study Dhanani et al. (2022), to help lawyers prepare their cases for trial. By adding legal domain information to Word2Vec embeddings, it suggests the pre-learned word embedding-based legal document recommendation paradigm. Utilizing computer clusters as well as technologies such as MapReduce with Spark, it presents a distributed version to overcome scalability concerns with large legal datasets. According to corresponding test results, the non-distributed framework outperforms the traditional Doc2Vec-based methodology in terms of efficiency as well as accuracy metrics. This work addresses scalability queries with distributed computing approaches in addition to incorporating domain-specific expertise into embedding processes for legal data analysis. Additionally, a comprehensive examination of the efficacy of transformer-based frameworks in automating the processing procedure of legal documents is absent from the study Nguyen et al. (2022), according to the observations. Though it recognizes the constraints of data as well as the intricacy of legal data, it ignores crucial adjustments that are needed to achieve the best output results and does not provide empirical proof of the effectiveness of Transformer methodologies. To close this gap as well as fully investigate the potential of Transformer frameworks in legal text processing, this research primarily aims to solve the deficiencies of previous studies by incorporating state-of-the-art approaches for automated legal data processing.

Research Tripathy et al. (2021) significantly examines pre-trained frameworks as well as embeddings in natural language processing by providing information on their utilizations as well as benefits and limitations in a range of tasks. This research frequently lacks a comprehensive evaluation of various pre-trained frameworks together with embedding procedures over a broad spectrum of NLP applications as well as datasets and architectures. While some studies concentrate on specific tasks or comparative assessments, they neglect to take practical efficacy as well as wider application into account. This gap results from the lack of an all-encompassing framework that assesses embeddings and pre-trained frameworks thoroughly according to a range of NLP tasks, going beyond mere comparisons. To fulfill this gap, this study offers a thorough as well as methodical analysis that evaluates the performance of several embedding procedures as well as pre-trained frameworks on a wide variety of NLP tasks with different datasets and architectures. Also, it is observed that, though there hasn't been much research done on BERT's adaptation to specialized domains, such as the legal domain, the literature Chalkidis et al. (2020) emphasizes BERT's amazing performance in a variety of different NLP tasks. Utilizing a variety of datasets, this corresponding paper assesses several methodologies for utilizing BERT models in legal tasks. It finds drawbacks in mindlessly adhering to prior pre-training as well as fine-tuning guidelines, with an emphasis on the necessity of a methodical study of approaches when utilizing BERT in specialized sectors. Adapting BERT with additional pre-training on domain-specific data and training BERT from scratch on domain-specific data, in addition to utilizing the original BERT are the three suggested processes in this corresponding study. The work also presents LEGAL-BERT, a family of BERT methodologies designed to help legal NLP research as well as applications in computational law together with legal technology, and suggests a wider hyper-parameter search space during fine-tuning for implemented downstream tasks.

The research De Martino et al. (2022) emphasizes how difficult it is to navigate the ever-changing legal landscape without sufficient technological assistance. It is observed that there have been several types of proposals for AI-based solutions to assist with legal activities. This corresponding work presents PRILJ, a revolutionary procedure designed to help legal experts prepare documents by finding paragraph regularities in court rulings accordingly. PRILJ significantly represents documents as well as paragraphs by utilizing an embedding-based procedure, which makes it possible to retrieve related paragraphs quickly and accurately. PRILJ has been experimentally evaluated by utilizing the EUR-Lex dataset as well as demonstrating its robustness against potential data noise in addition to confirming its efficacy in modeling a variety of legal themes as well as capturing semantic information appropriately. Moreover, it is observed that, due to complicated text patterns that cause false positives in concurrent frameworks, extracting legal entities from legal documents, particularly parties in contracts, remains a challenge. This work Sivapiran et al. (2023) presents a precise and significant procedure for extracting parties from contract agreements by utilizing contextual span representations. A dataset of one thousand annotated contracts is utilized in the corresponding study, which also improves a SQuAD 2.0 question-answering module by inserting changes to the normalization as well as dropout together with the encoder and activation layers, respectively. When compared to the state-of-the-art, baseline experiments indicate that the refined model performs better. The hybrid strategy, which combines dropout layers in addition to normalization and 24 encoder layers, produces the greatest and most accurate results, incrementing the exact match score to 0.942 by 6.2 percent, respectively.

It is observed that, due to categorization bias in earlier scoring-based procedures, answering legal questions presents difficulties, particularly in judicial exams. The relationships between reference volumes as well as questions and multiple-choice responses are frequently not successfully utilized by these frameworks. A BERT and attention-based paradigm for court examinations is proposed in this paper Chen et al. (2023) as a solution to this query. BERT is utilized for encoding questions in addition to multiple-choice answers and retrieved articles. The BM25 algorithm is significantly utilized to extract pertinent articles from legal books. Articles, as well as questions and replies, are deeply correlated with one another thanks to the attention-focused process implemented. The model's efficacy is highlighted by experimental results that depict improved precision above baselines. Furthermore, the corresponding model came in third in the judicial examination assignment of the Challenge of AI in Law 2022, demonstrating its strong performance accordingly.

# 3 Methodology

The accompanying section describes the corresponding structure as well as the methodology utilized for the Legal Recommendation Framework with respect to the Natural Language Processing-Biased Document Embeddings study. The following section's specific major objectives are to explain the purpose-driven recommendation process in depth, together with a detailed description of the datasets that were utilized, and go into detail about the various types of approaches that were utilized throughout this study.

The corresponding figure, 1 directs the implemented research process in this specified study. The research process concerning this research study varies in specific stages. The research study starts for the initial stage of the initial review of the literature, in which the initial review of the literature related to the corresponding research topic is conducted. After that, according to the next stage of the research process, the description of the research question and the preliminary suggestions are executed, as shown in the research process figure. Similarly, the next stage depicts the process of a thorough examination of the literature and problem definition. Together, the research process continues after the execution of the previous stage towards the stage of designing the frameworks and experiments, respectively. After the execution of these specified stages, the research process moves toward the Frameworks and the Conduct of Experiments stage, where the implementation and execution of experiments with respect to the framework are conducted. As shown in the research process figure, this specific stage is correlated with the previous two stages of examination of the literature and problem definition, as well as designing the frameworks and experiments accordingly. At the final stage, the research

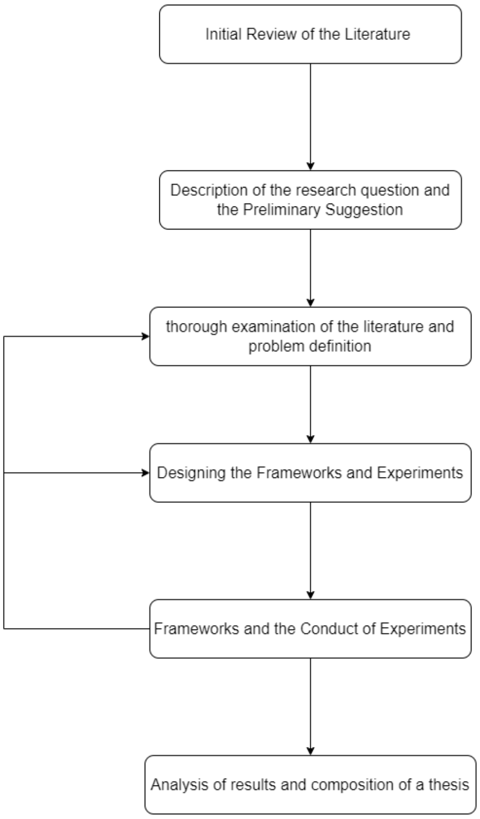


Figure 1: Research Process

process shifts to the analysis of results and composition of a thesis, where the complete analysis of results is conducted and, based on that, the composition of the thesis is implemented.

## 3.1 Proposed Methodology

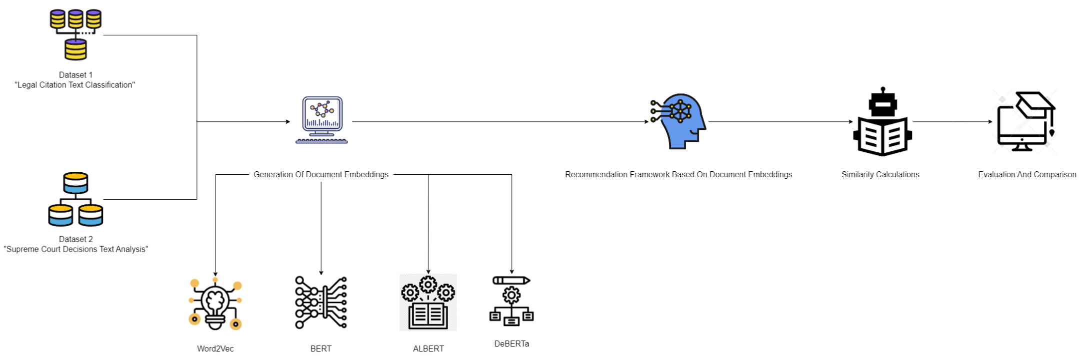


Figure 2: Proposed Methodology

The above figure, 2 concerning the research process, represents the proposed methodology of the corresponding research study. Obtaining two separate legal case datasets—the "Legal Citation Dataset" together with the "Supreme Court Decisions Text Analysis"—is the first step in the procedure. These respective datasets are designated for training as well as testing the recommendation framework, which is briefly described in the dataset section. To ensure consistency with normalization throughout the textual data, these datasets go through a thorough pre-processing step that cleans and standardizes the legal textual documents.

In the next step, four natural language processing procedures are specifically deployed in this research, as mentioned above, such as Word2Vec in combination with TF-IDF as well as Bidirectional Encoder Representations from Transformers (BERT) and the ALBERT textual processing model with the DeBERTa textual processing model, respectively. The Word2Vec, in combination with the TF-IDF model, first generates document embeddings for the datasets' legal textual citations accordingly. After this initial step, Bidirectional Encoder Representations from Transformers (BERT), together with the ALBERT textual processing model and the DeBERTa textual processing model, are fine-tuned on the implemented legal datasets to provide legal text-specific document embeddings.

After that, a specified framework for a recommendation framework is created by utilizing the generated document embeddings with the help of natural language processing procedures. Each respective natural language processing model's design, as well as implementation, are covered in the corresponding study, allowing the framework to provide legal textual citations that are based on the implemented legal textual citation datasets.

At the final stage, comparability measures between the respective suggestions made by each natural language processing procedure are calculated concerning the part of the corresponding research. To measure the specified similarity of the suggested legal textual citations, it makes the utilization of appropriate metrics, specifically cosine similarity, accordingly. Based on the specified recommendations that are generated for document embeddings, which are generated by each natural language processing procedure, each particular recommendation framework's performance is evaluated. The recommendations generated by Word2Vec in combination with TF-IDF, as well as Bidirectional Encoder Representations from Transformers (BERT) and the ALBERT textual processing model with the DeBERTa textual processing model, are compared as well as contrasted using a comparative analysis.

## 3.2 Dataset Description

As mentioned in the above section, the "Supreme Court Decisions Text Analysis" and "Legal Citation Dataset" datasets from Kaggle repositories are used as the appropriate data sources for the primary objectives of this specific study. For this corresponding study, most of the legal textual citations in the associated dataset are used. This corpus of information is essential to the diagnosis of diverse legal textual citation types across national boundaries as well as to the recommendation framework.

### 3.2.1 Dataset 1. Shivam Bansal, 2018. Legal Citation Text Classification, Kaggle, Version 1.

It is observed that, for legal citation analysis concerning this corresponding research, the "Legal Citation Text Classification" dataset by Shivam Bansal (2018) is a helpful resource. Because the textual data is tailored, particularly for legal citations, it becomes convenient to utilize it for the development and evaluation of natural language processing procedures as the basic building block in the architecture of the legal textual citation recommendation framework. Utilizing this dataset can help in the improvement of the procedures for performing legal research as well as learning more about citation trends and advanced legal information retrieval systems, respectively.

The respective Federal Court of Australia (FCA) has Australian court cases in the form of citations included in this corresponding dataset. The dataset contains all instances of Australian court cases from the following years: 2006, 2007, 2008, and 2009, respectively. Citation catchphrases, together with citation sentences as well as citation classes, are recorded for every textual citation document. Citation classes, which direct how the cases mentioned in this case were handled, are indicated in the specified form of textual representation.

#### 3.2.2 Dataset 2. Washington University in St. Louis (2021) Supreme Court Decisions: Text Analysis Kaggle. Version 1.

It is observed that, with the help of Washington University in St. Louis' "Supreme Court Decisions Text Analysis" dataset, scholars and data scientists can conduct the examination as well as draw conclusions from the textual data of Supreme Court rulings, respectively. With Version 1, it has become more comfortable to do tasks related to text analysis, spot trends, and support natural language processing and legal analytics research for the development as well as evaluation of the legal textual citation recommendation framework. This dataset gives authorization to do an investigation into the linguistic content of Supreme Court opinions to gain additional insight into how they could be applied to legal research and reasoning.

Moreover, when it comes to information about the United States Supreme Court, the Supreme Court database is proven as an authoritative resource for scholars as well as student journalists, and the general public. The corresponding database includes about two hundred variables pertaining to every case the Court has ruled between the respective years 1946 and 2015. Additionally, examples are the parties involved in the lawsuit, the court whose ruling the Supreme Court evaluated, the laws taken into consideration in the case, and the Justices' votes accordingly.

### 3.3 Methodology

This respective sub-section represents the details of different methodologies implemented in this study.

Specifically, to generate embeddings for legal citations within the datasets created in this study by utilizing the Word2Vec with tfidf features Arora et al. (2021). Utilizing the contextual implementation of words within the corpus as a basis for representation, the Word2Vec methodology helps in the generation of word embeddings. Simultaneously, the TF-IDF feature extraction procedure utilizes the frequency of terms within documents as well as their prevalence over the entire data collection to evaluate the relevance of each respective word. When combined, these corresponding procedures capture the contextual relevance as well as the importance of words in legal case collection, making it possible to generate meaningful numerical representations, which are also known as embeddings for the framework.

Additionally, in this work, the explicit utilization of legal datasets concerning a fine-tuning model is based on transformers, Bidirectional Encoder Representations from Transformers (BERT) Devlin et al. (2018). According to BERT's methodology, document embeddings are generated by thoroughly comprehending the word context in legal citation texts in a two-way mode. BERT significantly captures complex as well as subtle relationships within the corresponding legal citation text by undertaking the account words that immediately and obliquely surround a given word as well as phrase, respectively. This method enables the recommendation framework to be capable in every manner, together with contextually meaningful document embeddings.

Moreover, a specific efficient variant of BERT, known as ALBERT (A Lite BERT) Lan et al. (2019), is achieved by lowering the corresponding number of specified parameters without sacrificing the recommendation engine's performance. In this work, legal datasets are utilized to specifically train the ALBERT model to significantly generate document embeddings that effectively reflect the syntax as well as the meaning of legal citation language. With a prime focus on maintaining the quality of the document embeddings, its design strives for a more efficient proportion of model architecture, which accordingly makes it especially utilizable for legal text analysis in the recommendation framework.

Together with an extension of BERT known as Decoding-Enhanced BERT (DeBERTa) He et al. (2020), it primarily aims to improvise the decoding mechanism for improvised text comprehension. DeBERTa is specifically trained in utilizing legal datasets for the generation of document embeddings that are optimal for legal context interpretation. By optimizing the decoding procedure, DeBERTa significantly gains the capability to extract finer points as well as subtleties that are introduced in legal citation text, which is observed as appropriate for generating document embeddings specified to legal citation text understanding in the implementation of the recommendation framework.

Calculation of the absolute degree of similarity between each suggested legal citation textual generated significantly by Word2Vec by utilizing TF-IDF together with Bidirectional Encoder Representations from Transformers (BERT) as well as ALBERT and Decoding-Enhanced BERT (DeBERTa) is the corresponding next stage.

For the determination of the similarity of the sets of suggested legal citations generated by each natural language processing model to one another, a suitable similarity matrix, such as a cosine similarity matrix, is utilized. The respective cosine of the angle formed by the corresponding recommendation vectors is specifically computed by this metric, which directs how similar as well as close their directions are as compared to one another in a high-dimensional space.

Similarly to evaluating as well as comparing the significant effectiveness of the recommendation framework, similarity scores derived from these computations are utilized as a corresponding evaluation meter. To assess the efficacy and dependability of each implemented natural language processing model's recommendations, higher similarity scores accordingly indicate greater agreement as well as likeness between each recommendation offered by various implemented natural language methodologies.

# 4 Design Specifications

The following section describes the corresponding structure of the design specifications utilized for the Legal Recommendation Framework concerning the Natural Language Processing-Biased Document Embeddings study.

## 4.1 Implementation of Natural Language Processing Methods

Utilizing the Word2Vec natural language model in subjective conjunction with the TF-IDF (Term Frequency-Inverse Document Frequency) approach, the word embeddings for legal textual citations in both utilized datasets are generated. This implemented methodology tends to focus on the assessment of the generation of recommendations by producing word embeddings as well as evaluating the significance of terms in legal textual documents.

The utilization of Bidirectional Encoder Representations from Transformers (BERT), An approach based on transformers that have been significantly refined on utilized legal datasets allows for the bi-directional comprehension of contextual information accordingly. This corresponding methodology helps in the creation of document embeddings tailored to legal textual language comprehension while capturing the complex types of linkages observed in legal textual documents.

ALBERT, sometimes known as "A Lite BERT," is a specified performance-preserving parameter reduction procedure that is similar to BERT but tuned for the efficiency of the specific procedure. To develop document embeddings that are specifically designed to read the legal textual format more effectively, the natural language processing model ALBERT is trained on the utilized legal textual citation data collections. Whereas, expanding BERT's decoding mechanism for the improvement of textual comprehension is known as Decoding-Enhanced BERT (DeBERTa). Utilizing legal data as a training module, DeBERTa can provide significant document embeddings that are appropriately suited for comprehending legal complexities accordingly.

## 4.2 Creation of a framework for recommendations

The specified procedure for the development of a recommendation engine framework entails utilizing the corresponding document embeddings generated by the natural language methodologies that have been put into practice. The respective architecture of the framework will be optimized to make effective utilization of the document embeddings generated by the Word2Vec model together with Bidirectional Encoder Representations from Transformers (BERT) as well as ALBERT and DEBERTa. These document embeddings will be incorporated into the framework to help with recommendations, guaranteeing seamless interchange as well as efficient implementation of the unique semantic representations produced by each natural language model.

## 4.3 Computation of Similarity for Evolution

This stage entails calculating the specified similarity scores among the respective suggestions that are generated by every natural language model that has been put into practice. The prime goal is to do an objective assessment of how similar as well as comparable the recommendations made by Word2Vec using TF-IDF features together with Bidirectional Encoder Representations from Transformers (BERT) as well as ALBERT and DeBERTa are, utilizing appropriate specified similarity metrics such as cosine similarity. By calculating the mean similarity between the suggested legal textual citations, the corresponding similarity calculation procedure authorizes a comparative evaluation of the recommendations that are made by these natural language models. By making an appropriate assessment of the comprehension of the coherence as well as the consistency of the suggestions made by various implemented natural language models, this respective evaluation metric adds to the overall assessment of the effectiveness of the recommendation framework accordingly.

# 5 Implementation

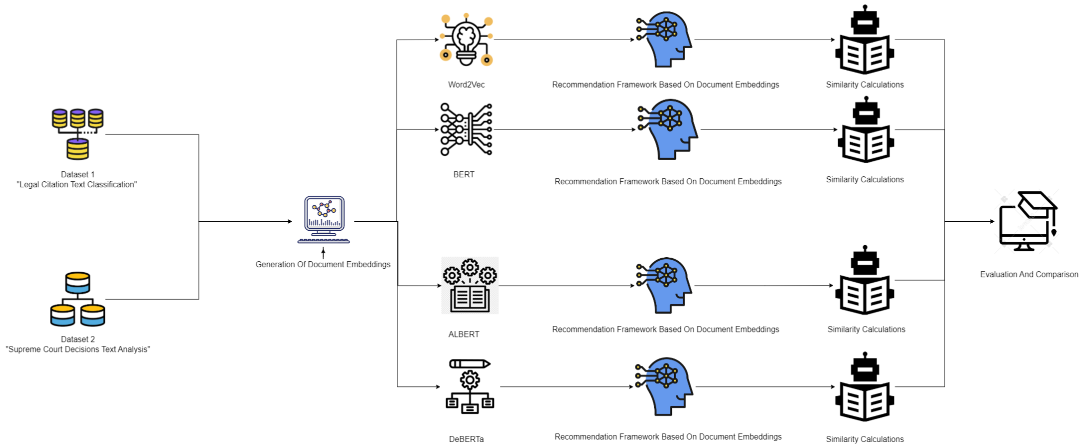


Figure 3: Implementation

The associated research study's implementation is depicted in the upper image 3. Utilizing corresponding pertinent libraries such as Gensim for Word2Vec and similarly Hugging Face Transformers for Bidirectional Encoder Representations from Transformers (BERT), as well as ALBERT and Decoding-Enhanced BERT (DeBERTa), the specified implementation entails merging four different natural language models such as Word2Vec with TF-IDF features together with Bidirectional Encoder Representations from Transformers (BERT), as well as ALBERT and Decoding-Enhanced BERT (DeBERTa). When utilized in conjunction with TF-IDF features, the Word2Vec model will assess word importance as well as identify semantic links in legal textual citation data. This will be trained for the generation of embeddings that are particularly suited to the subtleties of legal textual language. The legal textual citation data collections will be utilized to refine the pre-trained Bidirectional Encoder Representations from Transformers (BERT) as well as ALBERT and Decoding-Enhanced BERT (DeBERTa) natural language models, enabling them to better understand and generate document embeddings that are appropriate for the intricacies of legal textual language. In the next step, the legal textual citation data collection will be utilized by the Word2Vec model during training to generate the document embeddings that take TF-IDF weighting into account. Meanwhile, Bidirectional Encoder Representations from Transformers (BERT), as well as ALBERT and Decoding-Enhanced BERT (DeBERTa), will be fine-tuned to comprehend the legal domain and produce document embeddings that capture complex legal semantics. The prime goal of these corresponding implementations is to provide each natural language model with distinct words as well as document embedding, together with the conformation of a specified thorough comprehension of the legal textual context as well as linguistic subtleties, which are prerequisites for the creation of a successful legal textual citation recommendation framework.

The procedure of developing a recommendation engine entails constructing a framework that makes use of the document embeddings that are produced by each natural language model, in this case, Word2Vec with TF-IDF features together with Bidirectional Encoder Representations from Transformers (BERT), as well as ALBERT and Decoding-Enhanced BERT (DeBERTa). The strengths as well as subtleties that each natural language model's document embeddings capture are intended to be utilized by this respective framework. To propose legal textual citations from the data collection, algorithms will be developed that take proper advantage of the unique strengths of each natural language model. The primary aim is to utilize the distinct contextual comprehension as well as representations that are acquired by these natural language models to furnish precise with efficient suggestions for legal textual citations.

The experiment setup involves organizing experiments into two main sets, each comprising three experimental Jupyter notebooks. These sets represent varying sample sizes: 1000, 2000, and 3000 samples. The notebooks are designed to run experiments using the recommendation systems implemented for the respective datasets, generating recommendations based on the specified sample sizes. This setup aims to assess the recommendation systems' performance and effectiveness across different sample sizes to analyze their scalability and accuracy.

Utilizing Word2Vec with TF-IDF features together with Bidirectional Encoder Representations from Transformers (BERT), as well as ALBERT and Decoding-Enhanced BERT (DeBERTa), the implemented recommendation frameworks are run on datasets with varying sample sizes during the recommendation as well as similarity calculation phase, respectively. The specific similarity between each framework's recommendations is then calculated by utilizing measures such as cosine similarity. Which are based on the computed similarity scores as well as the recommendation frameworks' performance is evaluated for every data collection with the sample size throughout the assessment as well as analysis phase accordingly. The prime aim is to evaluate the efficacy of the framework by examining the precision as well as the coherence of the suggestions that are produced by the various natural language models on a range of specific sample sizes.

# 6 Evaluation

Natural language processing procedures have been included in legal document recommendation frameworks, leading to notable progress in their corresponding development. This section provides a summary of the evolutionary path toward the creation and improvement of legal textual citation recommendation frameworks. This evaluation analysis significantly explores the critical phases that have molded the sector of legal document recommendation frameworks, from the first implementations to the most recent approaches, including state-of-the-art NLP procedures.

## 6.1 Implementation Of Word2Vec, BERT, ALBERT, and DeBERTa Models Concerning Datasets Washington University in St. Louis (2021) Supreme Court Decisions: Text Analysis Kaggle, Version 1.

The sets of tests carried out utilizing the dataset on the Word2Vec, BERT, ALBERT, and DeBERTa models are shown in the next section. Supreme Court Decisions Text Analysis Kaggle Version 1 for Washington University in St. Louis, 2021.

The following table 1 shows the mean similarity scores as outcomes of the respective implemented experiments for Dataset Washington University in St. Louis (2021) Supreme Court Decisions: Text Analysis Kaggle Version 1, as discussed in the following subsections,

The following figures depict the results of the experiments conducted on four natural language models implemented concerning 1000, 2000, and 3000 samples, which contain four figures as recommendations with similarity scores together with a graphical representation of average similarity as well as similarity scores and density of similarity scores, respectively. Because of these resulting figures, the effectiveness of the recommendation framework implemented concerning natural language models is calculated.

| Natural Language Model | 1000 Samples | 2000 Samples | 3000 Samples |
|---|---|---|---|
| Word2Vec | 0.0 | 0.0 | 0.0 |
| BERT | 6.936242198944091 | 6.975235176086426 | 7.012186503410339 |
| ALBERT | 7.9408657312393185 | 8.172201323509217 | 8.231748628616334 |
| DeBERTa | 8.67090034484864 | 8.879297113418579 | 8.8809974193573 |

Table 1: Mean Similarity Scores

### 6.1.1 Experiments on Word2Vec Model with 1000, 2000, and 3000 samples

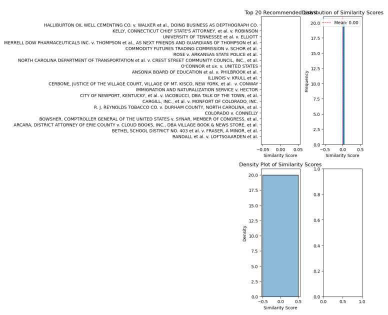

Figure 4: Experiment 1 Results Word2Vec Model 1000 Samples

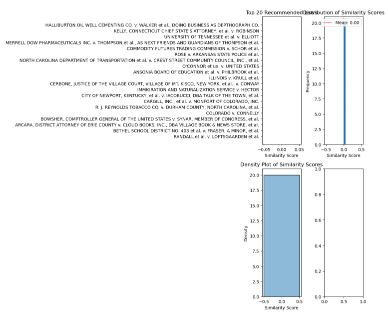

Figure 5: Experiment 1 Results Word2Vec Model 2000 Samples

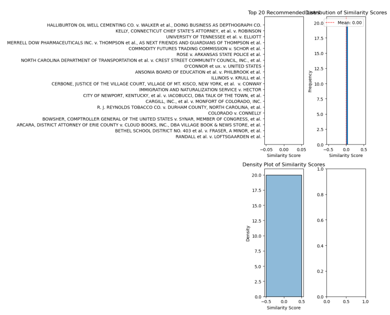

Figure 6: Experiment 1 Results Word2Vec Model 3000 Samples

The above figure 4 depicts the corresponding results of experiment number 1, which is conducted on the Word2Vec model with 1000 samples.

Concerning the input text "United States," the Word2Vec model generates recommendations using the first 1000 samples, which have null similarities between each recommendation generation and the actual text. This is observed to have resulted in a poor similarity score, despite the model using 1000 samples.

Similarly, the related findings of experiment number 1, which is carried out on the Word2Vec model with 2000 samples, are shown in Figure 5.

It is observed that the Word2Vec model generated recommendations for the input text "United States" using the first 2000 samples, which have null similarities between each recommendation generation and the actual textual format. This is why the model produced a poor similarity score with 2000 samples.

The above figure 6 depicts the corresponding results of experiment number 1, which is conducted on the Word2Vec model with 3000 samples.

Utilizing 3000 samples, it is observed that the Word2Vec model produced a poor similarity score because it generates recommendations based on the input text "United States" by using the first 3000 samples, which have null similarities between each recommendation generation and the actual legal textual data.

### 6.1.2 Experiments on BERT Model with 1000, 2000, and 3000 samples

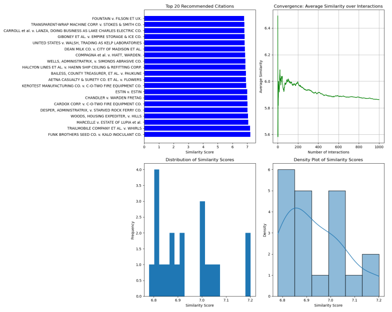

Figure 7: Experiment 2 Results BERT Model 1000 Samples

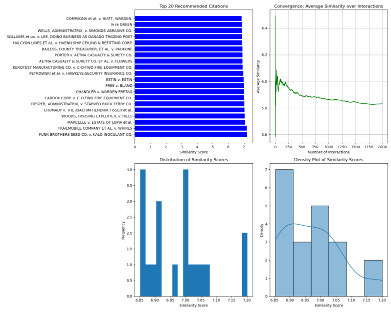

Figure 8: Experiment 2 Results BERT Model 2000 Samples

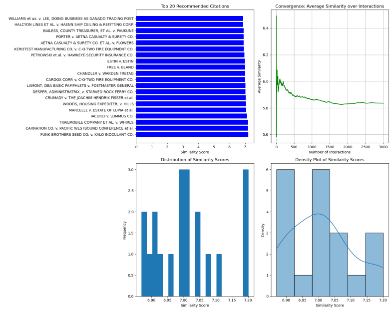

Figure 9: Experiment 2 Results BERT Model 3000 Samples

Figure 7 depicts the corresponding results of experiment number 2, which is conducted on the BERT model with 1000 samples.

It is observed that the BERT model reached a mean similarity score of 6.936242198944091 when applied to 1000 samples concerning the corresponding input text 'United States' for the suggestion of legal textual citations. The respective average similarity between the suggested legal citations as well as the corresponding ground truth citations is shown by this score accordingly. A greater similarity score means that the corresponding BERT model's recommendations are more contextually or semantically similar to the real legal textual citations. This finding significantly implies that, when compared to ground-truth legal textual citations, the BERT model, which was trained in addition to assessed on a thousand samples, shows a high degree of semantic closeness in its suggestions accordingly.

The related results of experiment number 2, which is carried out on the BERT model with 2000 samples, are shown in Figure 8.

It was observed after conducting the experiment that when the BERT model was utilized to recommend legal textual citations for 2000 samples, which are based on the input word "United States," the natural language model's mean similarity score was 6.9752351760864 26. The average similarity between the corresponding suggested legal textual citations, as well as the corresponding ground truth textual citations, is shown by this resulting score. This finding implies that, when compared to ground-truth legal textual citations, the BERT model, which was trained as well as assessed on 2000 samples, shows a high degree of semantic proximity in its suggestions. A greater similarity score means that the BERT model's recommendations are more contextually and semantically similar to the real textual legal citations.

The same results are shown in Figure 9. Experiment number 2 was carried out on the BERT model using 3000 samples.

Similarly, a mean similarity score of 7.0121865503410339 was obtained by the corresponding BERT model when it was applied to 3000 samples for the input text 'United States' to legally cite those samples. According to their separate ground truth textual citations as well as the suggested legal citations, this resulting score represents the average similarity between the recommendations and the actual text. When the BERT model's recommendations are more closely aligned with the actual legal textual citations in terms of meaning or context, the respective similarity score increases. Comparing the suggestions of the BERT model to the ground-truth legal textual citations, it appears that the respective model, which was trained and assessed on 3000 samples, demonstrates a high degree of semantic closeness.

### 6.1.3 Experiments on the ALBERT Model with 1000, 2000, and 3000 samples

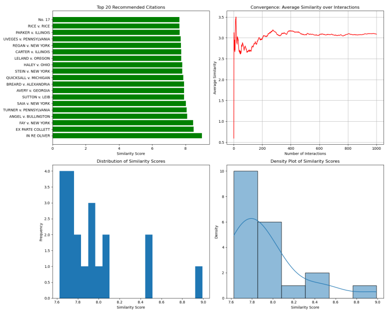

Figure 10: Experiment 3 Results ALBERT Model 1000 Samples

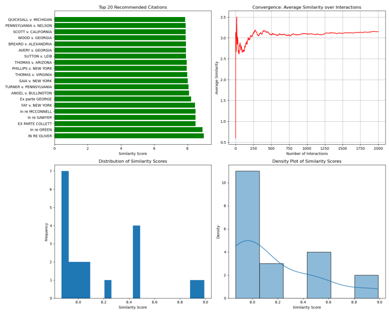

Figure 11: Experiment 3 Results ALBERT Model 2000 Samples

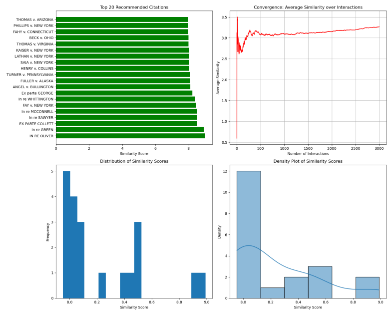

Figure 12: Experiment 3 Results ALBERT Model 3000 Samples

The equivalent results of experiment number 3, which is carried out on the ALBERT model with 1000 samples, are shown in Figure 10.

It is significantly observed that the ALBERT natural language model with 1000 samples yielded a mean similarity score of 7.9408657312393185 for the corresponding input text, "United States." This indicates that the recommended legal textual citations share a greater degree of semantic similarity with the input text than with the ground truth citations. The ALBERT model, which was trained on a thousand samples, was observed as appears to produce recommendations that are more closely aligned with the context or meaning of the respective input text, "United States," based on the higher similarity score. As a result, the suggested legal textual citations are more pertinent as well as contextually comparable to the input material provided, demonstrating the significant usefulness of the ALBERT model in this particular situation appropriately.

The corresponding findings of experiment number 3, which is carried out using the ALBERT model with 2000 samples, are shown in Figure 11.

Compared to the ground truth textual legal citations, the recommended legal textual citations are more semantically related to the corresponding input text, as seen by the mean similarity score of 8.172201323509217 for the ALBERT natural language model with 2000 samples about the input text "United States." With 2000 samples under its belt, the ALBERT model appears to have generated recommendations that are observed to be more closely aligned with the context or meaning of the input text, "United States," as evidenced by the greater similarity score accordingly. This suggests that the respective ALBERT model is working well in this situation because the suggested legal citations are more pertinent in addition to being contextually related to the input language that was provided.

The related outcomes of experiment number three, which is carried out on the ALBERT model with 3000 samples, are shown in Figure 12.

Similarly, the recommended legal textual citations are more semantically related to the input text than the ground truth legal textual citations, according to the ALBERT model implemented with 3000 samples, which yielded a mean similarity score of 8.231748628616334. The ALBERT model, which was trained on 3000 samples, appears to produce specific recommendations that are more closely aligned with the context as well as the meaning of the input text, "United States," based on the higher similarity score accordingly. As a result, the suggested legal citations are more pertinent in addition to being contextually comparable to the input material provided, demonstrating the usefulness of the ALBERT model in this particular situation.

#### 6.1.4 Experiments on the DeBERTa Model with 1000, 2000, and 3000 samples

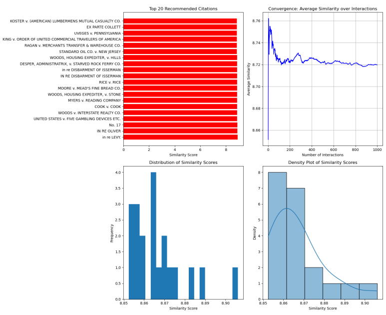

Figure 13: Experiment 4 Results DeBERTa Model 1000 Samples

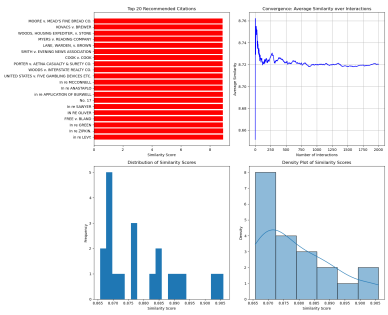

Figure 14: Experiment 4 Results DeBERTa Model 2000 Samples

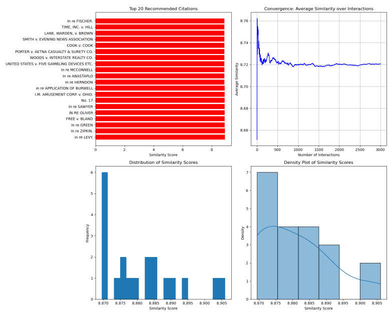

Figure 15: Experiment 4 Results DeBERTa Model 3000 Samples

The associated results of experiment number 4, which is carried out on the DeBERTa model with 1000 samples, are shown in Figure 13.

For the corresponding input text, "United States," the DeBERTa natural language model with 1000 samples yielded a mean similarity score of 8.67090034484864, indicating a high degree of semantic similarity between the recommended legal textual citations as well as the input text. This score significantly indicates that recommendations generally match the semantic context of the input text, "United States," quite well. A better fit in terms of semantic meaning is shown by a greater similarity score, which accurately suggests a stronger relevance together with the relationship between the suggested legal citations and the given entered text.

The same results are shown in Figure 14. Experiment number four is carried out on the DeBERTa model using 2000 samples.

The proposed legal textual citations had a high degree of semantic similarity with the corresponding input text, according to the implemented DeBERTa natural language model with 2000 samples, which yielded a mean similarity score of 8.879297113418579. This resulting score shows that recommendations generally match the semantic context of the entered input text, "United States," quite accurately and significantly. A better fit in terms of semantic meaning is shown by a greater similarity score, which accordingly suggests a stronger relevance as well as a relationship between the suggested legal textual citations with the given text.

The related results of experiment number 4, which is carried out using the DeBERTa model with 3000 samples, are shown in Figure 15.

The suggested legal textual citations show a high degree of semantic similarity with the corresponding input text, with a mean similarity score of 8.8809974193573 for the DeBERTa natural language model with 3000 samples about the input word "United States." Given the input word "United States," this score suggests that, generally speaking, the recommendations closely and accurately match the semantic context they convey. The more similarity points indicate a better fit in terms of semantic meaning as well as a higher level of relevance and correlation between the suggested legal citations and the text that is provided accordingly.

## 6.2 Implementation Of Word2Vec, BERT, ALBERT, and DeBERTa Models Concerning Dataset Shivam Bansal, 2018. Legal Citation Text Classification, Kaggle, Version 1

The following section represents the sets of experiments implemented on the Word2Vec, BERT, ALBERT, and DeBERTa models by utilizing the dataset Shivam Bansal, 2018. Legal Citation Text Classification, Kaggle, Version 1, respectively.

The following table 2 shows the mean similarity scores as outcomes of the respective implemented experiments concerning the dataset Shivam Bansal, 2018. Legal Citation Text Classification, Kaggle, Version 1, as discussed in the following subsections

The following figures depict the results of the experiments conducted on four natural language models implemented to 1000, 2000, and 3000 samples, which contain four figures as recommendations with similarity scores together with a graphical representation of average similarity as well as similarity scores and density of similarity scores, respectively. Because of these resulting figures, the effectiveness of the recommendation framework implemented concerning natural language models is calculated.

| Natural Language Model | 1000 Samples | 2000 Samples | 3000 Samples |
|---|---|---|---|
| Word2Vec | 0.0 | 0.0 | 0.0 |
| BERT | 6.846870279312133 | 6.984172058105469 | 7.039619207382202 |
| ALBERT | 4.325427639484405 | 4.465807580947876 | 4.62504243850708 |
| DeBERTa | 8.85757622718811 | 8.864254665374755 | 8.870682954788208 |

Table 2: Mean Similarity Scores

### 6.2.1 Experiments on Word2Vec Model with 1000, 2000, and 3000 samples

The related findings of experiment number 1, which is carried out on the Word2Vec model with 1000 samples, are shown in the above figure 16.

With the corresponding 1000 samples utilized, it is observed that the natural language Word2Vec model produced a poor similarity score. This is because the model utilizes the first 1000 samples to generate the respective recommendations concerning the input text "Brazil," which has null similarities between each resulting recommendation generation and the actual legal textual format.

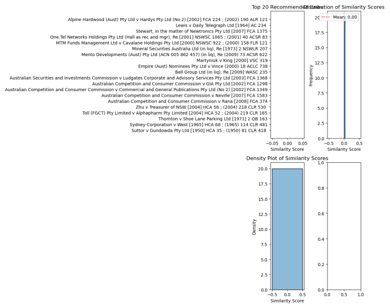

Figure 16: Experiment 1 Results Word2Vec Model 1000 Samples

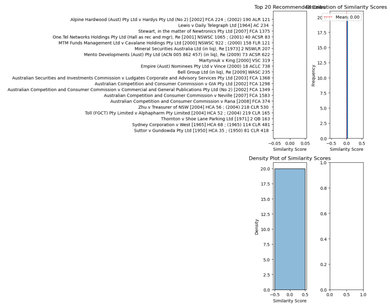

Figure 17: Experiment 1 Results Word2Vec Model 2000 Samples

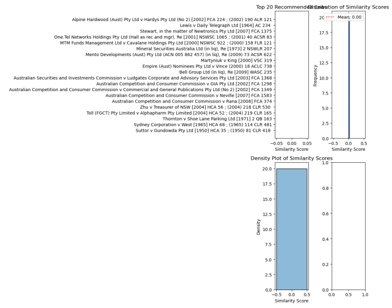

Figure 18: Experiment 1 Results Word2Vec Model 3000 Samples

Similarly, the findings of experiment number 1, which was carried out on the Word2Vec model with 2000 samples, are shown in the above figure: 17.

A low similarity score is observed when the Word2Vec model utilizes 2000 samples. This is because the model generates legal textual recommendations that are based on the implemented input text "Brazil" by utilizing the first 2000 samples, which have no similarities at all between each recommendation generation and the actual legal textual format.

The above figure 18 shows the relevant outcomes of experiment number 1, which is carried out using 3000 samples on the Word2Vec model.

In the last experiment, the Word2Vec model was observed to have produced a poor similarity score with 3000 samples utilized. This is because the model generates legal textual recommendations regarding the input text "Brazil" by utilizing the first 3000 samples, which have null similarities between one recommendation generation and the actual legal text.

### 6.2.2 Experiments on BERT Model with 1000, 2000, and 3000 samples

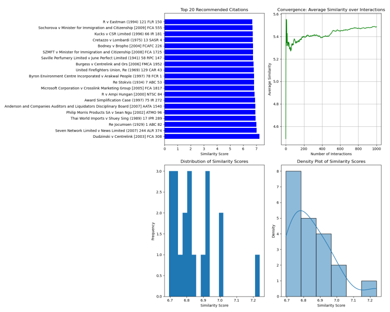

Figure 19: Experiment 2 Results BERT Model 1000 Samples

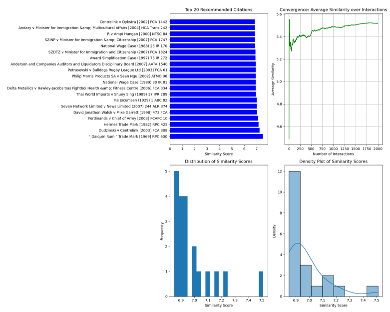

Figure 20: Experiment 2 Results BERT Model 2000 Samples

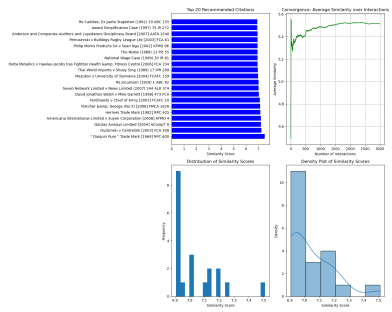

Figure 21: Experiment 2 Results BERT Model 3000 Samples

The related results of experiment number 2, which is carried out on the BERT model with 1000 samples, are shown in Figure 19.

It is observed that a corresponding mean similarity score of 6.846870279312133 was obtained by the BERT natural language model when it was applied to 1000 data samples for the input text 'Brazil' to produce legal textual citation recommendations. This score indicates how comparable the suggested legal textual citations, as well as the corresponding ground truth citations, are on average. An increased similarity score specifically indicates that the BERT model's recommendations are more closely aligned with the context as well as the meaning of the real legal textual citations. In comparison to the ground-truth legal textual citations, this result implies that the BERT natural language model, which was trained in addition to being assessed on 1000 samples, shows a high degree of semantic proximity in its suggestions.

The same results are shown in Figure 20. Experiment number 2 was carried out on the BERT model using 2000 samples.

As per the observations, the BERT natural language model reached a mean similarity score of 6.984172058105469. This was applied to 2000 samples of the implemented input text 'Brazil' to recommend legal textual citations. This specified score shows how comparable the suggested legal textual citations as well as the corresponding ground truth citations are on average. In comparison to the ground-truth legal textual citations, this result implies that the BERT natural language model, which was trained as well as assessed on 2000 samples, shows a high degree of semantic proximity in its suggestions. Hence, an increased similarity score indicates that the BERT natural language model's recommendations are more closely aligned with the context together with the meaning of the real legal textual citations.

The associated findings of experiment number 2, which is carried out on the BERT model with 3000 samples, are shown in Figure 21.

The BERT natural language model obtained a significant mean similarity score of 7.039619207382202 when applied to 3000 samples concerning the implemented input string 'Brazil' to recommend legal textual citations. This corresponding score depicts how comparable the suggested legal textual citations together with the corresponding ground truth legal textual citations are on average. In comparison to the ground-truth legal textual citations, this specified result implies that the BERT natural language model, which was trained with assessed on 2000 samples, shows a high degree of semantic proximity in its suggestions. An incremented similarity score indicates significantly that the natural language BERT model's recommendations are more closely aligned with the context as well as the meaning of the real legal textual citations.

### 6.2.3 Experiments on ALBERT Model with 1000, 2000, and 3000 samples

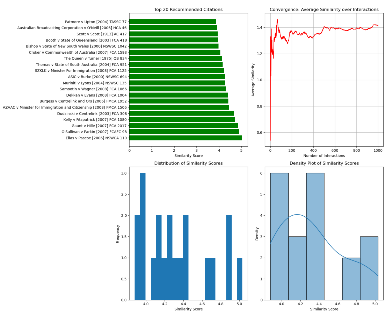


Figure 22: Experiment 3 Results ALBERT Model 1000 Samples

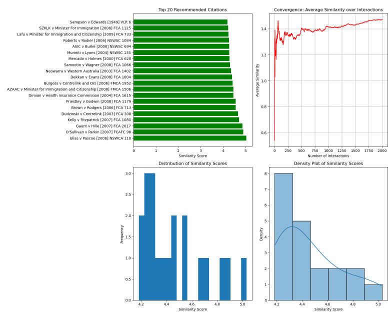


Figure 23: Experiment 3 Results ALBERT Model 2000 Samples

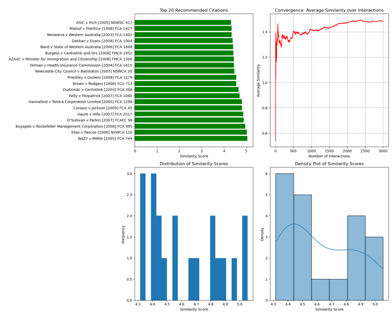


Figure 24: Experiment 3 Results ALBERT Model 3000 Samples

The similar findings of experiment number 3, which is carried out on the ALBERT model with 1000 samples, are shown in Figure 22.

Concerning the experimentation, the recommended legal textual citations are more semantically related to the implemented input text than the ground truth legal citations, according to the natural language ALBERT model with 1000 samples, which yielded a significant mean similarity score of 4.325427639484405 for the input string "Brazil." With a mediocre similarity score, the ALBERT model, which was trained on a thousand samples, likely produces recommendations that are not so accurately consistent with the context as well as the meaning of the input text, "Brazil." The outcome of the ALBERT model in this situation is demonstrated by the possibility that these suggested legal citations are less pertinent and additionally contextually related to the input language that was provided accordingly.

The related outcomes of experiment number three, which is carried out on the ALBERT model with 2000 samples, are shown in Figure 23.

In comparison to the ground truth legal textual citations, the recommended legal citations are less semantically comparable to the implemented input text, as indicated by the ALBERT natural language model's mean similarity score of 4.465807580947876 with the first 2000 samples for the input string "Brazil." This suggests that the ALBERT model is not working well in this situation because the suggested legal textual citations are observed as less pertinent and contextually related to the input language that was provided. With 2000 samples under its belt, the BERT natural language model appears to have generated recommendations that are less aligned with the meaning as well as the context of the input term "Brazil," as evidenced by this similarity score.

The equivalent results of experiment number 3, which is carried out on the ALBERT model with 3000 samples, are shown in Figure 24.

For the implemented input text "Brazil," the ALBERT natural language model with 3000 samples yielded a corresponding mean similarity score of 4.62504243850708, indicating that the suggested legal citations share a lower degree of semantic similarity with the input text than with the ground truth legal textual citations. The ALBERT model, which was trained on 3000 samples, appears to produce recommendations that closely match the meaning together with the context of the input text "Brazil," which is based on a lower similarity score. As a result, the suggested legal textual citations may be more pertinent as well as contextually comparable to the input material provided, demonstrating the less usefulness of the ALBERT natural language model in this specific situation.

#### 6.2.4 Experiments on the DeBERTA Model with 1000, 2000, and 3000 samples

The equivalent findings from experiment number 4, which was carried out on the DeBERTa model with 1000 samples, are shown in Figure 25.

It is observed that the suggested legal textual citations had a high degree of semantic similarity with the corresponding input text, according to the DeBERTa natural language model with 1000 samples, which yielded a mean similarity score of 8.85757622718811, respectively. This specified score indicates that recommendations often match the semantic context of the implemented input term "Brazil." The higher similarity score indicates a better semantic meaning match as well as suggests a stronger relevance with a stronger relationship between the suggested legal textual citations and the given text.

The corresponding findings of experiment number 4, which is carried out on the DeBERTa model with 2000 samples, are shown in Figure 26.

The DeBERTa natural language model with 2000 samples as well as the implemen-

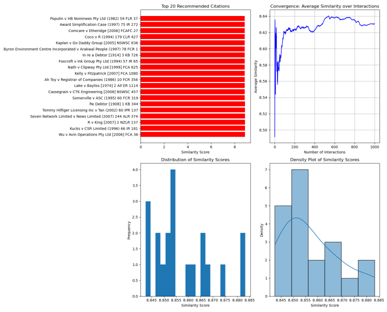

Figure 25: Experiment 4 Results DeBERTa Model 1000 Samples

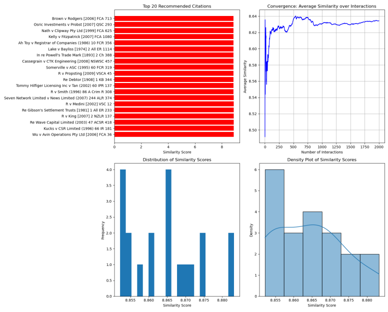

Figure 26: Experiment 4 Results DeBERTa Model 2000 Samples

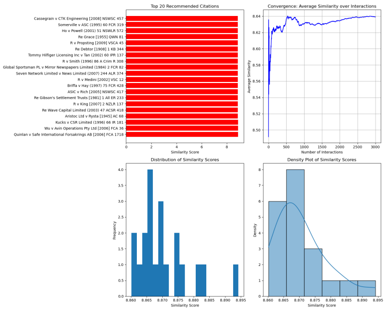

Figure 27: Experiment 4 Results DeBERTa Model 3000 Samples

ted input text "Brazil" yielded a significant mean similarity score of 8.864254665374755, indicating a high degree of semantic similarity between the input text together with the recommended legal textual citations. According to this resulting score, the recommendations often closely match the semantic context that the input text, "Brazil," conveys accordingly. A better fit in terms of semantic meaning is indicated by a greater significant similarity score, which accurately suggests a stronger relevance with the relationship between the suggested legal textual citations and the implemented textual data collection.

The associated outcomes of experiment number 4, which is carried out on the DeBERTa model with 3000 samples, are shown in Figure 27.

The respective proposed legal textual citations had a high degree of semantic similarity with the implemented input text, according to the DeBERTa natural language model with 3000 samples, which yielded a significant mean similarity score of 8.870682954788208. This corresponding resulting score shows that generally speaking, the recommendations closely match the semantic context that the entered input text, "Brazil," conveys accordingly. A greater similarity score appropriately indicates a better semantic meaning match together, suggesting a stronger connection with the correspondence between the suggested legal textual citations and the given textual data collections.

## 6.3 Discussion

Concerning the experiments, the outcome is that different implemented sample sizes and datasets have variable performance levels compared to the implemented natural language models. Early results showed that the Word2Vec natural language model could not capture specified semantic similarities between input text and legal citations, consistently producing a poor mean similarity score of 0.0 for both datasets and sample sizes. For both datasets, however, the mean similarity scores of the natural language models BERT, ALBERT, and DeBERTa significantly improved with varying implemented sample sizes. Above all, the natural language model DeBERTa consistently showed the greatest mean similarity scores, demonstrating its superiority over the other models in producing pertinent legal textual citations that align with the respective input text accordingly.

Additionally, the corresponding mean similarity scores for the natural language BERT, ALBERT, and DeBERTa models tended to get better as the implemented sample size increased, indicating that bigger sample sizes help in the improvisation of performance in producing significant suggestions that are more semantically matched. Furthermore, it is observed that the natural language model ALBERT's mean similarity scores were significantly lower than those of BERT as well as DeBERTa, respectively, indicating that it was less effective in the identification of semantic connections between the entered input texts together with the legal textual citations in the datasets. All things considered, the natural language model DeBERTa is the most successful model when it comes to selecting pertinent legal textual citations, regardless of sample sizes and datasets utilized. It consistently yields the greatest mean similarity scores compared to others.

To conclude the discussion aspect, it is studied that, in the production of document embeddings for legal textual citation datasets, four natural language processing models were utilized in the constructed legal recommendation framework: the Word2Vec natural language model with term frequency-invert document frequency, bidirectional encoder representations from transformers (BERT) together with ALBERT (a Lite BERT), as well as the DeBERTa (decoding-enhanced BERT). The Word2Vec natural language model, especially when combined with TFIDF, showed some effectiveness but not enough to capture the appropriate subtle legal textual meanings as well as contextual nuances, as evidenced by the mean similarity score of 0.0 it produced across the experimental sample sets, which are 1000, 2000, and 3000 samples. The main algorithm for producing legal textual citations became TF-IDF since it was better suited for legal textual material.

It is significantly observed that the natural language model DeBERTa (Decoding-Enhanced BERT) performed better in terms of similarity scores than Bidirectional Encoder Representations from Transformers (BERT) and ALBERT (A Lite BERT) in the comparison of transformer-based natural language models (Bidirectional Encoder Representations from Transformers (BERT) together with ALBERT (A Lite BERT) as well as the DeBERTa (Decoding-Enhanced BERT), indicating its stronger capacity to acknowledge the complexity of legal language. These implemented natural language models' ability to learn could be the cause of the observed performance discrepancies. DeBERTa's improvements in managing long-range dependencies probably contributed to its better performance, whereas ALBERT's lower parameter count may have hindered its capacity to record detailed legal context. These results highlight how crucial it is to choose models for recommendation frameworks that are tuned into the nuances of legal textual language to achieve the best results.

# 7 Conclusion and Future Work

Based on various implemented sample sizes as well as datasets, the performed experiments significantly demonstrate the differing performances of the natural language model Word2Vec in combination with TF-IDF with Bidirectional Encoder Representations from Transformers (BERT), ALBERT, and DeBERTa in proposing legal textual citations. Hence, it is observed that the natural language model Word2Vec's weak capability to capture semantic links is demonstrated by the fact that it consistently scored 0.0 mean similarity across all implemented datasets as well as sample sizes. On the other hand, the findings of the natural language models BERT as well as ALBERT and DeBERTa were more encouraging; in a particular aspect, the natural language model DeBERTa regularly produced the highest mean similarity scores, respectively. The natural language model, in addition to ALBERT and DeBERTa, demonstrated gains in suggesting more contextually appropriate legal textual citations as sample numbers grew. Concerning the comparison, DeBERTa, the natural language model, continuously performed better than the other implemented models, demonstrating its greater capability to match input textual data with legal textual citations. However, compared to the natural language models BERT and DeBERTa, the ALBERT model had comparatively lower mean similarity scores, indicating that it was not as good at capturing semantic links.

Future research can primarily focus on additional improvements as well as explorations to boost the effectiveness of the legal recommendation frameworks. First off, examining larger in addition to more varied legal textual citation datasets may help the natural language models better grasp the subtleties of legal textual language. Furthermore, adjusting specified hyper-parameters as well as training the natural language models more thoroughly could do improvisation for their performance. Additionally, testing with sophisticated language models other than Bidirectional Encoder Representations from Transformers (BERT) variations or looking into other pre-processing methodologies may yield more insights into enhancing the suggestion's precision. Furthermore, adding legal textual ontologies or domain-specific knowledge may tend to significantly improvisation of the natural language models' comprehension of legal circumstances. Finally, by utilizing the advantages of several types of natural models, investigating ensemble approaches as well as integrating many natural models may result in better recommendations.

# References


Arora, M., Mittal, V. and Aggarwal, P. (2021). Enactment of tf-idf and word2vec on text categorization, *Proceedings of 3rd International Conference on Computing Informatics and Networks: ICCIN 2020*, Springer, pp. 199–209.

Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N. and Androutsopoulos, I. (2020). Legal-bert: The muppets straight out of law school, *arXiv preprint arXiv:2010.02559* .

Chalkidis, I. and Kampas, D. (2019). Deep learning in law: early adaptation and legal word embeddings trained on large corpora, *Artificial Intelligence and Law* **27**(2): 171–198.

Chen, G., Luo, X. and Zhu, J. (2023). A legal multi-choice question answering model based on bert and attention, *International Conference on Knowledge Science, Engineering and Management*, Springer, pp. 250–266.

Cui, J., Shen, X. and Wen, S. (2023). A survey on legal judgment prediction: Datasets, metrics, models and challenges, *IEEE Access* .

De Martino, G., Pio, G. and Ceci, M. (2022). Prilj: an efficient two-step method based on embedding and clustering for the identification of regularities in legal case judgments, *Artificial Intelligence and Law* **30**(3): 359–390.

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805* .

Dhanani, J., Mehta, R. and Rana, D. (2022). Effective and scalable legal judgment recommendation using pre-learned word embedding, *Complex & Intelligent Systems* **8**(4): 3199–3213.

Dhanani, J., Mehta, R. and Rana, D. P. (2021). Legal document recommendation system: a dictionary based approach, *International Journal of Web Information Systems* **17**(3): 187–203.

Elwany, E., Moore, D. and Oberoi, G. (2019). Bert goes to law school: Quantifying the competitive advantage of access to large legal corpora in contract understanding, *arXiv preprint arXiv:1911.00473* .

Gan, M. and Jiang, R. (2013). Improving accuracy and diversity of personalized recommendation through power law adjustments of user similarities, *Decision Support Systems* **55**(3): 811–821.

He, P., Liu, X., Gao, J. and Chen, W. (2020). Deberta: Decoding-enhanced bert with disentangled attention, *arXiv preprint arXiv:2006.03654* .

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P. and Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations, *arXiv preprint arXiv:1909.11942* .

Mumcuoğlu, E., Öztürk, C. E., Ozaktas, H. M. and Koç, A. (2021). Natural language processing in law: Prediction of outcomes in the higher courts of turkey, *Information Processing & Management* **58**(5): 102684.

Nguyen, H.-T., Nguyen, M.-P., Vuong, T.-H.-Y., Bui, M.-Q., Nguyen, M.-C., Dang, T.-B., Tran, V., Nguyen, L.-M. and Satoh, K. (2022). Transformer-based approaches for legal text processing: Jnlp team-coliee 2021, *The Review of Socionetwork Strategies* **16**(1): 135–155.

Sil, R., Roy, A., Bhushan, B. and Mazumdar, A. (2019). Artificial intelligence and machine learning based legal application: the state-of-the-art and future research trends, *2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, IEEE, pp. 57–62.

Sivapiran, S., Vasantharajan, C. and Thayasivam, U. (2023). Party extraction from legal contract using contextualized span representations of parties, *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, pp. 1085–1094.

Tripathy, J. K., Sethuraman, S. C., Cruz, M. V., Namburu, A., Mangalraj, P., Vijayakumar, V. et al. (2021). Comprehensive analysis of embeddings and pre-training in nlp, *Computer Science Review* **42**: 100433.

Xiao, C., Hu, X., Liu, Z., Tu, C. and Sun, M. (2021). Lawformer: A pre-trained language model for chinese legal long documents, *AI Open* **2**: 79–84.

Zhong, H., Xiao, C., Tu, C., Zhang, T., Liu, Z. and Sun, M. (2020). How does nlp benefit legal system: A summary of legal artificial intelligence, *arXiv preprint arXiv:2004.12158* .