CBR-based Similar Case Retrieval from Judicial Records

By

Mehreen Gul

Fall-2018-MS-CS 276319 SEECS

Supervisor

Dr. Faisal Shafait

Department of Computing

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science in Computer Science (MS CS)

In

School of Electrical Engineering & Computer Science (SEECS) ,

National University of Sciences and Technology (NUST),

Islamabad, Pakistan.

(Aug 2022)

Online Printing Date & Time: Thursday, 04 August 2022 16:15:40

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis entitled "CBR-based Similar Case Retrieval from Judicial Records" written by MEHREEN GUL, (Registration No 00000276319), of SEECS has been vetted by the undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/M Phil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: ______________________

Name of Advisor: Prof. Dr. Faisal Shafait

Date: 02-Aug-2022

HoD/Associate Dean: ______________________

Date: ______________________

Signature (Dean/Principal): ______________________

Date: ______________________

Publish Date & Time: Thursday, 04 August 2022 , 10:06:25

Online Printing Date & Time: Thursday, 04 August 2022 16:15:53

Approval

It is certified that the contents and form of the thesis entitled "CBR-based Similar Case Retrieval from Judicial Records" submitted by MEHREEN GUL have been found satisfactory for the requirement of the degree

Advisor : Prof. Dr. Faisal Shafait

Signature: ______________

Date: 02-Aug-2022

Committee Member 1:Mr. Adnan Ul-Hasan

Signature: ______________

Date: 02-Aug-2022

Committee Member 2:Dr. Seemab Latif

Signature: ______________

Date: 03-Aug-2022

Committee Member 3:Dr. Rabia Irfan

Signature: ______________

Date: 03-Aug-2022

Publish Date & Time: Thursday, 04 August 2022 , 10:06:25

Dedication

To Abu

Online Printing Date & Time: Thursday, 04 August 2022 16:16:15

Certificate of Originality

I hereby declare that this submission titled "CBR-based Similar Case Retrieval from Judicial Records" is my own work. To the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics, which has been acknowledged. I also verified the originality of contents through plagiarism software.

Student Name: MEHREEN GUL

Student Signature: _______________

Publish Date & Time: Thursday, 04 August 2022 , 10:06:25

Acknowledgments

Glory be to Allah (S.W.A), the Creator, the Sustainer of the Universe. Who only has the power to honour whom He please, and to abase whom He please. Verily no one can do anything without His will. From the day, I came to NUST till the day of my departure, He was the only one Who blessed me and opened ways for me, and showed me the path of success. Their is nothing which can payback for His bounties throughout my research period to complete it successfully. After the Almighty Allah, I am very thankful to my parents, Ami, Abu for for supporting me all the way. I would also thank my siblings, my friends and colleagues for being there with me all the time and helping me specially through the tough times. I am thankful for Allah for blessing me so much.

Mehreen Gul

Contents

List of Figures

List of Abbreviations and Symbols

Abbreviations

SCP	Supreme Court of Pakistan
C.A	Civil Appeal
LMTC	Large-Scale Multi-Label Text Classification
STS	Semantic Text Similarity

Abstract

Courts require information technology that can handle the many ways that cases are handled since not all cases are handled in the same way. As a result, AI can be helpful for various sorts of court proceedings in various ways to decrease the case life cycle and cases pendency at the courts. To make legal information both understandable, actionable, and especially for providing right information at the right time, it has to be organised and given meaning. The judgement data is made available to the public for awareness and guidance. The massive volume of unstructured textual material have made it challenging for AI systems to give consumers and their search queries the best recommendations. Consequently, computing the similarity between numerical data points is the most crucial component of those systems. AI may assist those seeking for information, parties to a case, and judges with structured information. A system that can retrieve similar cases will not only lessens human effort but also reduce the amount of time it would take to retrieve similar cases. The study of machine learning and natural language processing (NLP) has advanced significantly since the widespread use of the Internet. Therefore we have implemented a Siamese neural network and a Large-Scale Multi-Label Text Classification having experiment with several neural classifiers to compute the judgment similarity.

Chapter 1

Introduction and Motivation

Documents are one of the fundamental sources of information recording, either on paper or in a digitalized form. It is essential to keep track of information that can be traced since the number of records is rapidly growing.

Organizations need specific insight to be extracted from the documents to visualize the association between various entities and to predict the outcomes. Lately, judgments have been digitized; therefore, building different systems can be helpful to process the work quickly for lawyers. In the legal domain, complete case cycle is based on case matter and laws used being used for conflict resolution. If laws have been extracted from the associated judgments, they can be utilized to find case similarities and give lawyers and judges access to a more specific dataset.

This chapter provides an introduction to our study as well as a description of the data used, challenges, etc.

1.1 Introduction

In Pakistan, the judgment datasets of all the major courts are widely available to the public. The judgments can be found on the official websites of the Pakistani courts, which include the High Courts as well as the Supreme Court of Pakistan. There are 2029 judgments that are available publicly on the website of the Supreme Court of Pakistan alone when accessed on 15-07-2022. This data continues to expand quickly due to an increase in the number of cases filed each year. It is now important to automate the court system by processing this large amount of data and getting insights from it.

Previously, when preparing for a lawsuit, lawyers had to not only cite legislation to support their position but also make reference to similar cases of judgment. It takes a long time to manually search through these cases. Thus, it is a very time-consuming process. This process can be automated by extracting relevant information from the judgments such as case details, laws, and case references. By using artificial intelligence (AI) and deep learning techniques, similar cases of judgment can be processed easily, which can save a great deal of time for lawyers and judges.

1.2 Legal System

Everything in the world of humans is legal; everything that is legal is "rule-based," and every rule is designed [26] to fit into the facts. It's difficult to fit the rules into the facts; this is when interpretation steps in to perform the difficult work. The human environment is composed of an expanding set of facts that are held together by a tangled web of rules; as it expands, it becomes more complex. By settling disputed facts, the judiciary upholds justice and helps to keep social order and balance in check. The judicial judgment process is intricate, abstract, and influenced by a variety of elements, including the background, culture, emotions, intuition, etc. of the judge. Judges break the law into pieces, bend and twist it to fit the circumstances using the interpretational tool, and have complete control over their own processes. Language serves as the channel via which facts and legal arguments are presented to judges.

However, the case life cycle also depends on other factors too. Before a case is presented to the Judge, the case lawyers, judge researcher etc. go through all the similar cases manually. It is a highly time consuming task to search for the relevant case in less time. For the very beginning till the case judgment is given, the whole process goes through this relevant cases retrieval, step many and hence increasing the case life cycle.

To ensure that legal information [42] is not only understandable but also useful for making decisions and retrieve similar cases, it needs to be organized and given context.

If legal information is augmented, AI can also provide advice and suggestions to those looking for information, to the parties involved in a case, and to judges. It will help all the judiciary stakeholders to get the right information at the right time with less effort.

Lawlor predicted that computers will eventually be able to analyze and forecast the results of judicial decisions in his foresighted research on the future use of information technology in the legal realm [1]. Lawlor contends that accurate forecasting of judges' behavior would need a scientific understanding of the manner in which the law and the facts affect the relevant decision-makers, i.e., the judges. More than fifty years later, improvements in natural language processing (NLP) and machine learning (ML) give us the means to automatically analyze legal documents in order to successfully develop predictive models of judicial outcomes.

The case life cycle starts by filling the case at the court. Then is case is prepared for presenting it to judges. For this purpose a lot of manual work is done such as case information and similar cases are retrieval. As the process is manual, the judges and lawyers go through a large bulk of cases to get the required information. After the case is presented, a case can have multiple hearings and then a final verdict is given that is used for future references. It's also possible that the case decision is being challenged by applying for appeal. Appeal is applied to the higher level of court only. So in the whole case life cycle, from the very beginning to the verdict announcement, the law stakeholders keep on searching for the relevant information.

1.3 AI at Judiciary of Pakistan

Artificial intelligence is the process by which computer systems can perform operations that ordinarily require human intelligence, such as speech recognition, visual perception, decision-making, and language translation.

Making decisions with AI's assistance can be a major factor in automating industrial processes. With sufficient data, AI has demonstrated the ability to understand insightful information that reveals underlying patterns and can predict with reasonable accuracy with new data. Many people refer to AI as the "new electricity" because it is quickly taking over as the foundation of every industry and our daily lives.

The use of AI in the legal system is assisting in providing timely justice to the public despite the world's painfully overburdened justice systems. Robot judges in Estonia and the Offender Profiling System (COMPAS) in the US are just two examples of the growing number of AI-enabled judicial support systems being used around the world. AI has already shown that it is a necessary component of a potential judicial system. While algorithms are nowhere near a substitute for human judgment, they can be a useful tool for gathering pertinent information from whole or incomplete older cases and recommending the likely outcome of a given petition.

More than 220 million people call Pakistan home, and every son of the soil has a fundamental right to receive justice. Fast justice for the country's common people is being hampered by a severe labor deficit in senior positions. Elite courts are making every effort to settle as many petitions and appeals as they can, but the number of cases still pending has consistently increased. All provincial high courts, including the high courts and in Islamabad, the supreme court, and others, have set up their own IT departments. To track the petition record and provide other facilitation services, these IT departments have developed their own solutions. To help the honorable judges at all levels, the IT section of the Sindh High Court has also developed several solutions. Although the IT department of different courts are working hard to facilitate the judges but it is the highly need of the time to incorporate AI in the present System.

1.4 Case Proceedings

The "case shelf-life" is the period of time required to process an issue from filing to resolution. The time for the distribution of the court's decision is not fixed. After extensive consultation with the relevant parties, including the Honorable justices of the Supreme Court of Pakistan, we have identified two areas where technology can help in achieving the desired goal:

(a) Judicial Support

(b) Case Management

The case management system was created by SCP's IT department as an internal information management system. A tracking number, a main category, and a subcategory are assigned to the case after it has been submitted. Additionally, it is decided to concurrently group the case with other comparable instances (so that the honourable court can process the bunch together). When preparing cases and searching for similar cases, among other things, the SCP's case management system also makes it simpler for attorneys to give Tag Lines—a summary of past rulings—and generalized Keywords (understandable to common people). The disadvantage of this method is that it seldom stores technical terms that might guide law enforcement personnel in finding a case.

1.5 Judgments

A court of law is a group of people with the power to hear and decide legal disputes in a variety of contexts, including criminal, civil, military, human rights, etc. When a lawsuit is filed in front of a court, the judge determines what actually occurred and what should be done about it. In a criminal case, the court has the authority to determine if the accused committed the crime, how to punish them, and how severe their sentence ought to be. In civil proceedings, the court may offer an amicable solution to any disagreements that could exist between the parties.

The conclusion reached by the courts in every particular case is made public as a judgment. A judgment is the court's conclusion following consideration of the rights and obligations of the parties involved in a legal action or proceeding. Establishing the parties' rights and obligations resolves the conflict between them.

A judgment contains all the pertinent details about how a case has been pursued. The people and organizations involved, those present at the hearings, the judges hearing the case, the date of the final hearing when the judgment is announced, some information regarding the reason for the case or suit, and any laws or precedent cases being cited for the specific case are all included in this information.

Based on the judgment content, it can be divided into five parts:

- Grounds
- Introduction
- Fact
- Rulings

- Conclusion

These parts are described in detail in Chapter 3 - Dataset. Although each judgment contains these same distribution but the judgment format varies from case to case and from judge to judge too. So there need to be some method to identify these parts.

The parties are given this judgment once it has been publicly revealed. The parties involved may either comply with the court's ruling or file an appeal with a higher court in opposition to it. The Supreme Court of Pakistan is the highest court in Pakistan from which a judgment may be appealed. It is possible to review the Supreme Court's ruling directly.

1.6 Similar Cases

Pakistan is regarded as a developing nation in the modern world, which implies that the systems that control various aspects of the nation are still in the process of growth. The Pakistani Judicial System is one of these systems. Pakistan currently has the fifth-highest population in the world and is projected to double in size over the next 20 years due to an ever-increasing pace of population growth. The entire situation places a significant burden on the judiciary, and we have seen that it has had a negative impact on society as evidenced by the rise in crime and corruption cases.

The Supreme Court of Pakistan is overworked and overburdened with cases. Comparatively fewer cases originate in Pakistan than in other international institutions and end up in the top courts. As a result, the Court always has a tonne of cases that are still pending. The Supreme Court had 44,658 cases outstanding as of May 15, 2020, according to the most recent data available [43], the biggest number ever in the history of the nation [39]. Every year, the Supreme Court receives between 14,000 and 16,000 petitions and appeals.

The High Courts also have a large number of cases pending. As of March 2019, there were 165,202 cases on the Lahore High Court's backlog, 88,972 cases in the High Court of Sindh, 29,455 cases in Peshawar High Court, and 6,158 cases in the High Court of Baluchistan. 1,461 cases were still pending in the Islamabad High Court as of March 2019.

At the present, the way the judicial system functions requires that every new case goes through a protracted documentation process. Finding all the similar cases that have already been decided on takes the greatest time, in essence. There are a variety of reasons why these "similar" cases are necessary, including but not limited to:

- To be judge-reviewed in order to uphold Supreme Court standards
- To be viewed by judges and lawyers in order to prepare
- To be used as a reference tool by court officials.

There are millions of cases and no digital records, manual laborers must sift through mountains of paperwork to locate cases that are similar to one another. The efficiency of the system is reduced this time.

We have built a system that accurately addresses the problem of locating similar cases using cutting-edge AI technology. The system's capabilities go beyond just locating similar cases, though; it also provides case data and even a parametric search tool that pulls similar cases using keywords and certain legislation. Another alternative is for a user to upload a pdf file of a new case application; the system will process it and can give you a summary, important details, and examples of related cases that are already in the database.

1.7 Problem Statement

In comparison to the pace at which new cases are filed, the current legal system takes a long time to resolve matters. The reason the judicial system is delayed is that, in contrast to the first world, we have not kept up with time and technology and still rely on tedious manual processes rather than digitizing the entire process and making it much faster. Numerous digital solutions can be used to meet our needs in the judicial process, but the only one that are simple and effective is the use of artificial intelligence to create **a system that retrieves similar records of cases** from the database based on actual content written inside the document rather than just finding similarities based on labeled entities like referenced laws, categories, and taglines.

Despite decades of research on court cases, there is still no robust system in place to handle similar judgment decisions in Pakistani courts. Dealing with similar cases in court manually requires **a lot of time and high expertise in the field**. Thus, there is a lack of a system to deal with similar cases and extract meaningful information within a short amount of time to facilitate lawyers and judges in courts.

The main challenges in similarity evaluation are:

- No dataset available
- Need to find some method that can be generalized for document of different categories
- No AI based algorithm available for finding legal documents similarity

1.8 Solution Statement

Developing a robust system to deal with similar cases in courts is the aim of this research work, which is based on artificial intelligence (AI). The solution we're putting forth needs a lot of data to train on, a lot of different system modules to perform various functions, and a lot of technical support to assist us in grasping the various jargon we encountered while developing the system. We require access to labeled data from the Supreme Court, which is essentially digital documents or scanned documents, each labeled in a group (or groups) of related documents. We require a variety of modules, including a pdf to-text converter, and an information extraction module to process the raw documents. To clarify the many procedures and jargon used in the legal system, we will need to contact a legal expert.

Millions of people's lives will be improved as a direct result of the problem being solved. The state has vowed to provide fast-tracked access to justice for those who previously had to wait a very long time for it to be served. When individuals are aware that there is an effective method for obtaining justice, they will be less likely to take the law into their own hands. There is a need to find some method that can be generalized for document of different categories to find the most similar documents.

1.9 Key Contribution

The purpose of this artificial intelligence-based research is to create a robust system to handle cases that are similar in court (AI). To solve the problem of similar cases in the courts, we will use AI, deep learning, and natural language processing techniques. For this purpose, a dataset is prepared from scraping the judgment data from Supreme court of Pakistan and an AI algorithm is developed to retrieve the similar case.

1.10 Upcoming Chapters

In the following chapters, the latter portion of this thesis document is structured.

Literature Review This chapter offers a glimpse into some of the important work that has been accomplished in the past on similar court cases. The dissertation's research is organized in this section.

Dataset This chapter provides the insights for understanding the dataset. The data acquisition, pre-processing, and significant information to be used for the current task.

Design and Methodology Our suggested AI models and methods for solving the issue are covered in this chapter. It breaks our strategy into various parts and gives an understanding of their technical specifics.

Implementation and Results The experiments and their findings are presented in this chapter. Additionally, it offers a thorough analysis of the findings along with specially chosen examples.

Conclusion The final concluding remarks are given in this chapter, along with information on the direction the research community should take moving forward.

Chapter 2

Literature Review

We looked at the literature to assess the algorithms in related issues as our solution wasn't for a highly specific scenario but rather was more broad and simplistic.

2.1 Traditional Approaches

We researched the literature to assess the algorithms for similar problems because our solution was not for a particular case, but rather a more general and basic one. The problem should be reviewed once again to ensure clarity: a vector space in which every court case is represented by a vector and similar cases are separated from each other by a smaller euclidean distance. First, we used a state-of-the-art algorithm by López-Sánchez et al [19]. They trained a network on the structure and content of click baits to utilize their algorithm to identify them. Therefore, we should examine the literature to spot clickbait since it is more appropriately aligned with our problem.

Numerous authors have tackled the problems of spam and false web content identification over the years. However, clickbaits are not always spammed or fraudulent, as Chakraborty et al [12] have noted. Instead, these are actual websites that serve sub-standard material. For this reason, other authors have attempted to solve the particular problem of automatic clickbait detection, including Chakraborty et al. [12], Chen Y et al. [10], and Potthast M et al. [13]. Unfortunately, the limitations of the various approaches and the lack of a common dataset for clickbait detection prevent direct comparisons of accuracy from being done very often. Because of this, the majority of the comparisons made in this section are on a qualitative level.

Recently, Potthast et al [13] presented a technique for automatically detecting clickbait. They concentrated on clickbaits on the social network Twitter, in contrast to other authors. Some of the features they used for classification, as a result, are only accessible in this situation (e.g. Twitter user name, mentions, and hashtags...). They assembled a corpus of 3,000 tweets from prominent Twitter publishers and assessed a number of traditional classifiers using various feature sets. The authors suggest utilizing n-gram characteristics collected from the textual contents of linked web pages, as in Biyani P et al [11]. The Random Forest Breiman L et al [3] classifier has the highest scoring model, with an F-1 score of 0.76. The addition of a modern sentiment analysis model, Socher R et al [8], as part of the feature extraction pipeline is one of Potthast's most important contributions.

Finding semantic similarity is done via case-based reasoning. In case-based reasoning, new cases are analyzed and resolved using the knowledge gained from previous cases. Though many scientists have worked in this field, Kolodner's [6] explanation of this technique was the best by far. Kolodner [6] provides a thorough overview of the CBR procedure and its applications in many industries.

According to Aamodt [2], a CBR-based system has four main steps.

- Retrieve the most similar cases
- Reuse the retrieved cases to solve a new problem
- Revise the decision if needed
- Retain the parts of the case for the new problem solution

Determining the similarities between the cases in a CBR-based system is the initial step, but it's crucial to determine how the cases will be represented before that. The conventional techniques employed for CBR representation models were described by López-Sánchez et al. in their publication [19]. These techniques are listed as follows:

- **Word Count Vector**: based on the number of times a word appears in the text, it translates it to vector space.

- **Term frequency-inverse document frequency (TF-IDF)**: combines TF and IDF to identify keywords in a variety of documents.

- **Latent semantic indexing**: analyzes the word semantic and its associations in a corpus by representing documents in low dimensions. It addresses the natural language's polysemy and synonymy properties.

- **A random projection**: Representation of a high-dimensional space in a feature vector with a low dimension. It generates the projection matrix using a random distribution, making it independent of the input data.

- **Latent Dirichlet allocation**: The document is portrayed as topic space using latent Dirichlet allocation. Topics are interpreted using words from papers, which are subsequently used to represent the texts. The text may include a variety of subjects.

2.2 TF-IDF

We used TF-IDF scores as a benchmark model to assess the semantic similarity of judicial documents. According to a study by Albitar S. et al. [9], it is possible to leverage semantic resources to increase the effectiveness of categorization by accounting for the meaning of the words used in text representation. As a result, the models that represent text can resolve some ambiguities as well as account for synonyms and word relationships. Many researchers, like Bloehdorn et al. [5], showed that applying semantics to text categorization increases its efficacy in particular domains, particularly when using semantic resources that are relevant to those domains. Using text-to-text semantic similarity metrics, they include semantics in the class prediction steps. In order to evaluate it in the context of text categorization, they created a new text-to-text semantic similarity measure (TF/IDF based), known as SemTFIDF. They also compare it to the well-known Cosine classical similarity measure, which is typically used in the Vector Space Model, as well as another text-to-text semantic similarity measure proposed in the literature (based on IDF).

The next and most significant addition, studied was a specialized case that was appropriately related to our problem, which was with legal documents. In order to enable sophisticated semantic search across the case law corpus and the graph-based matching of case descriptions onto case laws, D. Cavar et al. [16] describe a method for mapping facts and knowledge in legal texts, particularly case law opinions and holdings. Deep linguistic NLP components are crucial for knowledge graph development. They go over how these components' in-depth analyses enable processing of both the fundamental semantic links in legal documents as well as more sophisticated semantic and pragmatic features, such as implicatures and presuppositions.

2.3 Cosine Similarity

The next step is to specify which similarity measures (Euclidean distance, Cosine similarity, etc.) will be used to compute the case similarity after the cases have been represented in some vector form. On the basic CBR system, Mihajlovic et al. [31] investigated three distinct similarity metrics: Euclidean distance, Cosine similarity, and TS-SS similarity. Two separate datasets were used for the experiment. The same data pre-processing and feature extraction steps were applied to both the dataset and the query for case representation. The similar documents are then discovered using three different similarity metrics. The experiment's goal was to determine which similarity metric returned the most related papers.The outcomes demonstrated that cosine similarity outperforms ED and TS-SS in high dimensional space, leading to the conclusion that ED and TS-SS both experience what is known as the dimensionality curse. However, for datasets with larger textual documents, the performance of the measures needs to be analysed.

2.4 Jacrad Similarity

The two main categories of text similarity algorithms are syntactic and semantic [7]. The majority of syntactic analysis is centered on calculating and comparing the occurrence of characters or phrases in different texts. Semantic measurements offer methods for accounting for a word's surrounding words, or context, inside a text. Moodley et al. [32] evaluated three approaches from each category for this initial investigation. He chose to assess N-grams (N=5), Jaccard distance, and Term Frequency - Inverse Document Frequency (TF-IDF) [4] for syntactic metrics. For the Ngrams technique, it was discovered that the CDCN's similarity linkages and citation links overlapped more and more until N = 5. After that, the overlap begins to decrease (thus choosing N = 5). A technique for vectorising the CJEU case texts into document vectors is provided by TF-IDF and N-grams. A vector distance metric is required to determine how similar two papers are. He decided to compare these two approaches using the well-known cosine similarity distance metric. Stop words were the sole preprocessing that was done to the texts. The stop words that were eliminated were a combination of the following: 1) the set of all English stop words available in the Natural Language Toolkit Python library (https://www.nltk.org); 2) the set of words that are most frequently found in the case texts (those appearing in at least 90% of the documents); and 3) a selection of words and phrases that legal researchers identified as unique to the corpus. (e.g., “Court of Justice").

2.5 CNN

In order to combine deep learning techniques with the conventional CBR method and discover the best case representation in the form of a feature space at each stage of the CBR, López-Sánchez et al. [19] introduced a hybrid neural network called the Case-Representation Convolutional Network (CR-CNN). The CBR methodology was merged with Wor2vec word embeddings and deep metric learning techniques (i.e., DrLIM), enabling the generation of the appropriate case representation with high accuracy. By using the undersampling technique, the size of the Case-Base was also decreased.

After characteristics were extracted, [18] proposed an SVM classifier for predicting court decisions in criminal cases. Even though their method was more than 80% accurate, more study is required because it only covered a limited subset of the data from criminal proceedings. SVM, a method for supervised machine learning, was created by [15] to predict judgments by the French Supreme Court. The technique was integrated with an n-gram-based, conventional feature extraction framework. Additionally, n-fold cross-validation was carried out to examine the study’s findings. On the other hand, their model needs to be enhanced in order to analyze the balanced data. Deep learning (DL) models use word embedding and other complex feature engineering techniques to enhance the performance of ML methods. [14] proposed a neural network that uses attention to identify the charge and retrieve every possible item in every circumstance. Multiple techniques, including support vector machine learning (SVM), a neural network model (Bi-GRU), and stochastic gradient descent with positional labeling (SGDLP), were used to create the suggested model. This approach has to be improved because it cannot be applied to cases involving several defendants. The legal factual summary of the case may be used to automatically produce an official report and a judicial declaration. On the other hand, [23] created an encoder and decoder system based on LSTM. The collection consists of Chinese court documents. The strategy's scope was nonetheless constrained because it only gave priority to circumstances involving a single accusation and a single defendant, omitting those involving many accusations and accusers. However, the system's effectiveness might be improved by adopting more sophisticated techniques like reinforcement learning. The study [17] used a deep learning model that relied on neural networks to predict the outcomes of criminal cases based on official case factors. An attention mechanism and Bi-GRU with DL were used for this. The data set of 1000 verdicts from criminal proceedings was provided by the Thai Supreme Court. The model produced positive results with an accuracy of 66.67 percent. The model's efficiency can only be increased by including more attention processes, and the study's limitations include the fact that it only examined criminal occurrences.

2.6 Attention Model

The fundamental application of hierarchical attention models is the effective automatic reading and information extraction from case facts. They are distinct from the traditional sequence-to-sequence approach in two main aspects. The majority of hierarchical models have a two-tier architecture, with one tier dedicated to word attention vectors and the other to sentence attention vectors. Instead of passing the final hidden state of the encoding stage, the encoder transmits additional data to the decoder, which proves to be more advantageous than RNN. If we utilize an attention model without transformers, word embedding for a specific application takes longer to train. When using the pre-trained transformer model, we can avoid the necessity for labeled data for the word that traditional RNN requires.

A wide range of deep learning applications, including picture captioning, image production, language modeling, and translation, now leverage the attention method. As a result, in the judgment prediction model, important words are extracted from the lengthy document by giving them higher weight.

In order to classify documents, the hierarchical relationships of words and sentences in a document are modeled using a hierarchical attention model. "HAN" stands for Hierarchical Attention Network [29]. A deep neural network is used for document classification, specifically [36] and [33]. A HAN makes an effort to categorize a text using the information it can deduce about the document from its composite parts, or the phrases and words that make up the document. The "hierarchical" in HAN derives from the idea that this knowledge is constructed hierarchically, beginning with the use of words in a phrase and moving on to the use of sentences in a document [27].

2.7 BERT

In order to predict cases from the European Court of Human Rights, Chalkidis et al. [24] used Google's Bidirectional Encoder Representations from Transformers to develop a new HAN model (BERT). Numerous neural models, including BiGRU-Att, HAN, LWAN, BERT, and Hierarchical BERT, were evaluated using the suggested dataset. By performing (1) binary violation classification, (2) multilabel classification, and (3) case importance prediction using sentence scores, the study work has solved the limitations of the existing models. The paradigm for legal reading comprehension-based judgment prediction was created by Long et al. [30] and is based on comprehension questions rather than the text classification technique. Several challenging text inputs were used in the creation of this model. The AutoJudge framework was also created by them to include judgment prediction in legal texts. The performance of AutoJudge was superior to that of the standard model in terms of reliability and consistency. In Anand et al [38], classification and scoring of phrases in the training set based on how closely they resemble the human-written reference summary is accomplished using LSTM. By doing so, the problem of the prediction task's lack of labeled data is resolved.

Transformator models come in a variety of forms. BERT used to be the most effective NLP algorithm, but Google's XLNet has now surpassed it in effectiveness. A sentence from XLNet employs the idea of permutation language modeling. When compared to basic BERT, CamemBERT uses fewer parameters and is more commonly employed for legal jobs that need named entity recognition and Part Of Speech tagging. The Transformer over BERT(ToBERT) and Recurrence over BERT(RoBERT) approaches for classifying lengthy documents were proposed by Pappagari et al. [33] and outperformed pre-trained BERT. So, for upcoming projects including judgment predictions, we recommend employing HAN with a fine-tuned transformer model.

2.8 Embedding

Krithika Iyer [41] has explored whether the traditional NLP techniques have the same results for classifying legal documents as they have long-running sentences that don't follow the traditional/standard language linguistics and grammar patterns. Whereas the classification of long-sentenced documents is a fertile area of research. Krithika Iyer [41] carried out three distinct sets of experiments. For the first set, they used LDA with logistics regression; for the second, Doc2Vec with logistics regression; and Bert Neural Nets for the third one. The LDA model archives 0.133 and 0.47 for accuracy scores of 15 labels and 279 labels, respectively. For the second experimental set, paragraph embedding, doc2vec are used, which improves the overall accuracy to 0.48 and 0.63 for an accuracy score of 15 labels and 279 labels, respectively. The results showed the difficulty involved in classifying the legal documents. The result also showed that increasing the number of documents does not necessarily improve the accuracy of the system. Thus, it indicates that NLP techniques hold many challenges for analyzing and classifying legal documents.

A knowledge graph-based approach has been used by Cavar et al. [16] to classify legal documents. A system is presented to map the facts and knowledge from legal documents, which can be further used for semantic search and similar document retrieval. They have explained the cases where the traditional NPL fails in dealing with legal text and how we can deal with such problems. They used the deep linguistic NLP components for generating the knowledge graphs through which it is possible to understand the document's core semantic relations along with their advanced semantic and pragmatic properties, including implicatures and presuppositions. After that, they mapped the documents to the knowledge graphs and used the graph similarity to find the semantically similar/related documents.

Yinglong et al. [20] have presented an ontology based knowledge block summarization to compute the similarity of the documents to classify the Chinese judgment documents. Two ontologies, the top-level ontology and domain-specific ontologies are used to find the document semantics. Then these ontologies are merged to find the extra knowledge from the document. This ontology based semantic knowledge is then used to summarize the knowledge blocks. After that, Word Mover's Distance (WMD) similarity based on KNN document classification, is used to compute the similarity between the knowledge blocks instead of the whole document. The results show that the proposed method has higher accuracy then the original WMD approach.

In this model by Chalkidis et al. [24] to create L document representations, Ed utilizes one attention head per label.

$$a_{lt} = \frac{\exp(h_t^T u_t)}{\sum_t \exp(h_t^T, u_t)} \tag{2.8.1}$$

$$d_l = \frac{1}{T}\sum_{t=1}^{T} a_{lt} h_t \tag{2.8.2}$$

T is the number of tokens in the document, h_t is a context-aware representation of the t-th token, and u_l is a trainable vector that is used to calculate the attention scores for the l-th attention head. Alternatively, u_l can be considered as a label representation. Each head instinctively concentrates on various document tokens to determine if the corresponding label should be applied. To generate the probability of the associated label in this model, Dd uses L linear layers with sigmoid activations, each of which operates on a separate label-wise document representation d_l.

Tokenization and word segmentation are the first stages of natural language processing. Depending on the application we chose, the case fact description's raw information needs to go through some preprocessing. To combat the polysemy phenomenon in word representation, Mingjie Ling et al. [35] utilize the ELMO (Embeddings from Language Models) word embedding language model. According to Matthew E. Peters et al work [22], ELMO word embeddings are better at preventing polysemy phenomena than Skip-gram models and other word embeddings, such as word2vec and GloVe, which were previously popular. The main benefit of ELMO is that they have various word vectors for the same word in various contexts.

Chapter 3

Dataset

This chapter will provide information on the dataset utilised, its source, how it was created, and how it was pre-processed using court decisions. The SCP dataset is the subject of this study. Since it was created using publicly available Supreme Court of Pakistan rulings.

3.1 Data Understanding

In view of the project's current requirements, this stage largely focused on activities relating to data familiarisation, data quality analysis, and data insights. For the present project's collection of data is from SCP, SCP judgements that are either scanned or accessible in digital PDF format will be used. These judgments were then analyzed in order to better understand the different reporting styles being used. This helps in when formulating the guidelines for the data annotation that being used to annotate named entities (NER) and the relations that underlie them. The annotated data is then used by the AI systems for training.

For similar documents retrieval, it is important to understand that how the domain experts search through the documents and which parameters do they consider important in evaluating the similarity and how these parameter are filter there results from such a large bulk of data.

3.1.1 Current status at SCP

This section describes the case management system used by the Supreme Court of Pakistan and identifies opportunities for improvement. There are a total of 29 different case categories and more than 2,094 judgements on the Supreme Court of Pakistan's website [44]. On the Supreme Court of Pakistan website, it allows the users to search by case type, case number, case year, announcement date, citation, judge, parties' names, and keyword/tagline base; however, no field is necessary to do a search (shown in Fig. 3.1). For domain experts, the keyword/tagline section is more important. It relates to the main theme of the judgement; for example, the term education is added to a judgement if it has to do with education. Both general and technical keywords are acceptable.

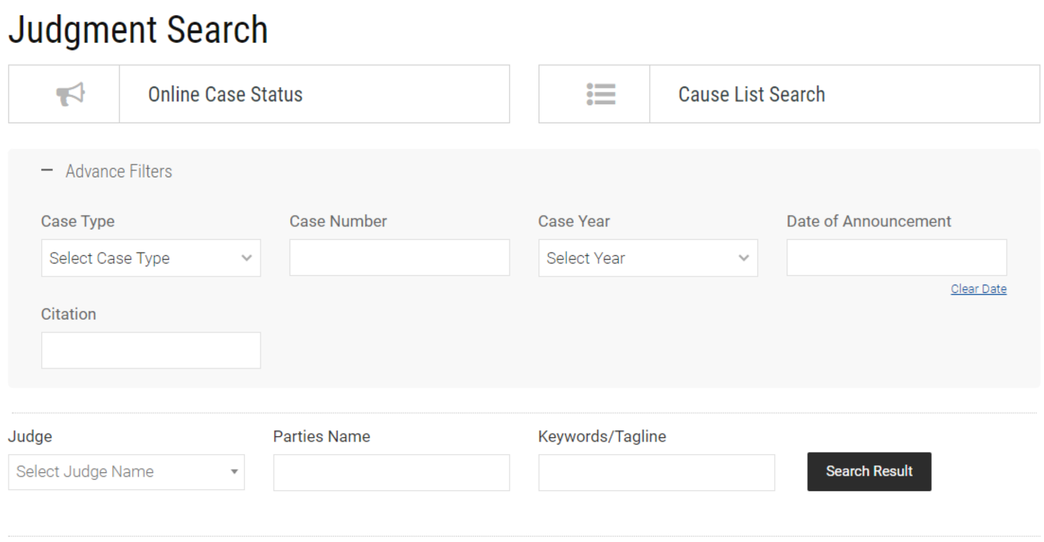


Figure 3.1: Supreme Court of Pakistan - Judgement Search

The taglines are added by the domain experts, and they refer to a sections of the constitution or another legislative principles that are mentioned in the judgement. The following information is provided in the results as a table: judgement date, case subject, case number, case title, author judge, tagline, and the PDF file of the judgement (shown in Fig. 3.2). The judgement summaries are currently written manually by a team of domain specialists of the Supreme Court of Pakistan. Before reaching the determined legal conclusion, they take into account the full verdict. They examine the bulk of cases to determine which ones are most relevant for similar case retrieval. All of these tasks are time-consuming, demanding, and need a lot of manpower.

The two primary sub-tasks of information extraction and relevant case retrieval are the core emphasis of this study. The initial prerequisite for doing IE is having a sizable dataset of annotated data that would be utilised to train the models. A benchmark dataset is also necessary for similarity evaluation in order to assess the level of quality of the cases that have been extracted.

We don't currently have any annotated datasets for the Supreme court of Pakistan. However, the Lahore High Court's reported judgements in the civil and criminal categories have been the subject of extensive study by the PU team for the NER task [28]. The most important and fundamental necessity for annotating a dataset is to create annotation guidelines. These guidelines outline the restrictions on what may and cannot be annotated.

3.1.2 Judgment Structure

The Supreme Court provides the judgments search facility on its website. Judgments can be searched on the basis of the judgment by case type, case no., year or party's name, subject/keyword, and date of the announcement. Judgments can also be searched via category. The form displays all the entries when searched on default values (no filters applied). It has no mandatory field for searching. There is no linkage to judgment for the cases searched for status on the portal or the previous cases that's been cited in the case.

Sr. No.	Case Subject	Case No	Case Title	Author Judge	Upload Date	Judgment Date	Citation(s)	Download
1	Suit for Declaration/Possession	C.A.53/2015	Abdul Khaliq (decd.) thr. LRs and another v. Fazal-ur-Rehman (decd.) thr. LRs and others	Mr. Justice Sajjad Ali Shah	23-07-2022	30-06-2022	N/A	PDF 156 KB
2	Suit for specific performance/.	C.A.342/2014	Allah Ditta & others v. Yaqoob Ali (decd.) thr. LRs.	Mr. Justice Mazhar Alam Khan Miankhel	15-07-2022	20-10-2020	N/A	PDF 1335 KB
3	Allotment/of House	C.A.1009/2010	Yar Muhammad & another v. Mst. Sameena Tayab & others	Mr. Justice Mazhar Alam Khan Miankhel	15-07-2022	07-12-2017	N/A	PDF 1878 KB

Figure 3.2: Supreme Court of Pakistan - Judgement Result

MetaData: Results show these data fields:

- Sr. No.
- Case Subject
- Case No
- Case Title
- Author Judge
- Upload Date
- Judgment Date
- Citation(s)
- File Size In Bytes
- Case File
- Year

There are a total of 2.094 judgments available on the website on 18/07/ [44] Having 555 being Civil Appeal judgments. Judgments are in the form of an HTML Table with a linked PDF. The PDF contains complete judgment. They are either a scanned or digital PDF, and needs pre-processing to utilized further.

A judgment is composed of mainly five parts.

- **Grounds** : parties, lawyers, judges, court and cases being addressed
- **Introduction** : describes whats the matter being addressed
- **Facts** : includes the references to other cases and laws
- **Rulings** : legislature being applied
- **Conclusion** : concluding the judgement and final decision

There are 176 different subjects, and each case is given a case subject defining the matter that is addressed in the case. Each case subject is further divided into sub-categories. For example, the service case are assigned a Service subject as it is the main matter being dealt in the case. In service, if there is some other specific matter, then the case will be assign the sub-subject category like promotion, absence without leave, etc. (See Fig. 3.3)

Main Categories	Sub Categories
Service	-
	Absence without Leave
	Absorption
	ACR. Expungement of Remarks
	Against reduction of penalty
	Against Reinstatment into Service
	Allotment of Accomodation
	Allowances
	Appointments
	Cancellation of Transfer Order
	Competency of Appeal before Service Tribunal
	Competency of Writ Petition
	Compulsory Retirement
	Confirmation
	Correction in Date of Birth
	Criminal and Departmental Proceedings
	Deputation
	Disciplinary Proceedings
	Discrimination
	Dismissal from Service
	Embezzlement - Corruption
	Employee Children Quota
	Enhancement in pay and allowances
	Enhancement of pay scale
	Fixation of pay
	Forfeiture of past service
	NOC for Study
	Others
	Pension
	Promotion
	Promotion Quota
	Railway Employees
	Recovery
Against Ex-Parte Orders	-
	Decree setting aside of
	Order Setting aside of
	Others
	Proceeding Setting aside of

Figure 3.3: Supreme Court of Pakistan - Case Subjects

We are using only Civil Appeals from SCP to perform our task. The reason for taking only SCP is that higher-level court cases have a high weightage for similarity evaluation (see Fig. 3.4). So for now we are just dealing with SCP data. We have taken only single category data as data across different categories will be very dispersed/versatile. Also, the data available on the court's website is not aligned with each other.

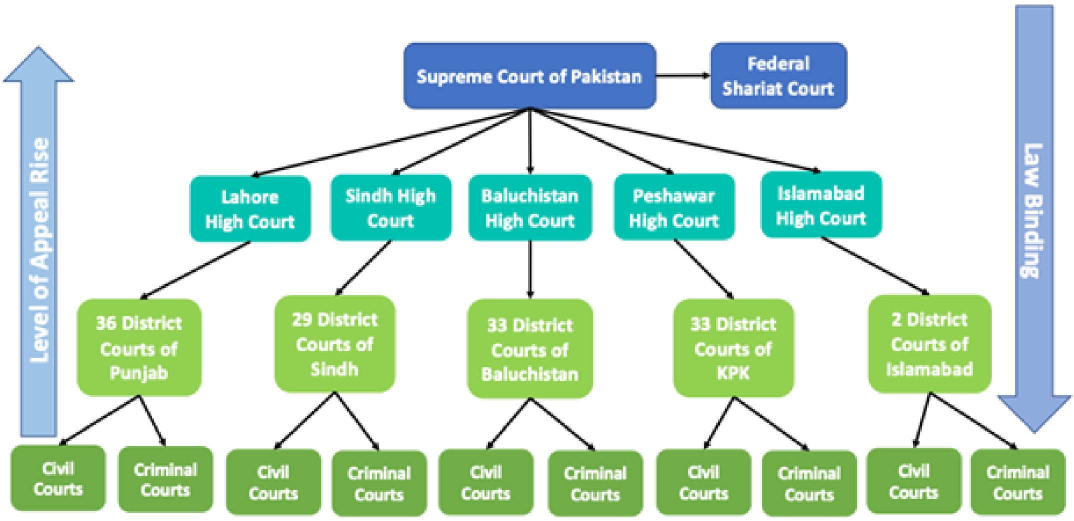


Figure 3.4: Court Hierarchy

3.2 Data Preparation

This phase is primarily concerned with the transformation of the data acquired in the previous step. This transformation is beneficial to be able to apply various Artificial Intelligence based approaches. In the context of the current project, this majorly includes the tasks of text extraction, data cleaning, data annotation, and feature engineering.

3.3 Data Acquisition

In Pakistan, the public has instant access to all of the major courts' judgment datasets. On the official websites of the Pakistani courts, including the High Courts and the Supreme Court of Pakistan, the judgments can be found. Only 2,094 judgments are accessible to the public on the Supreme Court of Pakistan's website. Due to a rise in the number of cases submitted each year, this data keeps growing significantly. Now more than ever, it's critical to automate the justice system by analyzing this large amount of data and drawing conclusions from it.

3.4 Text Extraction

The initial stage in creating an end-to-end information retrieval system from textual data was text extraction from documents. The judgements form the Supreme Court of Pakistan have been scrapped using BeautifulSoup python library.

The available documents comes in two formats.

- A digitally created PDF document
- A scanned document saved as a PDF document

Text extraction from both types of documents is different and bears unique challenges. Text Extraction from a digital document (pdf formed directly from a word document) is a task faced with multiple challenges like identification of headers, footers, page numbers, broken words, incorrect spellings, misplaced numerals, etc. These challenges need to be addressed before getting a clean document that can be used for AI algorithms. Scanned documents have several challenges like skew in scanning, page wraps, low quality of documents, non-textual parts, archived writing styles, etc. that make it challenging to extract text reliably.

3.5 Data Cleaning

Data quality is a key component in every data-science endeavour. The performance of the system is impacted by data noise. Therefore, the emphasis here is on cleaning up data prior to actually converting it into other formats that may be used for effective information extraction (IE) and similar document retrieval.

Metadata and the main body are the two basic components of legal decisions. The case number, details about the petitioner, judges, attorneys, and respondents, as well as other

IN THE SUPREME COURT OF PAKISTAN
(Appellate Jurisdiction)

PRESENT:
Mr. Justice Sajjad Ali Shah
Mr. Justice Amin ud Din Khan

'AFR'
Civil Appeals No. 53 & 54 of 2015.
(Against the order dated 26.12.2000 passed by the Peshawar High Court in CR No. 41 of 1995)

Abdul Khaliq (decd) thr. LRs. ... **Appellants** (In both)

Versus

Fazalur Rehman and others. ... **Respondents** (In both)

For the Appellant (s) (In both)	:	Mr. Tariq Mahmood, Sr. ASC Syed Rifaqat Hussain Shah, AOR
For the Respondents Nos. In CA 53/15: [1(i-iii), 5(Lrs), 6-11, 12(Lrs), 19(Lrs), 20-23, 33(Lrs), 35(Lrs)]:	.	Mr. Muhammad Munir Paracha, ASC ALONGWITH Mr. Zulfiqar Khalid Maluka, ASC.
For the Respondents Nos. In CA 54/15: [1-12, 19-23, 34-36]:	:	Mr. Muhammad Munir Paracha, ASC ALONGWITH Mr. Zulfiqar Khalid Maluka, ASC.
Other respondents:		Nemo
Date of Hearing	:	08.02.2022

Judgment

Sajjad Ali Shah, J. Leave was granted in these cases vide our order dated 21.1.2015 to consider whether the impugned judgment of the Peshawar High Court was inconsonance with the evidence led by the parties and the applicable law to the case.

2. The litigation in these cases is not only very old but has a very chequered background. The facts as pleaded and evident from the record are that the property/subject matter admittedly was owned by one Naaju who died somewhere around 1906 leaving behind one son by the name of Abdul Ghafoor and a daughter Mst. Roshnae. In accordance with the

Figure 3.5: Raw Judgement PDF

information, are typically included in metadata. On the other hand, the body includes the remarks, facts, assertions, and decisions made in related to the prevailing case. The body of a judicial decision may include graphics, tables, page numbers, headers, footers, and footnotes. All these information handlers are processed to display the information to consumers in a more legible fashion in sophisticated text formats like Word and PDF. However, these handlers may produce noise or trash values when extracting text.

So, following the manual removal of the corrupted or Urdu-language files, 515 Civil Appeal judgments were chosen to be utilized further. These documents are then pre-processed by removing the header, footers, page number and then proof read manually to remove/correct the broken and missing/misspelled words. The documents having images and tables are also filtered out. Sentence extraction, tokenization, paragraph identification, data normalisation, part-of-speech tagging, and lemmatization are among the most important tasks that were performed during pre-processing.

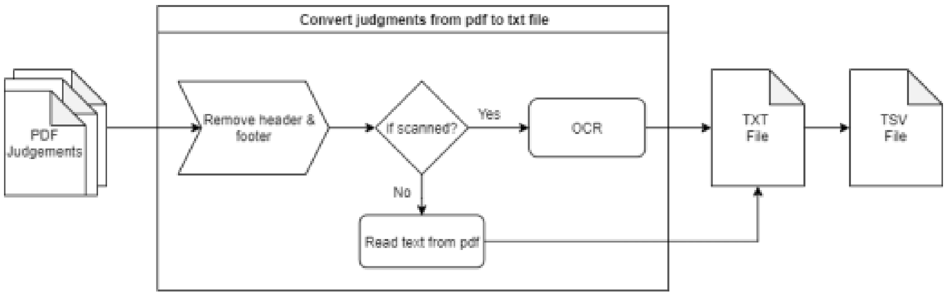


Figure 3.6: Data pre-processing

3.6 Data Annotation

Following the development of the annotation guidelines and tool, the dataset would then need to be annotated. The process of data annotation is iterative, therefore the annotation standards may need to be changed with time. The resulting files will be available for AI algorithms to model after data annotation.

3.6.1 Named Entity Identification

It takes a lot of time and effort to manually annotate a dataset. Many tools are available to aid with data annotation, which can increase productivity and decrease human error. One of the widely used tools for data annotation is DocAnno [47], which offers annotation capabilities for tasks including text classification, sequence labelling, and sequence to sequence activities. Users can generate labelled data with DocAnno [21] for tasks like sentiment analysis, named entity recognition, text summarization, etc. It takes text files as input and produces annotated data as JSON files.

The key challenge is, however, retrieving the necessary data from this unstructured legal language so it is important to know that which entities needs to be annotated. With the help of the legal experts we came to know that for similar judgement retrieval, the following entities play a significant role:

- Case Number
- Reference
- Reference Case
- Reference Court
- Appeal Case number
- Appeal Court

3.6.2 Annotation Guidelines

Using the specified items provided in 3.6.1 as a starting point, we settled on specific guidelines for the annotation of judgements. The labels for the listed entities and the criteria for their annotation are listed in the table 3.1.

Name Entity	Description
Case Number	These are the case numbers of the case being decided. In the Supreme Court of Pakistan often more than one cases are being decided in the same judgment. All those case numbers are to be labelled as CaseNo.
Reference	These are law references mentioned in a judgment. These might include Pakistan's law, Shariah Law, etc. Law references are included in a judgment in the paragraphs where the lawyers are using them for their defense and also where the judges mention them to justify their decision.
Reference Case	These are cases that are mentioned in a judgment to help decide a judgment. These can include older cases from Pakistan or any other country's courts that might involve a similar situation. These cases can also be referred to by the lawyers or the judges.
Reference Court	These are the courts that decided the cases that were referred to in the judgment; the courts that decided the cases being labelled as 'RefCase'.
Appeal Case No.	These are the cases that are being appealed against at the Supreme Court. The cases in the Supreme Court of Pakistan are often for appeals against judgments passed by any of the High Courts so an AppealCaseNo is essentially the case number of the case a High Court decided and that decision is now being challenged against in the Supreme Court.
Appeal Court	These are the courts that decided the case that is now being appealed against in the Supreme court; the courts that decided the cases being labelled as AppealCaseNo.

Table 3.1: Entities Description

3.7 Data Analysis

Law and legal practise generate a tremendous quantity of knowledge in the form of documents, charts, schemas, etc. The organisation and accessibility of these documents are crucial. After the basic data annotation is complete, the data has to be ready for AI-based models. To enable users to semantically mark material so that it may be discovered later, it is important to know the judgement structure and the process of evaluating the similarity between different documents. In different meetings with legal domain experts, we came to know that while searching for the similar judgements, the first thing they find common in the document is the case subject/case matter. The case subject refers to the area of the case its dealing (see Figure 3.3). It can be both generic or specific according to the case. After the case subject, the laws mentioned in the judgement plays the most important role. Domain experts then filter the cases by court of the case, i-e Supreme Court, high court etc. and the year of the judgement. The judgement of the higher level of court and of recent years are of higher priority. It is also needed to consider that the referred case decision is not over ruled/modified. The judge of the case is also important for filtering the cases for some domain experts.

3.8 Organize and structure Information

The format of the judgement available on the Supreme Court of Pakistan varies a lot. The judgement styles varies from judge to judge and from case to case. However, considering the content of the judgement, every judgement have the same components. The case starts with the details of the court, parties of the case, judges detail and cases being addressed. Following this, the first one or two paras of the judgement refer to the matter that is being addressed in the case. The next sections describes the rules and the reasoning followed by the grounds of the case. And finally the concluding para and the final decision of the case. So, the easy way to interpret the judgement is to go through the first an the last paras of the judgement.

Chapter 4

Design and Methodology

In this chapter we will discuss the methodology being used for retrieving the similar case retrieval from the court room data. In order to provide the cleaned data to AI algorithms, it is additionally labelled. After that, we combine the knowledge graph with the annotated data to represent the documents and identify the entity relationships, which are then used to determine how similar the documents are. To choose the best case representation, several models are integrated using a convolutional neural network. To acquire new features and adjust the system in accordance with the new situations, a case-based reasoning (CBR) technique is utilised.

At first, the approach used for similarity evaluation will be discussed then the baseline approaches, NLP and deep learning approaches. Different approaches have been used to complete the similarity evaluation task. Three approaches are used for computing the similar documents. These are broadly divided as following:

- **NER**: Used for extracting information from the judgement PDF file.
- **Semantic Text Similarity (STS)**: Computing the semantic similarity between the cases based on the case text
- **Document Classification**: used for classifying the documents on the basis of the case subject

The approach used for each of these task are discussed in details in the following sections.

4.1 BERT

As discussed in the previous chapter, for similar legal document retrieval, the legislation and previous cases mention in the case plays an important role. So we need name entity recognition to extract the laws and referred cases. So, in order to do Named Entity Recognition, we used BERT, a state-of-the-art model that has excelled in several NLP tasks.

The pre-trained model may be further trained and improved using domain-specific data in order to apply BERT for a particular NLP job. We used BERT-base for the NER's legal documents. A total of 12 layers/transformer blocks, a hidden size of 768, and 12 self-attention heads (L=12, H=768, A=12) make up the BERT-base.

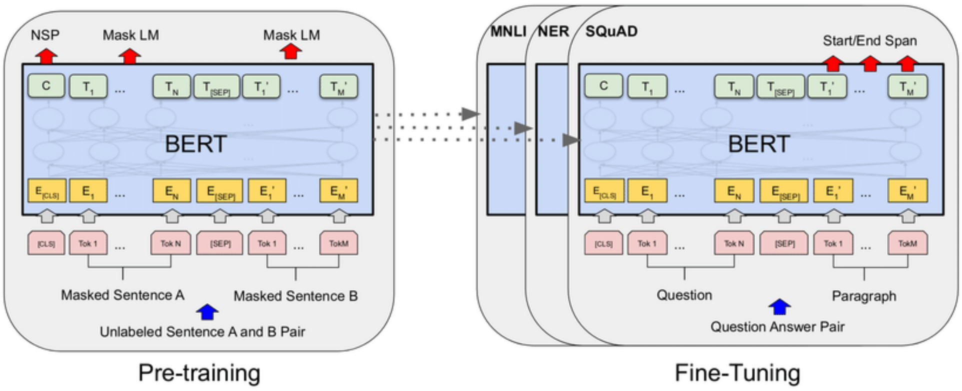


Figure 4.1: Pre-trained BERT Fine-tuning for NER

[40] used the labelled dataset developed using the Civil Appeal judgements from the Supreme Court of Pakistan for fine-tuning. Both generic named entities, such as Person, Organization, Location, etc., and domain-specific named entities, such as Law references, Case references, Case number, etc., were included in this dataset.

4.2 Siamese Neural Networks

Semantic Textual Similarity (STS) analysis is an important research topic in natural language processing and is applied in numerous applications, including question answering, document summarization, information retrieval, and information extraction.

Semantic textual similarity is the process of comparing two texts utilising both direct and indirect connections between them (STS). In the past, STS research has mostly focused on how similar short writings, such abstracts and product descriptions, are to one another. The usage of Siamese networks is common when determining a relationship or similarity between two things.

The intermediate neural network that was trained used a Siamese-like architecture with a case on each branch and a triplet loss function. A group of neural network architecture known as Siamese neural networks have two or more identical sub-networks. They are identical if their configuration, settings, and weights are the same. The two sub-networks replicate parameter updating. The neural networks learn to distinguish between two inputs rather than a model learning to categorise its inputs. It picks up on their similarities. Utilizing word embedding, we deployed Deep Siamese Bidirectional LSTM network to detect phrase and sentence similarity.

One model cell is represented by each block in the figure. The cell type is displayed on the first line, followed by the size of the cell on the second line, and the number of parameters utilised in the cell on the final line. The image clearly shows that we employed three input layers. The preprocessed data that has been transformed to int utilising word-to-int dictionary data serves as the input to "Input Layer 1" and "Input Layer 2." Whether the first two inputs are similar or not, the truth value serves as the input to "Input Layer 3."

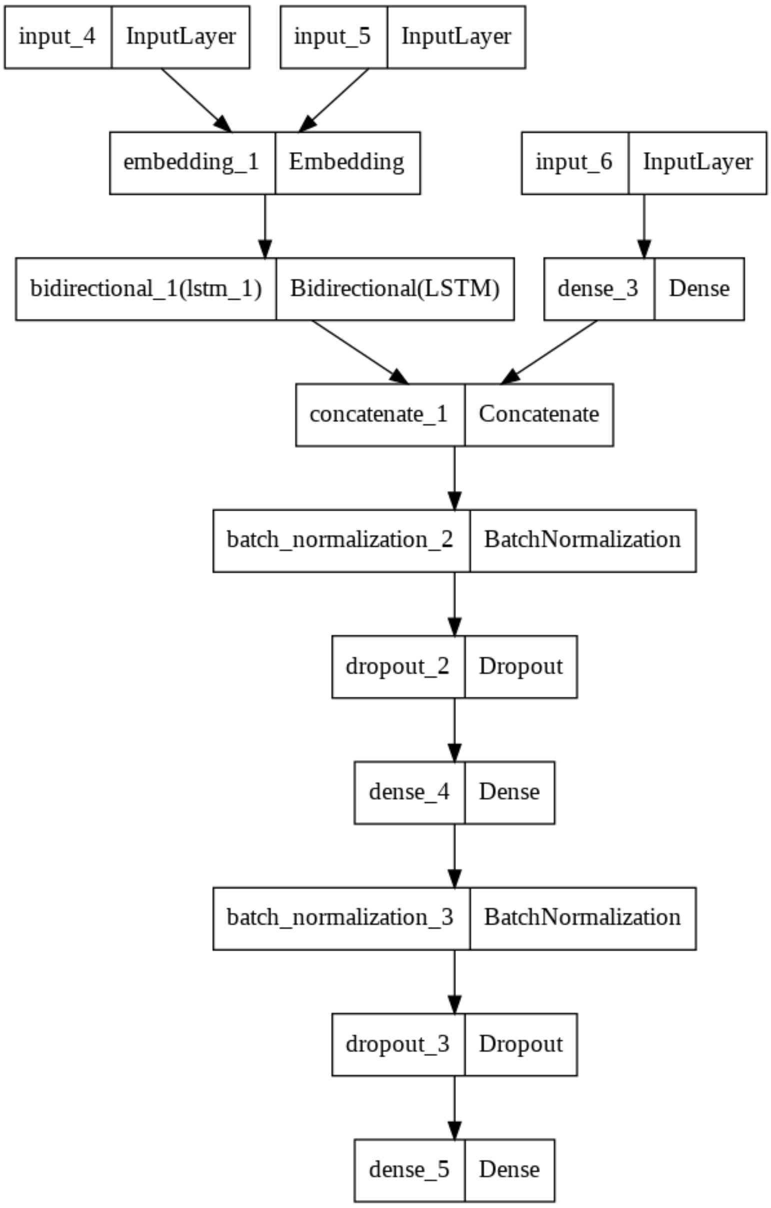


Figure 4.2: Model Architecture

The "Embedding Layer" then receives the inputs and transforms the sparse integers into dense vectors with 50 Dimension vectors each. Each word may be changed into a fixed-length vector with a predetermined size using the "Embedding Layer". Instead of only having 0s and 1s, the resulting vector is dense and contains actual values. Word vectors' fixed length and decreased dimensions enable us to express words more effectively. After that, we used a "Bidirectional LSTMs" cell to memorise the sequence's patterns. We found via testing that bidirectional is more effective than unidirectional in our situation. Bidirectional LSTMs are a development of conventional LSTMs that can enhance model performance for sequence classification issues.

The outputs from Bi-directional LSTMs and Dense 1 Layer are concatenated and passed into Dense Layers. The parameters used for Training are listed below:

Validation Split: 0.1

Drop Rate in LSTMs: 0.17

Drop Rate in Dense: 0.25

Number of LSTMs Units: 50

Number of Dense Units: 50

Activation Function: Relu

Loss: Binary Cross-Entropy

Optimizer: Nadam (Adam with Nesterov momentum)

4.3 Large-Scale Multi-Label Text Classification (LMTC)

The job of assigning to each document all the pertinent labels from a large set, generally including thousands of labels (classes), is known as large-scale multi-label text classification (LMTC) [25] [34].

The Label-Wise Attention Network, also known as CNN-LWAN here, was claimed to attain state of the art performance in LMTC on medical data in a research experiment using various neural classifiers on EURLEX57K. It was demonstrated that on EURLEX57K, a simplified BIGRU with self-attention performs noticeably better than CNN-LWAN. However, it achieved even better results on EURLEX57K by substituting a BIGRU for CNN-CNN LWAN's encoder. Additional improvements are achieved via context-sensitive ELMO embeddings and domain-specific WORD2VEC. As a result, it created stable baselines for EURLEX57K.

The experiment was performed with neural methods with the following components: (i) a token encoder (Ew), which makes token embeddings (wt) context-aware (ht); (ii) a document encoder (Ed), which converts a document into a single embedding; (iii) an optional label encoder (El), which converts each label into a separate embedding; and (iv) a document decoder (Dd Tokens are words unless otherwise noted, and Ew is a stacked BIGRU.

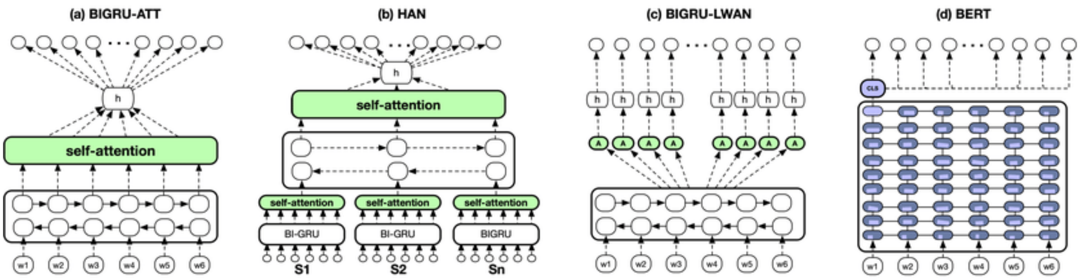


Figure 4.3: Model Architecture

We used various models and embedding techniques to classify the case matter, which can further be utilized for similarity evaluation of the documents by taking the laws applied in that case into account. The Meta data is available on the SCP website, and is used along with the text. The pre-processing techniques are used to make the hierarchy of the case subjects. Lately, the SCP website is assigning single case subjects, so our data is also single-labeled.

Chapter 5

Implementation and Results

This chapter includes a detailed description of the experiments we ran, along with the outcomes and performance data. A thorough analysis that builds on the findings is also included as a guide for future study.

5.1 Evaluation Metrics

The effectiveness of the developed systems is assessed using performance metrics. Multiple evaluation methods have been employed for semantic text similarity evaluation. The precision, recall, and F1-score will be used to assess NER models. Different evaluation metrics discussed in [37] are listed below:

5.1.1 Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1.1}$$

5.1.2 Precision

The proportion of relevant recommendations to all of the other recommendations is known as precision.

$$Precision = \frac{TP}{TP + FP} \tag{5.1.2}$$

5.1.3 Recall

The ratio of relevant recommendations to the total number of relevant items is known as recall.

$$Recall = \frac{TP}{TP + FN} \tag{5.1.3}$$

5.1.4 F1

F-score is a measure that combines precision and recall metrics.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{5.1.4}$$

5.1.5 Root Mean Square Error (RMSE)

When the rating prediction makes significant inaccuracies, it estimates a greater difference. It gives back a positive number.

$$\sum_{i=1}^{D}(x_i - y_i)^2 \tag{5.1.5}$$

5.1.6 Pearson

In order to assess linear relationships, a measure was introduced, which has since gained widespread acceptance in the area of statistics. The PCC formula yields a number between 0 and 1, where 0 denotes no correlation at all, 1 denotes a strong negative correlation, and 1 denotes a high positive correlation. The similarity between two users, U and V, is calculated using the following formula.

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} \tag{5.1.6}$$

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 (y_i - \overline{y})^2}} \tag{5.1.7}$$

5.1.7 Spearman

The Spearman correlation evaluates the monotonic relationship between two variables. In a monotonic relationship, if the value of the first variable changes then the value of the second variable changes as well, but without a constant rate (is not linear). Spearman rank-order correlation coefficient, named , can take a value between -1 and 1. It exists a strong relation between the Spearman correlation and Pearson Correlation Coefficient since is considered as the PCC between the rank variables. To calculate the Spearman rank correlation , we use the following formulas:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{5.1.8}$$

where

d = the pairwise distances of the ranks of the variables xi and yi .

n = the number of samples.

5.2 Experiments

5.2.1 NER

We have conducted experiments on 3 different datasets, including the CONLL2003 dataset, a dataset of civil judgments from the Lahore High Court, Pakistan, which was used by [28], and a dataset of civil appeal judgments from the Supreme Court of Pakistan, in order to assess the performance of BERT for named entity recognition for generic as well as legal named entities [40].

This is the dataset that was created utilising judgments from Pakistan's Supreme Court's Civil Appeal category. The files have a total of 14 listed entities and 214 Civil Appeal judgments. The construction of this dataset is specifically described in Chapter 3. The entities were labelled in the IBO format following the instructions provided in table 3.1. The Table 4.3 lists the names and counts of these labels.

Labels	Count	Labels	Count
B-per	4961	I-per	9749
B-loc	1703	I-loc	1163
B-org	3050	I-org	4934
B-caseno	1497	I-caseno	5151
B-resp	487	I-resp	3048
B-date	3850	I-date	1854
B-refCourt	306	I-refCourt	888
B-refCase	2301	I-refCase	32576
B-ref	4775	I-ref	32099
B-appealcourt	422	I-appealcourt	1778
B-appealcaseno	770	I-appealcaseno	3990
B-money	446	I-money	208
B-FIRno	23	I-FIRno	52
B-Approved	160	I-Approved	1

Figure 5.1: Labels count in the Supreme Court of Pakistan dataset

5.2.2 Siamese neural network

We have used the Supreme court of Pakistan pre-processed judgements as a dataset. The judgements are clean, without headers, footers, images, tables or any corrupted data. The baseline paper is using data that is created with the similarity score for all the documents. Whereas, the dataset we used, has been developed by ourselves keeping in mind the parameters mentioned in chapter 3. Due to the fact the creating the dataset for semantic text similarity is highly time consuming and in our specific problem, require a domain level expertise too, we have created a dataset mentioning 1 for similarity and 0 for not relevant. We have a total of 515 judgements that were used for training. A random split of 80-20 was made for training and validation data.

5.2.3 LMTC

The baseline model uses the European court data for legislation classification on the basis of the case concept. The dataset includes 57 thousand English EU legislative texts from the EUR-LEX portal that have been annotated with 4.3 thousand labels (concepts) from the European Vocabulary (EUROVOC). The dataset used is the JSON format multiple fields as shown in figure 5.2.

The dataset includes the following information:

- three (3) folders, each containing several JSON files (train, dev, test). Each JSON file includes data on a piece of legislation (EU Directive, Regulation, or Decision), as it has been made public on the Eur-Lex portal.
- The whole list of EUROVOC ideas (found in eurovoc en.json).

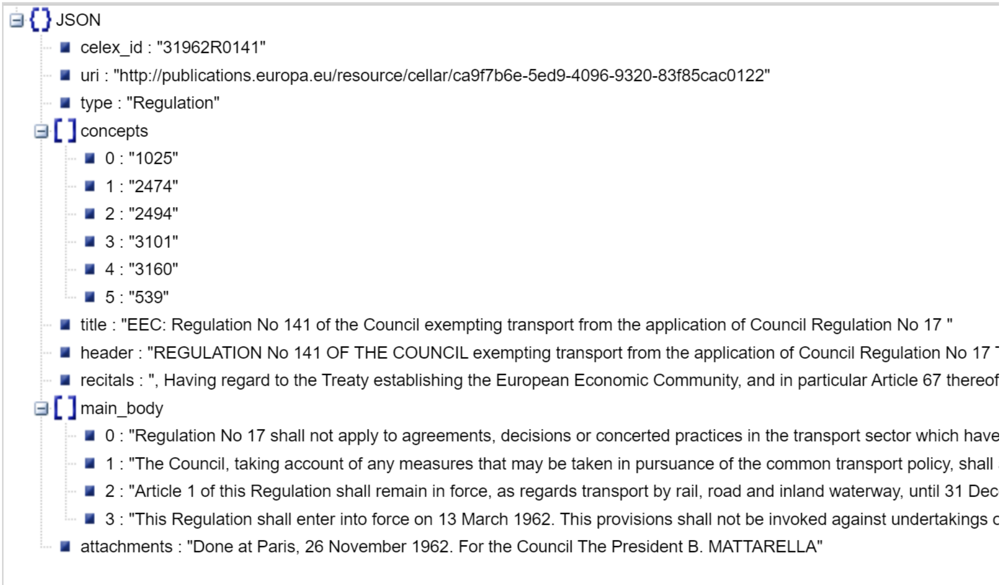


Figure 5.2: LMTC Dataset - JSON File

The following characteristics are included in each JSON file in the subsets (directories) of the dataset:

- **celex id**: The specific identification for each publication in CELLAR and Eur-Lex is the CELEX number.
- **uri**: The HTML-formatted Universal Resource Identifier (URI) for the legal act.
- **type**: Each legal act's kind (Directive, Regulation, Decision).
- **concepts**: A list of the EUROVOC concept IDs that apply to this provisions of the act.
- **title**: The title of the legislation.
- **header**: The legal act's heading.
- **recitals**: the legal act's recitals.
- **main body**: a list of the legal act's articles.
- **attachments**: The legal act's attachments

The header, recitals, body, and attachments of any legal act are sufficient to convey all of its contents.

A list of EUROVOC ideas have been tagged (annotated) on each page. The file eurovoc en.json additionally contains a description for each idea ID.

Each of the ideas in the JSON file that contains the whole set of EUROVOC concepts has the following properties:

- **concept id** The unique concept ID.
- **label**: The main label description.
- **alt labels**: If any alternative label descriptors are present, a list of them.
- **parents**: A list of each label's parents' IDs, if any.

Supreme court of Pakistan also lists the case subject in the judgment search results. Although the case subject sub-category is also allotted to the case but it is not available in the publicly accessible data on the website of Supreme Court. We have treated the case subject as case concept and classified the judgements. The model takes an input of the judgement body, case title, judges, and dates.

5.3 Results and Analysis

5.3.1 BERT

With BERT, we achieved an overall F1 score of 92.72 and an average F1 score of 91.08 for the Named Entities for the dataset consisting of judgments from the Supreme Court of Pakistan. The F1 scores achieved on this dataset for B-refCase is 98.82 and for B-ref is 92.17 [40]. The figures below shows the results.

The results shows significant improvement with the BERT base NER. The Figure 5.5 shows the comparison of results of BERT on SCP dataset and LHC dataset along with the results of the CRF on LHC dataset. The results for ref BERT perform 5% better than CRF. The difference in the refCase is due to the annotation difference in both the data set. LHC dataset annotated the refcases without the parties names whereas SCP dataset includes the parties names too.

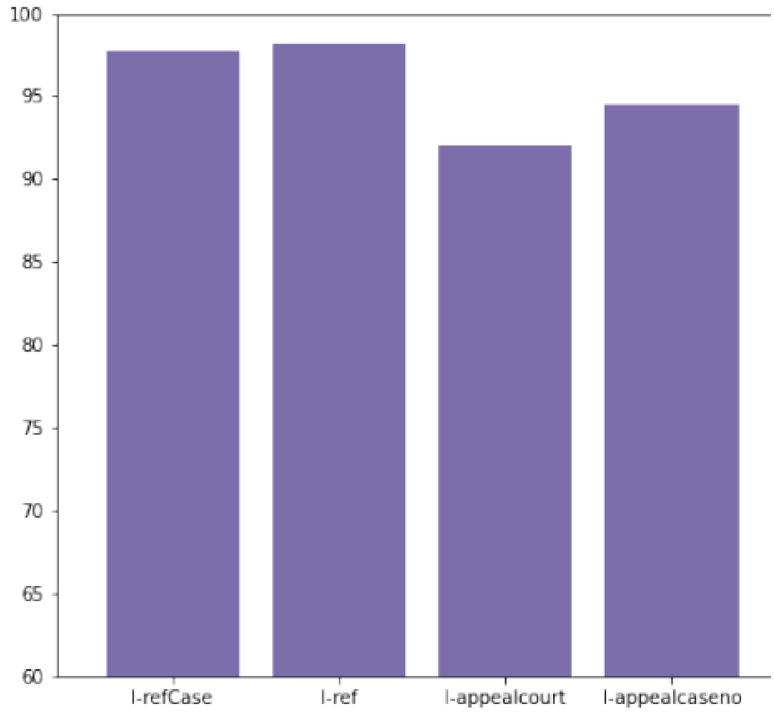


Figure 5.3: F1 score for I labels

Labels	F1-score	Labels	F1-score
B-refCourt	89.5	I-refCourt	87.54
B-refCase	98.82	I-refCase	97.75
B-ref	92.17	I-ref	98.21
B-appealcourt	90.77	I-appealcourt	91.97
B-appealcaseno	90.32	I-appealcaseno	94.37

Figure 5.4: F1 scores for individual labels

Labels	CRF model F1-score LHC	BERT F1 Score LHC	BERT F1 Score SCP
B-refCourt	97.16	92.31	89.5
I-refCourt	97.08	93.94	87.54
B-ref	87.26	91.02	92.17
I-ref	93.64	92.59	98.21
B-refCase	98.72	100.0	98.82
I-refcase	96.93	100.0	98.21

Figure 5.5: Comparison of BERT-base with CRF on LHC dataset

5.3.2 Siamese neural network

Train the neural network using a metric learning algorithm designed to minimize the distance between samples from the same category in the resulting feature space, while maximizing the distances between samples from different categories. At this stage we evaluate the model and the results are shown. In short we were able to get the final result as 13/14 which is more than 90 percent.

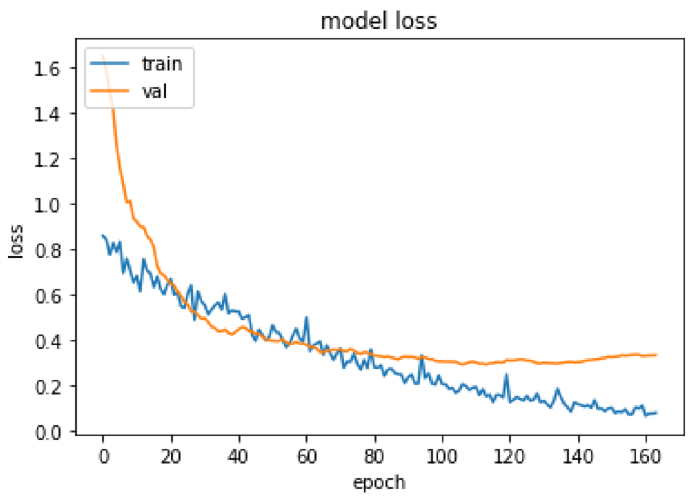


Figure 5.6: Model Accuracy

The baseline method results are shown in Figure 5.7.

Approach	τ	ρ	MSE
MALSTM (Baseline)	0.8822	0.8345	0.2286
LSTM	0.8831	0.8364	0.2195
GRU + Attention	0.8843	0.8372	0.2163
LSTM + Attention	0.8886	0.8386	0.2142
Bi-directional GRU	0.8896	0.8390	0.2125
GRU†	0.8901	0.8396	0.2112

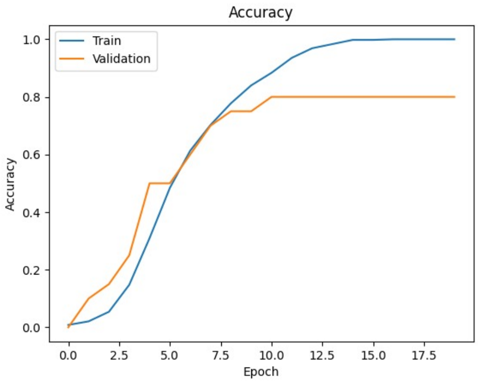
Figure 5.7: Evaluation Metrics for SCP dataset

The results show that the models performed better on the SCP dataset. The reason for this is because in the SICK dataset and STS benchmark dataset, the similarity is a score between the [1-5]. Where as in SCP dataset, the similarity is 1 for similar and 0 for non similar. Resulting in the better scores for the test data.

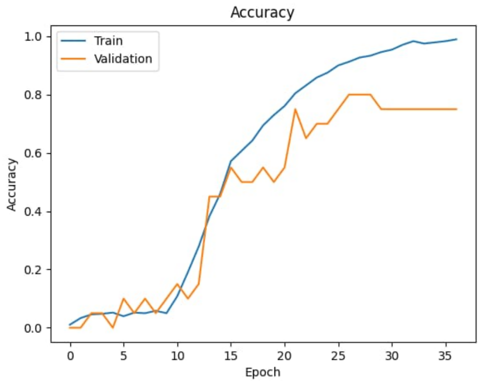
5.3.3 LMTC

We used various models and embedding techniques to classify the case matter, which can further be utilized for similarity evaluation of the documents by taking the laws applied in that case into account. The Meta data is available on the SCP website, and is used along with the text. The preprocessing techniques are used to make the hierarchy of the case subjects. Lately, the SCP website is assigning single case subjects, so our data is also single-labeled.

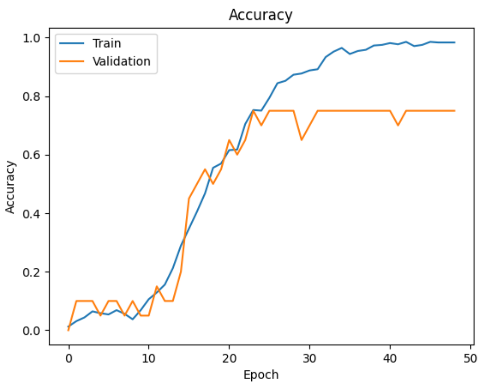
The results show that BERT base and BERT LAWN achieves the accuracy score of 80% and 77% respectively. Thus showing that the BERT-BASE performs better than BERT-LAWN. However, the paper shows that BIGRU performs better on the EURLEX57K dataset with the 76% score (See Figure 5.8 5.9). We achieved better results on our dataset with the same architectures but different dataset. We considered the case title, judge, the main body and dates of the case as input params.



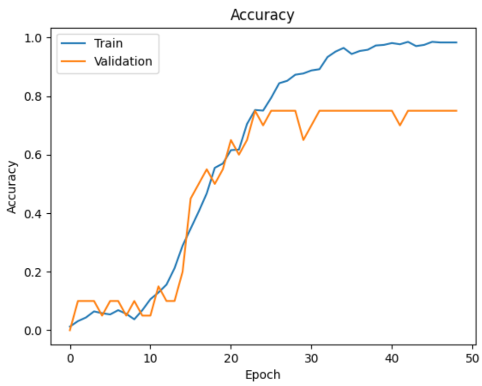
(a) BERT-Base Accuracy



(b) BERT-LAWN Accuracy

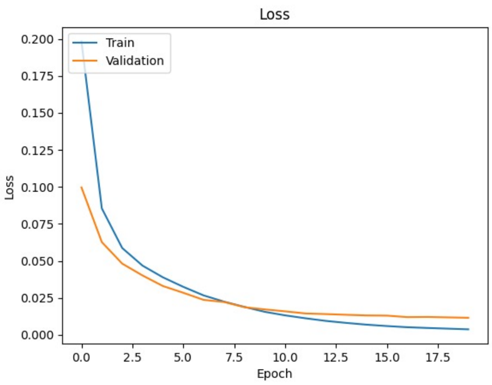


(c) BIGRU Accuracy

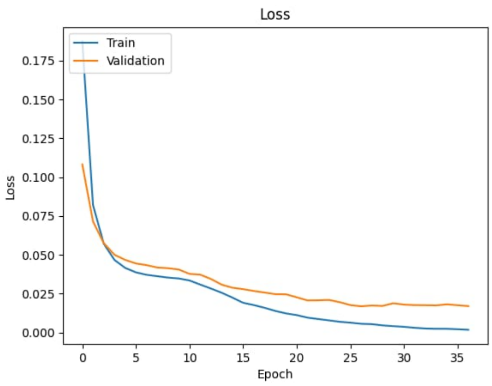


(d) LAWN BIGRU Accuracy

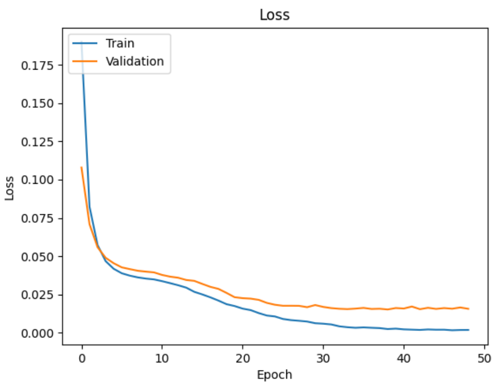
Figure 5.8: Model Accuracy



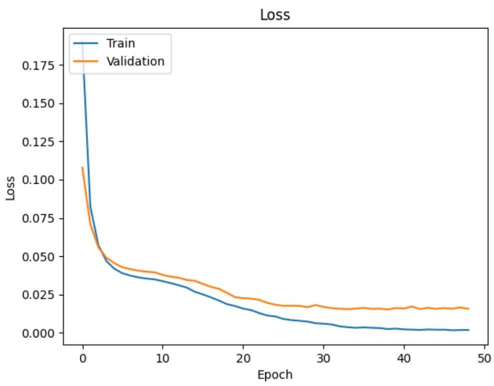
(a) BERT-Base Loss



(b) BERT-LAWN Loss



(c) BIGRU Loss



(d) LAWN BIGRU Loss

Figure 5.9: Model Losses

We have got the following recall and precision scores on SCP dataset (see Figure 5.10)

	Overall:			Few labels:			Zero label:		
	R@1	P@1	RP@1	R@1	P@1	RP@1	R@1	P@1	RP@1
BERT-LAWN	0.750	0.750	0.750	1.000	1.000	1.000	0.200	0.200	0.200
BERT-BASE	0.050	0.050	0.050	0.067	0.067	0.067	0.200	0.200	0.200
BIGRU	0.650	0.650	0.650	0.867	0.867	0.867	0.000	0.000	0.000
BIGRU-LAWN	0.050	0.050	0.050	0.067	0.067	0.067	0.200	0.200	0.200

Figure 5.10: Comparison of LMTC Results on SCP dataset

The results show that the BIGRU performs better on few labels, where as overall, BERT-LAWN perform better specially for Zero-label data set (with no occurrences in the training data).

Chapter 6

Conclusion

Massive volumes of textual information are produced by Pakistani courts in the form of proceedings and decisions. These documents are subsequently made available to the public to promote awareness and provide direction. It gets more and more challenging for a human to manually process the growing amount of accessible data. System developed based on machine learning and deep learning models to retrieve similar documents offers the possibility of reduce the time taken for relevant document searching from such a large and versatile data. Labeled data is necessary for the development of artificial intelligence based systems because it trains the algorithm to anticipate labels that may be used to extract critical information from a particular document.

To sum it up, our system is currently pretty similar to being desirable. Its current performance can also be enhanced, and it can be expanded with additional features. However, the efficiency of all the Pakistani law stake holders and the judiciary of Pakistan would increase straight away if they utilize this system immediately. Also we will get some insights, features importance and user feedback, that can be used to improve the system performance. The results will undoubtedly improve if we get access to additional labelled data and revisit the model design. One more feature that may be added to the system that, in our opinion, will be very useful is the ability to identify the text that was used to generate the prediction that the two papers are similar. Also if the dataset for similar document is provided, the system performance can be highly improved. This will be very beneficial in terms of confirming the behaviour of the model and laying the foundation for its improvement.

The performance of the system can also be improved if the system is assisted with the legal ontology and knowledge graphs. Both of these features needs legal experts for its creation as it requires a lot of domain knowledge and the case processing life cycle details.

We have got 92% F1 score for NER, 90% Accuracy for Semantic text similarity and 80% accuracy for document classification.

6.1 Future Work

The inclusion of judgements from other categories and case types in the training dataset is a noteworthy development for the current study. These judgments would includes the judgements from all other courts of Pakistan including the High courts of Pakistan rather than just only from the Supreme Court of Pakistan. A cases might have multiple matters/case subject being addressed. The results can be further improved by adding these multiple matter/case subject to the case meta data. Addition of knowledge graph and ontologies will also improve the results, and will be help specially dealing with different categories and case types. RNN (and GRU) based techniques are computationally expensive, particularly for lengthy texts. We want to research more effective computational techniques, such as dilated CNNs and Transformers. In an effort to overcome its length restrictions, we also want to test various hierarchical flavours of BERT.

Bibliography

[1] Reed C. Lawlor. "What Computers Can Do: Analysis and Prediction of Judicial Decisions". In: *American Bar Association Journal* 49.4 (1963), pp. 337–344. ISSN: 00027596, 21627975. URL: http://www.jstor.org/stable/25722338 (visited on 07/18/2022).

[2] Agnar Aamodt and Enric Plaza. "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches". In: *AI Communications* 7 (Aug. 2001), pp. 39–59. DOI: 10.3233/AIC-1994-7104.

[3] L Breiman. "Random Forests". In: *Machine Learning* 45 (Oct. 2001), pp. 5–32. DOI: 10.1023/A:1010950718922.

[4] Juan Enrique Ramos. "Using TF-IDF to Determine Word Relevance in Document Queries". In: 2003.

[5] Stephan Bloehdorn and Andreas Hotho. "Boosting for Text Classification with Semantic Features". In: vol. Vol. 3932. Aug. 2004, pp. 149–166. ISBN: 978-3-540-47127-1. DOI: 10.1007/11899402_10.

[6] Janet L. Kolodner. "An introduction to case-based reasoning". In: *Artificial Intelligence Review* 6 (2004), pp. 3–34.

[7] Wael H. Gomaa and Aly A. Fahmy. "Article: A Survey of Text Similarity Approaches". In: *International Journal of Computer Applications* 68.13 (Apr. 2013). Full text available, pp. 13–18.

[8] Richard Socher et al. "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank". In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle, Washington, USA: Association for Computational Linguistics, Oct. 2013, pp. 1631–1642. URL: https://aclanthology.org/D13-1170.

[9] Shereen Albitar, Sébastien Fournier, and Bernard Espinasse. “An Effective TF/IDF-based Text-to-Text Semantic Similarity Measure for Text Classification”. In: Oct. 2014. ISBN: 978-3-319-11748-5. DOI: 10.1007/978-3-319-11749-2_8.

[10] Yimin Chen, Niall J. Conroy, and Victoria L. Rubin. “Misleading Online Content: Recognizing Clickbait as "False News"”. In: WMDD ’15. Seattle, Washington, USA: Association for Computing Machinery, 2015, pp. 15–19. ISBN: 9781450339872. DOI: 10.1145/2823465.2823467. URL: https://doi.org/10.1145/2823465.2823467.

[11] Prakhar Biyani, Kostas Tsioutsiouliklis, and John Blackmer. “"8 Amazing Secrets for Getting More Clicks": Detecting Clickbaits in News Streams Using Article Informality”. In: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. AAAI’16. Phoenix, Arizona: AAAI Press, 2016, pp. 94–100.

[12] Abhijnan Chakraborty et al. “Stop Clickbait: Detecting and preventing clickbaits in online news media”. In: Aug. 2016, pp. 9–16. DOI: 10.1109/ASONAM.2016.7752207.

[13] Martin Potthast et al. “Clickbait Detection”. In: *ECIR*. 2016.

[14] Bingfeng Luo et al. “Learning to Predict Charges for Criminal Cases with Legal Basis”. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen, Denmark: Association for Computational Linguistics, Sept. 2017, pp. 2727–2736. DOI: 10.18653/v1/D17-1289. URL: https://aclanthology.org/D17-1289.

[15] Octavia-Maria Șulea et al. “Predicting the Law Area and Decisions of French Supreme Court Cases”. In: (Aug. 2017).

[16] Damir Cavar, Jon Herring, and Alison Meyer. “Law Analysis using Deep NLP and Knowledge Graphs”. In: 2018.

[17] Kankawin Kowsrihawat, Peerapon Vateekul, and Prachya Boonkwan. “Predicting Judicial Decisions of Criminal Cases from Thai Supreme Court Using Bi-directional GRU with Attention Mechanism”. In: *2018 5th Asian Conference on Defense Technology (ACDT)* (2018), pp. 50–55.

[18] Yi-Hung Liu and Yen-Liang Chen. "A two-phase sentiment analysis approach for judgement prediction". In: *Journal of Information Science* 44.5 (2018), pp. 594–607. DOI: 10.1177/0165551517722741. eprint: https://doi.org/10.1177/0165551517722741. URL: https://doi.org/10.1177/0165551517722741.

[19] Daniel López-Sánchez et al. "Hybridizing metric learning and case-based reasoning for adaptable clickbait detection". In: *Applied Intelligence* 48 (Sept. 2018). DOI: 10.1007/s10489-017-1109-7.

[20] Yinglong Ma, Peng Zhang, and Jiangang Ma. "An Ontology Driven Knowledge Block Summarization Approach for Chinese Judgment Document Classification". In: *IEEE Access* 6 (2018), pp. 71327–71338.

[21] Hiroki Nakayama et al. "doccano: Text annotation tool for human". In: *Software available from https://github.com/doccano/doccano* (2018).

[22] Matthew Peters et al. "Deep contextualized word representations". In: (Feb. 2018).

[23] Hai Ye et al. "Interpretable Charge Predictions for Criminal Cases: Learning to Generate Court Views from Fact Descriptions". In: *NAACL*. 2018.

[24] Ilias Chalkidis, Ion Androutsopoulos, and Nikolaos Aletras. "Neural Legal Judgment Prediction in English". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, July 2019, pp. 4317–4323. DOI: 10.18653/v1/P19-1424. URL: https://aclanthology.org/P19-1424.

[25] Ilias Chalkidis et al. "Large-Scale Multi-Label Text Classification on EU Legislation". In: *ACL*. 2019.

[26] Jewel Chanda. "A Scientific Judicial Perspective can solve many hurdles of practical application of AI 'expert system' for judicial decision making". In: 8 (July 2019), p. 16.

[27] Deqin Huang and Weizhen Lin. "A Model for Legal Judgment Prediction Based on Multi-model Fusion". In: *2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE)* (2019), pp. 892–895.

[28] Afnan Iftikhar, Syed Waqar Ul Qounain Jaffry, and Muhammad Kamran Malik. "Information Mining From Criminal Judgments of Lahore High Court". In: *IEEE Access* 7 (2019), pp. 59539–59547. DOI: 10.1109/ACCESS.2019.2915352.

[29] Shang Li et al. "Element-Aware Legal Judgment Prediction for Criminal Cases with Confusing Charges". In: *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)* (2019), pp. 660–667.

[30] Shangbang Long et al. "Automatic Judgment Prediction via Legal Reading Comprehension". In: *CCL*. 2019.

[31] Marko Mihajlović and Ning Xiong. "Finding the most similar textual documents using Case-Based Reasoning". In: *ArXiv* abs/1911.00262 (2019).

[32] Kody Moodley et al. "Similarity and Relevance of Court Decisions: A Computational Study on CJEU Cases". In: *Frontiers in Artificial Intelligence and Applications* 322 (Dec. 2019), pp. 63–72. DOI: 10.3233/FAIA190307.

[33] Raghavendra Pappagari et al. "Hierarchical Transformers for Long Document Classification". In: Dec. 2019, pp. 838–844. DOI: 10.1109/ASRU46091.2019.9003958.

[34] Ilias Chalkidis et al. "An Empirical Study on Large-Scale Multi-Label Text Classification Including Few and Zero-Shot Labels". In: Jan. 2020, pp. 7503–7515. DOI: 10.18653/v1/2020.emnlp-main.607.

[35] Mingjie Ling et al. "Hybrid Neural Network for Sina Weibo Sentiment Analysis". In: *IEEE Transactions on Computational Social Systems* 7.4 (2020), pp. 983–990. DOI: 10.1109/TCSS.2020.2998092.

[36] Kongfan Zhu et al. "Sequence Generation Network Based on Hierarchical Attention for Multi-Charge Prediction". In: *IEEE Access* 8 (2020), pp. 109315–109324. DOI: 10.1109/ACCESS.2020.2998486.

[37] Fethi Fkih. "Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison". In: *Journal of King Saud University - Computer and Information Sciences* (2021). ISSN: 1319-1578. DOI: https://doi.org/10.1016/j.jksuci.2021.09.014. URL: https://www.sciencedirect.com/science/article/pii/S1319157821002652.

[38] Deepa Anand and Rupali Sunil Wagh. "Effective deep learning approaches for summarization of legal texts". In: *J. King Saud Univ. Comput. Inf. Sci.* 34 (2022), pp. 2141–2150.

[39] *"Pendency of Cases in SC Hits Highest Level in Country's History." The Express Tribune, 7 Feb. 2020.* tribune.com.pk/story/2152445/pendency-cases-sc-hits-highest-level-countrys-history.

[40] *Information Extraction from Court Room Records.* Nida Ahmed, MS thesis, SEECS NUST.

[41] Krithika Iyer ksiyer. "Classification of Legal Text". In.

[42] *Reiling.* https://home.hccnet.nl/a.d.reiling/html/180627%20AI%20and%20justice%20for%20CEPEJ.pdf. Accessed: 2022-06-14.

[43] Reporter; The Newspaper's Staff. *"44,500 Cases Are Pending before SC." DAWN.COM, 28 May 2020.* www.dawn.com/news/1559757/44500-cases-are-pending-before-sc.

[44] *Supreme Court of Pakistan.* https://www.supremecourt.gov.pk. Accessed: 2022-06-14.