Proc. of the International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME) 16-18 November 2022, Maldives

Comparison of Transformer Models for Information Extraction from Court Room Records in Pakistan

Nida Ahmed[1], Seemab Latif[1], Rabia Irfan[1], Adnan Ul-Hasan[2], and Faisal Shafait[1,2]
[1] School of Electrical Engineering and Computer Sciences (SSECS),
National University of Science and Technology (NUST),
Islamabad, Pakistan
[2] Deep Learning Laboratory, National Center of Artificial Intelligence (NCAI),
Islamabad, Pakistan
{nahmed.mscs18seecs, seemab.latif, rabia.irfan, adnan.ulhassan, faisal.shafait}@seecs.edu.pk

***Abstract*—The legal domain has many opportunities when it comes to improvement and innovation through computational advancements. In Pakistan, as the number of reported judgments continues to grow at a rapid rate, it has become essential to process this massive chunk of data to better meet the requirements of the respective stakeholders. However, extracting the required information from this unstructured legal text is challenging. In this paper, we have compared different variations of BERT to see which would be more suited for a machine learning system that can automatically extract information from these publicly available judgments of the Supreme Court of Pakistan. A labelled dataset comprising of thirteen entities has been created using the publicly available legal judgments from the Supreme Court. Different pre-trained BERT models, namely $\text{BERT}_{\text{BASE}}$-uncased, $\text{BERT}_{\text{BASE}}$-cased and LegalBERT, are then further trained and fine-tuned on the created dataset for Named Entity with F1 scores of 92.47%, 94.72% and 92.51% respectively. The BERT models have been found to improve the F1 scores of previous studies on a dataset available from Lahore High Court, having smaller number of labels, with the F1 scores of 82.3%, 93.21% and 85.06%, respectively.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

A legal judgment is a document that contains the outcome of a legal case proceedings. It often times sets the precedence for future law suites and petitions. It can therefore be very beneficial to digitize these complex documents and index the information they contain for efficient future reference. The length and complexity of these legal judgments makes it difficult for human beings to acquire important information from them. The process not only becomes time consuming but also error-prone due to the large sizes of data that need to be examined to prepare a single case. Therefore, there is a dire need for an intelligent system that can extract the important information from these judgments. Information extraction can be one of the core systems on which other tasks could be dependent, for example, similar case retrieval system, data anonymization system, semantic search, etc.

Some of the challenges in the field of NER in the context of legal judgments include the types of named entities that

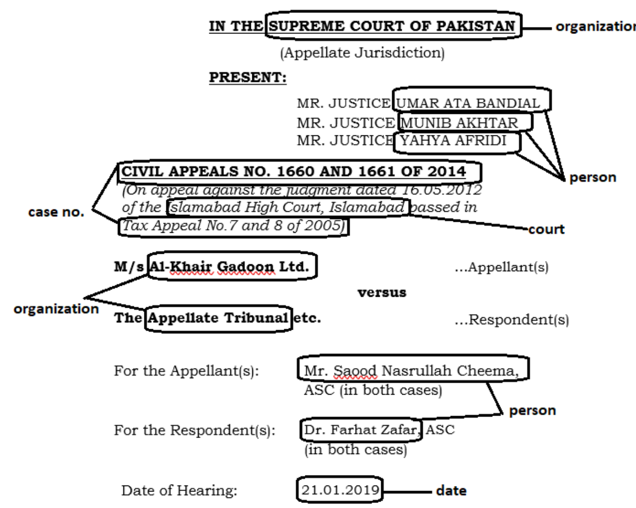

Fig. 1. Important information that needs to be extracted from a judgment includes the names of parties involved be it people or organizations, people attending, case number of both the current case and the case being appealed, dates of hearing, judges present, etc., alongside laws and referred cases that might be present in the body of the judgment.

may need to be extracted from court judgments. Examples of this include the different case and law references that are mentioned in judgments. Some of the important information can be seen in Figure 1 which shows the first page of a judgment for a Civil Appeal case, released by the Supreme Court of Pakistan. Another challenge is the same names often fall under different named entity labels depending on the context. For example, an organization's name can be based on a person's name, e.g., Gerry's foundation, which will confuse the NER system to extract it as a person's name (Gerry) or as an organization.

Very limited work has been done in Pakistan for judicial automation. There has been a recent interest among local researchers to use machine learning algorithms for legal entity extraction but the work has been limited to using conventional machine learning techniques like Conditional Random Fields (CRF), Hidden Markov Models (HMM), etc. Sharafat et al. [1] used civil judgments to extract 10 named entities and presented the results of three different models, namely, CRF, MEMM and HMM for the said dataset. Iftikhar et al. [2] used CRF, MaxEnt and TNT models for information extraction from criminal case judgments from Lahore High Court judgments. These methods only work under certain conditions, and the available dataset does not cover all the important information that may be present in a judgment.

There is a good opportunity to propose a more robust algorithm to employ the latest deep learning based entity extraction algorithms. Taking this into consideration, we compare different variations of BERT for a robust information extraction approach for court judgments. We have presented the results of different available variations of BERT on two different datasets based on the judgments of the Supreme Court of Pakistan (SCP) and Lahore High Court (LHC) containing entities that can help extract important information from the judgment text. It can be implemented in all court judgments regardless of the level of court as long as the case category remains the same.

The remainder part of the paper is organized in the following sections. Section II presents the background of several models that are designed for information extraction tasks. Section III discusses the models for the entity recognition tasks and describes our process for the dataset preparation. Section V presents the experiments and analysis of the results in detail. Finally, our conclusion and future work are described in the last section.

II. Related work

Legal domain is one of the specialized areas where NER is being used to extract important information from text including contracts, court records, case judgments, etc. The main reason NER is different for the legal domain is that there are domain specific named entities like court names, laws, and legislation and there is often an overlap in the named entities that fall under different labels.

Dozier et al. [3] proposed three different methods for Named Entity Extraction and Resolution including look-up methods, context rule based methods and statistical models. Leitner et al. [4] created a dataset consisting of German court decisions and used CRFs and Bi-LSTMs for Named Entity Recognition from said dataset. The proposed work was carried out under the European LYNX project which includes the development of a semantic platform for the creation of different document processing applications in the legal domain.

Skylaki [5] proposed the use of Pointer Generator Network for NER in the noisy text obtained from PDF files of US court judgments by formulating the NER task as a text-to-text sequence generation task and then training a pointer generator network to generate the entities in the document rather than labeling them. Wang et al. [6] proposed a Sequence Tagging Model (STM) that was created by combining an Inter-Dilated Convolution Neural Network (IDCNN) and a Bi-LSTM model. The model could be used for large scale data from Brazilian legal documents. The paper also compared the results of the model with IDCNN-CRF based model. Nuranti and Yulianti [7] proposed a method using the Bi-LSTM and CRF combination for the recognition of ten legal entities in Indonesian court decision documents. Chou and Hsing [8] used different text mining techniques on Chinese written judgments for criminal cases. Badji et al. [9] explored the effectiveness of different NER systems of Dutch Court rulings for the goal of information extraction and de-identification of entities in the published rulings.

Limited work has been done for judicial automation in Pakistan, including information extraction from judgments from courts of Pakistan. Iftikhar et al. [2] introduced the PULMS that is based on three algorithms, namely, CRF, MaxEnt and TNT, trained on manually annotated dataset to extract named entities from criminal case judgments and achieved the highest F1 score for the CRF model. Sharafat et al. [1] used three different models, CRF, MEMM, and HMM to extract 10 named entities from civil court proceedings and compared the results for different named entities with IBO tagging as well as IO tagging schemes with the highest F1 score of 86.62% using the CRF model.

With the introduction of pre-trained transformer based models like BERT [10], it is now possible to get state-of-the-art results by fine-tuning these models on a relatively smaller dataset. A version of BERT has also been released that has been pre-trained on legal documents and cases, called the LegalBERT [11]. In this study, we compare the results of three different pre-trained BERT models, namely, $\text{BERT}_{\text{BASE}}$-uncased, $\text{BERT}_{\text{BASE}}$-cased, and LegalBERT. We compare the results of these models on the dataset used by Sharafat et al. [1] and also on a dataset we created using Civil Appeal Judgments published by the Supreme Court of Pakistan (SCP). Details of these datasets have been mentioned in the upcoming sections.

III. Methodology

BERT (Bidirectional Encoder Representations from Transformers) was published by Devlin et al. [10], a team of researchers at Google AI Language. The key innovation of BERT is its bi-directionality enables it to understand the context of words depending on their surrounding words. We explore three different variations of BERT, a state-of-the-art model that has achieved exceptional results in different NLP tasks, for legal entity extraction from two different datasets of court judgments from the courts of Pakistan.

A. Dataset Creation

In our work, we not only used the dataset created by [1] that consists of civil judgments from the Lahore High Court of Pakistan, but also created our own dataset using Civil Appeal judgments from the Supreme Court of Pakistan (highest appellate court of the country). In order to create a dataset of the appropriate size, we decided to work with the judgments for the '*Civil Appeal*' case type as this category had the highest number of available cases on the Supreme Court website[1].

[1]https://www.supremecourt.gov.pk/

The relevant judgments were downloaded and then processed through a preprocessing pipeline consisting of check-sum validation. Furthermore, records containing documents in the Urdu language were filtered out and the remaining English distinct judgment count was 214. The length of each judgment varies from 3 to 40 pages. After consulting with law specialists, we decided on a list of 14 named entities for information extraction and created their annotation guidelines was created as shown in Table I. The judgments were then annotated using the open source annotation tool Doccano [12]. A dataset was then created from these annotated documents that followed the IOB (Inside, Outside, and Beginning) format.

B. Model Architecture

BERT's architecture consists of a multi-layer bidirectional transformer encoder based on the original implementation by Vaswani et al. [13]. BERT uses transformers, an attention mechanism having the ability to learn contextual relations between words in phrase(s).

$\text{BERT}_{\text{BASE}}$ has a total of 12 transformer blocks, Hidden size of 768 and 12 self attention heads (L=12, H=768, A=12) with the total number of parameters coming up to 110 million. $\text{BERT}_{\text{BASE}}$ is pre-trained on the BookCorpus dataset [14] that consists of $11,038$ unpublished books and the English Wikipedia (2,500M words). There are two variations of $\text{BERT}_{\text{BASE}}$ available, namely, $\text{BERT}_{\text{BASE}}$-uncased and $\text{BERT}_{\text{BASE}}$-cased where the only difference is the absence of word casing in the uncased version. LegalBERT [11] also has the same architecture as $\text{BERT}_{\text{BASE}}$ which has been trained on domain specific dataset consisting of $116,062$ documents of EU legislation, $61,826$ documents of the UK legislation, $19,867$ cases from the European Court of Justice, $12,554$ cases from HUDOC repository of European Court of Human Rights, $164,141$ cases from various courts in the USA as well as $76,366$ US Contracts from the EDGAR database.

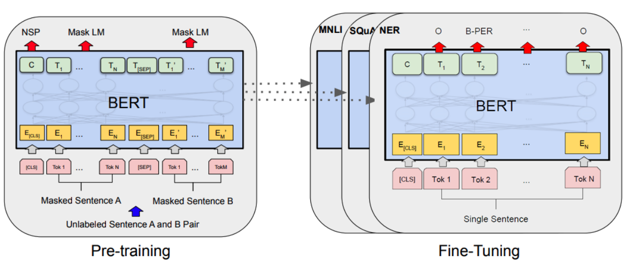

Fig. 2. To use BERT for specific NLP tasks like NER, the pre-trained BERT is fine-tuned using the dataset for said task; in our case, we fine-tune it using the dataset consisting of tokens and their entity labels.

Contrary to traditional models, transformers read the whole sequence of words concurrently and thus are considered bi-directional. BERT resolves constraint of uni-directionality with the use of the Masked Language Model (MLM) pre-training objective. In order to use BERT for a specific NLP task, the pre-trained model can be further trained and fine-tuned on domain specific data. For legal NER, we compared the results of $\text{BERT}_{\text{BASE}}$-uncased, $\text{BERT}_{\text{BASE}}$-cased, and LegalBERT.

TABLE I
LIST OF CHOSEN NAMED ENTITY LABELS SELECTED FOR INFORMATION EXTRACTION FROM COURT JUDGMENTS WITH THEIR DESCRIPTION AND EXAMPLES

Named Entity Label	**Description**	**Example**
Per	Name of a person; including judges, lawyers and others involved	Moazzam Ali Khan, Mian Muhammad Nawaz Sharif
Loc	Any location, name of a place or address	Islamabad, Lahore, DHA Karachi
Org	Name of an organization or business	NADRA, PMDC, Ziauddin Medical University
CaseNo	Assigned case number of the case being decided in the judgment	Civil Appeal 718 of 2007, CA 1366/2007
Resp	Name of respondent in the case be it an organization or a person	Najmul Hassan Kazmi, Ministry of Defense Rawalpindi
Date	Any complete date mentioned	12.07.1997, 2nd October, 2015
Refcourt	Court that passed a referred judgment	Calcutta High Court
Refcase	Any past judgment being referred to	Muhammad Akram V Altaf Ahmad (PLD 2003 Supreme Court 688)
Ref	Law references mentioned	Section 115 of the Code of Civil Procedure
Appealcourt	Court that passed the judgment that is being appealed against	High Court of Sindh, Karachi
Appealcaseno	Case number of the case being appealed against in the current case	WP.No. 11983/2005, Constitution Petition No. D-1807/1999
Money	Any monetary amount mentioned, includes fines, debts etc.	Rs.50,000, One lac Rupees.
FIRno	First Investigation Report (FIR) number that might be filed with the police	FIR.No.256/2013, FIR No. 324/12
Approved	Whether the judgment is approved for reporting	Approved

IV. Experiments

In order to find the best model for legal information extraction, we fine-tuned the previously mentioned three different $\text{BERT}_{\text{BASE}}$ models on two different datasets consisting of court judgments from the courts of Pakistan to evaluate the performance of these models for legal information extraction.

The hyper-parameters for these experiments were kept the same to have a more accurate comparison of each model's performance. Each of the models was trained for 30 epochs with a learning rate of $2e^{-4}$. The details of the datasets and their respective train-test split percentage for these experiments is given in Sections IV-A and IV-B.

A. LHC Dataset

The dataset used by Sharafat et al. [1] consists of 100 Civil proceeding judgments from Lahore High Court, Pakistan. The exact case type or category is not mentioned and the cases were chosen at random. The dataset contains a total of 10 named entities and is labelled using the IOB format. The annotation guidelines for the labels in the dataset are mentioned in [1]. The names of the labels and their counts are given in Table II. In order to have a fair comparison with the results published by Sharafat et al. [1], we used the same $90-10$ split where 90% of the data was used for training and 10% of the dataset was used for testing.

TABLE II
LABELS PRESENT IN THE LAHORE HIGH COURT DATASET AND THEIR RESPECTIVE COUNTS IN IBO FORMAT.

Labels	Count	Labels	Count
B-per	1,081	I-per	1,602
B-loc	255	I-loc	217
B-org	289	I-org	918
B-caseNo.	147	I-caseNo.	485
B-Misc.name	297	I-Misc.name	573
B-date	879	I-date	66
B-refCourt	475	I-refCourt	576
B-ref	422	I-ref	2,405
B-refCase	243	I-refCase	605
B-money	109	I-money	63

B. SCP Dataset

The dataset was created using 214 civil appeal judgments, after pre-processing, from the Supreme Court of Pakistan is annotated for 14 named entities labelled in the IOB format. This dataset contains some named entities that were not present in the Lahore High Court dataset e.g appealcase, appealcourt, etc. The names of the labels and their counts are given in Table III. This dataset was over twice the size of the LHC dataset so we used the standard $80-20$ test-train split where 80% of the data was used for training and 20% of the dataset was used for testing.

TABLE III
LABELS PRESENT IN THE SUPREME COURT DATASET AND THEIR RESPECTIVE COUNTS IN IBO FORMAT.

Labels	Count	Labels	Count
B-per	4,961	I-per	9,749
B-loc	1,703	I-loc	1,163
B-org	3,050	I-org	4,934
B-caseno	1,497	I-caseno	5,151
B-resp	487	I-resp	3,048
B-date	3,850	I-date	1,854
B-refcourt	306	I-refcourt	888
B-refcase	2,301	I-refcase	32,576
B-ref	4,775	I-ref	32,099
B-appealcourt	422	I-appealcourt	1,778
B-appealcaseno	770	I-appealcaseno	3,990
B-money	446	I-money	208
B-FIRno	23	I-FIRno	52
B-Approved	160	I-Approved	0

V. Results and Analysis

In order to measure the performance of the $\text{BERT}_{\text{BASE}}$ models, we calculate the F1 score for each named entity and average F1 score for each model.

A. LHC Results

TABLE IV
F1 SCORES OF CRF (AS PUBLISHED BY SHARAFAT ET AL. [1]), $\text{BERT}_{\text{BASE}}$-UNCASED AND $\text{BERT}_{\text{BASE}}$-CASED FOR INDIVIDUAL LABELS OF THE LHC DATASET.

Labels	CRF model F1-score [1]	$\text{BERT}_{\text{BASE}}$-uncased F1-score	$\text{BERT}_{\text{BASE}}$-cased F1-score	LegalBERT F1-score
B-per	94.28	94.21	97.94	94.35
I-per	96.72	96.97	96.58	98.08
B-loc	83.68	70.58	85.62	74.07
I-loc	65.79	71.26	72.73	80.49
B-org	75.04	36.36	70.43	51.85
I-org	84.11	75.00	87.50	84.21
B-caseNo.	94.06	83.33	96.29	82.35
I-caseNo.	97.57	92.80	98.63	80.00
B-Misc.name	70.64	85.71	88.89	73.91
I-Misc.name	71.91	81.81	95.24	81.89
B-date	97.49	99.46	100.0	98.4
I-date	84.0	-	100.0	-
B-refCourt	97.16	91.17	92.31	95.65
I-refCourt	97.08	92.85	93.94	97.61
B-ref	87.26	88.00	91.02	84.0
I-ref	93.64	96.49	92.59	96.51
B-refCase	98.72	98.03	100.0	75.47
I-refCase	96.93	100.0	100.0	81.65
B-money	91.15	93.33	100.0	85.71
I-money	55.10	100.0	100.0	100.0
Average	86.62	82.3	93.21	85.06

For the Lahore High court dataset used by [1], the highest published F1 score was 86.62% using a CRF model for Named Entity Recognition. The highest F1 score achieved by the CRF model for this dataset is 98.72 for the label of B-refCase. The lowest F1 score achieved was 55.10% for the label I-money and the second lowest was 65.79% for the label I-loc.

Using $\text{BERT}_{\text{BASE}}$-uncased we got an F1 score of 82.3%. This drop in F1 score is mainly due to the lower f1 score for the B-org label and the model not being able to identify the I-date entities. The main cause of the drop in the f1 score of the B-org was due to the model labelling the word 'The' in the names of an organization as 'O' (Other) due to the text being lower-cased. The model achieved the highest f1 score for the labels I-money and I-refcase.

With $\text{BERT}_{\text{BASE}}$-cased, we managed to achieve an F1 score of 93.21 which is considerably higher than the previously published results for this dataset. Using $\text{BERT}_{\text{BASE}}$-cased, we achieved the highest F1 score of 100 for the labels B-date, I-date, B-refCase, I-refCase, B-money, and I-money. The lowest F1 score observed was 70.43 for the label B-org. From a total of 20 labels, we achieved an F1 score greater than 90 for 15 of the labels.

Using LegalBERT, we achieved an f1 score of 85.06%. This is an improvement from the $\text{BERT}_{\text{BASE}}$-uncased model due to its pre-training dataset. We can see that this model could not correctly recognize any of the I-date entities either just like the other uncased model.

The comparisons of the F1 scores for all the individual entities in the dataset are as mentioned in Table IV. From these results, we can see that $\text{BERT}_{\text{BASE}}$-cased is better at recognizing most of the named entities than CRf models. $\text{BERT}_{\text{BASE}}$-cased also out performs its uncased variation as well as LegalBERT .

B. Supreme Court Results

Using $\text{BERT}_{\text{BASE}}$-uncased, we got an average F1 score of 92.47 for the Supreme Court dataset, with the labels B-FIRno, I-FIRno and B-Approved having the highest F1 score of 100 and most of the labels have an F1 score of over 80%.

With $\text{BERT}_{\text{BASE}}$-cased, we achieved an average F1 score of 94.72, with the labels B-FIRno, I-FIRno and B-Approved having the highest F1 score of 100 and most of the labels have an F1 score of over 90%.

Using LegalBERT, we achieved an average F1 score of 92.51% which is similar to the F1 score of $\text{BERT}_{\text{BASE}}$-uncased. When comparing the results of the two models, we can see that both the models have slightly lower F1 scores for the B-refCourt and I-refCourt labels as well as the B-appealcourt and I-appealcourt labels. This is due to the models confusing the court names that are refcourts to be appealcourts. It is important to note here that these two labels have a huge amount of overlap when it comes to the names of courts that are being labelled as either. Therefore, while was expected that that these labels would have a slightly lower F1 score both, the models have performed better than expected in recognizing these entities. Similarly, the labels B-resp and I-resp also have some overlap when it comes to the entities as the respondents in any given case can be a person, group of people or an organization.

The drop in the F1 scores of $\text{BERT}_{\text{BASE}}$-uncased and LegalBERT for the B-org and B-loc labels in this Supreme Court dataset is also similar to the Lahore High Court dataset, though not as extreme for the B-org label due to the fact that the word ‘The’ in the sentence was not labelled as B-org unless it was officially registered as the name of the organization.

The highest F1 scores achieved for this dataset were 100% for the entities ‘B-FIRno.’ and ‘I-FIRno.’. The lowest F1 score achieved was 87.54% for the entity ‘I-refCourt’. Only two labels have the F1 score below 90% which are ‘B-refCourt’ and ‘I-refCourt’. From Table V, we can see that BERT performs well on the Supreme Court Dataset that we created as well.

TABLE V
F1 SCORES OF $\text{BERT}_{\text{BASE}}$-UNCASED AND $\text{BERT}_{\text{BASE}}$-CASED FOR INDIVIDUAL LABELS OF THE SUPREME COURT DATASET.

Labels	$\text{BERT}_{\text{BASE}}$-uncased F1-score	$\text{BERT}_{\text{BASE}}$-cased F1-score	LegalBERT
B-per	92.47	93.67	93.09
I-per	96.65	96.49	96.83
B-loc	86.33	93.46	89.83
I-loc	90.90	94.90	91.38
B-org	85.56	98.88	85.89
I-org	87.95	90.02	88.32
B-caseno	94.76	96.25	94.37
I-caseno	96.56	96.55	96.59
B-resp	86.67	90.27	90.23
I-resp	89.30	91.30	90.65
B-date	98.33	96.96	98.19
I-date	93.78	96.22	94.54
B-refCourt	79.64	89.5	78.27
I-refCourt	81.15	87.54	82.78
B-refCase	93.73	98.82	94.10
I-refCase	97.90	97.75	98.07
B-ref	90.39	92.17	87.89
I-ref	94.23	98.21	93.18
B-appealcourt	89.03	90.77	87.42
I-appealcourt	93.85	91.97	89.83
B-appealcaseno	90.15	90.32	90.00
I-appealcaseno	93.79	94.37	94.43
B-money	95.00	96.77	95.62
I-money	98.70	95.20	96.20
B-FIRno	100	100	100
I-FIRno	100	100	100
B-Approved	100	100	100
I-Approved	-	-	-
Average	92.47	94.72	92.51

C. Analysis

From the previous sections, we can see that BERT, especially $\text{BERT}_{\text{BASE}}$-cased, achieves desirable results for information extraction from judgments from Pakistan’s courts. Based on its pre-training, BERT determines the context of a given word based on its surrounding words which means it can better predict the named entities in new data. This context based understanding of a given word also has a drawback that is very visible in the results where the model has some confusion between the labels of location and organization. This is much more visible in these datasets because the names of organizations and locations are oftentimes very similar as organizations can be named after people or locations, etc. This means that in a sentence/phrase ‘Person A from XYZ’ the model is not able to perfectly predict if XYZ is a location or an organization if it has not been trained on the dataset for that might have similar entities/words.

The two uncased models, $\text{BERT}_{\text{BASE}}$-uncased and LegalBERT don’t perform as well as $\text{BERT}_{\text{BASE}}$-cased, even with the pre-training using legal dataset in the case of LegalBERT. This is more noticeable in the ‘B-’ (Beginning) labels of the entities which show that the casing could help the model identify these entities better as visible from the results for $\text{BERT}_{\text{BASE}}$-cased. By comparing the results of $\text{BERT}_{\text{BASE}}$-uncased and $\text{BERT}_{\text{BASE}}$-cased, it would be safe to assume that a cased version of LegalBERT would perform even better than $\text{BERT}_{\text{BASE}}$-cased on these or other similar datasets.

VI. Conclusion

Courts are producing a massive amount of textual data in the form of legal proceedings/judgments which are publicly available for the sake of guidance and awareness. In the current study, we fine-tuned pre-trained BERT models on a dataset consisting of 100 civil judgments from different categories, as used by Sharafat et al. [1]. Using $\text{BERT}_{\text{BASE}}$-uncased, we achieved an F1 score of 82.3% for this dataset and an F1 score of 85.06% using LegalBERT. With $\text{BERT}_{\text{BASE}}$-cased, we achieved an F1 score of 93.21% which is a considerable improvement from the previously published highest F1 score of 86.62% for the said dataset.

Furthermore, a total of 214 Civil Appeal judgments from the Supreme Court of Pakistan were labelled with fourteen named entities. The pre-trained BERT models were then fine-tuned on this dataset to achieve an F1 score of 94.72% using the $\text{BERT}_{\text{BASE}}$-cased model while the $\text{BERT}_{\text{BASE}}$-uncased and LegalBERT models achieved F1 scores of 92.47% and 92.51% respectively.

In comparison to the results that have been reported previously, the results of $\text{BERT}_{\text{BASE}}$-cased appear to be promising, though it is important to note that the current results are only for Civil Appeal judgments. By including other categories of judgments in the dataset, we can increase confidence in these results. For this, we can label judgments from all the other categories in the Supreme Court of Pakistan. There might be some Named Entities that might be category-specific that can also be added for the said categories.

A variety of systems can also be built using the results from NER systems. For example, the extracted named entities can be used for certain question answering systems where named entities are required, a knowledge base can be populated by using the relationships between entities, any personal information of any individuals that might be mentioned in the judgment can also be anonymized using the results of NER, which can also be a future work.

References

[1] S. Sharafat, Z. Nasar, and S. W. Jaffry, "Legal data mining from civil judgments," in *International Conference on Intelligent Technologies and Applications*. Springer, 2018, pp. 426–436.

[2] A. Iftikhar, S. W. U. Q. Jaffry, and M. K. Malik, "Information mining from criminal judgments of lahore high court," *IEEE Access*, vol. 7, pp. 59 539–59 547, 2019.

[3] C. Dozier, R. Kondadadi, M. Light, A. Vachher, S. Veeramachaneni, and R. Wudali, "Named entity recognition and resolution in legal text," in *Semantic Processing of Legal Texts*. Springer, 2010, pp. 27–43.

[4] E. Leitner, G. Rehm, and J. Moreno-Schneider, "Fine-grained named entity recognition in legal documents," in *International Conference on Semantic Systems*. Springer, 2019, pp. 272–287.

[5] S. Skylaki, A. Oskooei, O. Bari, N. Herger, and Z. Kriegman, "Named entity recognition in the legal domain using a pointer generator network," *arXiv preprint arXiv: 2012.09936*, 2020.

[6] Z. Wang, Y. Wu, P. Lei, and C. Peng, "Named entity recognition method of brazilian legal text based on pre-training model," in *Journal of Physics: Conference Series*, vol. 1550, no. 3. IOP Publishing, 2020, p. 032149.

[7] E. Q. Nuranti and E. Yulianti, "Legal entity recognition in indonesian court decision documents using bi-lstm and crf approaches," in *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. IEEE, 2020, pp. 429–434.

[8] S. Chou and T.-P. Hsing, "Text mining technique for chinese written judgment of criminal case," in *Pacific-Asia workshop on intelligence and security informatics*. Springer, 2010, pp. 113–125.

[9] I. Badji, O. CORCHO, and V. RODRÍGUEZ-DONCEL, "Legal entity extraction with ner systems," Ph.D. dissertation, Master's thesis, Universidad Politécnica de Madrid, 2018. 23, 57, 2018.

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

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

[12] H. Nakayama, T. Kubo, J. Kamura, Y. Taniguchi, and X. Liang, "doccano: Text annotation tool for human," 2018, software available from https://github.com/doccano/doccano. [Online]. Available: https://github.com/doccano/doccano

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[14] Y. Zhu, R. Kiros, R. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, and S. Fidler, "Aligning books and movies: Towards story-like visual explanations by watching movies and reading books," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 19–27.