EJELS
EUROPEAN JOURNAL OF EMPIRICAL LEGAL STUDIES

EJELS 1(2) (2024): 179–206
https://doi.org/10.62355/ejels.24862

RESEARCH ARTICLE

How Effective Is the Judiciary? Evidence on Correlation Between Cases' Characteristics and Probability of Appeal

Maciej Świtała*

Article History

Submitted 2 Jun 2024.
Accepted 12 Oct 2024.

Keywords

effectiveness, judiciary, probability of appeal, topic model

Abstract

This research proposes a way to assess judicial effectiveness, proxied by the probability of appeal of a decision. Focusing on the example of regional courts in Poland, it classifies cases based on their most accurate topic, creating a topic model on judgements. This classification is used to provide descriptive evidence on cases' characteristics and their correlation with a higher or lower probability of appeal. The obtained results indicate that topic-based groups that are more heterogeneous in the legal departments of the associated cases are more likely to be appealed.

1 Introduction

The efficiency and effectiveness of the judiciary are issues frequently considered in publications from across the disciplines of economics and law. Simply stated, the difference between the two concepts is that efficiency is a criterion readily applied by economists for assessing the performance of given units. It is strongly bound to compare the allocation of resources to the final output (Hicks 1939, Kaldor 1939, Broadway & Bruce 1984), i.e. aims at the optimal use of resources (Marciano et al. 2019). Effectiveness, in turn, appears to be used when considering to what extent a unit is successful in an aspect of concern. Specifically, it is often assessed through considerations of the system's success in meeting a set of general democratic assumptions with an emphasis on the courts' impartiality and independence as well as on delivering legally correct judgements (Salihu & Gholami 2018, Gozgor et al. 2019, Iqbal et al. 2019), bound with law enforcement (Mišćenić 2019), the business costs of crime, contracts enforcement, systems integrity, military interference, restrictions on real properties sales, and police reliability (Gozgor et al. 2019).

The popularity of assessing judicial performance, regardless of the exactitude of the concept applied, corresponds to the fact that a well-performing judicial system is an essential element of any economy. As long as it guarantees both the protection of property rights as well as the enforcement of contracts, it significantly contributes to the efficient production and distribution of goods and services (OECD 2013, Banasik et al. 2021). After all, such a judicial system reduces both risk in commercial transactions and transaction costs, and this

* Department of Data Science, Faculty of Economic Sciences, University of Warsaw, ms.switala@uw.edu.pl, https://orcid.org/0000-0002-8532-9539.

positively affects corporate growth in the given country (Giacomelli & Menon 2013, Garcia-Posada & Mora-Sanguinetti 2015, Bełdowski et al. 2020). Therefore, it is widely admitted that an effective judiciary allows for entrepreneurship and economic growth as well as fight poverty (Acemoglu et al. 2001, Glaeser et al. 2004, Rodrik et al. 2004, Marciano et al. 2019, Eklund et al. 2020). Also, obtaining a loan for a private business is easier in regions with a well-performing judicial system (Jappelli et al. 2005). Furthermore, an underperforming judicial system can cause meaningful loss in the legitimacy of a political system through undermining the citizenry's trust in the fundamental protection of individual freedoms (Voigt 2016, Magalhães & Garoupa 2020, Banasik et al. 2021). The overall performance of the judicial system also importantly determines a country's investment attractiveness as the capacity among different countries for increasing production correlates to judicial performance (Fusco et al. 2021).

In line with these issues, this research paper aims at contributing to the literature mainly by offering a new way to assess judicial effectiveness. The tabled approach is based on a machine learning algorithm called BERTopic (Grootendorst 2022) which enables grouping the judgements thematically. Next, their performance can be compared using a selected effectiveness measure or proxy. What is novel about this approach, compared to the state of the art, is that measuring effectiveness is here proxied by the probability of appeal of a decision.

As the research applies a machine learning algorithm for assessing the effectiveness of the judiciary, it meets the needs of the constantly developing world and follows trends recently observed in the literature. After all, multiple publications discuss the need to automate processes and use new artificial intelligence (AI) technologies by the courts (Buocz 2018, Deeks 2019, Morison & Harkens 2019, Re & Solow-Niederman 2019, Schmitz 2019, Coglianese & Dor 2020, Ulenaers 2020, Wachter et al. 2021). Also, AI has already begun to be used to study the workings of courts, judges, and the judiciary (Aletras et al. 2016, Virtucio et al. 2018, McKay 2020) as well as to process legal texts (Loza Mencia & Furnkrantz 2010, Maxwell & Schafer 2010, Kriz & Hladka 2018). When it comes to more specific literature, topic modelling keeps getting more attention in the literature applying quantitative analysis to legal problems (Lauderdale & Clark 2014, Carter et al. 2016, Livermore et al. 2017, Leibon et al. 2018, Carlson et al. 2020, Livermore et al. 2020, Luz De Araujo & De Campos 2020, Carlson et al. 2021, Dadgostari et al. 2021). In this aspect, this research aims to be listed as another contribution to the state-of-the-art.

Still, the main contribution to the literature is associated with the fact that this research uses a measure of judicial system workload and a proxy for its effectiveness that is consistently neglected in empirical studies. More specifically, the probability of appeal is calculated for the obtained thematic groups of judgements. This may spark discussion about unpopular but still interesting alternative measures of judicial performance.

More formally, the approach introduced provides descriptive evidence on cases' characteristics and their correlation with higher or lower probability of appeal, which can be considered a proxy for assessing the effectiveness of the judiciary. The author is aware that topic-based analysis does not enable formal testing of research hypotheses. However, some expectations for the correlation were stated: the thematic groups bound with the same historically established set of legal norms (e.g., civil, criminal, labour, family) should be characterised by a similar level of burden, i.e., engagement of the court's resources, and therefore effectiveness, and thematic groups including multiple legal issues (borderline cases, e.g., mixing concepts present in both civil and criminal law) should be the ones that are the least effective components of the judicial system. In other words, a positive correlation between "borderness" of topics and their probability of appeal is expected. The

expectations presented stem from the belief that the engagement of court resources in a given case should be largely determined by the nature of the proceedings, particularly differing substantially between civil and criminal cases. Consequently, cases with a "borderline" nature, combining diverse elements and therefore naturally more complex, should require greater use of court resources.

For the purpose of demonstrating the introduced method of assessing judicial effectiveness, the Polish judicial system was taken into consideration as one of the most interesting cases. The 2021 EU Justice Scoreboard (European Commission 2021), implementing the methodology developed by CEPEJ (2018), states that Poland is the country with the third biggest number of incoming cases per 100 inhabitants in the European Union. In addition, the statistics clearly demonstrate that this results more from the workload in civil and commercial cases rather than administrative ones. Furthermore, when it comes to the time needed to resolve a case, Poland is situated in the middle of the ranking of European Union countries. Nevertheless, this is mainly due to the efficiency of the administrative courts, which are the fourth fastest among the analysed countries. Moreover, the resolution rate (CEPEJ 2018) within the Polish judicial system is the fourth worst among the considered countries. In fact, in 2019, the Polish courts resolved fewer cases than came in. This obviously influences the number of pending cases per 100 inhabitants, which is one of the highest in comparison to the other countries considered. Again, a vast majority of the pending cases are non-administrative. What should be emphasised, however, is that the statistics briefly presented above are not the only specificities justifying a deeper consideration of Polish judicial performance. Poland is one of the post-communist countries and has succeeded in transforming itself from central planning to a market economy (Balcerowicz 2005). Also, the Polish judicial system is perceived as slow, inefficient, and ineffective in the literature (Kociołowicz-Wiśniewska et al. 2017, Siemaszko et al. 2019, Kruczalak-Jankowska et al. 2020, Bełdowski et al. 2020). All the above combined raises the question of the possible impact of certain legal solutions that serve as a response to the previous regime on the current judicial effectiveness, e.g. the broad indemnification of damages caused by unlawful actions of public authorities.

This paper is structured as follows: firstly, related work is presented. Secondly, methods, along with a brief description of the Polish judicial system, as well as data description is provided. Next, results are reported and discussed, followed by conclusions.

2 Related Work

2.1 The Measures of Judicial System Efficiency and Effectiveness

To elaborate more on the topic of assessing judicial performance, the most common concepts are efficiency and effectiveness. Pareto (1896) defined economic efficiency as a state such that no entity can be made better off without at the same time making any other unit worse off. Boadway & Bruce (1984) formalised the thoughts of Kaldor (1939) and Hicks (1939) concluding that some state is preferred over another if there is no possibility of costless redistribution that would lead to a superior allocation of resources according to the Pareto criterion. Therefore, it should be concluded that in the case of measuring any public system's efficiency, the allocation of resources between the units should be analysed. The more the system is adapted to the needs of society, the more efficient it is. Hence, efficiency can be measured by comparing the output obtained from the system with the inputs and resources used. It should be stressed that this idea is widely adopted for measuring overall judicial system performance (Charnes et al. 1978, Lewin et al. 1982, Kittelsen & Førsund 1992,

Elbialy & Garcia-Rubio 2011, Yeung & Azevedo 2011, Calvez & Regis 2007, Santos & Amado 2014, Smuda et al. 2015, CEPEJ 2018, Bełdowski et al. 2020, Giacalone et al. 2020, Fusco et al. 2021).

In contrast, effectiveness appears to be a far more general term. Effectiveness describes to what extent a system is successful in an aspect of concern. This clearly diverges from an economic view of performance, framed in terms of comparing input and output. In recent years, authors using effectiveness have avoided considering the system output-to-input relation (Salihu & Gholami 2018, Gozgor et al. 2019, Iqbal et al. 2019, Mišćenić 2019). This understanding of the concept of effectiveness seems to fit the claim that effectiveness is the judicial system's ability to satisfy the demand for justice (Marciano et al. 2019).

Both the judicial system's efficiency and its effectiveness are frequently debated in publications from the disciplines of economics and law. Not only does this indicate the importance of the analysed matter, but also undoubtedly justifies the diversity of the system's performance measures. When it comes to efficiency, to consider the impact of various factors on it, both parametric and non-parametric methods are applied, while it seems that the latter are used more frequently.

As for the non-parametric models, the most common approach appears to be to consider the number of resolved cases within a certain time horizon (as this is the general quantitative output delivered by the courts) and to compare it with judicial system input, i.e., available resources. This approach should be perceived as established by Lewin et al. (1982) following the general idea of measuring the decision-making units' efficiency as previously suggested by Charnes et al. (1978). This general concept of using linear programming techniques is called data envelopment analysis (DEA). In general, it is based on the Pareto optimality theorem. Transferring this to analysis of the efficiency of the judicial system, Lewin et al. (1982) suggested that single court performance can be measured with the number of resolved cases in a specified time period. An extensive survey on this particular matter was subsequently provided by Liu et al. (2013). The DEA approach was then applied and extended in many noteworthy papers (Kittelsen & Førsund 1992, Elbialy & Garcia-Rubio 2011, Yeung & Azevedo 2011, Santos & Amado 2014, Fusco et al. 2020, Bełdowski et al. 2020, Giacalone et al. 2020, Achenchabe & Akaaboune 2021).

What should be emphasised at this stage, the non-parametric approaches, despite their popularity, as Bełdowski et al. (2020) rightly pointed out, do not allow for strict determination of the factors that affect efficiency. In response to this issue, the literature proposes various parametric approaches that enable the testing of hypotheses about the statistical significance of individual parameters. Antonucci et al. (2014) enumerated the following multi-dimensional parametric methods used for evaluating the efficiency of units: ordinary least squares (OLS), corrected ordinary least squares (COLS), and stochastic frontier analysis (SFA) (Aigner et al. 1977, Meeusen & van den Broeck 1977), ultimately applying the latter for the evaluation of Italian judicial system. Remaining within the framework of this typology of parametric methods, Espasa & Esteller-More (2015) used a fixed-effect panel stochastic frontier model to analyse Catalonian first instance courts. Moreover, Bełdowski et al. (2020) proposed a combination of DEA with OLS, followed by incorporation of panel data approach and the aforementioned SFA, to assess the efficiency of Polish district commercial courts.

Also, an extensive review of the literature enables us to identify a well-established set of judicial efficiency measures that do not refer to a concept of efficiency described within the economic framework. Clearance ratio (CR) should be perceived as one that is somewhat derivative from the ideas presented above. This particular concept assumes that the performance of individual courts can be measured with a ratio of resolved cases to incoming cases within a specified period of time. This obviously corresponds to the concept of

measuring the output in relation to the inputs of the system (CEPEJ 2018). Another efficiency indicator suggested in the literature is disposition time (DT), which aims to measure the time needed for resolving a pending case under the current pace of work in the court. Mathematically, it is a ratio of the number of pending cases to the number of resolved cases at the end of some period. It is usually multiplied by the number of days in a year to simplify the interpretation of the measure (CEPEJ 2018). Another approach for measuring judicial efficiency is associated with the concept of "reasonable time" for judicial proceedings. CEPEJ (2018) suggests considering the criteria established by the European Court of Human Rights, which enable calculating the length of proceedings as well as assessing its reasonableness (Calvez & Regis 2007). Smuda et al. (2015) used the duration of proceedings in order to assess a court system's efficiency.

The effectiveness of a judicial system is often recognized as the extent to which the system performs satisfactorily. Many researchers take into account the system's success in meeting a set of general democratic assumptions with an emphasis on the courts' impartiality and independence as well as on delivering legally correct judgements (Salihu & Gholami 2018, Gozgor et al. 2019, Iqbal et al. 2019). Also, general legal effectiveness appears to be perceived in the literature as strongly bound with law enforcement (Mišćenić 2019). Additional system characteristics that are considered when measuring the effectiveness include: the business costs of crime, contracts enforcement, systems integrity, military interference, restrictions on real properties sales, and police reliability (Gozgor et al. 2019). Therefore, it should be emphasised that there are no well-established general effectiveness indicators for judicial systems in the literature. However, Marciano et al. (2019) claims that the well-established measures such as CR or DT should be perceived as capturing effectiveness, not efficiency.

Concluding the above, we may state that an efficient judicial system should feature certain traits. Specifically, an efficient system should involve as few judges as possible that resolve as many cases as possible and in as little time as possible. In contrast, the effectiveness of the judicial system is in most publications perceived as meeting democratic standards or producing the fairest judgements possible. In this paper, the research hypotheses, together with the applied system performance measure, are associated with the problem of judicial effectiveness. Still, this obviously corresponds to some components of the system's efficiency.

2.2 Probability of Appeal as a Judicial System's Effectiveness Proxy

The paper takes into consideration the probability of appeal as an effectiveness proxy. Surprisingly, no research specifically considering the probability of appeal as an indicator of a system's effectiveness has been conducted to date. Some authors' findings do however suggest that it could be introduced as a measure of the considered phenomena.

Kornhauser (1999) mentions that the probability of appeal is inextricably correlated with the correctness of the decision (which corresponds with understanding effectiveness as presented before). On the other hand, he points out that the probability of a correct court decision increases with the amount of resources assigned to that court (which can be perceived as a conclusion that efficiency implies effectiveness). In light of this, the probability of appeal should be perceived as strongly bound with system performance. Santolino (2010) is another author who in passing mentioned that the efficiency and effectiveness of the judicial system can be measured via the probability of appeal. Furthermore, some authors have provided empirical evidence for the probability of appeal being determined by the length of the court decision (Carree et al. 2010). This again connects this specific measure

with some of the well-established efficiency indicators cited earlier. It should also be discerned that the probability of appeal is perceived in the literature as strongly dependent on the effectiveness of the sanctions adjudged within the criminal cases (Billiet et al. 2014). At this stage, it should be noted that there are a few studies that mention the probability of appeal in general, without considering its relationship to the effectiveness of the judge, the court, or the judicial system. Rather, they focus on the outcome of the appeal (Samaha et al. 2020, Ash et al. 2022) and therefore should be considered in the context of the efficiency of lawyers representing the parties, with the caveat that this is just an example of how the discussed measure can be used.

As presented above, the probability of appeal can be deemed a measure of judicial performance in two different ways: taking into account either its efficiency or effectiveness. When it comes to the first concept, the probability of appeal is bound with the case-resolution process. Simply put, a proceeding starts when a formal letter initiates it. Then, the judges in a specialised department of a court of competent jurisdiction resolve the case. Obviously, this takes some time and generates costs. Furthermore, the parties to the proceedings have the right to appeal against the decision of the court. Every single appeal implies more judges involved in solving the case, and more time needed for resolution—and this generates additional costs for the system. Therefore, the share of appeals among the judicial systems' components can be labelled according to its efficiency.

On the other hand, the probability of appeal appears to be coherent with the judicial effectiveness concept. Similarly to the previous effectiveness measures suggested in the literature (Salihu & Gholami 2018, Gozgor et al. 2019, Iqbal et al. 2019, Mišćenić 2019), the probability of appeal takes into account an issue that can affect the system's performance. More specifically, it captures the workload differences among the system's components, and this makes it a promising proxy for its effectiveness. The more appeals, the more probable that some cases' resolutions were questionable, which can result from many different system issues, e.g., the specificity of some cases, their complexity, insufficiently compelling justifications, etc. Moreover, the higher the probability of appeal, the more loaded the system. The greater the load, the more difficult it is for the courts to perform well and render "fair" judgements. Therefore, the probability of appeal should be deemed a measure of the system workload that is a promising proxy for its effectiveness.

Labelling the probability of appeal an effectiveness proxy nevertheless does not mean that the right to appeal makes the judicial system ineffective. On no account should the right to appeal be undermined. Still, it is possible that in some groups of cases, the probability of appeal could be lowered without violating the judicial system's fundamental democratic standards.

3 Methodology, the Polish Judicial System, and Data

3.1 Methodology

When it comes to the methodology applied in this research, a topic modelling approach was used for the purpose of thematically grouping judgements. In general, this should be perceived as an unsupervised machine learning set of algorithms that aims at grouping pieces of text with respect to the co-occurrence of keywords and phrases. To be more specific, the study used the BERTopic algorithm (Grootendorst 2022). However, to fully understand its functioning, it is necessary to briefly outline the concept of topic modelling in general, as well as to present the evolution of the various approaches that led to the development of the state-of-the-art algorithm used in this work.

Many different topic modelling algorithms were suggested in the literature at the turn of the century (Paatero & Tapper 1994, Landauer & Dumais 1997, Landauer et al. 1998, Hofmann 1999) with the most prominent being Latent Dirichlet Allocation (LDA) (Blei et al. 2003). Practical importance of LDA was later emphasised by subsequent publications introducing its further specific modifications adapting this algorithm to different tasks, most often extending simple text clustering with additional conditions and variables (Rosen-Zvi et al. 2004, Li & McCallum 2006, Blei & Lafferty 2007, Boyd-Graber & Blei 2008, Lacoste-Julien et al. 2008, Blei et al. 2008, Mimno et al. 2009, Ramage et al. 2009, Rabinovich & Beli 2014, Bhadury et al. 2016, Chien & Lee 2017, Jansson & Liu 2017, Sharma et al. 2017, Shi et al. 2017, Qiang et al. 2017, Bai et al. 2018, Jin et al. 2018). However, both the recent adoption of deep learning (Larochelle & Lauly 2012, Cao et al. 2015, Chien & Lee 2017, Jansson & Liu 2017, Sharma et al. 2017, Bai et al. 2018, Jin et al. 2018, Bhat et al. 2020, Doan & Hoang 2021, Terragni et al. 2021, Wang & Yang 2020, Zhao et al. 2021, Mazzei & Ramjattan 2022) and promising usage of word embeddings (Arora et al. 2016, Das et al. 2015, Liu et al. 2015, Nguyen et al. 2015, Li et al. 2016, Moody 2016, Xun et al. 2016, Shi et al. 2017, Qiang et al. 2017, Bianchi et al. 2020, Dieng et al. 2020, Thompson & Mimno 2020) indicated the possibility of enhancing classical topic models performance. This resulted in the introduction of a state-of-art BERTopic algorithm (Grootendorst 2022). Also, at this point it should be emphasised that so far BERTopic was reported as outperforming LDA in a majority of empirical analyses (Abuzayed & AlKhalifa 2021, De Groot et al. 2022, Egger & Yu 2022, Hutama & Suhartono 2022, Sangaraju et al. 2022, Scarpino et al. 2022, Zankadi et al 2022, Ao et al 2023). The history of the development of the topic modelling approach, as well as the results of the empirical comparison of algorithms mentioned in the previous sentence, contributed to the final selection of BERTopic for use in this study.

BERTopic is based on an idea of transformer-based pre-trained language models, often named text embedding techniques. Simplifying, the embeddings are contextual representations that can be used for further natural language processing models. Embedded documents are represented in a vector space and can be compared semantically. Even though the diversity of these models is undeniable (Acheampong et al. 2021, Kalyan et al 2021), the most popular appear to be GPT (Radford et al. 2018) and BERT (Devlin et al. 2018). BERTopic is based on a variant of the latter, Sentence Bidirectional Encoder Representations from Transformers (Sentence-BERT) (Reimers & Gurevych 2019).

Noticeably high dimensionality of text embeddings should be perceived as a meaningful challenge to the effectiveness of their computation. Therefore, in aim of optimising the process, the dimensionality of the embeddings used in BERTopic is reduced. It is possible with already well-established algorithms such as PCA (Pearson 1901, Hotelling 1933) or t-SNE (Hinton & Roweis 2002, Van der Maaten & Hinton 2008). Still, Grootendorst (2022) recommends the UMAP algorithm (McInnes et al. 2018) as it appears as preserving more of both local and global features of high-dimensional data when projected to lower dimensions.

With the embeddings' dimensionality reduced, the next step involves clustering them with HDBSCAN algorithm (Ester et al. 1996, Campello et al. 2013). An undoubtful advantage of using this algorithm is that it enables modelling noise as outliers and not assigning it to the clusters. However, as emphasised by Grootendorst (2022), Allaoui et al. (2020) demonstrated that combining UMAP with k-means clustering (MacQueen 1967) also provides satisfactory results.

Yet another step involved in using BERTopic is assessing tokens' contributions into the topics. A class-based modification of the TF-IDF measure (Joachims 1996)—which is the product of term frequency in a document and the logarithm of the inverse document

frequency (the share of documents containing a term)—is applied which enables merging topics with similar words' importance.

Parameters of the BERTopic model built in this research were optimised with respect to topic coherence measures, i.e. UCI (Newman et al. 2010), UMass (Mimno et al. 2012), UCI-NPMI (Aletras & Stevenson 2013). All have similar interpretations. The higher values indicate that tokens representing the topic are frequently observed together in the corpus (Roder et al. 2015). Furthermore, the results obtained were confronted with the actual interpretability of the topics as assessed by experts.

When it comes specifically to the methodology adopted in this research, each judgement from the considered corpora was assigned to the most probable topic obtained. Next, within each topic, a distribution of assigned judgements among courts departments of certain types was calculated. This was later used for determining whether a single topic comes from a certain group of historically established legal norms. In other words, if a vast majority of judgements assigned to a particular topic originated from courts departments of the same type (e.g., civil), then it was assumed that this thematic group is a "civil topic". On the other hand, such an approach made it possible to identify topics that contain similar legal problems but that are associated with different groups of legal norms. For example, a topic where half of the assigned judgements originated in civil departments and the other half came from criminal departments was labelled a "borderline", "civil-criminal" topic. More details were presented in the results section using the example of regional court judgements in Poland.

The share of judgements against which an appeal was filed was also computed for each thematic group of judgements. This enabled comparing the workload between the topics and evaluating their possible impact on system effectiveness.

3.2 The Polish Judicial System

The Constitution of the Republic of Poland of 2 April 1997 states in Article 174 that the courts and tribunals pronounce judgements in Poland. The latter are specifically named in the Constitution: the Constitutional Tribunal along with the Tribunal of the State. Regarding the courts, Article 175 names the Supreme Court and thereafter enumerates only the remaining court types: the common courts, administrative courts, and military courts. Also, the next Article states that court proceedings should have at least two-stages. Furthermore, the Polish Constitution states that the organisational structure, jurisdiction, and procedure shall be specified by acts of Parliament. Still, the Constitution's Article 177 states that the common courts administer all matters not saved for other courts.

The Law on the system of common courts of 27 July 2001 in Article 1 enumerates three types of the common courts: regional courts, district courts, and appeal courts. According to acts of Parliament, the Code of the Civil Procedure, along with the Code of the Criminal Procedure, the regional courts are courts of first instance for all matters that were not saved for the district courts. Therefore, the district courts are designated first instance courts for, to put it intuitively, more serious cases. Most importantly, the jurisdiction of district courts includes cases, as far as those in which the rules of civil procedure apply, of non-material rights and jointly asserted property claims, as well as cases of property rights in which the value of the subject of the dispute exceeds 100,000 PLN (ca €23,000), with exceptions for both categories. In criminal cases, district courts have jurisdiction over crimes and certain misdemeanours. Also, the district courts resolve appeals filed against the regional courts' judgements. The appeal courts consider only the appeals against the first-instance district

courts' judgements. As of the end of 2022, there are 11 appeal courts, 46 district courts, and 318 regional courts in Poland.

The common courts are divided into departments. In the regional courts, the mandatory departments to be created are: civil and criminal. Other departments, i.e., family and juvenile, labour and social security, commercial, land registry, and enforcement, may be established by the Minister of Justice. Every case is assigned to a competent department based on a consideration of the case's characteristics and its connections with legal provisions.

In the Polish judicial system, judges are appointed by the President of the Republic of Poland, on the proposal of the National Council of the Judiciary. Until 2018, members of the Council were elected by assemblies of judges of courts of various types. Starting 2018, the competence to select the members of the Council is exercised by the Sejm, i.e. the lower chamber of the Polish Parliament.

3.3 Data

All data was obtained from the System of Analysis of Courts Decisions at *saos.org.pl*. The aim of this portal is to publish the content of judgements of both the extraordinary as well as common courts in Poland. The scope of the published judgements was determined by a panel of the Polish common courts' judges. As a result, no exempted or repetitive content shall be published there. Neither are the 'irrelevant' judgements, i.e., the ones that do not exhibit a substantial legal information. As specified in the relevant regulation of the Minister of Justice: the decision on which judgement will be published on the portal is made by the judge who participated in its issuance, or by an official in charge. Also, all judgements available on the portal are anonymized.

For the purposes of this research, all judgements of the Polish regional courts published at *saos.org.pl* were collected. The number of considered documents was 84,579. They were issued in a time horizon from 27.05.2007 to 27.12.2022 (state as of 13.01.2023; some judgements appear to be published with a delay).

A set of standard preprocessing operations of the textual data was performed. Punctuation, special signs, one-letter words, as well as Arabic and Roman numbers were removed from the corpus. Then, texts were tokenized. Also, Polish words that had no information value were removed (listed in a dictionary provided within the Python programming language library named stop_words). Furthermore, courts' names and cases' symbols as well as Polish names and surnames were excluded (list provided within the *Morfeusz* program mentioned below). Finally, words were reduced to their root form with a lemmatization operation. As far as the Polish language processing is quite challenging, a tool that was originally prepared for linguistics was applied – the *Morfeusz* program (Woliński 2014).

After all the aforementioned operations, 167,199 unique tokens remained in the corpus. As bigrams and trigrams were introduced into the corpus, it made the overall number of unique tokens equal to 5,499,475. Tokens observed only once in the corpus were removed as useless in any further analysis which made the number of the unique tokens 3,708,686. Finally, when optimising parameters of the topic model, it appeared that removing tokens that appear at least once in at least 75% of the analysed texts improves model performance. After this operation, 3,708,640 unique tokens remained.

It should be noted at this point that at the initial stage of the research, models based on the LDA algorithm (Blei et al. 2003) were also constructed since this is a well-established alternative to the state-of-the-art BERTopic algorithm. The LDA results turned out to be much

less interpretable than the ones obtained with BERTopic. Also, initially attempts were made to construct a topic model by applying the BERTopic method on plain text, i.e. without markup but still including punctuation, stopwords, numbers, and most importantly with non-lemmatized words. This approach, despite being recommended by Grootendorst (2022), also provided far less interpretable and informative results than what was obtained after more extensive text preprocessing. In particular, BERTopic applied on plain text turned out to group judgements with respect to the location of the court. Perhaps judges in different courts use different formulas when drafting justifications. Nevertheless, the purpose of building the model was to group the texts thematically.

What should also be mentioned to justify the extensive preprocessing, is that the multilingual version of BERTopic was used in the analyses. Such a decision was made in view of the fact that the Polish-language versions of BERT, despite of being constantly developed (Dadas, Perełkiewicz & Poświata 2020, Kłeczek 2020, Rybak et al. 2020, Mroczkowski et al. 2021) are still not easily available. Specifically, the libraries available for Python do not involve the Polish version of BERT but stops with a multilingually pre-trained variant. Pre-training of BERT using the author's own resources appears to be disproportionately costly to the purpose of the study. Furthermore, the ideal would be to train BERT not necessarily on any Polish texts but on Polish legal texts, or even judgements of Polish common courts. Unfortunately, the most extensive database of such texts, i.e. the System of Analysis of Courts Decisions, includes less than 85,000 common courts judgements which appears to be an absurdly small number compared to corpora originally used by BERT's funders.

4 Results and Discussion

4.1 Topic Model

The built BERTopic model on a corpus of 84,579 Polish regional courts' judgements pronounced in 27.05.2007-27.12.2022 optimised with respect to the topic coherence measures, i.e. UCI (Newman et al. 2010), UMass (Mimno et al. 2011), and UCI-NPMI (Aletras & Stevenson 2013), resulted in obtaining 80 topics. Given the large number of topics obtained, it should be considered inexpedient to present and discuss every single one in the paper. However, it should be emphasised that all the obtained topics were characterised with both informative and interpretable keywords that allowed an expert assignment of the labels. The most frequent topics are presented in Table 1. As modelling involved processing Polish language, for the purpose of presentation, the most prominent tokens describing each topic were translated into English.

The topic that was reported as the most popular among the judgements in the analysed time horizon was associated with third party civil liability insurance. The next most prominent ones were connected to: traffic offences, termination of employment contract, alimony and securitization. The sixth thematic motive most frequently observed in the judiciary was similarly to the first one associated with third party liability insurance. Still, differences in both the most prominent keywords describing the topics as well as the weights assigned to the tokens were noticeable. Therefore, the topics were not merged together. The seventh most frequently observed thematic motive involved appeals against decisions of the social security institution. The next two were reported as associated with criminal law, i.e. labelled: drunk driving, misappropriation. The tenth most prominent topic seemed to be bound to the problem of victim's contribution to damage size in the form of renting a replacement vehicle more expensive than offered by the insurance institution.

Table 1. The top 10 most frequent topics obtained from the BERTopic model.

Topic index	Expertly assigned label	The most prominent tokens with obtained weights
1	Third-party liability insurance	civil liability (0.0032), suffering (0.0024), liquidation (0.0023), cervical (0.0022), replacement vehicle (0.0021), favour inc (0.0021), defendant incorporated (0.002), insurance liability (0.002), third-party liability insurance (0.002), liquidation proceedings (0.002)
2	traffic offences	blame (0.0155), accused (0.0117), speed (0.009), manoeuvre (0.006), offence (0.0057), pedestrian (0.0057), offence (0.0055), roadway (0.0051), traffic safety (0.0047), crossroads (0.0043)
3	termination of employment contract	employment relationship (0.0069), termination employment contract (0.0036), certificate employment (0.0032), prior notice employment contract (0.0031), employment prior notice (0.0028), terminate employment contract (0.0027), employment contract prior notice (0.0025), annual leave (0.0024), annual (0.0024), overtime (0.0023)
4	alimony	maintenance (0.0134), alimony (0.0131), maintenance obligation (0.009), child support (0.0078), minor claimant (male) (0.0062), earning opportunities (0.0058), mother minor (0.0053), minor claimant (female) (0.0047), representative (0.0047), favour minor (0.0046)
5	securitization	securitisation (0.0158), investment fund (0.0139), standardise (0.0114), securitization found (0.0101), securitization fund investment (0.0101), standardise securitisation (0.0101), standardise securitisation fund (0.01), investment closed-end (0.0089), closed-end investment fund (0.0089), securitisation fund (0.0078)
6	third-party liability insurance	civil liability (0.0032), favour incorporated (0.0025), favour incorporated company (0.0025), gtc (0.0024), liquidation (0.0023), vehicle repair (0.0023), replacement vehicle (0.0023), insurance liability (0.002), third party liability insurance (0.002), suffering (0.002)
7	appeal against the decision of the social security institution	pension (0.016), sickness benefit (0.0152), pension authority (0.0142), incapacity work (0.0102), social security institution (0.0097), insured (male) (0.0093), social department (0.0087), social security department (0.0087), insured (female) (0.0068), applicant (female) (0.0065)
8	drunk driving	alcohol intoxication (0.0145), milligram (0.0142), state intoxication (0.0138), exhale (0.0108), exhale air (0.0102), driving ban (0.0093), alcohol exhale (0.0093), alcohol exhale air (0.0089), land (0.008), land traffic (0.008)
9	misappropriation	misappropriation (0.0057), seizure (0.0051), purpose misappropriation (0.005), pln damage (0.005), base penal code (0.0043), commit theft (0.0042), accuse act (0.0041), break in (0.0038), explain defendant (0.0038), code criminal procedure (0.0034)
10	replacement vehicles (victim's contribution to damage size)	replacement vehicle (0.0046), rent vehicle (0.0038), rent replacement vehicle (0.0034), claimant company (0.0032), claimant company limit (0.0029), liability amount (0.0025), limit liability amount (0.0025), defendant company (0.0024), pln period day (0.0022), liability company (0.0022)

Note: Topic index corresponds to the position in terms of frequency of the topic in the corpus.

The other topics, i.e. those not presented in Table 1, were foremost bound to: evictions, beatings, improper performance of the tourist service, cumulative penalties, non-payment of credit, perpetual usufruct, counteracting drug addiction, leading to an unfavourable disposition of property, housing communities disputes, fiscal crimes, death threats, promissory notes, appeals against decisions of medial commissions on disability, inheritance, gambling law, default judgements, teacher card regulations, marital property separation, animal protection, long-term health impairment as a result of a traffic collision, non-payment of alimony crime, and servitudes.

4.2 Similar Topics

Having obtained the topics described by their keywords representation, a consideration of shares of judgements from different courts departments assigned to the topics was conducted. Also, the appeals share for every single topic was calculated. First of all, the focus was on verifying whether the detected thematic groups, within which judgements were predominantly issued in departments with identical specialisation, are characterised by similar probability of appeal. This created a need for grouping the obtained thematic groups with respect to the departments where the judgements assigned to the topics originated.

As the well-established approaches for performing such division – clustering of cases with respect to departments shares and pairwise Pearson's Chi2 tests (Pearson 1900) – did not provide any satisfactory results, basic descriptive statistics analysis was incorporated. The topics were divided in respect to the court departments of the biggest share. As a result, civil, criminal, labour, and family groups of topics were established. It is noteworthy that the judgements made in the civil and commercial departments were grouped together, i.e. as civil topics, as the rules of the division of cases between them are based primarily on the presence of entrepreneurs on both sides of the dispute.

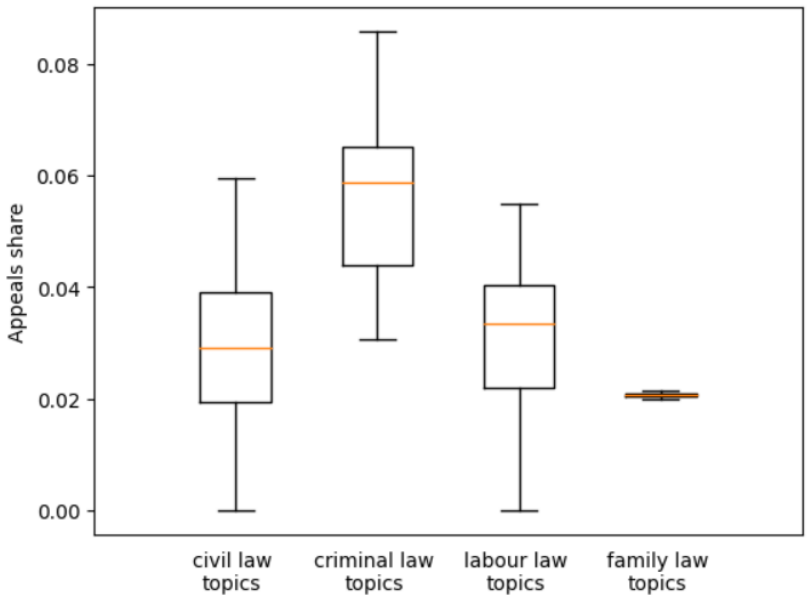

Figure 1. Distributions of appeal share between the thematic groups of judgements.

Next, distributions of topics' appeal shares between the groups were compared as presented in Figure 1 (all figures were generated with a Python programming language library named matplotlib). It is clearly visible that the appeal share distribution within criminal topics differs from the one within civil topics. Still, topics bound to mainly labour law and social insurance characterise themselves with an appeals share distribution similar to the civil ones. This seems to be due to the fact that labour law remains closely linked to civil law. In particular, the Polish Labour Code provides that, in matters not regulated therein, the provisions of the Civil Code shall apply accordingly, if they do not contradict the principles of labour law.

It was also formally tested if the presented distribution (see Figure 1) statistically differ from each other (see Table 2). The well-established Mann-Whitney (Mann & Whitney 1947) and Kruskal-Wallis (Kruskal & Wallis 1952) tests were applied. In a nutshell, these non-parametric statistical methods, here particularly useful due to the limited number of topics obtained, enable comparing two independent groups to verify if their distributions statistically differ from each other. Technically, both are based on calculation of sums of ranks per group. For both tests, hypotheses of dependence between civil and criminal law were rejected. This was also the case when comparing criminal and labour law topics. However, there was no statistically significant difference between civil and labour law topics' appeal shares distributions. In other words, distributions of the probability of appeal in thematic groups of judgements do not statistically differ between civil and labour law groups of topics. Still, they do differ when comparing distribution of the values of interest in criminal topics with civil and labour law groups.

Taking into account the above, it appears that thematic groups of court cases bound to the same historically established sets of legal norms are indeed characterised with similar probability of appeal as distribution of the probability of appeal among the topics connected to civil law does not statistically differ from the distribution observed for the labour law topics. Oppositely, the difference between civil and criminal topics, as well as labour law and criminal ones, is noticeable, i.e. the empirical probability of appeal is clearly higher in the second group.

Table 2. Results of statistic tests for equality of topics appeal shares distributions

Name of the test	Distributions compared	Test statistic
Mann-Whitney	civil vs. criminal	106.50 ***
	civil vs. labour law	179.00
	criminal vs. labour law	147.00 **
Kruskal-Wallis	civil vs. criminal	27.35 ***
	civil vs. labour law	0.15
	criminal vs. labour law	9.45 **

Note: Distribution of appeals share in family topics was not involved as this group involved only two topics. Significance note: *** - 0.1% significance, ** - 1% significance, * - 5% significance, . - 10% significance.

4.3 Borderline Topics

Next, it was considered if the detected thematic groups, within which decisions were made in court departments of different specialisations, are characterised by a higher probability of appeals than the topics in which cases were predominantly decided in departments of homogeneous characteristics. In order to verify this, it was necessary to identify the borderline topics, or in other words, to measure the topics' "borderness". It was decided to analyse the ratio of the two largest shares of judgements from departments of a given type (further referred to as "borderness measure") for each topic:

$$B_i = \frac{ts_{i,2}}{ts_{i,1}}$$

where B_i stands for the "borderness" of topic i, whereas $ts_{i,j}$ denotes j-th largest share of judgements from departments of a given type in all the documents assigned to topic i.

The measure introduced was intended to capture the extent to which judgements from the dominant department within the topic exceeded those from the most competitive type. However, the ratio involves dividing second largest share with the largest share. Theoretically, the largest share can take values from (0, 1]. Therefore, the way how the measure was computed enabled avoiding values of infinity. The lower the "borderness measure", the more a topic is bound with only one historically established set of legal norms. The value of 0 means that all the judgements grouped into separate topics were issued in departments of the same type.

Within the groups of obtained topics, Pearson's (Pearson 1895) and Spearman's (Spearman 1904) correlation measures between appeal shares and "borderness measure" were calculated (see Table 3). Both coefficients take values in the range [-1, 1], and positive values suggest a positive relationship between the analysed quantities, while negative values indicate the opposite. The Spearman's coefficient was tried as, in contrast to Pearson's, it checks if the variables increase or decrease in an orderly manner, and therefore not necessarily linearly. It is also more resistant to outliers.

Correlation was relatively weak (Pearson's coefficient of 0.0435, Spearman's coefficient of -0.0048, both statistically non-significant) when all the topics were taken into account. As for the topics concerning only the matters of criminal law, Pearson's coefficient suggested weak positive correlation (coefficient of 0.0177, non-significant) whereas Spearman's pointed out negative correlation (coefficient of -0.2176, non-significant). However, the correlation was reported as positive and noticeably higher compared to the previous examples when considering only the topics representing labour law disputes (Pearson's coefficient of 0.3538, non-significant; Spearman's coefficient of 0.7714, statistically significant assuming 10%-significance level). Finally, the correlation reported for the bound between appeals share and "borderness measure" in case of civil law topics was statistically significant and of a substantial strength when using both Pearson's (0.3036) and Spearman's coefficient (0.2603). Therefore, it should be concluded that there is a positive correlation between civil topics "borderness" and share of appeals against judgements assigned to these topics (see Figure 2).

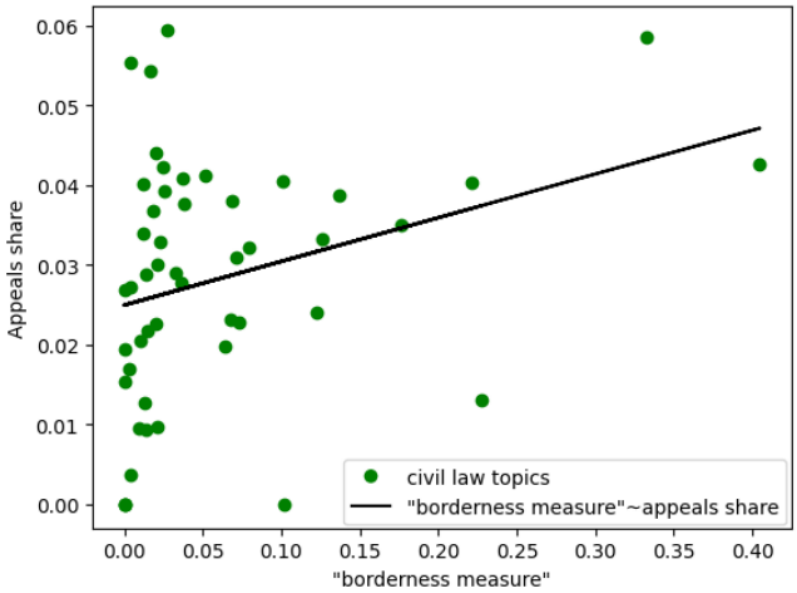

Figure 2. Relationship between appeals share and "borderness measure" in case of the topics concerning civil law disputes.

The highest "borderness measure" value was observed for the topic concerning disputes related to termination of employment contracts (see Table 4). This is indeed a thematic motif on the edge of civil and labour law – 39.61% of judgements assigned to this topic were made in civil departments. Appeals share in the considered topic was reported to be 0.0390 which is noticeably higher than mean appeals shares for civil topics (0.0282) and labour law topics (0.0297). Notably, this should be named an example of "borderness" having a positive impact on probability of filling an appeal. This seems reasonable when one considers the nature of the rulings made under this topic - these are indeed cases at the intersection of civil law and labour law, and thus potentially more complicated. Specifically, these are likely to be cases concerning the assertion of higher-than-determined compensation for unjustified or labour law-infringing terminations of employment contracts. More precisely: according to the regulations in force in Polish law, in the labour and social security law departments, the courts determine whether a termination of an employment contract is unjustified or violates specific provisions and, according to the employee's demand, declare the termination ineffective, reinstate the employee to work (both only in the case of a contract of indefinite duration) or award compensation (the only option for employee in case of a contract of definite duration). According to labour law regulations, the compensation is limited in its amount. However, it is not difficult to imagine a situation in which the extent of the damage exceeds the limit of compensation. In such a case, the employee may claim compensation for the remaining, i.e. surplus part of the damage under the general rules of civil law. Another trial will then take place, this time already in the civil department of the court. Nevertheless, the topic under discussion may also apply to different cases, e.g. termination of the employment contract by the employee motivated by bullying. As the latter is relatively difficult to be proven, it may be less risky to sue for infringement of moral rights. This is considered in the civil department.

Positive relationship between the "borderness" and probability of appeal was also observed in the case of the second most "borderline" topic – concerning non-payment of alimony crime. The "borderness" of it appears to be clear as this is an intuitive mix of criminal

and family law provisions. Its appeals share was 0.0426 which is higher than the average obtained for the family law topics (0.0207), but lower compared to the average for criminal topics (0.0564).

The topic with the third highest "borderness measure" value involves cases concerning lease of transmission facilities and the possibility of establishing a transmission servitude. Appeals share was 0.0585 which was reported higher than mean appeals shares for civil topics (0.0282) and labour law topics (0.0297). Interestingly, the lease of part or all of an enterprise can have the same effects as the takeover of an enterprise which is the subject of labour law regulations. Hence the relatively high proportion of judgements issued in the labour law departments in this topic (24.86%).

Table 4. Top 10 topics with the lowest ratio of two largest shares of judgements from departments of a given type for each of the topics obtained.

Topic index	Expert assigned label	Within-topic share of judgements issued in department: civil	criminal	labour	family	Appeals share	Borderness measure
57	termination of employment contract	0.3961	0.0000	0.6039	0.0000	0.0390	0.6559
41	non-payment of alimony crime	0.6702	0.0000	0.0585	0.2713	0.0426	0.4048
42	lease agreement, transmission servitude	0.7477	0.0018	0.2486	0.0018	0.0585	0.3325
49	health care, lawsuits against hospitals	0.7987	0.0195	0.1818	0.0000	0.0130	0.2276
44	promissory note loans	0.7976	0.0205	0.1762	0.0050	0.0403	0.2209
18	lease and perpetual usufruct of premises	0.7425	0.1038	0.1311	0.0226	0.0349	0.1766
32	teacher card regulations	0.1409	0.0206	0.8351	0.0034	0.0550	0.1687
46	excise tax crimes	0.1399	0.8601	0.0000	0.0000	0.0570	0.1627
14	improper performance of the tourist service	0.8761	0.0034	0.1196	0.0009	0.0387	0.1365
28	promissory note loans	0.8523	0.0211	0.1078	0.0181	0.0332	0.1264

The next topic on the list is associated with lawsuits against hospitals. Surprisingly, appeals share in this topic was reported to be relatively low, 0.0130. This is lower than average for different groups of topics (0.0282 for civil law, 0.0297 for labour law). "Borderness" in this group of cases comes from the fact that hospitals can obviously be employers. Therefore, 18.18% of judgements assigned to this topic originated from labour law departments whereas 79.87% were issued in civil departments. Two out of ten topics characterised with the highest "borderness measure" values were connected to a problem of promissory note loans. For these two groups of judgements, appeal shares were: 0.0403, 0.0332 (again higher than average for civil - 0.0282, and labour topics - 0.0297). The relatively high shares of assigned judgements originated in labour law departments in those that seem to be bound with a problem of a promissory note. The Polish Supreme Court recently held that when

issued by an employee a promissory note cannot be used as security for the employer's claims against that employee.

Another topic within the top ten highest values of the "borderness measure" is associated with the lease and perpetual usufruct of premises. Appeals share in this topic (0.0349) was higher than average in civil (0.0282) and labour law (0.0297) topics. Judgements assigned to this topic originated in different courts departments: 74.25% in civil, 13.11% in labour law and social insurance, 10.38% in criminal, and 2.26% in family law department. The legal protection of tenants, landlord-tenant disputes, as well as the determination of the landlord-tenant fee, are all governed by civil law. At the same time, however, it should be noted that the individual housing situation of a party to the civil proceedings may be bound with criminal or labour law too. In particular, it will be relevant when determining the amount of a fine in criminal proceedings. As far as labour law is concerned, employee housing still exists in Poland. When considering family law, the housing situation will be relevant in the division of the joint property of spouses.

The topic labelled "teacher card regulations" was another one characterised by a relatively high value of "borderness measure". Appeals share in this case was 0.0550, again higher than the reference values considered (0.0349 for civil topics, 0.0282 for labour law topics). As for this topic's "borderness", it is probably a similar case to the very first one described in this section, i.e. employees may claim compensation for the remaining, surplus part of the damage under the general rules of civil law.

Yet another "borderline" topic is connected to excise tax crimes. Appeals share for this topic (0.0570) was higher than both the average for civil topics (0.0349) and the average for criminal topics (0.0564). Obviously, the majority of cases assigned to this topic were resolved in criminal departments of regional courts (86.01%). Still, 13.99% came from civil departments. This can be due to compensation for decisions of tax authorities.

11.96% of judgements assigned to the topic labelled as "improper performance of the tourist service" originated from labour law departments which appears to be connected to the cases considering matters concerning leave granted to the employee by the employer. The appeals share of 0.0387 was again higher than the reference values mentioned above.

5 Conclusions and Recommendations

This research has aimed to introduce a new method for assessing judicial effectiveness based on a machine learning algorithm called topic modelling. The idea of measuring a judicial system's performance was demonstrated with an example dataset in the form of regional court judgements in Poland. Indeed, the suggested approach enabled the identification of meaningful insights into the matter discussed.

Topic modelling provides both interpretable and informative thematic groups of judgements. Considering the intra-topic distribution of the judgements between courts' departments, the thematic groups were linked with the historically established sets of legal norms, i.e., civil, criminal, family and juvenile, labour and social insurance law. Also, such an approach, combined with an introduced measure of the topics' "borderness", makes it possible to identify thematic groups of judgements concerning legal problems associated with more than one set of legal norms. This, along with examples found in the judgements of the Polish regional courts, yielded an intuitive grasp of the border topics.

The probability of appeal, labelled a measure of workload and a proxy for effectiveness, was calculated and compared between the obtained thematic groups. It appears that the topics that are homogenous, i.e., originate from the same set of legal norms, are characterised by a similar probability of appeal. Therefore, it can be concluded that such

thematic groups of cases place a similar burden on the system. More specifically, criminal topics are characterised by noticeably higher probability of appeal than civil ones.

As mentioned before, some topics clearly associated with more than one historically established set of legal norms were also identified. Formally, correlation between the introduced "borderness measure" and probability of appeal was reported positive. For the topics with assigned judgements primarily issued in civil departments, Pearson's correlation coefficient was 0.3036 and statistically significant. Therefore, it should be concluded that "borderness" is bound with higher probability of appeal. Treating the latter as a measure for workload and proxy for effectiveness, it suggests a need for further analyses to specific topics in judiciary as employees may claim compensation for the remaining, surplus part of the damage under the general rules of civil law, non-payment of alimony crime or the lease of part or all of an enterprise leading to its takeover.

The paper struggles to build an intuition for interpreting probability of appeal as a proxy for effectiveness which requires a broader discussion to ground it in the literature. I argue that it is possible that in some groups of cases, the probability of appeal could be lowered to increase the effectiveness of the judiciary. This can be achieved with many different system changes. One example might be more compelling justifications of court decisions regarding the most problematic legal issues. Reforming certain aspects of procedural law can also help improve how courts justify their decisions. Changing the regulations of expert opinions can also affect the probability of appeal. Obviously, none of these suggestions would violate the democratic right to appeal; rather, they help to convince the parties that the judgement is as just as possible. Also, the suggested improvements would, with a high probability, affect the system's efficiency. If fewer appeals are filed, then fewer judges would be involved, lower costs would be generated, and resolving cases would take less time.

On the other hand, the research can be perceived as providing empirical evidence on correlation between judgements heterogeneity, i.e. their "borderness", with probability of filling an appeal. This evidence can be expanded with additional analysis, e.g. investigating heterogeneity in defendants or other case characteristics, to shed light on why some cases are handled less effectively than others. Furthermore, this study takes into account only one determinant of the system's workload, and others should be included in the future.

When it comes to more technical enhancement of the presented analysis in the future, it would be beneficial to repeat the analysis given an expanded sample of judgements. Moreover, the topic model prepared could be tried on subsamples of the considered dataset. The dynamics of the topic's structure in time and space can also significantly contribute to the current state-of-the-art. What should be also considered is applying either spatial econometrics or supervised machine learning algorithms predicting and explaining the workload of the judiciary with different characteristics, together with a set of independent variables based on the textual data.

References

Abuzayed, A., & Al-Khalifa, H. (2021). BERT for Arabic topic modeling: An experimental study on BERTopic technique. Procedia computer science, 189, 191–194, DOI: 10.1016/j.procs.2021.05.096.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. American economic review, 91(5), 1369–1401, DOI: 10.1257/aer.91.5.1369.

Acheampong, F. A., Nunoo-Mensah, H., & Chen, W. (2021). Transformer models for text-based emotion detection: a review of BERT-based approaches. Artificial Intelligence Review, 54, 5789–5829, DOI: 10.1007/s10462-021-09958-2.

Achenchabe, Y., & Akaaboune, M. (2021). Determinants of Judicial Efficiency in Morocco. Open Journal of Business and Management, 9(5), 2407–2424, doi: 10.4236/ojbm.2021.95130.

Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. Journal of econometrics, 6(1), 21–37, DOI: 10.1016/0304-4076(77)90052-5.

Aletras, N., & Stevensson, M. (2013). Evaluating topic coherence using distributional semantics. In Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Long Papers, 13-22, URL: aclanthology.org/W13-0102.pdf.

Aletras, N., Tsarapatsanis, D., Preoţiuc-Pietro, D., & Lampos, V. (2016). Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. PeerJ computer science, 2, e93, DOI: 10.7717/peerj-cs.93.

Allaoui, M., Kherfi, M. L., & Cheriet, A. (2020). Considerably improving clustering algorithms using UMAP dimensionality reduction technique: A comparative study. In International conference on image and signal processing, 317–325, DOI: 10.1007/978-3-030-51935-3_34.

Antonucci, L., Crocetta, C., & d'Ovidio, F. D. (2014). Evaluation of Italian judicial system. Procedia Economics and Finance, 17, 121–130, DOI: 10.1016/S2212-5671(14)00886-7.

Ao, Z., Horváth, G., Sheng, C., Song, Y., & Sun, Y. (2023). Skill requirements in job advertisements: A comparison of skill-categorization methods based on wage regressions. Information Processing & Management, 60(2), 103185, DOI: 10.1016/j.ipm.2022.103185.

Arora, S., Li, Y., Liang, Y., Ma, T., & Risteski, A. (2016). A latent variable model approach to pmi-based word embeddings. Transactions of the Association for Computational Linguistics, 4, 385–399, DOI: 10.1162/tacl_a_00106.

Ash, E., Chen, D. L., & Galletta, S. (2022). Measuring Judicial Sentiment: Methods and Application to US Circuit Courts. Economica, 89(354), 362–376, DOI: 10.1111/ecca.12397.

Bahl, L. R., Jelinek, F., & Mercer, R. L. (1983). A maximum likelihood approach to continuous speech recognition. IEE transactions on pattern analysis and machine intelligence, 2, 179–190, DOI: 10.1109/TPAMI.1983.4767370.

Bai, H., Chen, Z., Lyu, M. R., King, I., & Xu, Z. (2018). Neural relational topic models for scientific article analysis. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 27–36, DOI: 10.1145/3269206.3271696.

Balcerowicz, L. (2005). Post-communist transition: Some lessons. In IEA occasional paper, 127, URL: papers.ssrn.com/sol3/papers.cfm?abstract_id=676661.

Banasik, P., Metelska-Szaniawska, K., Godlewska, M., & Morawska, S. (2021). Determinants of judges' career choices and productivity: a Polish case study, European Journal of Law and Economics, 53, 81–107, DOI: 10.1007/s10657-021-09688-4.

Bełdowski, J., Dąbroś, Ł. & Wojciechowski, W. (2020). Judges and court performance: a case study of district commercial courts in Poland. European Journal of Law and Economics, 50, 171–201, DOI: 10.1007/s10657-020-09656-4.

Bhadury, A., Chen, J., Zhu, J., & Liu, S. (2016, April). Scaling up dynamic topic models. In Proceedings of the 25th International Conference on World Wide Web, 381–390, DOI: 10.1145/2872427.2883046.

Bhat, M. R., Kundroo, M. A., Tarray, T. A., & Agarwal, B. (2020). Deep LDA: A new way to topic model. Journal of Information and Optimization Sciences, 41(3), 823–834, DOI: 10.1080/02522667.2019.1616911.

Bianchi, F., Terragni, S., Hovy, D., Nozza, D., & Fersini, E. (2020). Cross-lingual contextualized topic models with zero-shot learning. arXiv preprint, DOI: 10.48550/arXiv.2004.07737.

Billiet, C. M., Blondiau, T., & Rousseau, S. (2014). Punishing environmental crimes: An empirical study from lower courts to the court of appeal. Regulation & Governance, 8(4), 472–496, DOI: 10.1111/rego.12044.

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. The annals of applied statistics, 1(1), 17–35, DOI: 10.1214/07-AOAS114.

Blei, D. M., McAuliffe, J. D., Platt, J. C., Koller, D., Singer, Y., & Roweis, S. (2008). Supervised topic models advances. In Neural Information Processing Systems, 20, 121–128, URL: proceedings.neurips.cc/paper/2007/file/d56b9fc4b0f1be8871f5e1c40c0067e7-Paper.pdf.

Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet allocation. Journal of machine learning research, 3, 993–1022, URL: jmlr.org/papers/volume3/blei03a/blei03a.pdf?ref=http://githubhelp.com.

Boadway, R., & Bruce, N. (1984). A general proposition on the design of a neutral business tax. Journal of Public Economics, 24(2), 231–239, DOI: 10.1016/0047-2727(84)90026-4.

Boyd-Graber, J., & Blei, D. (2012). Multilingual topic models for unaligned text. arXiv preprint, DOI: 10.48550/arXiv.1205.2657.

Buocz, T. J. (2018). Artificial Intelligence in Court. Legitimacy Problems of AI Assistance in the Judiciary. Retskraft–Copenhagen Journal of Legal Studies, 2(1), 41–59, URL: static1.squarespace.com/static/59db92336f4ca35190c650a5/t/5ad9da5f70a6adf9d3ee842c/1524226655876/Artificial+Intelligence+in+Court.pdf.

Calvez, F., & Regis, N. (2007). Length of court proceedings in the member states of the Council of Europe based on the case law of the European Court of Human Rights. Council of Europe Publishing, 2nd edition, URL: marinacastellaneta.it/wp-content/uploads/2013/01/Rapport_2012_16_en.pdf.

Campello, R. J., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In Pacific-Asia conference on knowledge discovery and data mining, 160–172, DOI: 10.1007/978-3-642-37456-2_14.

Cao, Z., Li, S., Liu, Y., Li, W., & Ji, H. (2015). A novel neural topic model and its supervised extension. In Proceedings of the AAAI Conference on Artificial Intelligence, 29(1), 2210–2216, DOI: 10.1609/aaai.v29i1.9499.

Carlson, K., Dadgostari, F., Livermore, M. A., & Rockmore, D. N. (2021). A multinetwork and machine learning examination of structure and content in the United States code. Frontiers in Physics, 8, 676, DOI: 10.3389/fphy.2020.625241.

Carlson, K., Livermore, M. A., & Rockmore, D. N. (2020). The problem of data bias in the pool of published US appellate court opinions. Journal of Empirical Legal Studies, 17(2), 224–261, DOI: 10.1111/jels.12253.

Carree, M., Günster, A., & Schinkel, M. P. (2010). European antitrust policy 1957-2004: an analysis of commission decisions. Review of Industrial Organization, 36(2), 97–131, DOI: 10.1007/s11151-010-9237-9.

Carter, D. J., Brown, J., & Rahmani, A. (2016). Reading the High Court at a distance: topic modelling the legal subject matter and judicial activity of the High Court of Australia, 1903-2015. The University of New South Wales Law Journal, 39(4), 1300–1354, URL: opus.lib.uts.edu.au/bitstream/10453/63528/1/394-2.pdf.

CEPEJ – European Commission for the Efficiency of Justice. (2018). European judicial systems: Efficiency and quality of justice. Strasbourg: Council of Europe, URL: rm.coe.int/rapport-avec-couv-18-09-2018-en/16808def9c.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. Advances in neural information processing systems, 22, 1–9, URL: proceedings.neurips.cc/paper_files/paper/2009/file/f92586a25bb3145facd64ab20fd554ff-Paper.pdf.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. European journal of operational research, 2(6), 429–444, DOI: 10.1016/0377-2217(78)90138-8.

Chien, J. T., & Lee, C. H. (2017). Deep unfolding for topic models. IEEE transactions on pattern analysis and machine intelligence, 40(2), 318–331, DOI: 10.1109/TPAMI.2017.2677439.

Coglianese, C., & Dor, L. M. B. (2020). AI in Adjudication and Administration. Brooklyn Law Review, 86, 791–838, URL: brooklynworks.brooklaw.edu/cgi/viewcontent.cgi?article=2272&context=blr.

Dadas, S., Perełkiewicz, M., & Poświata, R. (2020). Pre-training polish transformer-based language models at scale. In Artificial Intelligence and Soft Computing: 19th International Conference, ICAISC 2020, Zakopane, Poland, October 12–14, 2020, Proceedings, Part II 19 (pp. 301–314). Springer International Publishing, DOI: 10.1007/978-3-030-61534-5_27.

Dadgostari, F., Guim, M., Beling, P. A., Livermore, M. A., & Rockmore, D. N. (2021). Modeling law search as prediction. Artificial Intelligence and Law, 29, 3–34, DOI: 10.1007/s10506-020-09261-5.

Das, R., Zaheer, M., & Dyer, C. (2015). Gaussian LDA for topic models with word embeddings. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 795–804, DOI: 10.3115/v1/P15-1077.

De Groot, M., Aliannejadi, M., & Haas, M. R. (2022). Experiments on generalizability of BERTopic on multi-domain short text. arXiv preprint, DOI: 10.48550/arXiv.2212.08459.

Deeks, A. (2019). The judicial demand for explainable artificial intelligence. Columbia Law Review, 119(7), 1829–1850, URL: jstor.org/stable/26810851.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. Journal of the American society for information science, 41(6), 391–407, DOI: 10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint, DOI: 10.48550/arXiv.1810.04805.

Dieng, A. B., Ruiz, F. J., & Blei, D. M. (2020). Topic modeling in embedding spaces. Transactions of the Association for Computational Linguistics, 8, 439–453, DOI: 10.1162/tacl_a_00325.

Doan, T. N., & Hoang, T. A. (2021). Benchmarking neural topic models: An empirical study. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, 4363–4368, URL: aclanthology.org/2021.findings-acl.382.pdf.

Egger, R., & Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. Frontiers in sociology, 7, 886498, DOI: doi.org/10.3389/fsoc.2022.886498.

Eklund, J., Levratto, N., & Ramello, G. B. (2020). Entrepreneurship and failure: two sides of the same coin? Small Business Economics, 54, 373–382, DOI: 10.1007/s11187-018-0039-z.

ElBialy, N., & Garcia-Rubio, M. A. (2011). Assessing judicial efficiency of Egyptian first instance courts: A DEA analysis. MAGKS joint discussion paper series in economics, 19, 1–28, URL: econstor.eu/bitstream/10419/56541/1/657907855.pdf.

Espasa, M., & Esteller-More, A. (2015). Analyzing judicial courts' performance: inefficiency vs. congestion. Revista de Economía Aplicada, 23(69), 61–82, URL: redalyc.org/pdf/969/96945385004.pdf

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In KDD Proceedings, Vol. 96, No. 34, 226–231, URL: cdn.aaai.org/KDD/1996/KDD96-037.pdf?source=post_page.

European Commission. (2021). The 2021 EU Justice Scoreboard. Brussels: European Commission, URL: commission.europa.eu/document/c6121790-3c0a-4b98-b49a-adc7cc9cd7c6_en?prefLang=pl.

Fusco, E., Laurenzi, M., & Maggi, B. (2021). Length of Trials in the Italian Judicial System: An Efficiency Analysis by Macro-Area. Justice System Journal, 42(1), 78–105, DOI: 10.1080/0098261X.2020.1852985.

Garcia-Posada, M., & Mora-Sanguinetti, J. S. (2015). Does (average) size matter? Court enforcement, business demography and firm growth. Small Business Economics, 44(3), 639–669, DOI: 10.1007/s11187-014-9615-z.

Giacalone, M., Nissi, E., & Cusatelli, C. (2020). Dynamic efficiency evaluation of Italian judicial system using DEA based Malmquist productivity indexes. Socio-Economic Planning Sciences, 72, 100952, DOI: 10.1016/j.seps.2020.100952.

Glaeser, E. L., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2004). Do institutions cause growth?. Journal of economic Growth, 9, 271–303, DOI: 10.1023/B:JOEG.0000038933.16398.ed.

Gozgor, G., Lau, C. K. M., Zeng, Y., & Lin, Z. (2019). The effectiveness of the legal system and inbound tourism. Annals of Tourism Research, 76, 24–35, DOI: 10.1016/j.annals.2019.03.003.

Giacomelli, S., & Menon, C. (2013). Firm size and judicial efficiency: evidence from the neighbour's court. Bank of Italy Temi di Discussione (Working Paper), 898, URL: citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=e78423bdd5089050494680 9e125a68952ef9ff01.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint, DOI: 10.48550/arXiv/2203.05794.

Hicks, J. R. (1939). The foundations of welfare economics. The economic journal, 49(196), 696–712, DOI: 10.2307/2225023.

Hinton, G. E., & Roweis, S. (2002). Stochastic neighbor embedding. Advances in neural information processing systems, 15, 1–8, URL: proceedings.neurips.cc/paper_files/paper/2002/file/6150ccc6069bea6b5716254057a194ef-Paper.pdf.

Hofmann, T. (1999). Probabilistic latent semantic indexing. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, 50–57, URL: dl.acm.org/doi/pdf/10.1145/312624.312649.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of educational psychology, 24(6), 417, DOI: 10.1037/h0071325.

Hutama, L. B., & Suhartono, D. (2022). Indonesian Hoax News Classification with Multilingual Transformer Model and BERTopic. Informatica, 46(8), 81–90, DOI: 10.31449/inf.v46i8.4336.

Iqbal, M. I., Susanto, S., & Sutoro, M. (2019). Functionalization of E-Court System in Eradicating Judicial Corruption at The Level of Administrative Management. Jurnal Dinamika Hukum, 19(2), 370–388, DOI: 10.20884/1.jdh.2019.19.2.2510.

Jansson, P., & Liu, S. (2017). Distributed representation, LDA topic modelling and deep learning for emerging named entity recognition from social media. In Proceedings of the 3rd Workshop on Noisy User-generated Text, 154–159, DOI: 10.18653/v1/W17-4420.

Jappelli, T., Pagano, M., & Bianco, M. (2005). Courts and banks: Effects of judicial enforcement on credit markets. Journal of Money, Credit and Banking, 37(2), 223–244, URL: jstor.org/stable/3838925.

Jin, M., Luo, X., Zhu, H., & Zhuo, H. H. (2018). Combining deep learning and topic modeling for review understanding in context-aware recommendation. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 1605–1614, DOI: 10.18653/v1/N18-1145.

Joachims, T. (1997). A probabilistic analysis of the Rocchio algorithm with TFIDF for text categorization. In ICML, 97, 143–151, URL: citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c52eb66e23b201cb44f567cbb270feadca532c9a.

Kaldor, N. (1939). Welfare propositions of economics and interpersonal comparisons of utility. The economic journal, 49(195), 549–552, DOI: 10.2307/2224835.

Kalyan, K. S., Rajasekharan, A., & Sangeetha, S. (2021). Ammus: A survey of transformer-based pretrained models in natural language processing. arXiv preprint, DOI: 10.48550/arXiv/2108.05542.

Kittelsen, S. A., & Førsund, F. R. (1992). Efficiency analysis of Norwegian district courts. Journal of Productivity Analysis, 3(3), 277–306, https://doi.org/10.1007/BF00158357.

Kłeczek, D. (2020). Polbert: Attacking polish nlp tasks with transformers. In Proceedings of the PolEval 2020 Workshop, pp. 79–88, URL: 2020.poleval.pl/files/poleval2020.pdf#page=79.

Kociołowicz-Wiśniewska, B., & Pilitowski B., & Burdziej, S. (2017). Ocena polskiego sądownictwa w świetle badań. Raport Fundacji Court Watch Polska, URL: monitorkonstytucyjny.eu/wp-content/uploads/2019/09/ocena_polskiego_sadownictwa_w_swietle_badan_vol_2.pdf.

Kornhauser, L. A. (1999). Appeal and supreme courts. In Encyclopedia of Law and Economics. Edward Elgar Publishing Limited, DOI: 10.4337/9781782540472.00007.

Kriz, V., & Hladka B. (2018). Czech legal text treebank 2.0. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation, 2387–2392, URL: aclanthology.org/L18-1713.pdf.

Kruczalak-Jankowska, J., Maśnicka, M., & Machnikowska, A. (2020). The relations between duration of insolvency proceedings and their efficiency (with a particular emphasis on Polish experiences). International Insolvency Review, 29(3), 379–392, DOI: 10.1002/iir.1392.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. Journal of the American statistical Association, 47(260), 583–621, URL: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=684ee7383ae4cc10a3b1d002f3cc97851521adc4.

Lacoste-Julien, S., Sha, F., & Jordan, M. (2008). DiscLDA: Discriminative learning for dimensionality reduction and classification. Advances in neural information processing systems, 21, 1-8, URL: proceedings.neurips.cc/paper_files/paper/2008/file/7b13b2203029ed80337f27127a9f1d28-Paper.pdf.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory and acquisition induction, and representation of knowledge. Psychological review, 104(2), 211, DOI: 10.1037/0033-295X.104.2.211.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. Discourse processes, 25(2–3), 259–284, DOI: 10.1080/01638539809545028.

Larochelle, H., & Lauly, S. (2012). A neural autoregressive topic model. Advances in Neural Information Processing Systems, 25, 1–9, URL: proceedings.neurips.cc/paper/2012/file/b495ce63ede0f4efc9eec62cb947c162-Paper.pdf.

Lauderdale, B. E., & Clark, T. S. (2014). Scaling politically meaningful dimensions using texts and votes. American Journal of Political Science, 58(3), 754–771, DOI: 10.1111/ajps.12085.

Leibon, G., Livermore, M., Harder, R., Riddell, A., & Rockmore, D. (2018). Bending the law: geometric tools for quantifying influence in the multinetwork of legal opinions. Artificial Intelligence and Law, 26, 145–167, DOI: 10.1007/s10506-018-9224-2.

Lewin, A. Y., Morey, R. C. & Cook, T. J. (1982). Evaluating the administrative efficiency of courts. Omega, 10(4), 401–411, DOI: 10.1016/0305-0483(82)90019-6.

Li, W., & McCallum, A. (2006). Pachinko allocation: DAG-structured mixture models of topic correlations. In Proceedings of the 23rd international conference on Machine learning, 577–584, DOI: 10.1145/1143844.1143917.

Li, C., Wang, H., Zhang, Z., Sun, A., & Ma, Z. (2016). Topic modeling for short texts with auxiliary word embeddings. In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, pp. 165–174, DOI: 10.1145/2911451.2911499.

Liu, Y., Liu, Z., Chua, T. S., & Sun, M. (2015). Topical word embeddings. In Proceedings of the AAAI Conference on Artificial Intelligence, 29(1), 2418–2424, DOI: 10.1609/aaai.v29i1.9522.

Liu, J. S., Lu, L. Y., Lu, W. M., & Lin, B. J. (2013). A survey of DEA applications. Omega, 41(5), 893–902, DOI: 10.1016/j.omega.2012.11.004.

Livermore, M. A., Beling, P., Carlson, K., Dadgostari, F., Guim, M., & Rockmore, D. N. (2020). Law Search in the Age of the Algorithm. Michigan State Law Review, 5, 1183–1239, URL: uvalaw-scholarship.s3.amazonaws.com/2020.5_Livermore_FINAL.pdf.

Livermore, M. A., Riddell, A. B., & Rockmore, D. N. (2017). The Supreme Court and the judicial genre. Arizona Law Review, 59, 837–901, URL: ariddell.org/papers/livermore-rockmore-riddell-judical-genre-59arizlrev837.pdf.

Loza Mencia, E., & Furnkrantz, J. (2010). Efficient multilabel classification algorithms for large-scale problems in the legal domain. In Semantic Processing of Legal Texts, 23–32, DOI: 10.1007/978-3-642-12837-0_11.

Luz De Araujo, P. H., & De Campos, T. (2020). Topic modelling brazilian supreme court lawsuits. In Legal Knowledge and Information Systems, 113–122, DOI: 10.3233/FAIA200855.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, 1(14), 281–297, URL: cs.cmu.edu/~bhiksha/courses/mlsp.fall2010/class14/macqueen.pdf.

Magalhães, P. C., & Garoupa, N. (2020). Judicial performance and trust in legal systems: Findings from a decade of surveys in over 20 European Countries. Social Science Quarterly, 101(5), 1743–1760, DOI: 10.1111/ssqu.12846.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics, 18(1), 50–60, URL: jstor.org/stable/2236101.

Marciano, A., Melcarne, A., & Ramello, G. B. (2019). The economic importance of judicial institutions, their performance and the proper way to measure them. Journal of Institutional Economics, 15(1), 81–98, DOI: 10.1017/S1744137418000292.

Maxwell, T., & Schafer, B. (2010). Natural language processing and query expansion in legal information retrieval: challenges and a response. International Review of Law, Computers & Technology, 24(1), 63–72, DOI: 10.1080/13600860903570194.

Mazzei, D., & Ramjattan, R. (2022). Machine Learning for Industry 4.0: A Systematic Review Using Deep Learning-Based Topic Modelling. Sensors, 22(22), 8641., DOI: 10.3390/s22228641.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint, DOI: 10.48550/arXiv.1802.03426.

McKay, C. (2020). Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making. Current Issues in Criminal Justice, 32(1), 22–39, DOI: 10.1080/10345329.2019.1658694.

Meeusen, W., & van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. International economic review, 435–444, DOI: 10.2307/2525757.

Mimno, D., Wallach, H., Naradowsky, J., Smith, D. A., & McCallum, A. (2009). Polylingual topic models. In Proceedings of the 2009 conference on empirical methods in natural language processing, 880–889, URL: aclanthology.org/D09-1092.pdf.

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In Proceedings of the 2011 conference on empirical methods in natural language processing, 262–272, URL: aclanthology.org/D11-1024.Pdf.

Moody, C. E. (2016). Mixing dirichlet topic models and word embeddings to make lda2vec. arXiv preprint, DOI: 10.48550/arXiv.1605.02019.

Mimno, D., & McCallum, A. (2012). Topic models conditioned on arbitrary features with dirichlet-multinomial regression. arXiv preprint, https://doi.org/10.48550/arXiv.1206.3278.

Mišćenić, E. (2019). The Effectiveness of Judicial Enforcement of the EU Consumer Protection Law. In Balkan Yearbook of European and International Law 2019, 129–153, DOI: 10.1007/16247_2019_8.

Morison, J., & Harkens, A. (2019). Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making. Legal Studies, 39(4), 618–635, DOI: 10.1017/lst.2019.5.

Mroczkowski, R., Rybak, P., Wróblewska, A., & Gawlik, I. (2021). HerBERT: Efficiently pretrained transformer-based language model for Polish. arXiv preprint, DOI: 10.48550/arXiv.2105.01735.

Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence. In Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics, 100–108, URL: aclanthology.org/N10-1012.pdf.

Nguyen, D. Q., Sirts, K., & Johnson, M. (2015). Improving topic coherence with latent feature word representations in map estimation for topic modeling. In Proceedings of the Australasian Language Technology Association Workshop 2015, 116–121, URL: aclanthology.org/U15-1014.pdf.

OECD (2013). Judicial performance and its determinants: a cross-country perspective, OECD Economic Policy Paper, 5, DOI: 10.1787/5k44x00md5g8-en.

Paatero, P., & Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics, 5(2), 111–126, DOI: 10.1002/env.3170050203.

Pareto, V. (1896). Cours d'économie politique: professé à l'Universjté de Lausanne (Vol. 1). F. Rouge.

Pearson, K. (1895). VII. Note on regression and inheritance in the case of two parents. Proceedings of the royal society of London, 58(347–352), 240–242, DOI: 10.1098/rspl.1895.0041.

Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50(302), 157–175, DOI: 10.1080/14786440009463897.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin philosophical magazine and journal of science, 2(11), 559–572, DOI: 10.1080/14786440109462720.

Qiang, J., Chen, P., Wang, T., & Wu, X. (2017). Topic modeling over short texts by incorporating word embeddings. In Advances in Knowledge Discovery and Data Mining: 21st Pacific-Asia Conference, PAKDD 2017, Jeju, South Korea, May 23–26, 2017, Proceedings, Part II 21, 363–374, DOI: 10.1007/978-3-319-57529-2_29.

Rabinovich, M., & Blei, D. (2014). The inverse regression topic model. In International Conference on Machine Learning, 199–207, URL: proceedings.mlr.press/v32/rabinovich14.pdf.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. Preprint 1–12, URL: mikecaptain.com/resources/pdf/GPT-1.pdf.

Ramage, D., Hall, D., Nallapati, R., & Manning, C. D. (2009, August). Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In Proceedings of the 2009 conference on empirical methods in natural language processing, pp. 248–256, URL: aclanthology.org/D09-1026.pdf.

Ramos, J. (2003). Using tf-idf to determine word relevance in document queries. In Proceedings of the first instructional conference on machine learning, 242(1), 29–48, URL: citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b3bf6373ff41a115197cb5b30e57830c16130c2c.

Re, R. M., & Solow-Niederman, A. (2019). Developing artificially intelligent justice. Stanford Technology Law Review, 22, 242–289, URL: law.stanford.edu/wp-content/uploads/2019/08/Re-Solow-Niederman_20190808.pdf.

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084, DOI: 10.48550/arXiv.1908.10084.

Roder, M., Both, A., & Hinnenburg, A. (2015). Exploring the space of topic coherence measures. In Proceedings of the eight ACM international conference on Web search and data mining, 399–408, DOI: 10.1145/2684822.2685324.

Rodrik, D., Subramanian, A., & Trebbi, F. (2004). Institutions rule: the primacy of institutions over geography and integration in economic development. Journal of economic growth, 9, 131–165, DOI: 10.1023/B:JOEG.0000031425.72248.85.

Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2012). The author-topic model for authors and documents. arXiv preprint, DOI: 10.48550/arXiv.1207.4169.

Rybak, P., Mroczkowski, R., Tracz, J., & Gawlik, I. (2020). KLEJ: Comprehensive benchmark for Polish language understanding. arXiv preprint, DOI: 10.48550/arXiv.2005.00630.

Salihu, H. A., & Gholami, H. (2018). Mob justice, corrupt and unproductive justice system in Nigeria: An empirical analysis. International Journal of Law, Crime and Justice, 55, 40–51, DOI: 10.1016/j.ijlcj.2018.09.003.

Samaha, A. M., Heise, M., & Sisk, G. C. (2020). Inputs and Outputs on Appeal: An Empirical Study of Briefs, Big Law, and Case Complexity. Journal of Empirical Legal Studies, 17(3), 519–555, DOI: 10.1111/jels.12263.

Sangaraju, V. R., Bolla, B. K., Nayak, D. K., & Kh, J. (2022). Topic Modelling on Consumer Financial Protection Bureau Data: An Approach Using BERT Based Embeddings. arXiv preprint, DOI: 10.48550/arXiv.2205.07259.

Santolino, M. (2010). Determinants of the decision to appeal against motor bodily injury judgements made by Spanish trial courts. International Review of Law and Economics, 30(1), 37–45, DOI: 10.1016/j.irle.2009.09.002.

Santos, S. P., & Amado, C. A. (2014). On the need for reform of the Portuguese judicial system – Does Data Envelopment Analysis assessment support it? Omega, 47, 1–16, DOI: 10.1016/j.omega.2014.02.007.

Scarpino, I., Zucco, C., Vallelunga, R., Luzza, F., & Cannataro, M. (2022). Investigating topic modeling techniques to extract meaningful insights in Italian long COVID narration. BioTech, 11(3), 41–55, DOI: 10.3390/biotech11030041.

Schmitz, A. J. (2019). Expanding access to remedies through E-court initiatives. Buffalo Law Review, 67, 89–163, URL: digitalcommons.law.buffalo.edu/cgi/viewcontent.cgi?article=4724&context=buffalolawreview.

Sharma, D., Kumar, B., & Chand, S. (2017). A survey on journey of topic modeling techniques from SVD to deep learning. International Journal of Modern Education and Computer Science, 9(7), 50, DOI: 10.5815/ijmecs.2017.07.06.

Shi, M., Liu, J., Zhou, D., Tang, M., & Cao, B. (2017). WE-LDA: a word embeddings augmented LDA model for web services clustering. In 2017 ieee international conference on web services (icws), 9–16, DOI: 10.1109/ICWS.2017.9.

Siemaszko, A., Ostaszewski, P., Klimczak, J., & Włodarczyk-Madejska, J. (2019). Sense of security among residents of Warsaw. Survey results. Law in action, 38, 140–158, DOI: 10.32041/pwd.3809.

Smuda, F., Bougette, P., & Hüschelrath, K. (2015). Determinants of the duration of European appellate court proceedings in cartel cases. Journal of Common Market Studies, 53(6), 1352–1369, DOI: 10.1111/jcms.12259.

Spearman, C. (1904). The proof and measurement of association between two things. The American Journal of Psychology, 15(1), 72–101, URL: digamoo.free.fr/spearman1904a.pdf.

Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. (2012). Exploring topic coherence over many models and many topics. In Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning, 952–961, URL: aclanthology.org/D12-1087.pdf.

Terragni, S., Fersini, E., Galuzzi, B. G., Tropeano, P., & Candelieri, A. (2021). OCTIS: Comparing and optimizing topic models is simple!. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, 263–270, DOI: 10.18653/v1/2021.eacl-demos.31.

Thompson, L., & Mimno, D. (2020). Topic modeling with contextualized word representation clusters. arXiv preprint, DOI: 10.48550/arXiv.2010.12626.

Ulenaers, J. (2020). The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?. Asian Journal of Law and Economics, 11(2), DOI: 10.1515/ajle-2020-0008.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(11), 2579–2605, URL: jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf.

Virtucio, M. B. L., Aborot, J. A., Abonita, J. K. C., Avinante, R. S., Copino, R. J. B., Neverida, M. P., Osiana, V. O. (2018). Predicting decisions of the philippine supreme court using natural language processing and machine learning. In 2018 IEEE 42nd annual computer software and applications conference (COMPSAC), 2, 130–135, DOI: 10.1109/COMPSAC.2018.10348.

Voigt, S. (2016). Determinants of judiciary efficiency: A survey. European Journal of Law and Economics, 42(2), 183–208, DOI: 10.1007/s10657-016-9531-6.

Wachter, S., Mittelstadt, B., & Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. Computer Law & Security Review, 41, 105567, DOI: 10.1016/j.clsr.2021.105567.

Wang, X., & Yang, Y. (2020). Neural topic model with attention for supervised learning. In International Conference on Artificial Intelligence and Statistics, 1147–1156, URL: proceedings.mlr.press/v108/wang20c/wang20c.pdf.

Woliński, M. (2014). Morfeusz reloaded. In Proceedings of the ninth international conference on language resources and evaluation, LREC, 1106–1111, URL: citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=a6115f4593e7da10d5d990c42ee5c242a7c56058.

Xun, G., Gopalakrishnan, V., Ma, F., Li, Y., Gao, J., & Zhang, A. (2016). Topic discovery for short texts using word embeddings. In 2016 IEEE 16th international conference on data mining (ICDM), 1299–1304, DOI: 10.1109/ICDM.2016.0176.

Yeung, L. L., & Azevedo, P. F. (2011). Measuring efficiency of Brazilian courts with data envelopments analysis (DEA). IMA Journal of Management Mathematics, 22(4), 343–356, DOI: 10.1093/imaman/dpr002.

Zankadi, H., Idrissi, A., Daoudi, N., & Hilal, I. (2022). Identifying learners' topical interests from social media content to enrich their course preferences in MOOCs using topic modeling and NLP techniques. Education and Information Technologies, 1–18, DOI: 10.1007/s10639-022-11373-1.

Zhao, H., Phung, D., Huynh, V., Jin, Y., Du, L., & Buntine, W. (2021). Topic modelling meets deep neural networks: A survey. arXiv preprint, DOI: 10.48550/arXiv.2103.00498.