

ISSN: 2543-6821 (online)

Journal homepage: <http://ceej.wne.uw.edu.pl>

Maciej Świtała

# Predicting the Amount of Compensation for Harm Awarded by Courts Using Machine-Learning Algorithms

**To cite this article**

Świtała, M. (2024). Predicting the Amount of Compensation for Harm Awarded by Courts Using Machine-Learning Algorithms. Central European Economic Journal, 11(58), 214-232.

**DOI:** 10.2478/ceej-2024-0015

 To link to this article: <https://doi.org/10.2478/ceej-2024-0015>



Maciej Świtala University of Warsaw, Faculty of Economic Sciences, Długa 44/50, 00-241 Warsaw, Poland  
corresponding author: ms.switala@uw.edu.pl

# Predicting the Amount of Compensation for Harm Awarded by Courts Using Machine-Learning Algorithms

---

## Abstract

The present study aims to explain and predict the monetary amount awarded by courts as compensation for harm suffered. A set of machine-learning algorithms was applied to a sample of decisions handed down by the Polish common courts. The methodology involved two steps: identification of words and phrases whose counts or frequencies affect the amounts adjudicated with LASSO regression and expert assessment, then applying OLS, again LASSO, random forests and XGBoost algorithms, as well as a BERT approach to make predictions. Finally, an in-depth analysis was undertaken on the influence of individual words and phrases on the amount awarded. The results demonstrate that the size of awards is most strongly influenced by the type of injury suffered, the specifics of treatment, and the family relationship between the harmed party and the claimant. At the same time, higher values are awarded when compensation for material damage and compensation for harm suffered are claimed together or when the claim is extended after it was filed.

## Keywords

compensation amount | harm | machine learning | Polish courts | prediction

## JEL Codes

K150

---

## 1. Introduction

Discussion of the potential use of machine-learning algorithms in legal sciences is recently causing widespread discourse. Primarily, the operation of the judiciary is analysed. Specifically, a number of researchers aim at predicting the decisions of the courts (Aletras et al., 2016; Katz, Bommarito & Blackman, 2017; Medvedeva, Vols & Wieling, 2020; Sulea et al., 2017). This paper attempts to follow the trend observed in the literature as it aims at explaining and predicting the amount of money awarded as compensation for harm suffered with the use of machine-learning algorithms.

When it comes to determining the amount of compensation for harm, it is in fact an abstract valuation of what amount compensates for negative experiences. Thus, the issue is particularly difficult for both lawyers and judges. The former are not infrequently forced to extrapolate the extent of harm on the basis of the literature, case law, and their own experience to precisely determine the value of the subject matter of the dispute as it determines the amount of the court

fee and influences the burden of court costs on the parties. As for the judges, they in turn face the difficult task of assessing the reasonableness of the amount of such a claim. Moreover, in pursuance to standards of the democratic rule of law and specific law provisions (Articles 327<sup>1</sup> § 1, 328 § 1 of the Polish Code of the Civil Procedure), they should draft convincing reasons for each judgement. Also, I believe that reasoning of a poor quality can lead to more appeals and place a higher burden on the judicial system and therefore negatively affect its efficiency and effectiveness. Finally, public opinion may question the awarded amount of compensation. and its assessment can affect the legitimacy of a political system.

Machine-learning algorithms were applied, as they easily capture potential nonlinear relationships in the data. The econometric algorithms, even though they also make it possible to capture nonlinearity, require prior assumptions on the type of relationship between the variables (i.e., by specifying interactions between them or transforming variables). With no assumptions made (i.e., looking for them, for example, by trying all possible interactions and different transformations,

the number of variables substantially increases). Specifically in this research, where the number of independent variables was higher than the number of observations, it would make it impossible to successfully build the classic model.

As for the research hypotheses, it should be expected that machine-learning algorithms enable more accurate prediction of the amount of compensation awarded for harm suffered than do econometric alternatives. Moreover, I expect that the words and collocations whose occurrence in the body of the judgement will most strongly affect the amount awarded should be those contained in the body of the key provisions of the law. In the Polish legal order, harm is compensable only on the basis of specific provisions or contracts. As for the former, those include specific words and phrases, such as “health”, “medical treatment”, “work”, “death”, “family bound”, and “personal rights”. At a higher level of detail, I also expect words related to the above or detailing, e.g., defining the specific injuries and names of family members. Finally, I expect that certain procedural steps affect the amounts awarded. Those can be, in particular, the transformations on the object and subject sides of the process. There is no doubt that the extension of the claim (i.e., the demand for a higher amount after a proceeding has started), should imply the award of higher amounts. After all, such an adjustment of the claim should be made in cases where the evidence obtained later suggests a higher win.

Despite the undoubted practical and social relevance of the issue discussed, there are very few studies in the literature on the use of machine-learning tools for the prediction of the amount of compensation specifically for harm suffered (Dal Pont et al., 2023; Hsieh, Chen & Sun, 2021; Torres, Guterres & Celestino, 2023; Yeung, 2019). Therefore, the main contribution of this research is that this paper is the first to explain and predict monetary amounts awarded as compensation for harm suffered by applying machine-learning algorithms to a data set that is not limited to judgements pronounced in specific types of cases but involves a distinctively heterogeneous set of cases. Also, it is the first to use textual data covering multiple courts from across a country’s legal system. Furthermore, this research takes into consideration the Polish judicial system, which implies using a unique data set.

This data set included judgements of Polish common courts published through the System for Analysis of Court Decisions [in Polish, *System*

*Analizy Orzeczeń Sądowych (SAOS)*]. All available 5,348 judgements handed down from July 30, 2010 to December 6, 2022 in which compensation for harm was awarded were analysed.

As for the independent variables used to explain and predict the amounts awarded as compensation for harm suffered, two procedures for creating them were tried: counts of words and phrases in the body of a judgement and TF-IDF statistics describing those (Joachims, 1997). It was decided to automatically select variables that affect the modelled quantity with LASSO regression (Tibshirani, 1996). Next, the potential predictors thus obtained were still subjected to an expert review. This was primarily to avoid including in the final model the count or frequency of words and collocations indicating information that was not known before filing the first letter to the court. The aim was to minimize look-ahead bias—hence, the data set consisted of judgements, not lawsuits. I strongly believe this is the best solution I was able to adopt due to limited data availability.

The variables identified as affecting the dependent variable were considered further by applying different predicting algorithms. Specifically, LASSO (Tibshirani, 1996), random forests (Breiman, 2001), extreme gradient boosting (Chen et al., 2015; Friedman, 2001), the BERT algorithm (Devlin et al., 2018), as well as linear regression were applied. The latter was involved to compare machine learning with classical econometric algorithms as it appears to be a well-established practice in the literature (Chlebus, Dyczko & Woźniak, 2021).

The Polish perspective was considered because it does not differ much from most European countries in terms of obligation law. This can be perceived as a far-reaching result of the wide adoption of law provisions common in the Roman law (Wołodkiewicz & Zabłocka, 2014) as well as the influence of foreign legislation to shape the Polish Code of Obligations and therefore indirectly the currently in force Polish Civil Code (Brzozowski, 2021). At the same time, however, the Polish judicial system appears to be unique in its ineffectiveness in comparison to other European Union countries (Bełdowski, Dąbroś & Wojciechowski, 2020; European Commission, 2021; Kruczalak-Jankowska, Maśnicka & Machnikowska, 2020). Furthermore, Poland is one of the postcommunist countries that has succeeded in transforming itself from central planning to a market economy (Balcerowicz, 2005), and this fact creates certain minor idiosyncrasies in Polish obligation law. All this makes any consideration

of the Polish legal system undeniably unique, yet similar when it comes to the legal basis.

The remaining part of the article is structured as follows: firstly, principles on compensation for harm suffered in the Polish judicial order and the literature on compensation amount prediction are outlined. Next, the data used and methods applied are concisely reviewed. Afterward, the results are discussed. The last section summarizes the main conclusions.

## 2. Related Work

### 2.1. Compensation for Harm Suffered in the Polish Judicial Order

The titles from which the fulfilment of civil liability obligations arise include two types of liability: indemnity and non-indemnity liabilities. As for the indemnity liability regimes, they include liability under a number of titles, including non-fulfilment or improper fulfilment of an obligation (in Latin, *ex contractu*), commission of a tort (in Latin, *ex delicto*) and others (Kaliński, 2021a, 2021b). The main feature of indemnity liability is the intent to compensate for damage. There is no statutory definition of damage in the Polish legal system and the doctrine used to define it imprecisely (Kaliński, 2021b; Radwański, Olejniczak & Grykiel, 2022). Still, it seems indisputable to call damage the impairment that the damaged party suffers in the sphere of their property or personality as a result of events that violate their autonomy (Kaliński, 2021a).

The literature deals with material and non-material damage. The criterion for this division is the sphere of occurrence of the effects of the damaging event (Kaliński, 2021b; Radwański et al., 2022). The obligation to compensate for material damage is the rule. In the case of non-material damage, if it results from the violation of personal rights, it can be compensated under a specific provision (usually Articles 445, 446 § 4, 446<sup>2</sup>, or 448 of the Polish Civil Code). Any non-material damage deriving from another type of violation shall be subject to compensation only under the agreement of the parties (Kaliński, 2021a). In addition, the Polish legislator introduced (but did not define) the term “harm” (in Polish, *krzywda*). The author of this paper, following the interpretation expressed by Kaliński (2021a), is inclined to interpret that harm means non-material

damage recognized in personal rights and is therefore compensable under a specific provision.

Key to the considerations made in this paper is Article 445 of the Polish Civil Code. The court may award the injured party an appropriate sum as monetary compensation for the harm suffered. Such empowerment of the court is limited to the cases listed in the preceding article (i.e., bodily injury causing disorder of health, as well as the total or partial loss of earning capacity of the injured party). In addition, compensation may be awarded in cases of deprivation of liberty, inducement by deception, rape, or abuse of a relationship of dependence to submit to a lewd act. The court may also award the immediate family members of a person who died as a result of bodily injury or disorderly conduct by tort under Article 446 § 4 of the Polish Civil Code. Also, the immediate family members of a person who suffered a bodily injury or infliction of a disorder of health of a severe and permanent nature and cannot establish or continue a family relationship can be awarded an appropriate sum as monetary compensation for the harm suffered as specified in Article 446<sup>2</sup> of the Polish Civil Code. Finally, an appropriate sum may also be awarded to the one whose personal good has been violated (i.e., under Article 448 in conjunction with Articles 23 and 24 of the Polish Civil Code).

Polish jurisprudence and literature agree that compensation for non-material damage has primarily a compensatory function (i.e., its primary purpose is to compensate the injured party for his or her negative experiences). Other functions, such as preventive or punitive, are less important (Kaliński, 2021a; Kryla-Cudna, 2018; Radwański et al., 2022).

The literature lists a number of factors influencing the size of the compensation for harm to be awarded. The type of personal good that has been violated is indicated, as well as the intensity and extent of the violation. In addition, the literature points to the importance of the degree of fault of the perpetrator, and sometimes to the individual financial and personal situation of the victim (Kaliński, 2021a; Radwański et al., 2022). Also, the subsequent behaviour of the responsible person may affect the extent of the compensation (Safjan, 2020). The standard of living of society also appears to be relevant to the size of the amount awarded, but only in the sense that in economically developed countries higher sums are compensated (Kaliński, 2021a).



## 2.2. Predicting the Amount of Compensation for Harm

So far, the problem of predicting the amount of compensation specifically for harm suffered has not been the subject of much research in the literature. Yeung (2019) introduced a variant of the BERT algorithm (Devlin et al., 2018), which was fine-tuned to enhance its performance on German legal texts. The algorithm was compared with its alternatives from a number of aspects. Most importantly, a regression model was built that aimed at predicting the amount of compensation to be awarded by courts. The German Legal BERT introduced by the author outperformed competitive approaches despite a linear regression model based on TF-IDF (Joachims, 1997). Nonetheless, the author did not distinguish between material and non-material damage.

Hsieh et al. (2021) focused on prediction of the amount awarded as compensation for non-material damage by the Taiwan Taichung District Court. As for the results, random forests outperformed KNN and CART. What seems a substantial limitation in comparison to this article, the authors considered only the cases of mental suffering due to fatal car accidents. Also, they did not analyse cases that involved more than one entity on the part of the defendant. Similarly, Dal Pont et al. (2023) also focused on the issue of predicting amounts awarded as compensation for non-material damage. Working on the example of the Brazilian legal system, the authors analysed cases issued in the State Special Court at the Federal University of Santa Catarina. The introduced pipeline involved multiple machine-learning algorithms with the best-performing XGBoost algorithm. It should be emphasized that the sample consisted of court decisions resolving “daily and minor conflicts” in which customers sued airlines on the basis of the Brazilian Code of Consumer Protection.

Torres et al. (2023) aimed at predicting compensation amounts awarded for both material and non-material damage. The authors considered a classification problem (i.e., decoded awarded amounts to categories of “low”, “medium” and “high”). Surprisingly, multinomial logistic regression outperformed the random forests algorithm as well as naive Bayes and support vector machine approaches. What should be noted, a rather thematically homogeneous sample of lawsuits was considered, as all the cases involved suing airlines. Also, the data did not include the texts of the judgements pronounced, but variables describing the factual state.

Related to the prediction of amounts of compensation received specifically for harm suffered, studies on monetary amounts adjudicated as compensation awarded due to other types of damage can be found. As for prediction of material damage, it appears not a frequent area of interest, as changes in property are relatively easy to calculate or value. Still, it sometimes appears difficult to determine the extent of property damage. As a result, in practice parties to a contract usually agree to contractual penalties (liquidated damages). Alshboul et al. (2022a, 2022b) aimed at predicting those in cases concerning highway construction projects. With its use of a broad set of machine-learning algorithms, it should be labelled one of the few studies considering, despite substantial legal differences, matters similar to those discussed in this article. Also, a mention of studies on punitive damages should be made. This legal institution is adopted primarily in the common law countries and is substantially different from compensation for harm suffered, primarily because of its repressive and deterrent character (Andrych-Brzezińska, 2020; Kochanowski, 2019). Still, when referring to empirical legal studies on compensation, Eisenberg et al. (2006, 2010, 2015) should be mentioned as adopting the econometric models with the aim of predicting the punitive damages.

Beyond the above, many studies on predicting court decisions as a whole have been carried out. These are usually simplified to considering classification problems (i.e., it is predicted if a court deems a certain factual state to be a violation of a certain provision or not). Also, structured data sets with variables describing case characteristics are most commonly used. Operating only on the texts of judgements should be perceived as an exception. When it comes to more recent literature, US Supreme Court decisions have been analysed (Katz et al., 2017), along with the European Court of Human Rights case law (Aletras et al., 2016; Medvedeva et al., 2020; Valvoda, Cotterell & Teufel, 2023) and judgements pronounced in France (Sulea et al., 2017), Germany (Waltl et al., 2017), the Philippines (Virtucio et al., 2018), UK (Strickson & La Iglesia, 2020), Turkey (Mumcuoğlu et al., 2021). Also noteworthy are the latest extensive literature reviews on the subject by Cui, Shen & Wen (2023) and by Medvedeva, Wieling & Vols (2023).

To summarize the above, the issue of the prediction of the amount of compensation awarded by the court for harm suffered is rarely considered in the literature. Instead, the outcome of the trial in general is most often predicted. This indicates that there is a research gap,

one this article aims to fill. As far as machine-learning algorithms that are typically used for prediction, it is not a rule that these approaches are better than classical econometric methods. In particular, Yeung (2019) and Strickson & La Iglesia (2020) obtained better results using the latter. Nonetheless, the superiority of machine learning over classical modelling approaches is to be expected, primarily because of the ability of the former to take non-linear relationships into account without previous assumptions on their shape or exact integration between the independent variables.

When it comes to explaining the variability of the amounts of compensation awarded for harm suffered, this also appears marginalized in the empirics. Thus, expectations as to the influence of the number or frequency of particular words and phrases in the body of a judgement on the amount of compensation to be awarded are formulated in this study based on the non-empirical studies constituting the Polish legal literature. Most of the doctrine takes the view that compensation for harm should correspond to the extent of the harm suffered. It is therefore most sensible to construct a typology of situations in which a claim for harm is justified. Undeniably, this should coincide with the hypotheses of the special legal provisions to which the claims must refer. Thus, the counts and frequencies of the words and phrases used therein are expected to have the strongest influence on the amount awarded. Of course, words detailing them are also expected, which has to do with the general nature of the content of the legal provisions. The hypothesis of an effect on the amount of non-material damage compensation to be awarded of subjective and objective procedural transformations is instead an overly logical but expert conjecture of the author.

## 3. Data and Methods

### 3.1. Data

Data used in this research were obtained from the System of Analysis of Courts Decisions, at the web address [saos.org.pl](http://saos.org.pl). This portal was established to publish the content of judgements of extraordinary and common courts in Poland. The scope of the judgements published there was determined by a panel of the common courts judges. No exempted, deemed, or irrelevant content is published there. However, at the same time, there is no formal obligation to share texts of judgements on the site. As a result, the portal

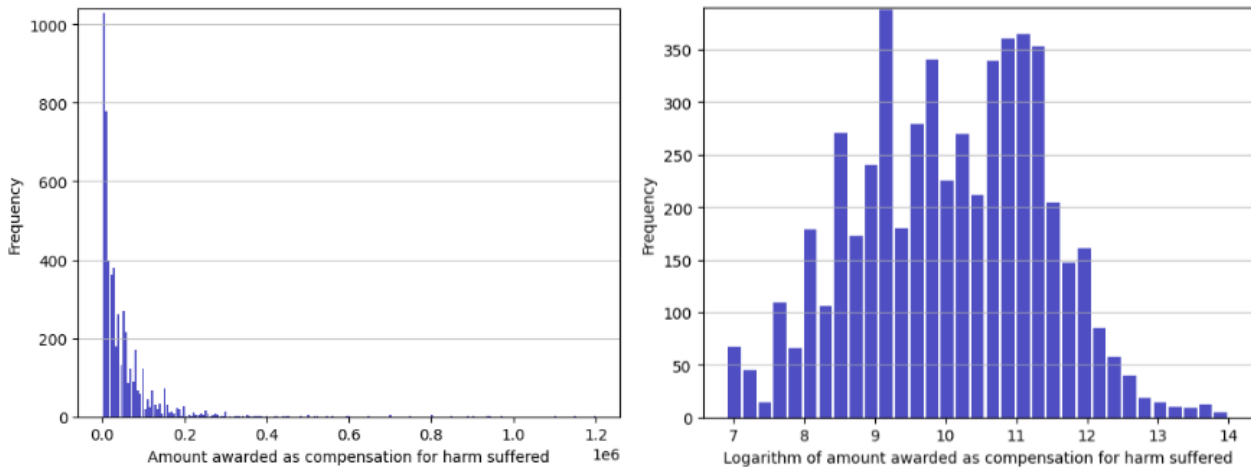
publishes not every portion of designated content. This undoubtedly represents a burden on data sources in terms of their representativeness. Still, to the best knowledge of the author, it is the most comprehensive source of data available. Technically, all judgements of the Polish common courts published at [saos.org.pl](http://saos.org.pl) were collected. Next, the sample was narrowed to judgements that awarded any compensation for harm suffered. The number of these was 5,348. They were issued from July 30, 2010 to December 6, 2022 (state as of January 13, 2023 as some judgements appear to be published with a delay).

In the sample, values of compensation awarded for harm suffered varied between 1,000 and 1,200,000 PLN. On average, the Polish courts award 51,489 PLN. The median was noticeably lower (i.e., 25,000 PLN), which demonstrates a noticeable right-sided skew of the distribution of the dependent variable (Fig. 1, generated with Python *matplotlib* library). The variable to be predicted was log-transformed, as this can bring relationships closer to the linear and therefore make them easier to analyse. It also reduces the possible effect of outliers on the predictions and turns multiplicative relationships into additive ones, which also allows the models to benefit.

## 3.2. Methods

### 3.2.1. Data Preprocessing

As the amount of compensation to be awarded depends primarily on the extent of harm suffered, sentences (i.e., parts of judgements involving settlement of a case on its merits) were removed (i.e., only statements of reasons were analysed). Also, non-informative elements such as punctuation, special signs, and one-letter words were removed. Next, texts were tokenized, which resulted in obtaining 141,555 unique tokens. All were reduced to their original form with a lemmatization operation performed with the *Morfeusz 2* program (Kieraś & Woliński, 2017) which, to the best of my knowledge, currently allows the most accurate lemmatization of the Polish language. After lemmatization, 52,628 unique tokens remained. Next, Polish stop words were removed using a dictionary provided within the Python library *stop\_words*. This reduced the number of unique tokens to 52,542. With bigrams and trigrams introduced, it made the overall number of unique tokens 7,387,673. Still, to enable a reasonable time for processing data, words



**Figure 1.** Distribution of Amounts Awarded as Compensation for Harm Suffered and Their Logarithms

and phrases that were observed in less than 1% of the analysed judgements were removed, which reduced the number of tokens to 55,264. Next, both counts and TF-IDF statistics (Joachims, 1997) were calculated for all tokens, and those were further considered as independent variables in different models.

### 3.2.2. Feature Selection

For both approaches to constructing the independent variables for the later models, the number of variables was noticeably high (i.e., equal to the number of tokens that remained after all preprocessing operations). Using all of them would prevent effective optimization of the model parameters and expose the algorithms to overfitting. For this reason, it was decided to semi-automatically select variables that affect the amount awarded for harm suffered. Firstly, LASSO regression (Tibshirani, 1996) was performed. With the regularization applied, some coefficients are shrunk and others are equated to zero. As a result, LASSO retains only the most prominent features in the model. It is also worth noting that variables were standardized before applying LASSO. In the next step, the potential predictors obtained with LASSO were subjected to an expert review in order to reduce possible look-ahead bias. More specifically, tokens that involved information unknown before the lawsuit was filed were removed. I believe it was the best solution for the problem of limited data availability in the considered case. The texts of lawsuits are not available. To exemplify words and collocations removed expertly, those specifying the type of adjudicating court were, in particular, not used in further analyses.

After all, in Polish conditions indicating whether a district or regional court is competent often depends on the value of the subject matter of the dispute that stems from the lawyer's initial approximation of the amount of compensation for harm.

As for the approach involving counts of n-grams, at first LASSO regression recommended 1,077 tokens having influence on values awarded by courts. Then, those were assessed expertly, which made the final number of the tokens 531. When considering TF-IDF statistics as independent variables, LASSO named 395 tokens, so that expert assessment ended up with 297 tokens TF-IDF identified as the most prominent predictors.

### 3.2.3. Applied Machine-Learning Algorithms

The data set was randomly split into training and test subsets at a ratio of 80:20. When training different models, a fivefold cross-validation procedure was applied in order to prevent the model from overfitting. Optimization of model parameters was performed based on a grid search approach with respect to root mean squared error measure.

As for machine-learning algorithms applied to the selected features, once again the LASSO regression (Tibshirani, 1996) was used. The repeated use of this approach, this time specifically for the prediction of compensation, results from the fact that the expert removal of some of the tokens after the previously described automatic selection of variables undeniably influences the estimates obtained.

More flexible algorithms, based on the regression trees concept, were also applied. Their selection was made in such a way as to use different approaches that deal well with potential nonlinear dependencies (i.e., in contrast to econometric tools, do not require prior assumptions on the shape of nonlinear dependencies). As the most common and competing ideas in recent machine learning are bagging (Efron, 1979) and boosting (Kearns & Valiant, 1989), representatives of both (i.e., random forests (Breiman, 2001) and XGBoost (Chen et al., 2015; Friedman, Hastie, & Tibshirani, 2000; Friedman, 2001), respectively, were considered).

The idea behind random forests is to repeatedly build regression trees (Breiman et al., 1984) (i.e., a structure of sequential data partitions conditioned on regressors' values and leaves representing the average values of the dependent variable in subsets obtained after all splits in a path). In a random forest, every single tree is built on a different bootstrap sample, and each split considers only a random subset of regressors. This represents a simplification aimed at speeding up the process of building a large number of trees. The final prediction of the value of the explanatory variable for a given observation is the average prediction from all the trees constructed. What should be addressed, according to Breiman (2001), is that random forests do always converge—and this appears to solve possible overfitting problems.

Intuitively, XGBoost first involves constructing a simple regression tree model. Then another tree is constructed, though this time it is designed to reduce errors made by the previous tree and the final one. Intuitively, boosting enables assigning higher weights to the subsets that were wrongly predicted in the previous iteration. More technically, yet not expanding much on the rather complex mathematical foundations (Friedman, 2001), boosting aims at sequential application of an algorithm to reweighted data (Friedman et al., 2000). Final prediction is obtained by calculating weighted sum from all the constructed trees. XGBoost should be named fine-tuned as to the efficiency of computing and scalable implementation of the briefly explained framework (Chen et al., 2015).

The BERT algorithm (Devlin et al., 2018) was also considered. Technically, BERT trains deep bidirectional representations from unlabeled text, whereby it conditions context in layers. In simple terms, BERT involves two stages: pretraining and fine-tuning. In the former, the model is trained

on unlabeled data over different pretraining tasks. Pretrained BERT is a language model that can be easily fine-tuned. As a result, it can be applied relatively easily for a range of tasks. Specifically, in the case of predicting the amount awarded as compensation, it is possible to add an additional output layer enabling consideration of a regression task.

BERT, despite being a modern approach to natural-language processing, nevertheless has a number of drawbacks that can potentially affect its use. Its complexity affects its explainability (Devlin et al., 2018). Moreover, BERT's structure was initially built for different purposes than regression tasks. Numerous studies confirm the superiority of BERT over feature-based approaches in classification problems (Balagopalan et al., 2020; González-Carvajal & Garrido-Merchán, 2020; Mumcuoğlu et al., 2021), but very few consider comparing BERT and other algorithms' performance in regression tasks (Yeung, 2019). Moreover, even though Polish versions of BERT are being developed all the time, to stop with the most recent publication from Mroczkowski et al. (2021), those are still not easily available, e.g. the Python library *transformers* does not involve the Polish version of BERT but stops with a multilingual variant. Also, pretraining of BERT in the use of its own resources appears to be disproportionately costly. Furthermore, the ideal would be to train BERT on judgements of Polish courts, which does not appear possible due to poor data availability. Therefore, a multilingual variant of BERT was adopted in the research.

The well-established measures were used to assess the quality of the prediction made by each model. Precisely, these were: root mean squared error, mean absolute percentage error, and their median-based variants. Root mean squared error informs how much on average the prediction is wrong in units of measurement of the dependent variable. Mean absolute percentage error uses percentage values. Median measures are less sensitive to error outliers.

The study also applies tools that allow in-depth investigation of the influence of independent variables on the modelled value (i.e., explainable machine learning (Adadi & Berrada, 2018; Arrieta et al., 2020; Gunning & Aha, 2019). Specifically, an approach based on game theory named SHAP, which can be used for explaining any machine-learning model, was used. Simply stated, it calculates the importance value of each feature (Shapley, 1952). Moreover, the partial dependence idea was considered. This can be intuitively named the expected target response,



expressed as a function of input features (Hastie et al., 2009). It can be used, in particular, to assess the strength and direction of influence of individual variables on the modelled value.

## 4. Results and Discussion

### 4.1. Models Estimated

To be precise, the optimal alpha parameters in LASSO models were: 0.01 and 0.00001, with token counts and TF-IDF statistics tried as independent variables, respectively. Optimization of random forest algorithms for token counts ended with a maximum tree depth of 10 and 1,000 estimators to be considered in a single split. For TF-IDF, those were 10 and 300, respectively. As for XGBoost, in the case of token counts, optimal parameters were: maximum tree depth of 3, eta equal to 1, and gamma equal to 0. In the case of TF-IDF statistics, those were: 4, 1, and 0, respectively.

Prediction error measures obtained on the test set are presented in Table 1. The lowest values of all measures used were observed for the models based on the random forest algorithm. Judging from the measures describing errors in units of measurement of the dependent variable (i.e., root mean squared error and root median squared error), the one based on TF-IDF statistics was found to be the best. In contrast, as for the measures expressed in percentage scale (i.e., mean absolute percentage error and median absolute percentage error), the random forest model based on token counts was reported with the best performance. BERT was reported to have the worst performing model for all measures except the root mean squared error, as it was optimized with respect to this one.

It is hard to deem the predictive quality of models satisfactory. Moreover, it is hard to compare their quality with the literature. As mentioned before, there are few studies that are sufficiently similar to compare the results (i.e., those that consider compensation for non-material damage as a regression problem). Yeung (2019) did not distinguish between material and non-material damage. Also, he presented results only in terms of  $R^2$  measure. Finally, his study was rather about assessing his fine-tuned BERT's performance. Dal Pont et al. (2023) ended up with providing a root mean squared error below 2,000, which is a far better result than was obtained in this research. However, Dal Pont et al. (2023) considered the problem of predicting the amounts awarded in fairly

homogeneous factual situations (i.e., cases involving airlines as defendants). In addition, in that study, the authors analysed only 928 cases issued in the State Special Court resolving “daily and minor conflicts”. This truncation of the sample can potentially affect positively the quality of the prediction obtained. This study considers prediction on a sample from the entire common judiciary in terms of the topics in dispute.

Given the incomparability of the results obtained with the body of literature on the subject, it is important to consider the quality of the models constructed on purely praxeological grounds. The resulting measures of prediction error on the test set are undoubtedly high. At the same time, however, it should be borne in mind that the court, when deciding on compensation for harm suffered, is in each case bound by the claimant's demand. In other words, the court cannot award more than the amount requested in the claim. Thus, the question arises as to whether the model overestimations should indeed be considered errors. If the court has not awarded the entire amount claimed, the overestimation of the model should not be regarded as an error. With models' overestimates assumed not to be errors (Table 2), for the mean error measures in four different cases, four different algorithms were reported as the best-performing. Nevertheless, when it comes to median error measures, BERT was reported with zero values. This means that more than half of the models' predictions were in fact overestimates. These results allow us to build some intuition as to how much the constraint on the courts by the claimant's demand may affect the error rates presented in Table 1.

As the random forests algorithm was reported better than OLS in every standard configuration (Table 1), there is no reason to reject the research hypothesis, which stated machine learning's supremacy over econometric alternatives. In support of this, with token counts used as the independent variables, in three out of four considered measures, XGBoost outperformed LASSO, and LASSO outperformed OLS. However, with TF-IDF adopted as the independent variables, LASSO outperformed OLS, and OLS outperformed XGBoost.

### 4.2. Influence of Tokens on Compensation Amount

The token “family” was reported as having the strongest influence on the compensation amount (Figure 2, generated with Python library *shap*; tokens

**Table 1.** Error Measures Obtained on a Test Set With Different Algorithms Applied

Algorithm applied	Predictors	Root mean squared error	Mean Absolute percentage error	Root median squared error	Median absolute percentage error
OLS	Token counts	308,207.87	112.32	11,747.55	52.79
	TF-IDF	100,743.48	94.59	11,759.07	49.25
LASSO	Token counts	287,706.24	109.68	11,474.52	52.64
	TF-IDF	95,976.82	93.46	11,435.79	48.64
Random forests	Token counts	75,268.12	86.43	11,064.28	47.16
	TF-IDF	74,674.55	88.59	10,518.75	48.13
XGBoost	Token counts	94,271.95	105.27	12,654.46	52.18
	TF-IDF	192,230.66	102.79	12,821.45	51.48
Multilingual BERT	-	95,665.94	426.45	38,062.44	88.03

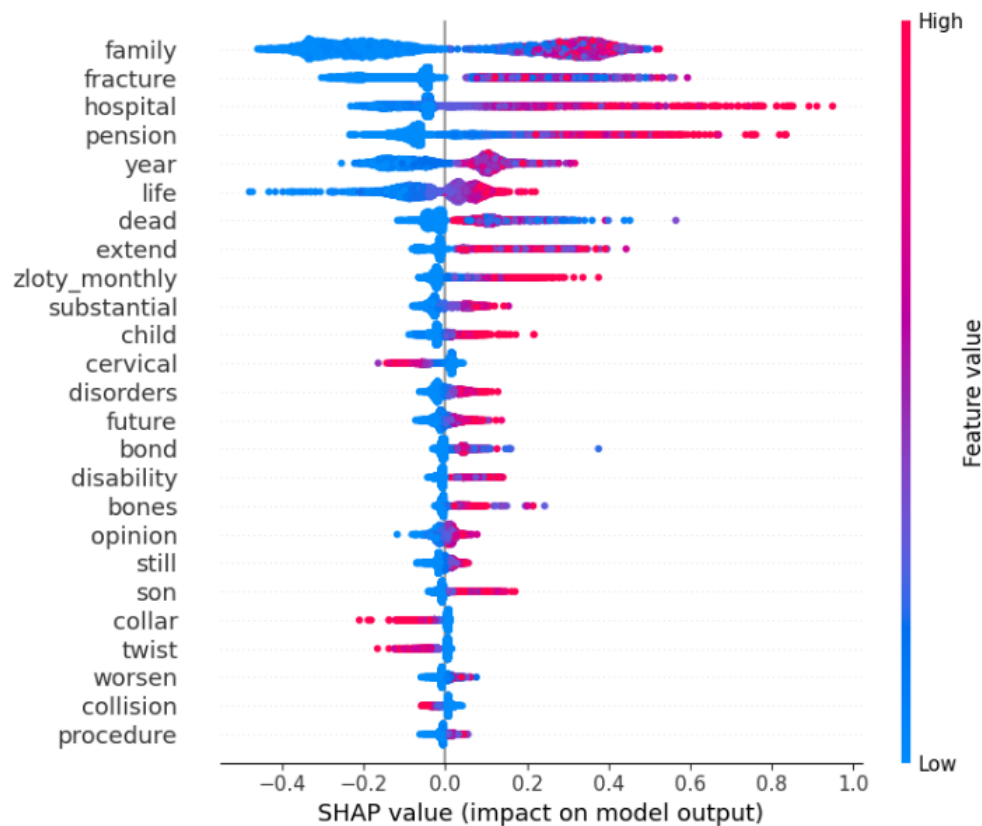
**Table 2.** Error Measures Obtained on a Test Set With Different Algorithms Applied and Models' Overestimates Assumed as Not Errors

Algorithm applied	Predictors	Root mean squared error	Mean Absolute percentage error	Root median squared error	Median absolute percentage error
OLS	Token counts	61,829.04	23.92	1,024.79	5.29
	TF-IDF	67,563.77	21.83	178.39	1.08
LASSO	Token counts	62,019.89	23.83	1,061.22	5.57
	TF-IDF	68,215.32	21.78	129.66	0.69
Random forests	Token counts	74,939.24	23.11	1,729.41	8.01
	TF-IDF	73,939.24	22.80	1,741.93	9.02
XGBoost	Token counts	71,414.42	23.83	1,069.24	7.93
	TF-IDF	65,657.99	24.13	1,209.66	7.68
Multilingual BERT	-	82,212.22	29.04	0.00	0.00

in this section were translated for the purpose of presentation). Its presence reflects the practical importance of the regulation of Articles 446 § 4 and 446<sup>2</sup> of the Polish Civil Code. The former specifically states that the court may award compensation for harm suffered to the immediate family members of a person who died as a result of bodily injury or disorderly conduct by tort. The latter refers to bodily injury or infliction of a disorder of health of a severe, permanent nature and at the same time requires the consequences of the inability to establish or continue a family relationship. Among the top 25 most impactful tokens, there are ones more associated with this

particular regulation (tokens such as „life“, „dead“, „child“, „bond“, and „son“).

Some of the most influential tokens indicate the type of health damage suffered, thus linking to Article 445 in conjunction with Article 444 of the Polish Civil Code, but also Article 448 in conjunction with Articles 23 and 24 of the Polish Civil Code, as health constitutes a personal good in the Polish legal order. Those tokens are: „fracture“, „cervical“, „bones“, and „twist“. Also, a number of tokens refer to situations where damage has occurred, e.g. „collision“, as well as implications of the above: „hospital“, „disorders“,



**Figure 2.** Distribution of Amounts Awarded as Compensation for Harm Suffered among the 25 Most Impactful Features

“disability”, “collar” (“orthopaedic collar”), “worsen”, and “procedure” (“medical procedure”).

Some of the tokens also appear to indicate whether the effects of an event have been visible for an extended period of time (“year”, “still”). At the same time, some of the strongest influences on the amount of compensation awarded for harm suffered had tokens of an evaluative nature, e.g., “substantial”. Moreover, the token “opinion” seems to originate from “expert opinion”, which can be consulted by the court.

Among the 25 most influential tokens, one indicating an extension of the claim was also observed: “extend”. In short, in situations where the amount indicated in the lawsuit becomes obsolete for various reasons, it is permissible to amend it. The influential nature of these may suggest that filing a claim for a certain amount and then extending the claim affects the amount awarded.

In the Polish legal system, nothing prevents the accumulation of claims (i.e., claiming both compensation for damage and compensation for harm suffered in a single claim). In view of this, among the tokens affecting the latter, some were observed that relate to the former. According to Article 444 § 2 of

the Polish Civil Code, if the injured person has lost all or part of his or her earning capacity, or if his or her needs have increased or his or her future prospects have decreased, he or she may demand an appropriate pension from the person liable for compensation for the damage. A few tokens appear to stand for this provision: “pension”, “zloty\_monthly” (*zloty* stands for the Polish currency), and “future”. While the origin of the above tokens can be easily explained, the interpretation of their influence on the modelled volume is interesting. Indeed, it appears that the amount of compensation for harm is influenced by whether compensation for injury is claimed in the same lawsuit.

As for the direction of the effect of the number of individual tokens on the amount of compensation for harm suffered awarded by the court, it appears that higher amounts are awarded in the case of facts involving breach of family bonds, death, fractures, the long-lasting visible effects of the damaging event such as staying in hospital, disability, accidents involving children, etc. The relatively lower amounts are awarded in case of cervical injuries requiring using orthopaedic collars, twisting, or violating personal

rights. Also, it seems that the accumulation of claims (i.e., claiming both compensation for damage and compensation for harm suffered in a single action), affects the awarded amounts of compensation for harm suffered positively. It also appears that the amount of compensation awarded for harm suffered is positively influenced by the fact that the claim has been extended.

Expanding on the question of the direction and strength of the influence of the individual variables, the partial dependence plots were also analysed (Figure 3, generated with Python library *scikit-learn*). As the model in question was constructed based on explanatory variables being token counts, the interpretation of the graphs is relatively straightforward – the horizontal axis of each plot stands for the number of times a certain token is observed in the body of judgements, whereas the vertical axis, in simple terms, depicts the average compensation amount awarded in judgements with a certain number of the analysed token's occurrence. It is clearly visible that for different tokens, a different occurrence number is enough for a maximum increase in amount awarded, with other factors remaining unchanged. For example, the tokens “family”, “fracture”, “dead”, “extend”, and “disability” need to be used at least two times in a body of a judgement to increase compensation substantially. The tokens “year”, “bond”, and “worsen” require three occurrences to meet the maximum. Similar shapes of the partial dependence plots were observed in the case of the tokens “life”, “zloty\_monthly”, “substantial”, “disorders”, “bones”, and “still”. For some tokens it appears that the higher occurrence, the better – “hospital”, “child”, “future”, “opinion”, “son”, and “procedure”. Obviously, the occurrence of some tokens lowers the amounts awarded by the court. Those are “cervical”, “collar”, “twist”, and “collision”.

Figure 3 also shows the use of token counts to build model safeguards against the use of certain words in a context where they are not factual descriptions, but, for example, quotations of specific legal provisions. Adequate use of a word or phrase seems to approximate that it describes the dominant theme. Specifically, it is noteworthy that, in some cases, the partial dependence plots do not stabilize after a certain word count is reached, and look basically like depicting monotonic increases in adjudicated amounts with more occurrences of particular words and collocations. Interestingly, these tokens (i.e., “hospital”, “child”, “future”, “opinion”, “son”, and “procedure”, either describe family members or can be considered

to closely correspond to certain quantitative variables. For example, the number of occurrences of the token “hospital” can approximate the number of hospital admissions of an individual. Still, what should also be addressed, the shape of the graphs at their ends on the right may (though not necessarily) be observed due to the relatively small number of observations with such high values of the independent variables.

What appears to be a particularly practical issue is at what number of tokens in question is the highest amount of compensation for the harm suffered awarded. In view of this, the analyses of the partial dependence plots were extended as shown in Table 3. For each token, it was determined at what counts its PDP plot reaches a minimum and a maximum. In this way, it was possible to determine the maximum by which changes in the count of a given token could, on average, affect the amount of compensation to be awarded, with other factors remaining constant.

The highest difference between the minimum and maximum value of the PDP plot was observed for the “pension” token. The average amount awarded in a judgement without the occurrence of this token was 21,027.80 PLN. The fivefold occurrence of this token in the content of the judgement was associated with an increase in the amount awarded to an average of 38,725.99 PLN (i.e., by 17,698.18 PLN), with other factors remaining unchanged. A similar increase in magnitude was observed between the non-occurrence of the “family” token (an average of 16,532.26 PLN was adjudicated) and the sixteen-fold occurrence of the same (an increase by 17,468.12 PLN). The third largest change was observed for the “hospital” token. Its non-occurrence was associated with an average award of compensation of 20,559.30 PLN. On the other hand, 20 occurrences of this token in the content of the decision resulted in awarding 14,768.65 PLN more. The fourth and final token whose counts result in a change of more than 10,000 PLN was the “fracture” token. This was an increase of 10,818.46 PLN.

As for research hypotheses, it was expected that the words and collocations whose occurrence in the body of the judgement will most strongly affect the amount awarded should be those contained in the body of the key provisions of the law or related facts detailing them. The description of the results attached above clearly demonstrates the relationship of the most influential keywords with specific legal provisions [i.e., Article 444 (in conjunction with 445), 446, 446<sup>2</sup>, and 448 of the Polish Civil Code]. Precisely, I mean tokens related to family ties (apparently a violation



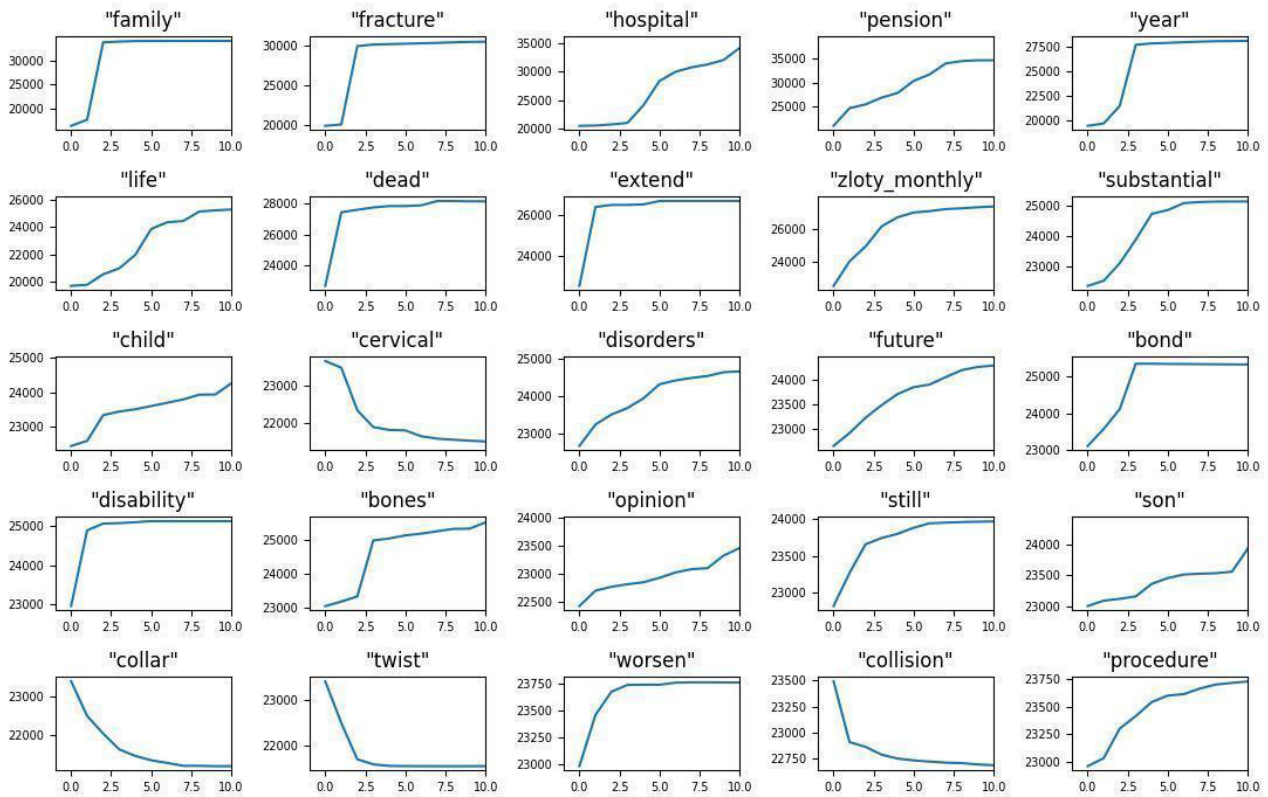


Figure 3. Partial Dependence Plots for the Most Impactful Features

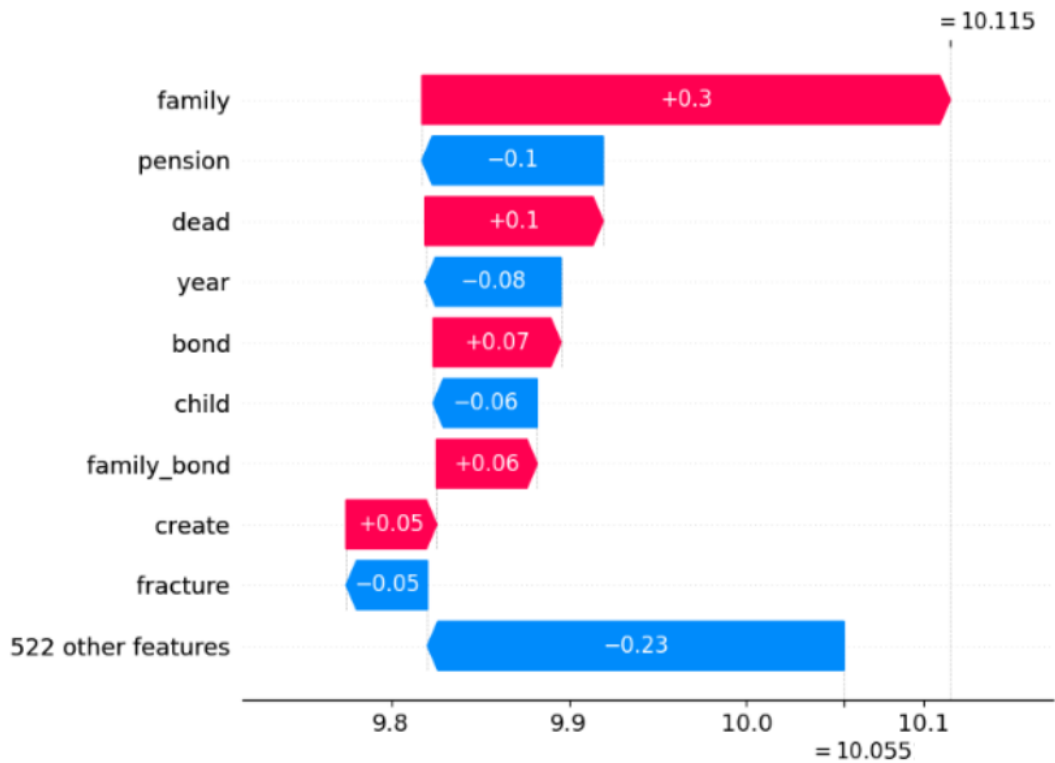


Figure 4. Example Model Prediction: Judgement of the Regional Court in Gliwice Dated August 31, 2014, File Number I C 1946/14

**Table 3.** Tokens Characterized by the Highest Differences Between Maximum and Minimum Value of Their Partial Dependence Plots

<b>Id</b>	<b>Token</b>	<b>Average awarded compensation when token does not occur</b>	<b>Minimum awarded compensation</b>	<b>Maximum awarded compensation</b>	<b>Difference between maximum and minimum awarded compensation</b>
1	pension	21,027.80	21,027.80 (0)	38,725.99 (5)	17,698.18
2	family	16,532.26	16,532.26 (0)	34,000.38 (16)	17,468.12
3	hospital	20,559.30	20,559.30 (0)	35,327.96 (20)	14,768.65
4	fracture	19,852.62	19,852.62 (0)	30,671.07 (25)	10,818.46
5	year	19,473.51	19,473.51 (0)	28,048.15 (10)	8,574.65
6	life	19,712.14	19,712.14 (0)	25,905.34 (36)	6,193.20
7	dead	22,645.50	22,645.50 (0)	28,181.21 (7)	5,535.71
8	zloty_monthly	22,557.29	22,557.29 (0)	27,708.96 (29)	5,151.67
9	extend	22,570.76	22,570.76 (0)	26,658.50 (6)	4,087.74
10	substantial	22,346.35	22,346.35 (0)	25,140.15 (19)	2,793.80
11	bones	23,056.96	23,056.96 (0)	25,536.80 (25)	2,479.84
12	child	22,439.85	22,439.85 (0)	24,908.14 (39)	2,468.29
13	bond	23,108.54	23,108.54 (0)	25,423.49 (24)	2,314.95
14	cervical	23,675.45	21,368.98 (22)	23,675.45 (0)	2,306.47
15	disorders	22,655.92	22,655.92 (0)	24,947.08 (17)	2,291.17
16	collar	23,419.00	22,944.30 (11)	23,419.00 (0)	2,264.04
17	disability	22,944.30	22,944.30 (0)	25,129.94 (5)	2,185.64
18	family_bond	23,262.10	23,262.10 (0)	25,430.35 (10)	2,168.25
19	twist	23,417.04	21,541.70 (7)	23,417.04 (0)	1,875.34
20	future	22,640.79	22,640.79 (0)	24,386.69 (22)	1,745.90

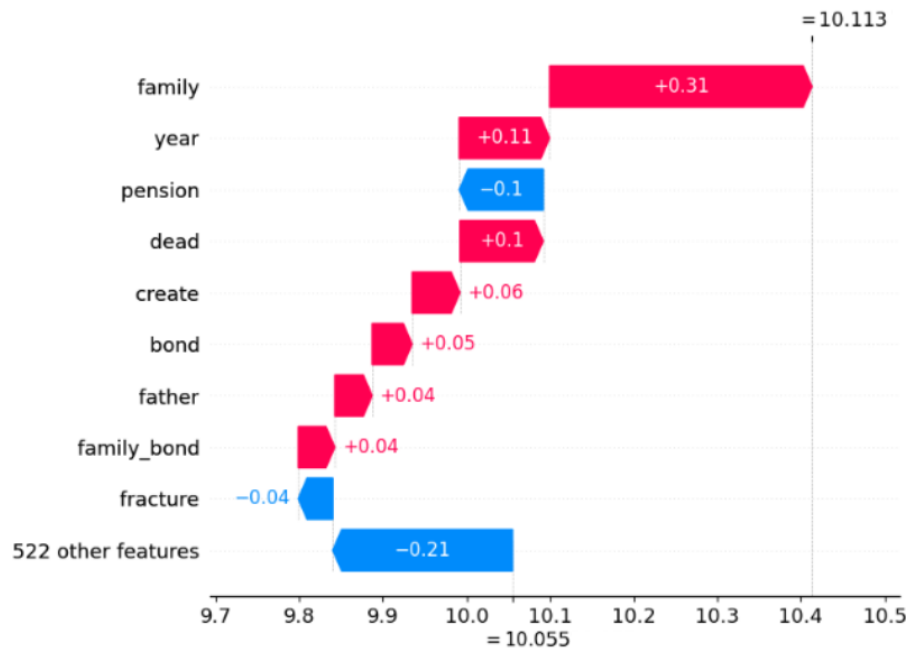
Note. The values in brackets correspond to the number of token occurrences at which a given amount is awarded on average.

of Articles 446 § 4 and 446<sup>2</sup>) or hospitalization and specificities of the injury suffered as well as specifying whether the injurious event led to death. What also appears in line with the expectations formulated at the beginning is a token describing one of possible procedural transformations. That is, extension of the claim was observed among the most impactful ones. Thus, it should be concluded that there are no grounds for rejecting the aforementioned research hypotheses. Interestingly, what also has a positive influence on the amount of compensation to be awarded for harm is claiming both compensation for damage (or indemnity pension from Article 444 § 2 of the Polish Civil Code) and compensation for harm suffered.

It appears that explaining the considered phenomena would be even more compelling with examples of judgements illustrating how the model works and how its predictions compare with actual judicial decisions.

Therefore, two random judgements were drawn from the test set for the purpose of demonstration (Figures 4–5, generated with Python library *shap*).

In the first referred case (Figure 4), the claimants requested a monetary amount of 12,000 PLN as compensation for harm suffered. The actual state involved the occurrence of a traffic accident that resulted in the death of a sister of the claimants. The court awarded 12,000 PLN. The model predicted 24,710.91 PLN (exponential of 10.115 as presented in Figure 5). It appears that what has primarily driven the model for such an overestimation was the number of occurrences of the token “family”. It is worth repeating that courts in Poland cannot award more than what was requested in the claim. As a result, overestimation here does not necessarily mean the model’s error, but rather underestimation of the value of the claim by the claimants.



**Figure 5.** Example Model Prediction: Judgement of the Regional Court in Zambrów Dated September 24, 2015, File Number I C 335/15

The second reviewed case (Figure 5) involved requesting 51,000 PLN. The facts concerned the death of the claimant's father as a result of being hit by a car while cycling. The court awarded 26,000 PLN. The model predicted 33,289.59 PLN. This value was obtained primarily due to relatively high counts of the tokens "family", "year", "dead", "create", "bond", "father", and "family\_bond". Interestingly, in this case, even before the court proceeding started, the insurer of the accident perpetrator paid 9,000 PLN to the claimant as compensation for harm suffered. It was not directly included in the model, as it was based only on token counts. It makes the model's underestimation even lower. However, at the same time, it clearly demonstrates possible limitations of the presented methodology.

## 5. Conclusions

In summary, this research aimed to explain and predict the amount of money awarded as compensation for harm suffered using machine-learning algorithms. The Polish common courts were considered as an example. The counts and TF-IDF statistics of the tokens present in the judgements were used as explanatory variables of the modelled quantity. The massive number of variables made semi-automatic selection of regressors necessary. LASSO regression followed by

an expert adjustment to minimize look-ahead bias was adopted. The selected features were used in models of different types (i.e., OLS, again LASSO, random forests, XGBoost, and BERT algorithm).

The best results (i.e., the lowest prediction error on the test set) was obtained for the random forests algorithm. Very few studies could be labelled sufficiently similar to compare the prediction accuracy. What appears to be the most problematic matter is that the court deciding on the amount of compensation for harm in each case cannot award more than the amount requested in the claim. Thus, the question arises as to whether the model overestimations should be considered errors. With this in mind, further in-depth research seems necessary to more closely assess the predictive quality of the models.

As for expected feature importance, it indeed seems that the words and collocations whose occurrence in the body of the judgement most strongly affects the amounts awarded are those contained in the body of the key provisions of the law and related ones detailing them. Specifically, higher amounts are awarded in the case of facts involving breach of family bonds, death, fractures, and the long-lasting visible effects of the damaging event, such as staying in hospital. Also, it appears that extension of the claim positively affects the amount of compensation awarded for harm. Interestingly, the obtained results

show that claiming for both compensation for damage and compensation for harm implies higher amounts awarded for the latter. Finally, it appears that some token proxy for specific quantitative variables, as their partial dependence plots did not stabilize with a certain word count reached. For example, the number of occurrences of the token “hospital” can approximate the number of hospital admissions of an individual.

When discussing the practical application of the methodology presented in this article, it should be considered a useful complementary tool to help lawyers and judges make decisions. Of course, it would be a mistake to use it to completely automate adjudication. The reasons addressed in the extensive literature on this matter, to stop with referencing only few most recent papers from Contini (2020), Xu (2022), and Said et al. (2023), are mainly ethical impediments and possible bias in fairness of automatic resolution of disputes. Still, lawyers can consider performing similar analyses in their everyday work, as they often need to extrapolate the extent of harm in order to precisely determine the value of the subject matter of the dispute because it implies both the amount of the court fee and the burden of court costs on the parties. Moreover, courts can consider applying the presented methodology to help judges to determine the appropriate monetary value that should be adjudicated for harm suffered. Nevertheless, the individual nature of the harm suffered dictates that in each case one should expect an expert adjustment of the amount awarded by the judge.

Even though a practical application of the presented methodology would not be more than a complementary tool, it is still a substantial involvement in the legal field. While it should not compromise the integrity of the judgments, the public perception of this phenomenon may be different. This could possibly violate public trust in the judicial system. Furthermore, legal professionals can perceive even a slightly automated adjudication with noticeable mistrust. The very introduction of such a solution could heavily affect the jurisprudence of compensation and therefore requires extensive consultation.

In the future, I plan to extend the analyses with information on the initially claimed amount that would allow a more reliable assessment of the quality of the prediction. In addition, clearer rules for selecting variables for the model should be established. It would also be worthwhile to consider the prediction of the amount of compensation for harm as a discontinuous, ordinal variable.

## References

- Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
- Aletras, N., Tsarapatsanis, D., Preoțiu-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. <https://doi.org/10.7717/peerj-cs.93>
- Alshboul, O., Alzubaidi, M. A., Mamlook, R. E. A., Almasabha, G., Almuflih, A. S., & Shehadeh, A. (2022a). Forecasting Liquidated Damages via Machine Learning-Based Modified Regression Models for Highway Construction Projects. *Sustainability*, 14(10), 5835. <https://doi.org/10.3390/su14105835>
- Alshboul, O., Shehadeh, A., Mamlook, R. E. A., Almasabha, G., Almuflih, A. S., & Alghamdi, S. Y. (2022b). Prediction Liquidated Damages via Ensemble Machine Learning Model: Towards Sustainable Highway Construction Projects. *Sustainability*, 14(15), 9303. <https://doi.org/10.3390/su14159303>
- Andrych-Brzezińska, I. (2020). Punitive damages: Czyli o odszkodowaniu karnym w prawie amerykańskim oraz Europejskiej Debacie na temat funkcji odpowiedzialności odszkodowawczej. *Transformacje Prawa Prywatnego*, 4, 5–54. <https://journals.law.uj.edu.pl/TPP/article/view/519/252>
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bannet, A., Tabik, S., Barbado, A., ... Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and challenges Toward Responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Balogopalan, A., Eyre, B., Rudzicz, F., and Novikova, J. (2020). To BERT or Not To BERT: Comparing Speech and Language-based Approaches for Alzheimer’s Disease Detection. arXiv preprint. <https://doi.org/10.48550/arXiv.2008.01551>
- Balcerowicz, L. (2005). Post-COMMUNIST Transition: Some Lessons. *IEA Occasional Paper*, 127. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=676661](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=676661)
- Beldowski, J., Dąbros, 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. <https://doi.org/10.1007/s10657-020-09656-4>



- Breiman, L. (2001). Random Forests *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. (1984). *Classification and Regression Trees*. Chapman & Hall/CRC. <https://doi.org/10.1201/9781315139470>
- Brzozowski, A. (2021). In A. Brzozowski, J. Jastrzębski, M. Kaliński, E. Skowrońska-Bocian (Eds.). *Zobowiązania: Część ogólna* (4th ed.), Wolters Kluwer.
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... & Zhou, T. (2015). *Xgboost: Extreme gradient boosting* (R package version 0.4-2) [Computer software]. <https://cran.ms.unimelb.edu.au/web/packages/xgboost/index.html>
- Chlebus, M., Dyczko, M., & Woźniak, M. (2021). Nvidia's Stock Returns Prediction Using Machine Learning Techniques for Time Series Forecasting Problem. *Central European Economic Journal*, 8(55), 44-62, <https://doi.org/10.2478/ceej-2021-0004>
- Contini, F. (2020). Artificial intelligence and the transformation of humans and technology interactions in judicial proceedings. *Law, Technology and Humans*, 2(1), 4–18, <https://doi.org/10.5204/lthj.v2i1.1478>
- Cui, J., Shen, X., & Wen, S. (2023). A Survey on Legal Judgment Prediction: Datasets, Metrics, Models and Challenges. *IEEE Access*, 11, 102050–102071. <https://doi.org/10.1109/ACCESS.2023.3317083>
- Dal Pont, T. R., Sabo, I. C., Hübner, J. F., & Rover, A. J. (2023). Regression Applied to Legal Judgments to Predict Compensation for Immaterial Damage. *PeerJ Computer Science*, 9, e1225. <https://doi.org/10.7717/peerj-cs.1225>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint. <https://doi.org/10.48550/arXiv.1810.04805>
- Eisenberg, T., Eisenberg, T., Wells, M. T., & Zhang, M. (2015). Addressing the Zeros Problem: Regression Models for Outcomes With a Large Proportion of Zeros, With an Application to Trial Outcomes. *Journal of Empirical Legal Studies*, 12(1), 161–186, <https://doi.org/10.1111/jels.12068>
- Eisenberg, T., Hannaford-Agor, P. L., Heise, M., LaFountain, N., Munsterman, G. T., Ostrom, B., & Wells, M. T. (2006). Juries, Judges, Juries, and Punitive Damages: Empirical Analyses Using the Civil Justice Survey of State Courts 1992, 1996, and 2001 Data. *Journal of Empirical Legal Studies*, 3(2), 263–295. [https://scholarship.law.cornell.edu/lsrp\\_papers/30/](https://scholarship.law.cornell.edu/lsrp_papers/30/)
- Eisenberg, T., Heise, M., Waters, N. L., & Wells, M. T. (2010). The Decision to Award Punitive Damages: An Empirical Study. *Journal of Legal Analysis*, 2(2), 577–620, <https://doi.org/10.1093/jla/2.2.577>
- Elfron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1), 1–26. <https://doi.org/10.1214/aos/1176344552>
- European Commission. (2021). *The 2021 EU Justice Scoreboard*. Publications Office of the European Union. [https://commission.europa.eu/system/files/2021-07/eu\\_justice\\_scoreboard\\_2021.pdf](https://commission.europa.eu/system/files/2021-07/eu_justice_scoreboard_2021.pdf)
- Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5), 1189–1232. <https://www.jstor.org/stable/2699986>
- Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive Logistic Regression: a Statistical View of Boosting (With Discussion and a Rejoinder by the Authors). *The Annals of Statistics*, 28(2), 337–407. <https://doi.org/10.1214/aos/1016218223>
- González-Carvajal, S., & Garrido-Merchán, E. C. (2020). Comparing BERT Against Traditional Machine Learning Text Classification. arXiv preprint. <https://doi.org/10.48550/arXiv.2005.13012>
- Gunning, D., & Aha, D. (2019). DARPA's Explainable Artificial Intelligence Program. *AI Magazine*, 40(2), 44–58. <https://doi.org/10.1609/aimag.v40i2.2850>
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Vol. 2, 1–758). Springer.
- Hsieh, D., Chen, L., & Sun, T. (2021). Legal Judgment Prediction Based on Machine Learning: Predicting the Discretionary Damages of Mental Suffering in Fatal Car Accident Cases. *Applied Sciences*, 11(21), 10361. <https://doi.org/10.3390/app112110361>
- Joachims, T. (1997). A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization. In *ICML 97*, 143–151. <https://dl.acm.org/doi/10.5555/645526.657278>
- Kaliński, M. (2021a). *Szkoda na mieniu i jej naprawienie*. (3rd ed.). C. H. Beck.
- Kaliński, M. (2021b). In Brzozowski, A., J. Jastrzębski, M. Kaliński, & E. Skowrońska-Bocian

(Eds.). *Zobowiązania: Część ogólna* (4th ed., chap. 3, 8). Wolters Kluwer.

Katz, D. M., Bommarito, M. J., & Blackman, J. (2017). A General Approach for Predicting the Behavior of the Supreme Court of the United States. *PloS One*, 12(4), e0174698. <https://doi.org/10.1371/journal.pone.0174698>

Kearns, M., & Valiant, L. G. (1989). Cryptographic Limitations on Learning Boolean Formulae and Finite Automata. In *Proceedings of the Twenty-first Annual ACM Symposium on Theory of Computing (STOC '89)*. (pp. 433–444). Association for Computing Machinery. p. 433–444. <https://doi.org/10.1145/73007.73049>

Kieraś, W., & Woliński, M. (2017). Morfeusz 2: Analizator i generator fleksyjny dla języka polskiego. *Język Polski*, 97(1), 75–83. <https://www.ceeol.com/search/article-detail?id=528784>

Kochanowski, M. (2019). Rozważania na temat represyjnych i prewencyjnych elementów odpowiedzialności odszkodowawczej na przykładzie instytucji odszkodowania karnego (punitive damages) w świetle orzecznictwa Sądu Najwyższego oraz Trybunału Konstytucyjnego. *Studia Prawa Publicznego*, 1(17), 83–100.

Kruczalak-Jankowska, J., Maśnicka, M., & Machnikowska, A. (2020). The Relation between Duration of Insolvency Proceedings and their Efficiency (with a Particular Emphasis on Polish Experiences). *International Insolvency Review*, 29(3), 379–392. <https://doi.org/10.1002/iir.1392>

Kryła-Cudna, K. (2018). *Zadośćuczynienie pieniężne za szkodę niemajątkową powstałą wskutek niewykonania lub nienależytego wykonania umowy*. C. H. Beck.

Medvedeva, M., Vols, M., & Wieling, M. (2020). Using Machine Learning to Predict Decisions of the European Court of Human Rights. *Artificial Intelligence and Law*, 28, 237–266. <https://doi.org/10.1007/s10506-019-09255-y>

Medvedeva, M., Wieling, M., & Vols, M. (2023). Rethinking the Field of Automatic Prediction of Court Decisions. *Artificial Intelligence and Law*, 31(1), 195–212. <https://doi.org/10.1007/s10506-021-09306-3>

Mroczkowski, R., Rybak, P., Wróblewska, A., & Gawlik, I. (2021). HerBERT: Efficiently Pretrained Transformer-based Language Model for Polish. arXiv preprint. <https://doi.org/10.48550/arXiv.2105.01735>

Mumcuoğlu, E., Öztürk, C. E., Ozaktas, H. M., & Koç, A. (2021). Natural Language Processing in

Law: Prediction of Outcomes in the Higher Courts of Turkey. *Information Processing & Management*, 58(5), 102684. <https://doi.org/10.1016/j.ipm.2021.102684>

Radwański, Z., Olejniczak, A., Grykiel, J. (2022). *Zobowiązania: Część ogólna* (15th ed.). C. H. Beck.

Safjan, M. (2020). Art. 445 [Zadośćuczynienie pieniężne]. In Pietrzykowski, K. (Ed.). *Kodeks cywilny: Komentarz Art. 1-449<sup>o</sup>*, (10th ed., Vol. 1). C. H. Beck.

Said, G., Azamat, K., Ravshan, S., & Bokhadir, A. (2023). Adapting Legal Systems to the Development of Artificial Intelligence: Solving the Global Problem of AI in Judicial Processes. *International Journal of Cyber Law*, 1(4). <https://irshadjournals.com/index.php/ijcl/article/view/49>

Shapley, L. S. (1952). A Value for N-Person Games. RAND Corporation <https://www.rand.org/pubs/papers/P295.html>

Strickson, B., & De La Iglesia, B. (2020, March). Legal Judgement Prediction for UK Courts. In *Proceedings of the 3rd International Conference on Information Science and Systems* (pp. 204–209). <https://doi.org/10.1145/3388176.3388183>

Sulea, O. M., Zampieri, M., Vela, M., & Van Genabith, J. (2017). Predicting the Law Area and Decisions of French Supreme Court Cases. arXiv preprint. <https://doi.org/10.48550/arXiv.1708.01681>

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>

Torres, G. D. O., Guterres, M. X., & Celestino, V. R. R. (2023). Legal Actions in Brazilian Air Transport: A Machine Learning and Multinomial Logistic Regression Analysis. *Frontiers in Future Transportation*, 4, 1070533. <https://doi.org/10.3389/ffutr.2023.1070533>

Valvoda, J., Cotterell, R., & Teufel, S. (2023). On the Role of Negative Precedent in Legal Outcome Prediction. *Transactions of the Association for Computational Linguistics*, 11, 34–48. [https://doi.org/10.1162/tacl\\_a\\_00532](https://doi.org/10.1162/tacl_a_00532)

Virtucio, M. B. L., Aborot, J. A., Abonita, J. K. C., Avinante, R. S., Copino, R. J. B., Neverida, M. P., ... & Tan, G. B. A. (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)*. (Vol. 2, 130–135. IEEE. <https://doi.org/10.1109/COMPSAC.2018.10348>

Waltl, B., Bonczek, G., Scepankova, E., Landthaler, J., & Matthes, F. (2017). Predicting the Outcome of Appeal Decisions in Germany's Tax Law. In *Electronic Participation: 9th IFIP WG 8.5 International Conference, ePart 2017, St. Petersburg, Russia, September 4-7, 2017, Proceedings*, 9, 89–99. Springer. [https://doi.org/10.1007/978-3-319-64322-9\\_8](https://doi.org/10.1007/978-3-319-64322-9_8)

Wołodkiewicz, W., & Zabłocka, M. (2014). *Prawo rzymskie: Instytucje*. (6th ed., chap. 1). C. H. Beck.

Xu, Z. (2022). Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities. *Applied Artificial Intelligence*, 36(1). <https://doi.org/10.1080/08839514.2021.2013652>

Yeung, C. M. (2019). *Effects of Inserting Domain Vocabulary and Fine-Tuning BERT for German Legal Language* (Master's thesis, University of Twente). <https://essay.utwente.nl/80128/>