Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Artificial intelligence in judicial adjudication: Semantic biasness classification and identification in legal judgement (SBCILJ)

Kashif Javed, Jianxin Li [*]

School of Law, Zhengzhou University, Zhengzhou, 450001, Henan, China

ARTICLE INFO

Keywords:
Artificial intelligence
Judicial adjudication
Semantic biasedness
Legal judgment
CAIL dataset
Natural language processing

ABSTRACT

History reveals that human societies have suffered in terms of social justice due to cognitive bias. Semantic bias tends to amplify cognitive bias. Therefore, the presence of cognitive biases in extensive historical data can potentially result in unethical and allegedly inhumane predictions since AI systems are trained on this data. The innovation of artificial intelligence and its rapid integration across disciplines has prompted questions regarding the subjectivity of the technology. Current research focuses the semantic bias in legal judgment to increase the legitimacy of training data. By the application of general-purpose Artificial Intelligence (AI) algorithms, we classify and detect the semantics bias that is present in the Chinese Artificial Intelligence and Law (CAIL) dataset. Our findings demonstrate that AI models acquire superior prediction power in the CAIL dataset, which is comprised of hundreds of cases, compared to a structured professional risk assessment tool. To assist legal practitioners during this process, innovative approaches that are based on AI may be implemented inside the legal arena. To accomplish this objective, we suggested a classification model for semantic bias that is related to the classification and identification of semantic biases in legal judgment. Our proposed model legal field uses the example of categorization along with the identification of the CAIL dataset. This will be accomplished by identifying the semantics biases in judicial decisions. We used different types of classifiers such as the Support Vector Machine (SVM), Naïve-Bayes (NB), Multi-Layer Perceptron (MLP), and the K-Nearest Neighbour (KNN) to come across the preferred results. SVM got 96.90 %, NB has 88.80 %, MLP has 86.75 % and KNN achieved 85.66 % accuracy whereas SVM achieved greater accuracy as compared to other models. Additionally, we demonstrate that we were able to get a relatively high classification performance when predicting outcomes based just on the semantic bias categorization in judicial judgments that determine the outcome of the case.

1. Introduction

The global judiciaries have been plagued by a backlog of cases, resulting in a decline in justice. Therefore, it has been seen as a significant objective of sustainability [1]. AI has entered its innovative year after renewed interest from the business community. Therefore, AI is promising to solve numerous problems that are currently being faced by adversarial systems (common law) and inquisitorial systems (civil law), including the backlog of cases, delays in the administration of justice, and misuse of discretionary power by the judges as being subjective in their decisions across the globe. Like the other innovative techniques of AI, machine learning

[*] Corresponding author.
E-mail address: jianxinlee@zzu.edu.cn (J. Li).

https://doi.org/10.1016/j.heliyon.2024.e30184
Received 5 February 2024; Received in revised form 20 April 2024; Accepted 22 April 2024
Available online 26 April 2024

(ML) is also capable of helping the legal domain solve the above-stated issues. Che et al. [2] proposed a medical vision-language pre-training with frozen language models and latent space geometry optimization (M-FLAG) that uses a frozen language model for training stability and efficiency while also introducing a novel orthogonality loss to harmonize the latent space geometry. Zhongwei et al. [3] propose Unifying Cross-Lingual Medical Vision-Language Pre-Training (Med-UniC) to combine multimodal medical data from English and Spanish, the two most widely spoken languages. We propose Cross-lingual Text Alignment Regularization (CTR) to explicitly integrate medical report semantics from diverse language communities.

Judicial systems stand for maintaining a calm, progressing, and just society, so it is in dire need to use AI techniques to solve administrative issues [4,5]. “Rule-based” systems have been utilized in the operation of the administration of legal processes ever since the seventh decade of the twentieth century. Until recently, the field of law has placed some imitations on the use of rule-based reasoning, case-based reasoning, and machine learning. This is because a deep understanding of settled legal principles is necessary to effectively utilize these methods in the legal domain [6]. One cannot ignore the prospect that the capability of machine learning models to acquire rules from huge datasets could lead to the elimination of cognitive biases that are inherent to human beings, as well as an improvement in the precision of decision-making. In every culture, words and phrases of language are associated with specific situations and carry their relative meanings. These words may have different meanings when translated into another language, and they may even have different ways of understanding. This semantic bias may trigger the cognitive bias when judges decide the cases. This is a possibility of subjectivity that cannot be denied. The models built by Machin Learning use the available big data that carries semantic bias. Such bias data will cause bias results as it acquires rules from this data [7].

The monetary domains, sentencing, and criminal recidivism (COMPASS and JSORRAT–II) usage for recidivism in criminal matters are three of the most significant areas in which machine learning has been utilized to assist with decision-making in the legal profession. These three areas are among the most crucial ones. Some people have expressed their worries about the likelihood that the improper implementation of artificial intelligence in a variety of fields could lead to the formation of decisions that are influenced by prejudice [6]. In addition to this, they have the propensity to inherit these norms as a consequence of the data biases that were brought about by discriminatory behaviors that occurred in the past. Because machine learning models tend to inherit these rules, this is a consequence that has arisen as a result of this tendency, along with the prospect of implementing automated judgments that are biased against specific communities or groups of people who are members of minority groups [8,9]. Protected or sensitive factors, which are those that pertain to certain groups and may include things like gender, color, nationality, or religion, should, in an ideal world, not have any influence on the outcome of the decision about computational fairness. Because these characteristics are associated with particular groups.

Throughout their respective histories, the fields of law and criminology have made extensive use of language analysis as a means of information collecting. Particularly in the field of forensic linguistics, for instance, text classification has been utilized in a great number of different situations. In the present day, we can automate a major percentage of the analysis that was formerly carried out manually, such as in the case of the Unabomber terrorist attack [10]. It is currently possible to acquire access to computer programs that employ a technique known as "machine learning" to assess whether or not a person is masculine or feminine [11].

In the context of this work, we explore the prospects of mechanized information extraction and language analysis to facilitate statistical study in the legal domain, with a particular focus on legal judgments because these are used as judicial precedents that are legally binding, like the rules, and also termed Judge-made law. Moreover, we explore the potential of utilizing natural language processing techniques to mechanically forecast semantic bias in judicial rulings using the Chinese AI and Law Challenge datasets (CAIL, 2018) [12]. The dataset is based on the information gathered from the official website of the Wenshu courts, where we can find the legal judgments of the Chinese courts. While applying machine learning models, computers are taught to make predictions about legal case outcomes by analyzing the quantitative properties of words and phrases, as well as their connections, which are extracted from unorganized legal data in raw form that has been obtained from physical legal documents generated by different courts [13]. The most reliable data in the legal field is either obtained from enacted statutes or legal rulings. Statutes once enacted are not altered again and again, but the legal rulings have many new spectrums daily, as each day is a new day and each case is a new case. The most important aspect of every decision can be identified by using this method to get accurate predictions, which will aid in finding the word that has the utmost impact on the execution of the legal rulings because the gentle and proper execution of the ruling is as important as the rulings themselves, as we say that justice seems to be done [14].

Our proposed model aims to support legal professionals in addressing complex legal rulings from various perspectives. Efficiently managing the workload involves organizing the necessary information, such as plaint, written statement, oral and documentary evidence, and reasoning provided by attorneys. This helps to address the backlog of cases and streamline the legal research process. In addition, the model will accurately classify and identify the data and relevant laws for judicial verdicts, while also excluding any semantic bias. To achieve this goal, we developed an automated system that can classify legal judgments that are based on machine learning techniques. This system initially plays the role of organizing legal judgments and afterward evaluates the capability and precision of the legal classification model using (SVM), (NB), (MLP) and (KNN) classifiers. These classifiers are used to predict the most precise output and the same are used in the context of legal rulings.

Machine learning techniques have made it possible to use computational methods for conducting quantitative analysis of the language employed in a court case. Subsequently, this analysis can be used to train the computer to make predictions regarding the court’s decision. If the results can be accurately predicted, it would be possible to analyze the words that had the greatest impact on the legal decision. This could assist in determining the important factors for judicial decisions. As a limitation, it is obvious that when addressing the forecasting of legal decisions, we are solely referring to the available data and methodologies employed. This forecasting is in a generalizing way and is quite not confined to dealing with specific cases or statutes.

1.1. Motivation of the study and contribution

The judiciary is commonly referred to as the third branch of government in a state. The judiciary has a crucial role in upholding the rule of law and exercising its authority in judicial review, which is vital for the overall welfare of society. Legitimate legal decision-making is the core objective of the rule of law. It provides an executable solution for the conflicts between the parties. It decides the rights and obligations as per settled norms and enacted laws [15]. The notion of judicial discretion empowers judges to render decisions based on the principles of fairness, utility, and societal certainty. However, sometimes, the discretion of the judges leads to biased decisions. The first and foremost requirement of the legal judgment that concludes the rights and duties is to be fair, accurate, and free of subjectivity. Like many other types of biasedness, semantic bias is one of the types that has importance in a legal judgment, as the judgment carries words and phrases in writing.

The inherent subjectivity and ambiguity in language, known as semantic bias, presents a formidable obstacle to decision-making processes in the court system. Injustices and societal inequities can result from this prejudice when it shows up in the interpretation of legal texts, the creation of judgments, and the execution of decisions. The phenomenon known as semantic bias takes place when legal judgments and legal orders are influenced more by the symbolic meanings and emotional connotations of words used in the judgment for logical reasoning than objective facts. This semantic bias occurs due to the words themselves, as they have positive and negative meanings, their framing, and the structure of the judgment and stereotype writing. The use of biased language in legal documents or during court proceedings might result in unfair punishment or prejudice against particular groups of people, which would be against the legal maxim that justice seems to be done [16,17]. Semantic bias leads to the unfair implementation of decisions and a biased repository of judgments. Recent advances in machine learning have opened up exciting new possibilities for reducing the impact of semantic bias on automated decision-making. The investigation of machine learning methods to reduce semantic bias is driven by the urgent necessity to tackle systematic inequities in decision-making procedures. Hence the question arises of how machine learning techniques mitigate semantic bias?

Literature reveals that research so far has found regularities, irregularities, legalities, and illegalities in judicial decisions. To our knowledge, no work has been done by scholars regarding the finding of semantic bias in judicial judgments by using any kind of dataset. So, our work on finding semantic bias in legal judgment is novel. The legitimacy of AI requires finding solutions for all probable biases in judgment, as legal judgments will be used as raw data, and this raw data will be used for the training of models and systems to help in judicial decisions. The judicial precedent (former authoritative reported judgment) is of great value in legal adjudication [15]. Any AI system to help in judicial adjudication requires using judicial precedents as raw data to find regularities that might help with future predictions and solutions to conflicts. If this raw data contains any kind of bias that will result in bias as well, that would become a question mark for the legitimacy of AI in adjudication. Some studies discovered a distinct disparity in the frequency with which male-related, female-related, gender-related, bias-related, race-related, and religious-related keywords appeared in the original documents, expert-generated summaries, and model-generated summaries. After careful observation, a study determined that the general domain LLMs, such as ChatGPT and Davinci, have generated a little higher percentage. Hence, finding the biases that are already there in the data collection, algorithms, and architectures is an important step to do right now. Therefore, identifying the semantic bias that exists in court precedents is the primary emphasis of this research. This research is useful for making the use of artificial intelligence in adjudication more legitimate. This study aims to aid in creating more open, responsible, and equitable legal systems that respect the values of justice and equality for all by utilizing cutting-edge algorithms and data-driven strategies. It also opens up opportunities for academics to further contribute to the complete elimination of semantic bias in judicial precedents before using the raw data for training models.

Referring to an experiment that demonstrated how the meaning and context of a sentence can influence its interpretation motivated this research [18]. So, Semantic Biasness Classification and Identification in Legal Judgement (SBCILJ) is a kind of documentation by uses (SVM), (NB), (MLP) and (KNN) classifiers to transfer each document in vector form to preserve the semantic information. Therefore, documents with semantic bias are located near each other in the legal judgment document. This research has made some important contributions, which are emphasized as follows:

- We used the Chinese AI and Law (CAIL) dataset, the first large-scale Chinese legal dataset for judgment prediction.
- All of the documents are pre-processed using NLP techniques that are considered mainstream.
- We represent them in a multidimensional semantic feature space, through algorithms by SVM, NB, MLP, and KNN classifiers. Classifier to predict and classify the most accurate result.
- The simplest equation that would distinguish biased data from each other according to the classification with the least degree of error is one that we establish.
- We achieved a relatively high classification performance (average accuracy of 86 %) when predicting outcomes based only on the semantic biases of the judges who tried the case.

After this, we shall delve into previous endeavors that applied automatic analysis to the field of law. The application of ML to the classification of legal texts is elaborated upon in Section 3. Section 4 describes the data used in the experiments. We report the outcomes of four experiments that we conducted for this study in Section 5. In Sections 6 and 7, we analyze the findings and formulate conclusions, respectively.

2. Review of literature

In the subsequent sections, we provide a concise overview of previous studies that assist in the retrieval of legal material and the detection of anomalies in legal case judgments through the use of various methodologies, such as clustering-based techniques.

2.1. Algorithmic impartiality

Algorithmic fairness, a component of the research in Accountability, Fairness, and Transparency focuses on addressing unfairness occurring inside algorithmic systems. In this study, we examine the concept of group fairness, which refers to a process or decision being deemed fair if it avoids any kind of discrimination based on an individual's affiliation with a protected group. Despite of advancement of AI systems, there remains a lack of comprehension regarding the fundamental features of AI systems. We face challenges in determining whether an AI system is fair or if it is unintentionally perpetuating biases. The ways to detect and address these issues to ensure that AI systems align with societal decision constraints is a question mark [19]. According to Katsaros et al. [20] consumers' perceptions of procedural justice do not need to conflict with algorithmic decision-making. Based on existing empirical research in this sector, we believe that the antecedents for procedural fairness can be included in the algorithmic decision-making processes used by content moderation platforms. There is no assurance that the algorithm is "fair." There are many definitions of fairness in the literature, with a minimum of 21 mentioned. Multiple research studies provide an overview of different definitions of algorithmic fairness, such as those by Refs. [4,21]. In addition, Chouldechova et al. [22] have shown that some group fairness criteria are incompatible with one another. Data-driven automated decision-making systems are growing. If not audited properly, these models can harm people, especially marginalized ones. The proliferation of such systems in our daily lives highlights the necessity of examining model biases. Group fairness approaches compare groups based on a sensitive attribute and assess model prediction evidence [23]. Fairness is a concept rooted in ethics and law, driven by values. For example, the United States has implemented laws to prohibit discrimination in various areas including religion, race, and sex. In addition, it prohibits discrimination as to national origin in the Civil Rights Act, act the Immigration and Age discrimination act. Likewise, the European Convention on Human Rights (Article 14) prohibits any form of discrimination based on various factors such as “sex, race, color, language, religion, political or another opinion, national or social origin, association with a national minority, property, birth, or another status”. Implementing and evaluating fairness is challenging since it is centered around value concepts rather than a technical aspect of ML models. As machine learning models are being utilized in legal systems, the issue of fairness is gaining significant attention [24]. While discrimination probably happens at several stages and throughout the process, in this case, we only discuss fairness in judicial decisions; other forms of fairness, such as process fairness, are not included. Reliable metrics show that there is a noticeable difference in the outcomes of people from different groups in our particular case, which amply demonstrates the discrimination. But meeting just one metric guarantees groups' equivalence in terms of that metric alone. Therefore, the concept of fairness is heavily influenced by specific circumstances and personal beliefs.

2.2. Network-based approaches

One of the challenging problems in legal case documents is classifying the similarity between two different approaches, i.e. text-based and citation-based approaches. In this context, citation recommendation and prior-case retrieval are two famous applications [25]. Network-based approaches mainly create a citation network by using referential information. The similarity score is then determined by analyzing the direct or indirect citations that were used in the previous step. Recent studies have explored network-based methods from related fields, such as analyzing the citation network of scholarly articles [6]. To extract visual qualities, Heng et al. [26] proposed a Convolutional Neural Network (CNN) based on hybrid pooling. Using either the maximum or average pooling functions, this technique modifies the CNN model's pooling layers at random. The effectiveness of CNN-based feature extraction models is increased by this technique. Using either the maximum or average pooling functions, this technique modifies the CNN model's pooling layers at random. The CNN-based feature extraction model may operate more efficiently as a result of this technique. Previous similar cases strongly influence the current case. Document similarity is often assessed as a whole. Because legal case materials are voluminous, it may be helpful to simply compare the key concepts or summaries. The methodologies used to compare court case papers are the topic of this survey. This study divides these efforts into citation-, content-, and summary-based methodologies. A comprehensive survey of legal document summarization methodologies was also conducted [27,28]. Kumar et al. [29] Examined the similarity scores among judgments by utilizing co-citation analysis and bibliographic coupling for the precedent citation network. Dhanani, Mehta, and Rana introduce a cutting-edge Legal Document Recommendation System (LDRS) that utilizes graph clustering to group similar judgments and identify relevant ones within those clusters [27,30]. Koniaris et al. [31] utilized network statistical and structural information, such as degree, to capture the resemblance among legal documents from the EU. Nonetheless, one of the most important factors determining how successful network-based techniques are is the network's degree of connectedness. Many legal cases often rely on a narrow range of precedents, statutes, and laws, resulting in a fragmented citation network [30].

2.2.1. Text-based approaches

Methods that rely on textual evidence make an effort to record how similar the judgments are in terms of vocabulary or meaning. The cosine similarity score is commonly used in mainstream methodologies to quantify the similarity among the document vectors. Recently, legal document similarity analysis has used a lot of different vectorization methods, like TF-IDF, LDA, Word2Vec, and

Doc2Vec, to quickly turn text into real-valued vectors with set lengths. There are many examples of using the type of modeling known as semantic vector space to find similarity indexes. Similarly, Kumar et al. [29] put forth the strategy, namely TF-IDF, for Indian legal judgments for constructing vector space, but these appear to be high-dimensional vectors. Ayden et al. [32] create language representations for text-based medical records, model numerical health data with the Gated Recurrent Unit (GRU), and automatically mix the two streams to forecast ICD-9 codes for the intensive care unit. We go over the preprocessing and classification procedures, demonstrating that our proposed two-stream model outperforms other cutting-edge studies in the literature. While Nanda et al. [33] consider LDA-based top modeling to find similarity indexes, this is not suitable for long texts, i.e., legal judgments that are too long in common law legal systems because this judgment carries detailed insight and prudence of the judge who decides the case. Therefore, examining the similarity of legal documents by employing topic modeling based on LDA, may not be optimal for lengthy textual documents like judgments [34]. Doc2Vec, Word2Vec, and another shallow NN-based embedding [35,36] Explore the semantic vector space with a focus on contextual information, which plays a crucial role in preserving the semantic relationships between words or documents. Sagathadasa et al. note that Word2Vec and lexical relevance were utilized as a process to identify domain-specific semantic similarities. Dipankar et al. [37] used Word2Vec and lexical significance to identify semantic similarities within domains and built a "risk ometer" system that uses Doc2Vec and supervised machine learning to assess the legal contracts' risk. Likewise [38], Doc2vec shown superior performance to TF-IDF, LDA, and Word2Vec in the empirical investigation that was carried out by Mandal et al. This was determined by the findings of the human expert similarity score [30].

2.2.2. *Hybrid approaches*

Using both textual and referential information, Kumar et al. [39] tried a hybrid strategy to improve performance. Using prior decisions and the idea of "paragraph links" to create a citation network. A paragraph link exists between two different judgments if the cosine similarity score between the TFIDF vectors of two paragraphs from different judgments is higher than the threshold. This indicates that the two judgments are related to each other. When compared to separate methods, the performance is noticeably better. Cluster analysis was carried out by Raghav et al. [40] on Indian legal papers utilizing a citation network and paragraph linkages. For US Supreme Court opinions, Leibon et al. [41] have also used network-based methods and text representation approaches like LDA. This study uses sophisticated methodologies, including Fuzzy AHP, Fuzzy DEMATEL, and Logistic regression (LR) models, to build a novel strategy for mapping groundwater potential zones (GWPZs). GWPZ was conducted by integrating hydrologic, soil permeability, morphometric, topographical distribution, and anthropogenic elements into 27 distinct criteria utilizing multi-criteria decision models [42]. There is a way to find relevant legal documents that Sugathadasa et al. [43] suggested. It combines a text-based strategy, like Text Rank [44], for finding the resemblance between different sentences with adopting a strategy known as a network-based strategy, like Node2Vec [45], for node entrenching.

2.3. *Retrieval of legal information*

The use of ICTs, which generate vast quantities of digital information, has recently transformed the legal sphere. To address this, researchers have looked into more effective retrieval methods. One of these methods is knowledge extraction. The other one is the natural language processing method. In addition to it machine learning theoretical model is also used. Moreover, non-monotonic deontic logic is also famous for exploring by lawyers. Along with other techniques rule-based techniques, and expert systems also play their pivotal role, in facilitating the exploration of such archives by attorneys and judges [46].

Information retrieval (IR) methods can aid in the retrieval of legal information from preexisting databases containing legal documents. Indeed, they have already been employed profitably for numerous purposes in literature. When it comes to analyzing legal case judgments, NLP-based approaches are generally considered superior. This is because these approaches combine data-driven methods with embedding models, which makes them able to directly identify legal concepts and jurisprudence without any hindrance thereto or different kinds of representations, such as tagged feature-value pairs or some kinds of logical predicates [47,48]. Ashley proposes a query-based system that retrieves legal information by systematically evaluating many factors that determine the sentences' interpretive value. They amalgamate characteristics into a composite metric that takes into consideration several facets. The task formulation, data set assembly, and extensive task analysis form a strong basis for implementing a learning-to-rank approach [49]. In their study, Wei et al. [50] detail the case of a patient who experienced severe hemodynamic abnormalities 3 h after undergoing treatment due to embolization of the LAAO device into the left ventricular outflow tract, including a ruptured mitral valve and subsequent extensive mitral regurgitation. Urgent surgery was required to remove the device and repair the mitral valve as a result. Furthermore, an examination of prior studies concerning surgical methodologies employed in the removal of dislodged left atrial appendage occludes was conducted.

In recent times, the big data paradigm and the presence of advanced technologies for big data analytics are causing legal authorities to lean towards publishing court case papers on their internet databases. However, scholars in the field of artificial intelligence are grabbing hold of this chance to improve already-conducted research and make a novel contribution to the field of legal informatics because there are some complexities in modern information technology that are even rising regularly. The importance of this study is due to its wider scope, i.e., legal information systems, legal drafting with the help of computers, and data banks of judicial decisions. The legal informatics concept prevailed in 1963 when Wolfgang Baade advanced the concept of "jurimetrics," and later on in 1969, Losano's concept was "luscibernetic". Using a machine learning technique [51], automated the annotation of common law report sentences with facts and principles pertaining to the law.

To determine if a sentence is a principle, fact, or neutral text, the suggested method uses a feature selection stage in conjunction with an NBM classifier. This study depended on a very short number of reports. Besides the number of reports, the results were positive.

Shao et al. [52] presented a model of the two-stage system. The model showed the results in enhanced relevance estimation because it used the technique to leverage the attention on the user end. Devin et al. [53] started the study in 2019 [53]. The fine-tuned version was observed to have the best performance, and the model was trained from the ground up using texts that were relevant to the area. Chalkidis et al. (2020) note that LEGAL-BERT (Bidirectional Encoder Representations from Transformers) presents the textual content of the legal information obtained from different legal documents by using embedding techniques. In addition, BURT uses a technique to come across the basic relation between the words and phrases that are available in the digital textual form. Moreover, the same result was found when different forms of BERT like “general-purpose pre-trained” [54], “domain-specific” or even “fine-tuned” [54] BERT were compared as to their result regarding classification.

Similarly, legal papers that are based on the Bag-of-Concept (BoC) approach can be expressed using BoLC-Th, which is an acronym that stands for Bag of Legal Concepts Based on Thesaurus. One of the distinctive characteristics of this method is that it generates weighted histograms of concepts by taking into account the proximity of each word to the synonym that corresponds to it in a thesaurus. This technique generates vectors that are most discriminative by emphasizing the words that are more appropriate to the context in which they are displayed [55]. In a competition on the extraction of legal information related to the legal domain, embedding techniques were used by the participants to sort out the multidimensional problems of the related field. The BERT-PLI technique seems to be most relevant regarding the extraction of legal information [56]. This technique makes use of BERT to capture the semantic associations at the paragraph level. Subsequently, it makes use of the aggregated interactions between paragraphs to determine the significance between two instances. A dataset that is connected to the legal industry is used to fine-tune the BERT model that is located in BERTPLI [56], the same as in LEGAL-BERT. Although other studies have looked into the application of embedding techniques, the authors of this study believe that the approach that is described here is a novel proposition. This empirical evaluation shows that it is much more robust to the presence of data noise compared to baseline and state-of-the-art solutions. This method is the first of its kind to use a 2-step clustering approach to investigate legal domain text at the whole legal judgement and passage levels to identify their regularities.

3. Fundamentals of machine learning

When we talk about an area of study within the realm of artificial intelligence known as machine learning focuses on the creation of a specialized operational model that takes into account the required characteristics [57]. It is a technique of analyzing the available data that contributes to the automation of the analytical and methodological construction of models. If we come to the basic operation it is pertinent to say that the fundamental operation of machine learning technology is dependent on the pattern of a vast amount of data, because the training of the model requires a variety of data to come across the sequences thereby [58,59]. After then, the obtained patterns by use of ML are categorized and can be utilized in a variety of tasks. The model is trained to produce the most appropriate outcome possible. This idea makes it possible to tackle some problems by classifying them into two major categories: classification difficulties and regression challenges. With machine learning, a system may automatically learn from its past mistakes and get better at what it does, all without human intervention or code. ML can be classified into four distinct categories: semi-supervised learning, reinforcement learning, supervised learning, and unsupervised learning. These classifications are based on the learning techniques that are used and how they are applied. Every learning technique applied to a model has its pros and cons. In supervised learning, we use a labeled dataset where we are aware of input and output. We use linear regression, KNN, and SVM. Mostly, it is used for predicting the possible outcome. There are two main types of supervised learning: classification and regression. Whereas, in classification, the output data is discrete, and in regression, the output is a continuous value. In unsupervised learning, the data is unlabeled, like we have only input in our data but not the corresponding output. In this kind of learning, the data is grouped based on similarity in the features provided in the data, and these types of groups are known as clusters. There are many other types of unsupervised learning, including clustering and association.

In semi-supervised learning, the data is in mixed form, as some of the data has only input but no output, and some of the data has both input and output. In this way, we use a variety of data that increases the level of accuracy of the results.

Reinforcement learning is reward-based learning in which we have agents. These agents perceive the environment, and based on these perceptions, action is taken. If the action of the agent is correct, then we reward the agent, and in the case of an incorrect action, the agent gets a penalty. In this way, agents learn which action has to be performed and from which action they need to abstain. Like in the case of legal judgments, when we make a model and train that model by reinforcement learning for bias classification if our model classifies the bias correctly from the legal judgment, we can award the agent (that is, in fact, the model we trained) a “good rating," and if it fails to classify the bias, we can award a “bad rating".

3.1. Machine learning for semantic biasness classification

When it comes to the realm of law, there are a variety of approaches that may be taken while dealing with legal matters. Systematizing the data and automating the process are the primary goals of the majority of the procedures that are carried out. Several different approaches to the automatic processing of legal documents have been studied by the authors in this section. Different techniques of AI are used to classify legal documents and later use them as big data for future prediction and solving current problems. It is humanly not possible to read all of the cases to get the guidelines for some recent problems. Therefore, finding the correlations and similarities between different legal judgments provides a basis for deciding new cases easily. This would be done to assist legal professionals in balancing their overwhelming workload of pending cases. Through the implementation of this strategy, the law will become more approachable, manageable, predictable, intelligible, and beneficial. Because it is written in natural languages, the majority of legal information is completely unstructured.

Consequently, to process this legal information, a variety of methodologies established through the use of NLP have been developed. Current research aims to construct an automated system to find the semantic bias in legal judgments that has a strong tendency to increase cognitive bias. As we know, the basic right associated with each technology is its fare usage. If technology is harmful to society, it can never be allowed to save society from the harm of technology. AI uses big data, as in the case of legal judgment. When we train some models on big data of decided judgments, having certain biases will result in biases of high magnitude. For this purpose, the supervised learning technique has been used so far to get better results. An automated legal classification system for legal judgment is based on supervised learning techniques. The data provided to the system is legal judgments that have already been decided by the courts, i.e., the CAIL dataset. The computer recognizes the different patterns of judgment as an initial training phase. Afterward, the data is served to the system to train it. This is done to evaluate the performance of the proposed machine-learning-based categorization model. The performance that is produced from the correctness of the legal categorization model is used to substantiate the case information, which is then used to structure the case information. The other main objective of the system is to dig out the categories of the paragraph, like the details of the basic reason for litigation, supportive evidence, relevant law, arguments forwarded by the parties, prudence applied by the judges, the basis of the decision, and final verdict, because, until the system categories the basic divisions of the paragraphs, it would not be possible and appropriate to catch the semantic bias in the judgment.

Afterward, techniques of ML are applied to get the results as depicted in Fig. 1. To get more accuracy, we apply SVM, NB, MLP, and KNN classifiers for forecasting as well. It is necessary to offer the machine learning program a case that does not contain the judgment (during the "testing phase") to evaluate its performance. The programmer is then required to provide the judgment that is most likely to occur. The program uses the data that it determined to be significant during the training phase to make this decision, also known as "classification".

3.2. Dataset

Our investigations are carried out using a dataset that was obtained from the China Judgements Online 1 online database. This section begins with an explanation of how we acquire such datasets, is followed by a discussion of a set of baseline systems that we compare with, and finishes with a presentation of the findings. Existing research in the field of LJP frequently conducted experiments using the CAIL dataset [12] or pulled task-specific sentences from available judgment documents to generate their dataset [60]. On the other hand, none of them includes all five of the responsibilities that we require. Additionally, these datasets have been subjected to some pre-processing steps, which makes it challenging to create a mapping between each instance and the initial court judgment document. In light of this, we generate a dataset on our own using the same approach as described in Ref. [12]. In the first place, we distinguished between cases that had several defendants and those that did not, since various defendants correspond to different outcomes of trials. Aside from the fact that the top 102 law articles in Chinese Criminal Law are not pertinent to certain charges, we also filtered these designations. Through this method, we were able to obtain a dataset that had 225,843 criminal judgment papers, which included 200 charges and 183 law articles. To construct the test set, we chose 12,810 cases at random and made certain that it included all possible categories of charges and legal articles. Moreover, we chose 12,634 cases to form our validation set, while the

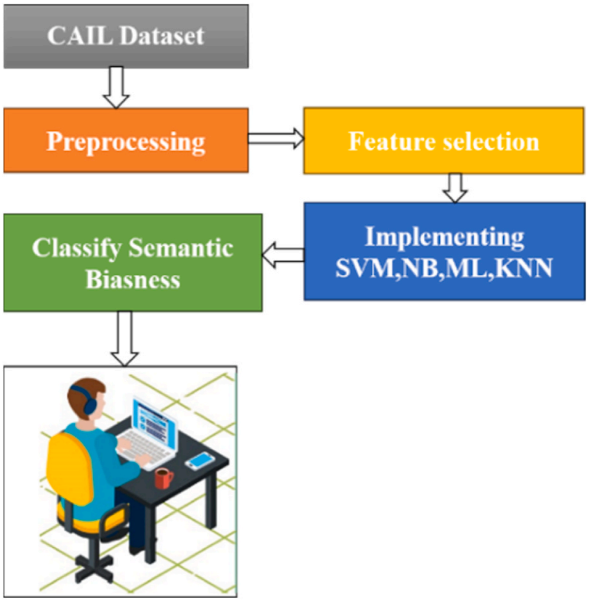

Fig. 1. The architecture of the proposed paradigm.
Source: Own extracted.

remaining cases were used to construct our training set. An average of 1.06 charges are involved in each case, and each case refers to 1.14 different sections of legislation; yet, there is only one term of punishment. We provide more detailed statistics in Table 1.

A Feature Extraction

The n-gram and the terms can be defined as two primary forms of combined feature extraction methods that are utilized in the process of text classification.

n-gram: For the n-gram extraction method, a window of length n is utilized to move across the entirety of a corpus [61]. Following that, we obtain all of the sets of consecutive words or characters that are contained within each window. To reduce the amount of ambiguity that is associated with individual words, the n-gram algorithm is designed to obtain the composite features that emerge continually. The n-grams known as bigram and trigram are frequently utilized. Nevertheless, the impact of the structure of the text, which includes things like punctuation and stop words, is not taken into account.

terms: With an n-gram, composite features are solely extracted based on their co-occurrence, regardless of the order and position of the member terms [62]. A term set, on the other hand, is fundamentally different from an n-gram. Alternatively, term sets can be described as arbitrary combinations of words that are paired together in the lexicon. There is, however, a difficulty that arises from this combination, and that is an explosion of combinations even for two-term sets combinations. When a vocabulary size of n is considered, this indicates that there will be 2n different sorts of pairings.

B Feature Selection

To improve the performance of a text classifier and simultaneously minimize the feature dimension, feature selection is a technique that is frequently utilized. During the process of feature selection, the score of each feature is often decided by a generic criterion. Following this, the top N features are selected from the feature subset (where N is a number that has been obtained through experimentation). Chi-square is a well-known statistical method that has been useful in determining the degree to which individuals can differentiate themselves from one another [63]. The formula for this method is as follows in equation (1).

$$Z^2(f_t) = \frac{N(wz - xy)^2}{(w+y)(x+z)(w+x)(y+z)} \tag{1}$$

In this equation, the variables w and y represent the number of documents that contain f_t in the positive and negative classes, respectively, while the variables x and z represent the number of documents that do not contain f_t. in the positive and negative classes, respectively. The sum of the documents that make up the training set is denoted by equation $N = w + x + y + z$. A technique that is based on Z^2 was proposed by Ref. [64] regarding the evaluation of term sets as given below in equation (2).

$$\widetilde{Z}^2\left(f_{t,v}\right) = \frac{N(\widetilde{w}\widetilde{z} - \widetilde{x}\widetilde{y})^2}{(\widetilde{w}+\widetilde{y})(\widetilde{x}+\widetilde{z})(\widetilde{w}+\widetilde{x})(\widetilde{y}+\widetilde{z})}, \tag{2}$$

where f_{tv} is the abbreviation for the two-term set that is composed of the terms f_t and f_v, $\widetilde{w}$ $\widetilde{y}$ respectively The numbers indicate the number of documents that contain both or either of f_t and f_v in the positive and negative classes, and $\widetilde{x}$, $\widetilde{z}$ are the number of documents that not contain any of f_t and f_v in the positive and negative classes, respectively. That a portion of the members is also capable of communicating information is indicated by this.

C Classification

Numerous methods are used for processing legal documents and other related material for the sake of classification. Researchers have made efforts to automate the system for this classification by using different kinds of processes. Currently, our focus is on the use of ML and DL algorithms within the argument-based prediction system for legal judgments. Because the majority of the legal documents filed in trial courts are of an unstructured kind, they must be structured appropriately. In most of the judiciaries, the process of automation and digitalization of legal documents like patents, written statements, different types of evidence, and arguments is still incomplete. First of all, we need to prepare soft copies of all legal documents. For this purpose, we use the technique of supervised learning, where we label the documents properly to get optimal results. To calculate the outcome of our suggested legal model that is based on arguments, data is submitted to the system despite finalizing the results being determined. Another name for this phase is the testing phase, and it is responsible for producing reliable results. The flowchart for classification is shown in Fig. 2.

Table 1
CAIL dataset Statistics presenting words and exceptional words.

	Statement	Law	Charge	Punishment	Opinion	Content
No. of words	119.46	3	7.26	6.41	148.28	137.90
No. of Exceptional words	225,843	183	200	207	225,843	183

Source: own extracted from CAIL dataset [12].

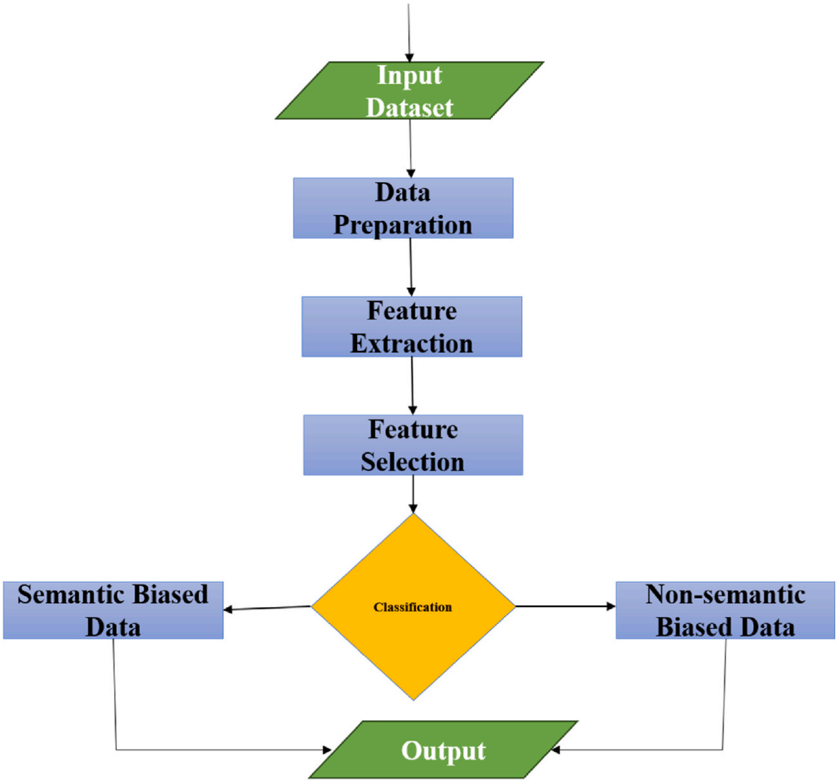

Fig. 2. Classification model that is proposed to be semantically biased.
Source: Own extracted.

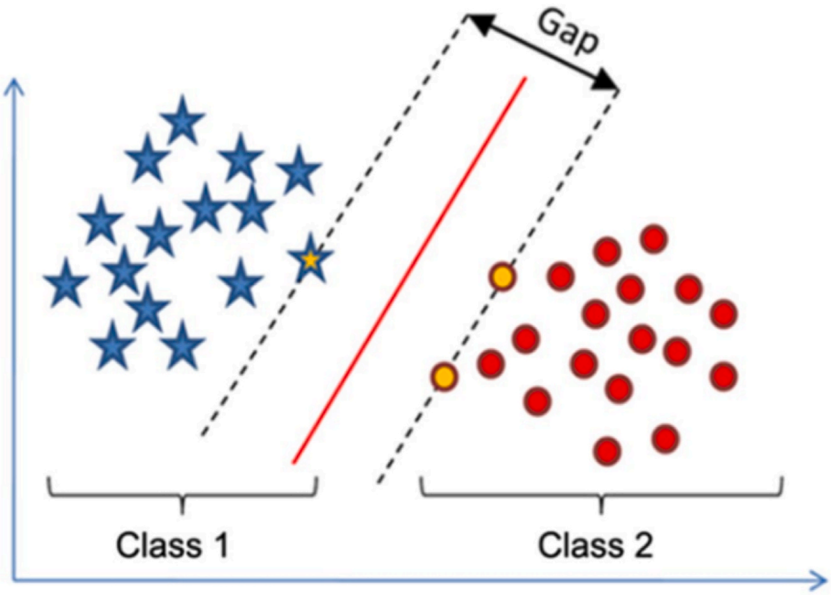

Fig. 3. Svm diagram.
Source: take from Ref. [65].

3.3. Support Vector Machine

The SVM is an additional example of a supervised machine-learning technique. Although it may be used for both classification and regression, it is typically employed for problems that are associated with the classification. Through the utilization of kernels, it is utilized in both linear and non-linear processes. The SVM can identify linear separation and functions most well in situations where there are a limited number of points for a large number of observations. Additionally, it is well known that SVM can be learned throughout the entire world. Essentially, SVM can learn the linear threshold function in its most fundamental application. On the other hand, with the assistance of an easy plugin that is suited for the kernel, they can upgrade their function to an alternative future function. On the other hand, SVMs are extremely computationally intensive and theoretically complex. This is also a type of classifier that employs an N-dimensional hyperplane to classify a collection of points or data separately. Essentially, a hyperplane is nothing more than an n-dimensional line that serves the purpose of dividing data that belongs to two distinct classes. This particular classifier is used in situations involving regression as well as classification issues. To develop the classifier, we made use of the SVM module that is included in the Scikit-Learn package with Python. Presented in equation (3) is the statement that describes the creation of the classifier:

$$svmModel = svm.SVC(\,) \tag{3}$$

To adapt the inputs to the classifier, we have employed the fit () function from the svm, SVC () class. This function is responsible for feeding the inputs, which are data from the dataset, into the classifier that we have created. Equation (4) is an illustration of the statement that explains how the inputs are matched to the classifier:

$$svmModel.fit(XTrain, YTrain) \tag{4}$$

After adopting the inputs to the classifier model, we obtain the classes that are predicted by our classifier from the testing input data. This is the process of discovering the predictions or classes. In the svm, SVC () class, the predict () function is used to carry out the task of prediction. Equation (5) is a statement regarding the prediction of output classes:

$$Y \Pr edict = svmModel.pred(XTest) \tag{5}$$

An example of a system that is operational is depicted in Fig. 3. The algorithm that is used to separate the data will determine which hyperplane is the most effective. To differentiate between the two symbols, the hyperplane serves as the center line. In this case, the support vectors are the symbols that are closest to the lines. Fixing the hyperplane in such a way that the margin that is most likely to be attained concerning the points is the goal of the algorithm known as the Support Vector Machine. Following the completion of the training, a fresh and distinct case set is utilized for an evaluation of the effectiveness of the ML approach. Every case is examined to determine whether or not the accused individual is a criminal, and the results of this examination are then compared with the documents that were first produced by the court.

3.4. K-Nearest Neighbour

This algorithmic classification approach is straightforward. It yields a considerable measurement of extremely modest output in terms of cataloging precision. Furthermore, it is a direct classifier that is based on the class of objects closest to one another. On vast amounts of datasets, it functions most effectively. Each unknown test sample is assigned to a particular class in the KNN classifier, which is the class to which the majority of the sample's KNN belongs. To put it another way, it is determined by the one that has received the greatest sum of polls from its nearest neighbors, to whom the object will be assigned. There are a few different names for the KNN algorithm, which can be categorized as firstly memory-based reasoning, secondly instance-based learning, thirdly case-based reasoning, fourthly example-based reasoning, and finally lazy learning are the core topics of the discussion of current research. It is

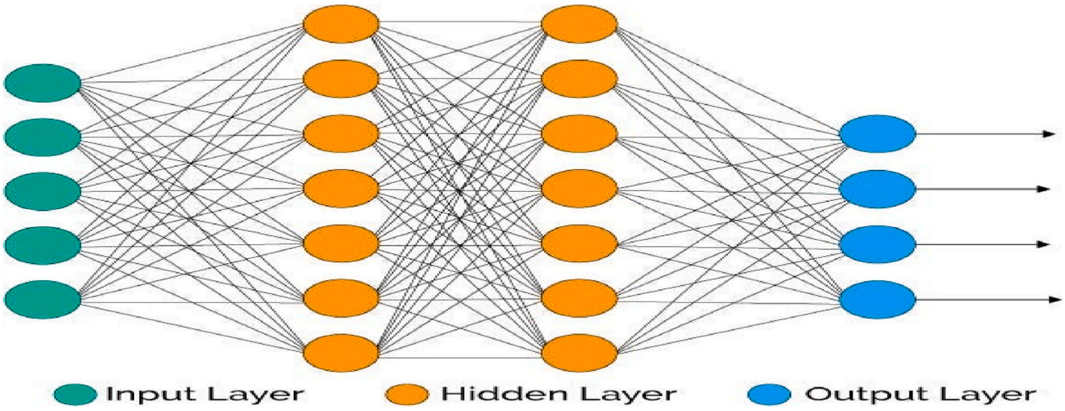

Fig. 4. The conceptual framework of our MLP model.
Source: Own extracted from MLP concept.

essentially non-linear and can discern between linear and nonlinear datasets, depending on the data. In our suggested legal model, we have replicated the words with things since the KNN Classifier implies that things that are similar to one another dwell close to one another. To create the classifier, we made use of the K Neighbors classifier module that is included in the Scikit-Learn package of Python. Details regarding this module are as follows in equation (6).

$$knnClassifer = KNeighborsClassifer(\ ) \tag{6}$$

Compatibility of the inputs with the classifier: To integrate data into the recently created classifier, we have employed the fit () function of the K Neighbors Classifier class, as outlined in equation (7).

$$knnClassiferfit(XTrain, YTrain) \tag{7}$$

The process of determining the predictions and classes begins with the insertion of inputs into the classifier model. Next, we use the predict () function of the K Neighbors Classifier class to determine the classes that our classifier has predicted. This function is also mentioned in equation (8).

$$Y\,\mathrm{Pr}\,edict = knnClassifier.pr\,edict(XTest) \tag{8}$$

3.5. Multi-layer perception

In the context of machine learning, an Artificial Neural Network (ANN) is a specific sort of model that mimics the organization of the human brain and its constituent neurons to classify incoming data sets within a machine. As can be seen in Fig. 4, the authors of this research have constructed a feed-forward network that is composed of multiple layers.

Four layers of density make up our model. The input layer, first hidden layer, second hidden layer, and output layer are the layers that are engaged in this process. The output layer is the last. A layer's nodes are connected to all of the nodes in the layer below them, and these connections are ordered in the following order:

3.5.1. Dense layer

A layer that is entirely connected means that every neuron is connected to every other neuron through their respective layers. This layer is a layer that is completely connected. All learning features that are derived from amalgamations of all aspects of the layers that came before it is utilized by it. The first dense layer is comprised of sixteen nodes that accept input in a single dimension with the shape (4, 1). The Rectified Linear Unit, alternatively known as ReLU, is the activation function that is utilized in this layer. Through the use of this activation function, the output is converted into the weighted sum of the inputs to the output nodes. In our model, the sum of nodes that are considered to be present in the second, third, and fourth dense layers is assumed to be 8, 4, and 1 accordingly. To obtain the class probability scores, the ReLU activation function is utilized in the second and third dense layers, whereas the SoftMax activation function is utilized in the fourth dense layer.

3.5.2. Dropout layer

To prevent overfitting, we have implemented a dropout layer in between each consecutive dense layer. This layer has a dropout rate of 0.2, which means that one out of every five inputs will be eliminated at random during each epoch.

It is well acknowledged that the Multilayer Perceptron (MLP) is among the most widely used supervised neural classifiers. At this point, a vast number of learning paradigms have been established, and in addition to that, they can carry out nonlinear mapping. Nonlinear activation functions are the substance that makes up MLP networks. Hidden layers and associated synaptic weights are responsible for carrying out this non-linear mapping when it comes to the MLP network. Iterative determination of the biases and weights of the MLP network is accomplished by the utilization of backpropagation, which is a general supervised optimization technique.

Following the acquisition of anticipated class labels, the aforementioned labels are subjected to analysis utilizing the cataloging information and precision mark, which are presented in tabular format below.

4. Experimental results

Our legal model which is based on the ML algorithm that we have proposed has been anticipated and performed utilizing Python 3.5. Additionally, AMDA Ryzen 5 with 16 GB of RAM has been utilized for the processing of the model, and it has been executed on the Windows 10 professional. To determine whether the accused individual is a criminal or a non-criminal, the final output is classified according to this distinction. Different classifiers, including SVM, NB, MLP, and KNN, have been utilized in this research. These classifiers have been utilized to analyze the data. To achieve higher levels of accuracy in classification, it has been suggested that an adequate and identifiable ensemble classifier be utilized. The Precision, recall, F1-score, and Accuracy are calculated from equations (9)–(12).

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$\mathrm{Rec}all = \frac{TP}{TP + FN} \tag{10}$$

$$F1 - score = 2 * \frac{\Pr ecision \times \text{Recall}}{\Pr ecision + \text{Recall}} \tag{11}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

Whereas **TP** = True Positive, **FP**=False Positive, **TN** = True Negative, and **FN**=False Negative.

The performance measures of the classifiers are presented in Table 2 and Fig. 5. As a performance metric, F1-score, Precision, and Recall have been utilized in this research project for evaluation. This was done to determine whether or not the machine learning models are feasible. Specifically, the F-1 score is the Harmonic Mean that is associated with precision and recall. The term "Precision" is the correct label that should be provided for the percentage that is relevant to the instance. The term "recall" refers to the precise identification of a specific label for the proportion that is associated with a case.

The accuracy of classification achieved by Machine Learning classifiers through the application of the K-Fold cross-validation technique is presented in Table 3 and Fig. 6. The solution to the problem of insufficient data for the formation of validation sets is the application of the cross-validation technique. The process of resampling has been carried out by k-fold cross-validation in this particular piece of research. To begin, the initial sample is divided into k independent subsets of equal size, which are denoted by the letters S1, S2, S3, …, Sk. This division is accomplished by a stratified split. Immediately following this, a k-times implementation of the training mockup test is carried out chronologically. In this case, the value of k has been determined to be 5, and on the foundation of this, all of the mockup sets have been separated into five parts. In light of this, the execution of the entire procedure is comprised of five epochs, each of which is comprised of a different combination of training and testing samples. In conclusion, the estimated average of five procedures has been completed to deliver accuracy and computing time on the estimate of the entire classifier recitals.

4.1. Compare to state-of-art

We have made the comparison of our models with different state of art methods such as DT, LR, CNN-BiLSM, BERT, LSTM, and AlexNet. The resulting accuracy of the last model AlexNet was 95.15 while our model got an accuracy of 96.90 which is high compared to the other state-of-the-art methods. It speaks volumes that our model is more accurate as compared to the rest models. The comparison is shown in Table 4.

4.2. Discussion

The context of a word is an essential component of its representation. In addition to lexical definitions and even explicit knowledge, context is responsible for the creation of meaning for any given notion. Consequently, the co-occurrence of a term with a situation that is generally positive or negative may promote the development of a subtle associative meaning. Adaptive goals could be served by this, such as supporting individuals with reading comprehension by assisting them in predicting the valence of concepts that are close to one another. On the other hand, it has the potential to imbue these statements with subtle affective tones, which can then be used to shape global appraisals of other individuals [72]. The GloVe word embedding was trained using a collection of text available on the internet.

The findings suggest that language itself contains recoverable and accurate imprints of our historical biases. These biases may be morally neutral, such as when it comes to insects or flowers; problematic, such as when it comes to race or gender; or even simply veridical, reflecting the status quo for the distribution of gender concerning careers or beginning names [73]. Thus, finding semantic bias in judicial decisions is optimal before we use it as raw data to learn models that could be helpful for judicial assistance in adjudication. Our model classifies semantic bias by categorizing bias into three different types i.e. i) Positive and negative connotation of words, ii) referencing phrases out of context, and iii) stereotype bias. Different judgments that contain semantic biases are given in Table 5.

In the first case, during the writ application, the petitioner asserts that the ad hoc judge made a mistake by applying the incorrect legal standard. In the written reasons for judgment, the court stated that no evidence indicated a major and objective bias against any of the attorneys or parties engaged in the case. However, the petitioner claims that the recently modified Article 151(B) necessitates a finding of a "substantial and objective basis" for recusal, emphasizing the contrast between "bias" and "basis." The petitioner contends that the terms have distinct implications, with "bias" being a customary cause for recusal and "basis" establishing a more comprehensive required ground. Although this may at first appear to be a semantic issue, the petitioner argues that the terms have unique implications. Because the petitioner asserts that the judge may have applied the more limited criterion of "bias" in an erroneous manner and

Table 2
Measurement of performance.

Sr. No.	Name of Classifier	Precision	Recall	F1-score
1	SVM	0.94	0.97	0.95
2	NB	0.88	0.93	0.87
3	MLP	0.86	0.88	0.84
4	KNN	0.85	0.87	0.82

Source: own extracted from confusion matrix using equations (9)–(12).

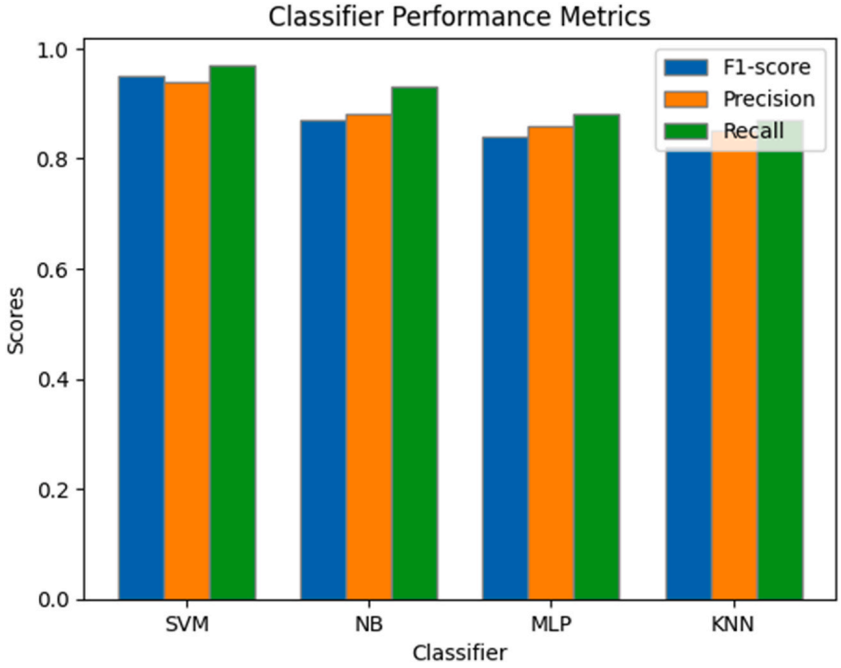

Fig. 5. The effectiveness of traditional machine learning routines.
Source: Own extracted from Table 2 data.

Table 3
The accuracy of classification.

Sl. No.	Classifier	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Accuracy %
1	SVM	93.80	95.90	96.23	95.95	94.55	96.90
2	NB	87.90	89.73	87.50	86.70	85.75	88.80
3	MLP	82.20	84.40	86.75	86.40	85.96	86.75
4	KNN	81.70	83.70	82.63	84.55	83.85	85.66

Source: own extracted from models experiment output.

neglected to take into consideration the more general grounds of other "bases" for recusal, it is necessary to conduct a de novo review of the judgment that was handed down by the trial court in light of the newly updated legal standard.

In the second cited case, as part of its effort to have the Tenth Counterclaim dismissed, the National Patent and Trademark Office (NPI) contends that it did not "remove" the statement that was made regarding biased contacts; rather, it says that the remark was only "omitted" from the relevant patent applications. The court categorizes the argument as a semantic issue, highlighting the fact that the core of GPS's claim is that the statement was not intentionally removed but rather that it was absent from the document. In addition, NPI asserts that the patent examiner, and not NPI, was the one who said that the previous art did not disclose any biased interactions. Nevertheless, the court makes notice of the fact that this response does not address the primary complaint against GPS. GPS asserts that the patent examiner was misled into believing that biasing contacts were a novel aspect in comparison to the previous art because the remark about biasing contacts was not included in the patent application during the examination process. Although the court does not dive into the merits of GPS's claim at this juncture, it does note that GPS has presented adequate details regarding the who, what, when, where, and how in support of its Tenth Counterclaim. This, in turn, suggests that the rejection of the counterclaim is not merited at this time.

In the other cases, the judge describes Mr. Miles's actions, and the judge's critical attitude is made clear by the selection of the labeling word distortion rather than the more neutral operation, for example. Similarly, the judge refers back to a document that was published by Jefferson County, which includes a subject that is involved in a dispute regarding racial discrimination in the allocation of children to public schools. A significant deal of detail is provided by the court regarding the drawbacks of the internal restrictions that Jefferson County has in place; the judge has stated it in terms that are quite general and imprecise; it fails to make clear; it indicates that without providing any further explanation. In addition, he captures these remarks through these ambiguities, which plainly show the judge's critical stance towards the briefing that Jefferson County presented as evidence against the defendant. However, distortion and ambiguity are not the sole lemmas via which the Court expresses its most explicit opinion when it comes to the matter at hand.

Similarly, "to ensure that the rights of children with disabilities and parents of such children are protected," is one of the aims that the Individuals with Disabilities Education Act (IDEA) seeks to accomplish. (B) to Section 1400(d) (1). In the language that was

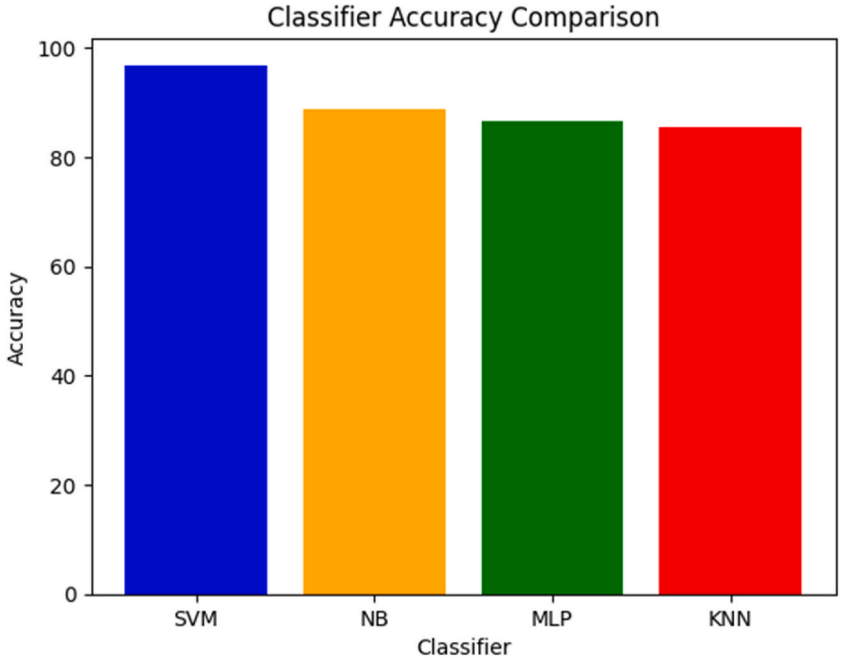

Fig. 6. The accuracy of traditional machine learning techniques is compared.
Source: Own extracted from Table 3 data.

Table 4
Comparison with different state-of-art methods.

Method	Year	Accuracy %
DT [66]	2018	85.00
LR [67]	2019	88.87
CNN-BiLSM [68]	2020	81.90
BERT [69]	2021	94.92
LSTM [70]	2022	84.60
AlexNet [71]	2023	95.15
SVM (Our)	–	**96.90**

Table 5
Classification of the semantic biasedness by model.

Ref.	Judgments	Type-1	Type-2	Type-3	Correction
[74]	"No evidence to reflect any substantial and objective bias toward any of the parties or attorneys involved …,"	×	×	✓	Replacement of "bias" with "basis."
[75]	"remove" the statement concerning biasing contacts but rather assert that the statement was merely "omitted" from the relevant patent applications.	✓	×	×	Replacement of “Omitted” as to “removal”
[76]	The judge clearly expresses his disapproval of Mr. Miles’s behavior, as evident from his use of the term "distortion" instead of a more neutral term like "operation".	✓	×	×	Distortion with operation
[77]	great detail," "ambiguities," and "negative standpoint" instead of This/these/that/those, or misapplication, and ill-defined test, error, etc.	✓	✓	×	These ambiguities with proper annotation
[78]	to guarantee that the rights of children with disabilities and the parents of such children are safeguarded throughout history.	×	✓	✓	Putting the rights of both
[79]	a Federal Employees’ Liability Act (FELA) plaintiff can reach a jury if he can demonstrate that his employer’s carelessness was even the tiniest cause of his injury.	×	×	✓	Use of “proximate’’ cause, slightest’’ cause.

Source: own extracted from references [74–79].

mentioned, the word "rights" denotes both the rights of the kid and the rights of the parents; otherwise, the grammatical structure would be completely incomprehensible. The other sections lend credence to this viewpoint.

5. Conclusion

It is undeniable that a fair society is essential for progress. The fairness of utilizing AI in legal judgments is crucial for its legitimacy. Given the potential for cognitive bias, the automation and digitalization of judicial systems in light of AI innovation pose a significant threat. For AI to be effectively used in the courtroom, it is crucial to train models using high-quality data, including ML, NLP, and LLM, while also minimizing any unnecessary data. Semantic bias is one of the important types of biases that are found in legal judgments because judges have discretionary powers to decide the cases. While preparing judgments, judges use words and phrases that can cause difficulties in understanding and implementation of the judgments. It can also cause distress in society and breach of fundamental rights, as these judgments have to be used as data warehouses so if in case of any bias carried by this judgment might result in an aggravated form of bias when used as a warehouse of AI modes. This task involves categorizing the data and preparing it for training by the standards of the legal system. In the present study, we have carried out some experiments based on Machine Learning algorithms. These experiments comprised analyzing the language of the judgments that were made using the Chinese AI and Law (CAIL) dataset to categorize the semantic biases if there were any particular semantic biases. To determine the semantic connection among the legal datasets, classification, and identification have been carried out in an organized manner. For semantic classification, four different classifiers such that SVM, NB, MLP, and KNN have been utilized and analyzed for the classification of semantic biases, which are substantially larger than in other circumstances, our models perform significantly better than those of other models. We have conducted thorough experimental evaluations, which have demonstrated that the created method is accurate and efficient when used with real data. As a result, this method can be considered a helpful tool in real-world scenarios, including situations in which massive collections of legal documents need to be analyzed. Not only have our suggested machine learning algorithms been able to outperform four existing state-of-the-art competitor systems, but they have also been able to achieve many superior performances when the amount of noise in the data is increased. However, it will be necessary to do additional studies to determine how these systems could be enhanced by employing a more sophisticated legal and linguistic understanding.

Funding

National Social Science Foundation project "Research on the Construction of Governance System in Cyberspace in the New Era" project No.19BFX019 and National Social Science Fund major project "Big Data Sovereignty Security Guarantee System construction Research" phased results (project number: 21&ZD168)

Data availability statement

Some or all data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request.

CRediT authorship contribution statement

Kashif Javed: Methodology, Formal analysis. **Jianxin Li:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge Government of China for awarding the projects of National Social Science Foundation project and National Social Science Fund major project that have funded this research. Authors are also thankful to school of law, Zhengzhou University for the opportunities provided for research.

References

[1] O. Nasir, R.T. Javed, S. Gupta, R. Vinuesa, J. Qadir, Artificial intelligence and sustainable development goals nexus via four vantage points, Technol. Soc. 72 (2023) 102171.
[2] C. Liu, et al., M-FLAG: medical vision-language pre-training with frozen language models and latent space geometry optimization, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2023, pp. 637–647.
[3] Z. Wan, et al., Med-unic: Unifying cross-lingual medical vision-language pre-training by diminishing bias, Adv. Neural Inf. Process. Syst. 36 (2024).
[4] R. Berk, H. Heidari, S. Jabbari, M. Kearns, A. Roth, Fairness in criminal justice risk assessments: the state of the art, Socio. Methods Res. 50 (1) (2021) 3–44.
[5] S. Goel, R. Shroff, J. Skeem, C. Slobogin, The accuracy, equity, and jurisprudence of criminal risk assessment. Research Handbook on Big Data Law, 2021, pp. 9–28.

[6] J. Zeleznikow, The benefits and dangers of using machine learning to support making legal predictions, Wiley Interdisciplinary Reviews: Data Min. Knowl. Discov. (2023) e1505.
[7] J.-F. Bonnefon, I. Rahwan, A. Shariff, The moral psychology of Artificial Intelligence, Annu. Rev. Psychol. 75 (2023).
[8] D. Pessach, E. Shmueli, A review on fairness in machine learning, ACM Comput. Surv. 55 (3) (2022) 1–44.
[9] A. Ignatiev, M.C. Cooper, M. Siala, E. Hebrard, J. Marques-Silva, Towards formal fairness in machine learning, in: Principles and Practice of Constraint Programming: 26th International Conference, CP 2020, Louvain-La-Neuve, Belgium, September 7–11, 2020, Proceedings 26, Springer, 2020, pp. 846–867.
[10] J.C. Oleson, A requiem for the Unabomber, Contemp. Justice Rev. (2023) 1–29.
[11] J. Perera, S.-H. Liu, M. Mernik, M. Črepinšek, M. Ravber, A graph pointer network-based multi-objective deep reinforcement learning algorithm for solving the traveling salesman problem, Mathematics 11 (2) (2023) 437.
[12] Xiao C., et al., "Cail2018: a large-scale legal dataset for judgment prediction,", arXiv preprint arXiv:1807.02478 (2018). pp.1–4.
[13] S. Greenstein, Preserving the rule of law in the era of artificial intelligence (AI), Artif. Intell. Law 30 (3) (2022) 291–323.
[14] F. Bell, L. Bennett Moses, M. Legg, J. Silove, M. Zalnieriute, AI Decision-Making and the Courts: A Guide for Judges, Tribunal Members and Court Administrators'," Australasian Institute of Judicial Administration, 2022.
[15] J. Wróblewski, Legal decision and its justification, Logique Anal. 14 (53/54) (1971) 409–419.
[16] People v. Rose, NY: county court, in: Misc. 2d, vol. 82, 1975, p. 429.
[17] A. Ollikainen, Asset Partitioning in the Trust, University of Oxford, 2018.
[18] K. Rayner, A. Pollatsek, J. Ashby, C. Clifton Jr., Psychology of Reading, 2012.
[19] J. Zhang, E. Bareinboim, "Fairness in decision-making—the causal explanation formula,", in: Proceedings of the AAAI Conference on Artificial Intelligence vol. 32, 2018, 1.
[20] M. Katsaros, J. Kim, T. Tyler, Online content moderation: does justice need a human face? Int. J. Hum. Comput. Interact. 40 (1) (2024) 66–77.
[21] F. Mehmood, E. Chen, M.A. Akbar, A.A. Alsanad, Human action recognition of spatiotemporal parameters for skeleton sequences using MTLN feature learning framework, Electronics 10 (21) (2021) 2708.
[22] A. Chouldechova, Fair prediction with disparate impact: a study of bias in recidivism prediction instruments, Big Data 5 (2) (2017) 153–163.
[23] F. Pfisterer, Algorithmic fairness, in: Applied Machine Learning Using Mlr3 in R, Chapman and Hall/CRC, 2024, pp. 316–324.
[24] S. Barocas, M. Hardt, A. Narayanan, Fairness and Machine Learning: Limitations and Opportunities, MIT Press, 2023.
[25] J. Sun, S. Huang, C. Wei, Chinese legal judgment prediction via knowledgeable prompt learning, Expert Syst. Appl. 238 (2024) 122177.
[26] Heng Q., Yu S., Zhang Y., A new AI-based approach for automatic identification of tea leaf disease using deep neural network based on hybrid pooling, Heliyon 10 no.5 (2024); e26465.pp.1-13.
[27] A. Trivedi, A. Trivedi, S. Varshney, V. Joshipura, R. Mehta, J. Dhanani, "Similarity analysis of legal documents: a survey,", ICT Analysis and Applications: Proceedings of ICT4SD 2 (2021) (2020) 497–506. Springer.
[28] F. Mahmood, K. Abbas, A. Raza, M.A. Khan, P.W. Khan, Three dimensional agricultural land modeling using unmanned aerial system (UAS), Int. J. Adv. Comput. Sci. Appl. 10 (1) (2019).
[29] S. Kumar, P.K. Reddy, V.B. Reddy, A. Singh, Similarity analysis of legal judgments, in: Proceedings of the Fourth Annual ACM Bangalore Conference, 2011, pp. 1–4.
[30] A. Mandal, R. Chaki, S. Saha, K. Ghosh, A. Pal, S. Ghosh, Measuring similarity among legal court case documents, in: Proceedings of the 10th Annual ACM India Compute Conference, 2017, pp. 1–9.
[31] M. Koniaris, I. Anagnostopoulos, Y. Vassiliou, Network analysis in the legal domain: a complex model for European union legal sources, Journal of Complex Networks 6 (2) (2018) 243–268.
[32] Ayden M.A., Yuksel M.E., Erdem S.E.Y., A two-stream deep model for automated ICD-9 code prediction in an intensive care unit, Heliyon10, no.4 (2024); e25960. pp.1-14.
[33] R. Nanda, K.J. Adebayo, L. Di Caro, G. Boella, L. Robaldo, Legal information retrieval using topic clustering and neural networks, in: COLIEE@ ICAIL, 2017, pp. 68–78.
[34] C. Guo, M. Lu, W. Wei, An improved LDA topic modeling method based on partition for medium and long texts, Annals of Data Science 8 (2021) 331–344.
[35] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, Adv. Neural Inf. Process. Syst. 26 (2013).
[36] F. Mehmood, E. Chen, T. Abbas, M.A. Akbar, A.A. Khan, Automatically human action recognition (HAR) with view variation from skeleton means of adaptive transformer network, Soft Comput. (2023) 1–20.
[37] K. Sugathadasa, et al., Synergistic union of word2vec and lexicon for domain specific semantic similarity, in: 2017 IEEE International Conference on Industrial and Information Systems (ICIIS), IEEE, 2017, pp. 1–6.
[38] D. Chakrabarti, et al., Use of artificial intelligence to analyse risk in legal documents for a better decision support, in: TENCON 2018-2018 IEEE Region 10 Conference, IEEE, 2018, pp. 683–688.
[39] S. Kumar, P.K. Reddy, V.B. Reddy, M. Suri, Finding similar legal judgements under common law system, in: Databases in Networked Information Systems: 8th International Workshop, DNIS 2013, Aizu-Wakamatsu, Japan, March 25-27, 2013. Proceedings 8, Springer, 2013, pp. 103–116.
[40] K. Raghav, P. Balakrishna Reddy, V. Balakista Reddy, P. Krishna Reddy, Text and citations based cluster analysis of legal judgments, in: Mining Intelligence and Knowledge Exploration: Third International Conference, MIKE 2015, Hyderabad, India, December 9-11, 2015, Proceedings 3, Springer, 2015, pp. 449–459.
[41] G. Leibon, M. Livermore, R. Harder, A. Riddell, D. Rockmore, Bending the law: geometric tools for quantifying influence in the multinetwork of legal opinions, Artif. Intell. Law 26 (2018) 145–167.
[42] C. Singha, K.C. Swain, B. Pradhan, D.K. Rusia, A. Moghimi, B. Ranjgar, Mapping groundwater potential zone in the subarnarekha basin, India, using a novel hybrid multi-criteria approach in Google earth Engine, Heliyon 10 (2) (2024) e24308.
[43] K. Sugathadasa, et al., "Legal document retrieval using document vector embeddings and deep learning,", Intelligent Computing: Proceedings of the 2018 Computing Conference 2 (2019) 160–175. Springer.
[44] A. Kumari, D. Lobiyal, Efficient estimation of Hindi WSD with distributed word representation in vector space, Journal of King Saud University-Computer and Information Sciences 34 (8) (2022) 6092–6103.
[45] A. Grover, J. Leskovec, node2vec: scalable feature learning for networks, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 855–864.
[46] C. Sansone, G. Sperlí, Legal information retrieval systems: state-of-the-art and open issues, Inf. Syst. 106 (2022) 101967.
[47] C. Xiao, Z. Liu, Y. Lin, M. Sun, Legal knowledge representation learning, in: Representation Learning for Natural Language Processing, Springer Nature Singapore, Singapore, 2023, pp. 401–432.
[48] Z. Sun, J. Xu, X. Zhang, Z. Dong, J.-R. Wen, Law Article-Enhanced Legal Case Matching: A Causal Learning Approach, 2023.
[49] J. Šavelka, K.D. Ashley, Legal information retrieval for understanding statutory terms, Artif. Intell. Law (2022) 1–45.
[50] W. Ye, et al., Emergent surgical retrieval of a left atrial appendage occluder migrated into the left ventricular outflow tract with secondary massive mitral regurgitation: a case report and literature review, Heliyon 10 (2024) e27112. PP. 1-8.
[51] G. De Martino, G. Pio, M. Ceci, PRILJ: an efficient two-step method based on embedding and clustering for the identification of regularities in legal case judgments, Artif. Intell. Law 30 (3) (2022) 359–390.
[52] Y. Shao, Y. Wu, Y. Liu, J. Mao, S. Ma, Understanding relevance judgments in legal case retrieval, ACM Trans. Inf. Syst. 41 (3) (2023) 1–32.
[53] Kenton, Jacob Devlin Ming-Wei Chang, and Lee Kristina Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of NAACL-HLT*, 2019. pp. 4171–4186.
[54] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, I. Androutsopoulos, LEGAL-BERT: the Muppets Straight Out of Law School, 2020 arXiv preprint arXiv: 2010.02559.

[55] W.M. Costa, G.V. Pedrosa, Legal information retrieval based on a concept-frequency representation and thesaurus, ICEIS 1 (2023) 303–311.
[56] Y. Shao, et al., BERT-PLI: modeling paragraph-level interactions for legal case retrieval, IJCAI (2020) 3501–3507.
[57] G. Manikandan, S. Abirami, A survey on feature selection and extraction techniques for high-dimensional microarray datasets, Knowledge Computing and its Applications: Knowledge Computing in Specific Domains II (2018) 311–333.
[58] R. Sil, D. Saha, A. Roy, A study on argument-based analysis of legal model, in: Innovations in Bio-Inspired Computing and Applications: Proceedings of the 11th International Conference on Innovations in Bio-Inspired Computing and Applications (IBICA 2020) Held during December 16-18, 2020 11, Springer, 2021, pp. 449–457.
[59] B.M. Pavlyshenko, Machine-learning models for sales time series forecasting, Data 4 (1) (2019) 15.
[60] H. Zhong, Y. Wang, C. Tu, T. Zhang, Z. Liu, M. Sun, Iteratively questioning and answering for interpretable legal judgment prediction, Proc. AAAI Conf. Artif. Intell. 34 (1) (2020) 1250–1257.
[61] R. Tesar, V. Strnad, K. Jezek, M. Poesio, Extending the single words-based document model: a comparison of bigrams and 2-itemsets, in: Proceedings of the 2006 ACM Symposium on Document Engineering, 2006, pp. 138–146.
[62] D. Badawi, H. Altınçay, Termset weighting by adapting term weighting schemes to utilize cardinality statistics for binary text categorization, Appl. Intell. 47 (2) (2017) 456–472.
[63] M.L. McHugh, The chi-square test of independence, Biochem. Med. 23 (2) (2013) 143–149.
[64] D. Badawi, H. Altınçay, A novel framework for termset selection and weighting in binary text classification, Eng. Appl. Artif. Intell. 35 (2014) 38–53.
[65] R. Sil, A. Roy, A novel approach on argument based legal prediction model using machine learning, in: 2020 International Conference on Smart Electronics and Communication (ICOSEC), IEEE, 2020, pp. 487–490.
[66] Y. Asim, A.R. Shahid, A.K. Malik, B. Raza, Significance of machine learning algorithms in professional blogger’s classification, Comput. Electr. Eng. 65 (2018) 461–473.
[67] E.L. Park, S. Cho, P. Kang, Supervised paragraph vector: distributed representations of words, documents and class labels, IEEE Access 7 (2019) 29051–29064.
[68] Y. Zhu, Y. Li, Y. Yue, J. Qiang, Y. Yuan, A hybrid classification method via character embedding in Chinese short text with few words, IEEE Access 8 (2020) 92120–92128.
[69] J. Case, A. Clements, The impact of sentiment in the news media on daily and monthly stock market returns, in: Data Mining: 19th Australasian Conference on Data Mining, AusDM 2021, Brisbane, QLD, Australia, December 14-15, 2021, Proceedings 19, Springer, 2021, pp. 180–195.
[70] S. Srivastava, R. Tiwari, R. Bhardwaj, D. Gupta, Stock price prediction using LSTM and news sentiment analysis, in: 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), IEEE, 2022, pp. 1660–1663.
[71] M.A. Wajid, A. Zafar, M.S. Wajid, A deep learning approach for image and text classification using neutrosophy, Int. J. Inf. Technol. (2023) 1–7.
[72] D.J. Hauser, N. Schwarz, How seemingly innocuous words can bias judgment: semantic prosody and impression formation, J. Exp. Soc. Psychol. 75 (2018) 11–18.
[73] A. Caliskan, J.J. Bryson, A. Narayanan, Semantics derived automatically from language corpora contain human-like biases, Science 356 (6334) (2017) 183–186.
[74] "Anderson v. Dean," in So. 3d vol. 346, in: La: Court of Appeals, 5th Circuit, 2022, p. 356.
[75] Dist. Court, in: NATIONAL PRODUCTS INC. v. INNOVATIVE INTELLIGENT, PRODUCTS, LLC, WD, Washington, 2021.
[76] "Leegin creative leather products v. PSKS, inc," in US, Supreme Court 551 (2007) 877.
[77] "PARENTS INV. IN COMM. SCH. v. Seattle School," in US, Supreme Court 551 (2007) 701.
[78] "Winkelman v. Parma city school dist," in US vol, Supreme Court 550 (2007) 516.
[79] "Norfolk southern ry. Co. V. Sorrell," in US vol, Supreme Court 549 (2007) 158.