Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Anomaly Detection in Cybersecurity Service Contracts

Pavol Sokol

pavol.sokol@upjs.sk

University of Pavol Jozef Šafárik

Laura Bachňáková Rózenfeldová

University of Pavol Jozef Šafárik

Dávid Varga

University of Pavol Jozef Šafárik

Simona Rudohradská

University of Pavol Jozef Šafárik

Regina Hučková

University of Pavol Jozef Šafárik

Research Article

Keywords: systematic content analysis, machine learning, anomaly detection, cybersecurity, contract

Posted Date: September 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10796774/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Anomaly Detection in Cybersecurity Service Contracts

Abstract

The increasing availability of legal documentation provides opportunities for applying empirical research methods that enable a comprehensive analysis of complex legal data. Systematic Content Analysis (SCA) offers a structured approach for examining legal texts, while recent advances in machine learning allow for the detection of patterns that deviate from expected behaviour. These approaches are also relevant in legal research, as the examination of legal data necessitates the exploration of the established legal practice and identification of possible deviations from it. This paper examines the possibility of anomaly detection in contracts through the employment of machine learning methods. The conclusions presented in this regard are formulated on a basis of a case study focusing on contracts concluded between public authorities and private service providers in the field of cybersecurity. In this regard, we propose an approach to the detection of anomalous contracts, providing a categorization of attributes that impact the determination of anomalies and describing their significance in the selected domain of cybersecurity contracts.

Keywords: systematic content analysis, machine learning, anomaly detection, cybersecurity, contract

1 Introduction

Law requires interpretation. The results achieved through it should be predictable, expecting similar outcomes in similar circumstances. However, the determination of what represents the 'standard' and what deviates from it in legal practice may prove difficult. In some instances, the identification of legal anomalies may be as simple as finding judicial decisions with outcomes differing from the established and accepted decision-making practice. This is, however, not easily done in relation to contracts that allow wider discretion in their formulation, especially if the specific contractual type is not regulated in the applicable legislation. The application of the legal principle of freedom to contract must also be considered, as it provides the contractual parties with the autonomy to decide all aspects of their contracts. In this regard, the contractual parties define themselves the 'standard' in contract formulation as well as the irregularities (anomalies, outliers), reflecting the specific realities of the given sector in which the contracts are concluded.

The increasing availability of legal documentation invites the possibility of implementing different methodological approaches into its interpretation. These include different empirical research methods allowing the comprehensive analysis of complex legal data. In this paper, we employ machine learning to achieve our main objective, which is **the analysis of anomaly detection in legal documentation**, specifically selected contracts concluded in the field of cybersecurity. The implementation of machine learning methods enables the detection of anomalies (outliers), which refers to the problem of finding patterns in data that do not conform to expected behaviour, identifying significant (and often critical) actionable information in a wide variety of application domains (Chandola et al. 2009).

To achieve the stipulated research objective, we specify the following research questions:

- What constitutes an anomaly in cybersecurity service contracts?
- Which attributes contribute to anomalous cybersecurity service contracts and which attributes characterize the standard (non-anomalous) cybersecurity service contracts?

To answer these research questions, we apply unsupervised outlier detection methods to contracts concluded in the field of cybersecurity. In parallel, we apply the method of systematic content analysis (SCA) that allows a structured empirical approach for analysing legal documents. SCA can be defined as a methodical, replicable technique for compressing many words of text into fewer content categories based on explicit coding rules (Weber 1990). It provides objective and scientifically verifiable findings resulting from the analysis of a variety of texts (Salehijam 2018). In comparison to conventional legal scholarship that analyses issues presented in a limited scope of selected cases, SCA works by analysing a larger group of similarly weighted cases to find overall patterns (Hall and Wright 2008). In addition to the analysis of the relevant case law, SCA can also be used to examine other forms of legal documentation covering a variety of legal subject-matters, including contracts.

As our case study, we examine selected legal issues, focusing on notions we consider especially important in contract formulation (determination of the contractual subject-matter, regulation and limitation of parties' liability for damages, determination of sanctions for contract non-performance and protection of information). The selected contracts cover agreements concluded between public authorities and private service providers, the subject-matter of which is the provision of various cybersecurity-related services. This dual perspective - combining outlier detection techniques with SCA – enables the identification of anomalous contracts, exploration of attribute categories influencing their classification, and distinguishing between standard contractual features and those contributing to anomalies.

This paper extends a previously published conference contribution [reference removed for anonymous review]. Compared to the original, it includes significantly enhanced content and broader coverage. The contribution of this paper lies in proposing an approach for detecting anomalous legal texts (contracts). Furthermore, the paper analyses possible categories of attributes relevant for the identification of such

anomalies. Finally, it provides a description of attributes that can be considered significant in contracts related to cybersecurity—both those typically present in standard contracts and those contributing to contractual anomalies.

This paper is organized into six sections. After the introduction, we present the related works in Section 2, focusing on contextual analysis in law and the detection of anomalies in contracts. Section 3 describes the applied methods, data sources, and preprocessing steps. Section 4 reports the overall results of anomaly detection, and presents the SHAP-based attribute analysis. Section 5 provides a detailed discussion of selected contracts. Finally, Section 6 summarizes the findings and outlines challenges and directions for future research.

2 Related works

Legal research has significantly benefited from the inclusion of empirical research methods traditionally employed in different research fields. The increasing employment of these research methods, including systematic content analysis (SCA), in legal research has been discussed in numerous studies (Baldwin and Davis 2003; Monahan and Walker 2011; Leeuw 2015). Hall and Wright (2008) discussed the possibility of content analysis forming a basis for a uniquely legal empirical methodology employed with the objective to analyse a set of legal documents, such as judicial opinions on a particular subject, through their systematic reading, recording of their consistent features and drawing inferences about their use and meaning with the purpose of generating objective, falsifiable, and reproducible knowledge about what courts do and how and why they do it. Oldfather et al. (2012) subsequently supported the incorporation of SCA into legal research, considering this method as reflecting both traditional legal scholarship's attention to texts and empirical legal studies' systematization. Salehijam (2018) explored the utility of applying SCA as a complementary tool in legal research, provided an overview of this method and its various stages and assessed the suitability and necessity of SCA by considering its limitations and benefits. Further studies confirmed the increasing practice of implementing this method in legal research allowing for the examination of larger quantities of legal documents, often focusing on judicial decisions, including the analysis of the case law of the European Court of Human Rights (Gerards et al. 2023), Court of Justice of the European Union (Schrepel 2024), case transcripts of rulings of national courts (Smit et al. 2018), or advocacy briefs submitted to the court (Evans et al. 2007).

Another group of research papers focuses on **the detection of anomalies in the judicial system and legal texts**.

Mukhsimbayev et al. (2020) raised the question of what constitutes an anomaly in the judicial system. They proposed two models for detecting anomalies in judicial practice and compared their effectiveness. Esposito et al. (2023) offer a different perspective on anomalous judicial processes, analysing events that prolong court proceedings using machine learning methods.

The detection of anomalies in legal texts is closely related not only to the application of law but also to the legislative process. Spreeuwenberg et al. (2001) focused on improving the quality of legislation and legal decisions through knowledge verification

and anomaly detection in legislation. An interesting domain of legal texts is asylum decisions. Kaltenhäuser et al. (2025) emphasize that the identification of outliers or anomalies may risk marginalizing individuals labeled as "outliers." They simultaneously call for a debate on designing algorithms with attention not only to accuracy but also to their societal consequences. Another field of anomaly detection in legal texts concerns contracts. Dechtiar et al. (2025) introduced the concept of "contract smells" – patterns signaling anomalies in contractual texts. Authors demonstrated their detection using large language models such as GPT.

Terms of Service (ToS) constitute a prominent application domain for automated contract analysis, particularly for identifying provisions that may disadvantage consumers. The Claudette system (Lippi et al. 2019) introduced supervised sentence-level detection and categorization of potentially unfair clauses in ToS. Subsequent work (Lagioia et al. 2019; Ruggeri et al. 2022) extended this line by employing memory-augmented neural networks that link classified clauses to expert-defined legal rationales, thereby providing natural-language explanations of the predictions. More recently, Panarelli et al. (2025) compared fine-tuned transformer models with different LLM prompting strategies for the same task. Unlike these clause-level approaches, our method applies unsupervised anomaly detection to complete contracts, where an anomaly does not necessarily indicate unfairness or a legal defect.

3 Methodology

This study employs systematic content analysis (SCA) to transform contractual text into a structured representation based on explicitly defined content categories. SCA is a systematic and replicable technique for coding textual material according to explicit coding rules (Weber 1990), reflecting the broader use of empirical methods in legal research (Burns and Hutchinson 2009). The unit of analysis is the complete contract, represented by a feature vector derived from coded contractual provisions and contextual attributes.

The coded attributes of each contract are subsequently analyzed using unsupervised anomaly detection to identify contracts that deviate from common patterns in the dataset. Finally, SHAP is applied to explain the resulting anomaly scores and to determine how individual attributes contribute to the classification of contracts as more typical or anomalous.

The analyzed contracts, coded dataset, source code, and generated results are publicly available in the accompanying repository [reference removed for anonymous review].

3.1 Data collection and dataset

The scope of this study is limited to the analysis of selected legal aspects of contracts concluded between public authorities and private service providers, the object of which is the provision of cybersecurity services. The analysis is based on a dataset of contracts obtained from the Slovak Central Register of Contracts – a public database of contracts that are obligatorily published pursuant to the Act No. 211/2000 Coll. on free access to information (Freedom of Information Act) that currently contains

Attribute	Description	Category	Count
Contract ID	Unique contract identifier	Basic	567
Effective date	Date of entry into force	Basic	567
Number_employees_mean	Employees of authority	Basic	567
Revenue_effective_norm	Authority revenues in year	Basic	567
Number of contractor employees	Employees of contractor	Basic	567
Turnover_effective_norm	Contractor revenues in year	Basic	567
Documentation	Cybersecurity documentation	Subject	169
Trainings	Training provisions	Subject	229
IT_monitoring	IT environment monitoring	Subject	32
Delivery_SW_HW	Delivery of SW/HW	Subject	36
Incident	Incident handling service	Subject	285
Consulting_services	Consultancy services	Subject	399
CISO_law	Cybersecurity manager	Subject	393
Security_audit_Act_29	Self-assessment/audit	Subject	86
Audit_law	Cybersecurity Act audit	Subject	46
Fixed_term	Contract with fixed term	Subject	303
Indef_term	Indefinite term contract	Subject	264
Payment_one_time	One-off payment	Subject	197
Payment_monthly	Monthly payment	Subject	407
Liability_limit_amount	Liability limit by amount	Liability	90
Limit_proven_damage	Diligence not limited	Liability	78
Failure_cooperate	Lack of cooperation	Liability	89
Failure_apply_sec_measures	Missing sec. measures	Liability	87
Incorrect_docs	Incorrect documentation	Liability	74
Other liability aspects	Other liability rules	Liability	139
Circumstance_excluding_liability	Excluding circumstances	Liability	64
Contract_penalty_agreed	Contractual penalty	Penalty	182
Penalty_private_nonperf	Non-performance penalty	Penalty	118
Penalty_state_delay_pay	Delay penalty (public)	Penalty	50
Penalty_other_breach	Other breaches penalty	Penalty	68
Priv_supplier_can_withdraw	Withdrawal right	Withdrawal	480
Material_breach_defined	Material breach	Withdrawal	111
Confidentiality_clause	Confidentiality clause	Confid.	516
GDPR_clause_present	GDPR/data protection	Confid.	252

Table 1 Contractual attributes with short descriptions, categories, and counts (number of values of 1 in binary data).

more than 5 million published contracts. The selection of contracts that formed the basis for our analysis was determined on the following basis: (1) inclusion of specific words in the contract title, specifically the use of the Slovak equivalents of the terms 'cyber' or 'cybersecurity' or their abbreviation 'KB'; and (2) determination of the service providers operating on the national market that included the specific suppliers (identified on the basis of the contracts from the previous point). The authors have analysed 567 contracts concluded in the period 23 July 2019 – 6 November 2025. The dataset created and analysed during the current study is available in the Mendeley Data repository [reference removed for anonymous review].

As the research presented is based on the analysis of contracts published in the Slovak Central Register of Contracts, this constitutes a natural methodological delimitation of the examined sample. Although extending the dataset to include

business-to-business (B2B) contracts concluded between private-sector entities would provide a more comprehensive insight into contractual practice in the field of cybersecurity, such an extension is not realistically feasible. Given the sensitive nature of contractual relationships and the information contained in these contracts, it cannot reasonably be expected that private entities would voluntarily disclose such contracts for research purposes. Consequently, the resulting dataset is primarily intended for non-commercial research, the analysis of contractual practices within the public sector, and the support of further scholarly research on the legal regulation of cybersecurity.

3.2 Data coding and attributes

The objective of the data coding stage was to represent each complete contract using a common set of structured attributes. The attributes were identified by legal experts during the manual review of the collected contracts. The contracts were subsequently coded manually using shared attribute definitions to ensure a consistent interpretation of the examined contractual provisions.

The coded dataset contained numerical and binary attributes, together with one nominal categorical attribute representing the type of contracting authority. Numerical attributes captured quantitative characteristics of the contracts or the contracting parties, such as prices, revenues, and employee counts. Binary attributes indicated the presence or absence of a particular contractual feature, with 1 representing its presence and 0 its absence. Accordingly, all contracts were represented using the same attribute structure, even when a particular contractual provision was not present. The nominal attribute describing the type of contracting authority was transformed into numerical features during the subsequent preprocessing stage.

The attributes were organized into six groups according to the contractual or contextual aspect they represented. Table 1 presents the attributes, their short descriptions, categories, and numbers of occurrences in the dataset. Missing numerical values and the transformation of the coded attributes into the final feature matrix are described in Section 3.3.

3.3 Data preprocessing

Before applying the anomaly-detection methods, the coded data were transformed into a numerical feature matrix. Most contractual attributes were binary and were encoded as 1 when the corresponding feature was present and as 0 when it was absent. The only nominal categorical attribute was the type of contracting authority. This attribute was transformed using one-hot encoding into seven binary features representing municipalities, state organizations, healthcare facilities, commercial companies, schools, professional chambers, and social care facilities.

The originally textual contract-duration attribute was converted into a single binary feature, with 1 representing a fixed-term contract and 0 representing an indefinite-term contract. The contract identifier and the effective date were retained as metadata for presenting and interpreting the results but were not included among the input features of the anomaly-detection model.

Employee counts were originally provided as intervals. Each interval was replaced by its midpoint; for example, the interval 10–19 employees was represented by the value 14.5. Missing values in four continuous attributes—the numbers of employees of the contracting authority and the supplier, the authority's revenue, and the supplier's turnover—were replaced by the median of the corresponding attribute calculated from the available observations. For the employee-count attributes, zero values were also treated as missing. Because median imputation would otherwise make imputed and originally observed values indistinguishable, a separate binary indicator was retained for the supplier's employee count. This feature records whether the original value was missing and subsequently imputed.

Monetary and percentage values were converted from their original textual representations into numerical values. Several continuous monetary and organizational attributes exhibited strongly right-skewed distributions. Skewness was quantified using the Fisher–Pearson coefficient of skewness (Joanes and Gill 1998), where values close to zero indicate an approximately symmetric distribution and positive values indicate right skewness. For these attributes, most contracts had relatively low or moderate values, whereas a small number of observations had substantially higher values, forming a long tail.

To reduce the dominance of extreme values in the continuous attributes, a signed `log1p` transformation was applied to variables with pronounced skewness. This transformation compresses large values while preserving their order, preventing a small number of extreme observations from dominating the anomaly scores. The transformed continuous attributes were subsequently scaled using `RobustScaler` (Pedregosa et al. 2011), as they represented quantities on different numerical scales, such as prices, revenues, employee counts, and percentages. This prevented attributes with large numerical ranges from dominating attributes measured on smaller scales during anomaly detection. Binary attributes, including one-hot encoded attributes, were not scaled.

Figure 1 illustrates this preprocessing procedure using the annual contract price including VAT. Before preprocessing, a small number of contracts had substantially higher annual prices than the majority of observations. After the signed `log1p` transformation and robust scaling, the long upper tail was compressed while the ordering of the contracts was preserved.

The resulting dataset contained 40 features for each of the 567 contracts. The scaled feature matrix was used in the subsequent analyses, while an aligned unscaled version was retained for displaying the original feature values in the explanatory visualizations.

3.4 Anomaly detection methods

Since no ground-truth labels identifying anomalous contracts were available, the task was formulated as unsupervised anomaly detection. Let $\mathcal{C} = \{C_1, C_2, \ldots, C_n\}$ denote the set of contracts. Following the coding and preprocessing stages, each contract C_i is represented by a numerical feature vector $\mathbf{x}_i \in \mathbb{R}^m$.

The anomaly-detection model assigns each contract an anomaly score

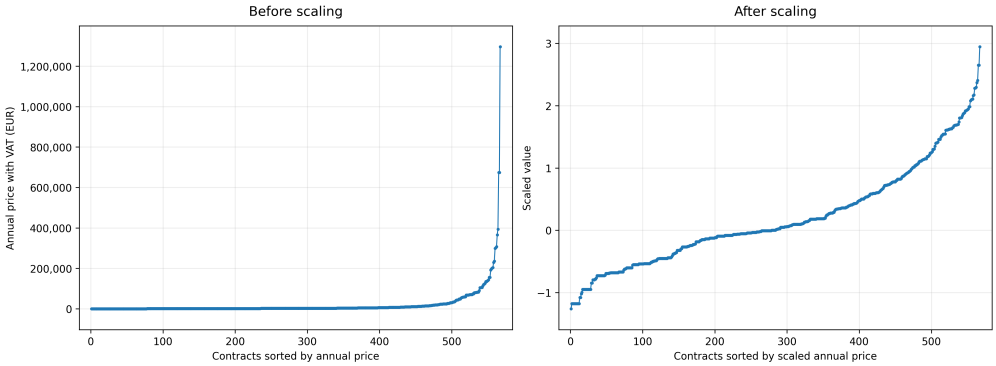


Fig. 1 Distribution of the annual contract price before preprocessing and after logarithmic transformation and robust scaling. Contracts are ordered according to the displayed feature values

$$f(\mathbf{x}_i) \in \mathbb{R}, \tag{1}$$

where higher values indicate a stronger deviation from the patterns observed in the dataset. A contract is designated as anomalous when its score exceeds the threshold τ determined by the assumed contamination level:

$$C_i \text{ is anomalous} \iff f(\mathbf{x}_i) > \tau. \tag{2}$$

Rather than relying on a single anomaly detection method, we adopted an ensemble approach. Individual unsupervised detectors rely on different assumptions about the structure of anomalous observations and may therefore produce substantially different rankings on the same dataset. We first compared a set of candidate detectors and selected a subset that was subsequently combined using the LSCP dynamic ensembleZhao et al. (2019).

3.4.1 Selection of base detectors

Because no ground-truth labels identifying anomalous contracts were available, the candidate detectors could not be compared in terms of predictive accuracy. They were therefore evaluated using two complementary criteria:

- **Agreement** examines whether different detectors identify the same contracts among the most anomalous observations. Higher agreement indicates that the detectors capture a shared anomaly signal, whereas lower agreement suggests that they may respond to different types of deviations. Agreement alone, however, does not demonstrate that the identified contracts are true anomalies.
- **Resampling stability** examines whether a detector produces similar top-ranked anomalies when fitted to slightly different subsets of the dataset. Higher stability indicates that the detector is less sensitive to small changes in the analyzed sample.

The candidate set consisted of PCA (Shyu et al. 2006), ECOD (Li et al. 2022), COPOD (Li et al. 2020), Isolation Forest (Liu et al. 2012), HBOS (Goldstein and

Dengel 2012), KNN (Angiulli and Pizzuti 2002), One-Class SVM (Schölkopf et al. 2001), CBLOF (He et al. 2003), and LODA (Pevný 2016).

Agreement between detectors was evaluated using their sets of the $k = 40$ highest-scoring contracts. The resulting top-k sets were compared using the Jaccard coefficient (Jaccard 1901):

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}, \tag{3}$$

where A and B denote the top-k sets produced by two detectors. Higher values indicate that the detectors identify a larger proportion of the same contracts as anomalous.

Detector stability was evaluated through repeated resampling of the dataset. For each resampled subset, the detector was refitted and its top-k contracts were identified. The mean pairwise Jaccard similarity between the resulting top-k sets was then used as the stability measure. Higher values indicate that small changes in the analyzed sample have a smaller effect on the contracts identified as anomalous. Agreement and stability therefore measure the consistency of the detectors rather than their predictive accuracy.

The level of agreement varied considerably across detector pairs. The strongest pairwise overlaps were observed between ECOD and COPOD, with a Jaccard coefficient of 0.74, and between Isolation Forest and KNN, with a coefficient of 0.63. PCA also exhibited moderate to strong agreement with ECOD and COPOD, whereas HBOS showed an overlap of approximately 0.60 with these detectors.

In the resampling analysis, PCA, ECOD, COPOD, and HBOS achieved the highest stability values, ranging approximately from 0.32 to 0.35. KNN and Isolation Forest achieved values of approximately 0.27–0.28, whereas One-Class SVM, CBLOF, and LODA showed lower stability values of 0.18, 0.11, and 0.06, respectively.

Based on these results, six detectors were retained for the ensemble: PCA, ECOD, COPOD, Isolation Forest, HBOS, and KNN.

3.4.2 LSCP ensemble

The selected detectors were combined using Locally Selective Combination in Parallel (LSCP), an unsupervised dynamic ensemble method (Zhao et al. 2019). Unlike static ensembles, LSCP may use a different subset of detectors for each contract. For the evaluated contract, it first identifies a local neighbourhood of contracts with similar feature values.

For each neighbouring contract, the highest standardized score produced by the base detectors is used to form a pseudo-target. Each detector is then evaluated according to the Pearson correlation between its scores in this neighbourhood and the pseudo-target.

The resulting correlations are divided into n_bins intervals. Detectors from the most populated interval are selected, and their standardized scores for the evaluated contract are averaged to produce its final anomaly score.

The LSCP ensemble used $n_bins = 4$ to avoid excessive fragmentation when selecting among six base detectors. The contamination parameter was set to 0.05,

corresponding to approximately the 5% highest-scoring contracts being designated as anomalous. All remaining parameters were left at their default values.

3.5 Explainability

To explain the contribution of individual attributes to the anomaly scores, we employed SHapley Additive exPlanations (SHAP) (Lundberg and Lee 2017). For each contract C_i, SHAP decomposes the LSCP anomaly score as

$$f(\mathbf{x}_i) = \mathbb{E}[f(X)] + \sum_{j=1}^{m} \phi_j(\mathbf{x}_i), \tag{4}$$

where $\mathbb{E}[f(X)]$ is the baseline anomaly score determined from the SHAP background data, and $\phi_j(\mathbf{x}_i)$ is the contribution of feature j to the anomaly score of contract C_i.

A positive SHAP value indicates that the corresponding feature shifts the anomaly score above the baseline, whereas a negative SHAP value shifts it below the baseline. SHAP therefore explains the continuous anomaly score $f(\mathbf{x}_i)$, not the binary designation obtained by comparing this score with the threshold τ.

Because LSCP does not provide a model-specific SHAP explainer, its decision function was explained using the model-agnostic *SamplingExplainer*. The background dataset was summarized by 30 k-means centroids, which provided a compact representation of the feature distribution and reduced the computational cost of the explanation procedure. SHAP values were computed for all 567 contracts using an approximation budget of 2,000 samples for each explained contract.

Both LSCP and the SHAP explainer operated on the preprocessed and scaled feature matrix. However, the corresponding unscaled feature values were retained separately and supplied as display values in the SHAP visualizations. Consequently, the SHAP contributions reflect the behaviour of the model on scaled inputs, while the values displayed in the plots remain interpretable in their unscaled form and, where applicable, in their original units.

Since *SamplingExplainer* approximates SHAP values through sampling, the additive relationship may be subject to minor numerical approximation error. The resulting explanations describe how individual features contribute to the model's anomaly score and should not be interpreted as causal effects or as direct evidence that a contract contains a legal defect.

4 Results

After running *SamplingExplainer* on the LSCP trained model, we obtained per-feature attributions that can be visualized in two complementary ways. Global views (e.g., beeswarm plot in Figure 3) show how all attributes contribute across the dataset, highlighting which features consistently drive anomaly detection.

Local views (e.g., waterfall plot in Figure 6) explain individual contracts, revealing how specific attributes push a given case toward or away from being anomalous. These explanations allow us not only to detect outliers but also to understand the underlying reasons for their anomalous character.

Before turning to feature-level explanations, it is useful to examine how LSCP scores distribute over time. Figure 2 shows the distribution of all analyzed contracts according to their effective dates (x-axis). The y-axis represents the resulting LSCP anomaly score. The top 5% most anomalous contracts (29 in total) are highlighted in orange; in the next section, we examine which attributes drive such contracts to appear anomalous.

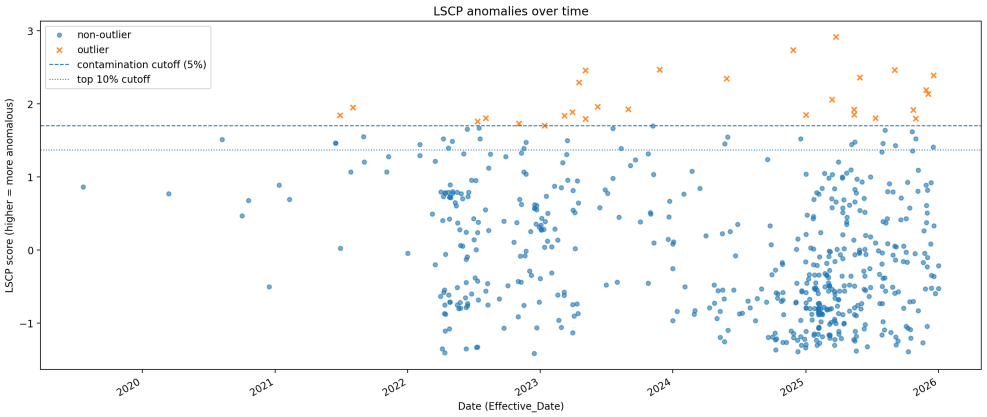


Fig. 2 Distribution of contracts by effective date (x-axis) and LSCP value (y-axis). The top 5% most anomalous contracts (29 in total) are highlighted in orange

4.1 Determination of contract anomality

A visual representation of all attributes is shown in Figure 3 in the form of a beeswarm plot. In this visualization, each dot corresponds to one (row, feature) pair. The x-axis represents the SHAP value $\phi_j(x)$, indicating how much a given feature shifts the LSCP score either upward (to the right, more anomalous) or downward (to the left, more normal). The color encodes the raw feature value for that instance (red = high value, blue = low value), while vertical jitter reflects the density of points.

This plot allows us to observe which features generally increase the anomaly score (right tail) or decrease it (left tail), and whether high or low feature values are responsible for these shifts.

Contracts selected for analysis and the collection of attributes determining the different categories of contractual obligations in chosen areas (determination of the contractual subject-matter, regulation and limitation of parties' liability for damages, determination of sanctions for contract non-performance and protection of information) establish a basis for the identification of what can be considered 'standard' in contractual regulation in the cybersecurity sector, and, correspondingly, which provisions may be classified as atypical or anomalous.

The analysis of the individual attributes representing the above-specified categories of contractual provisions allowed the identification of how these attributes and their

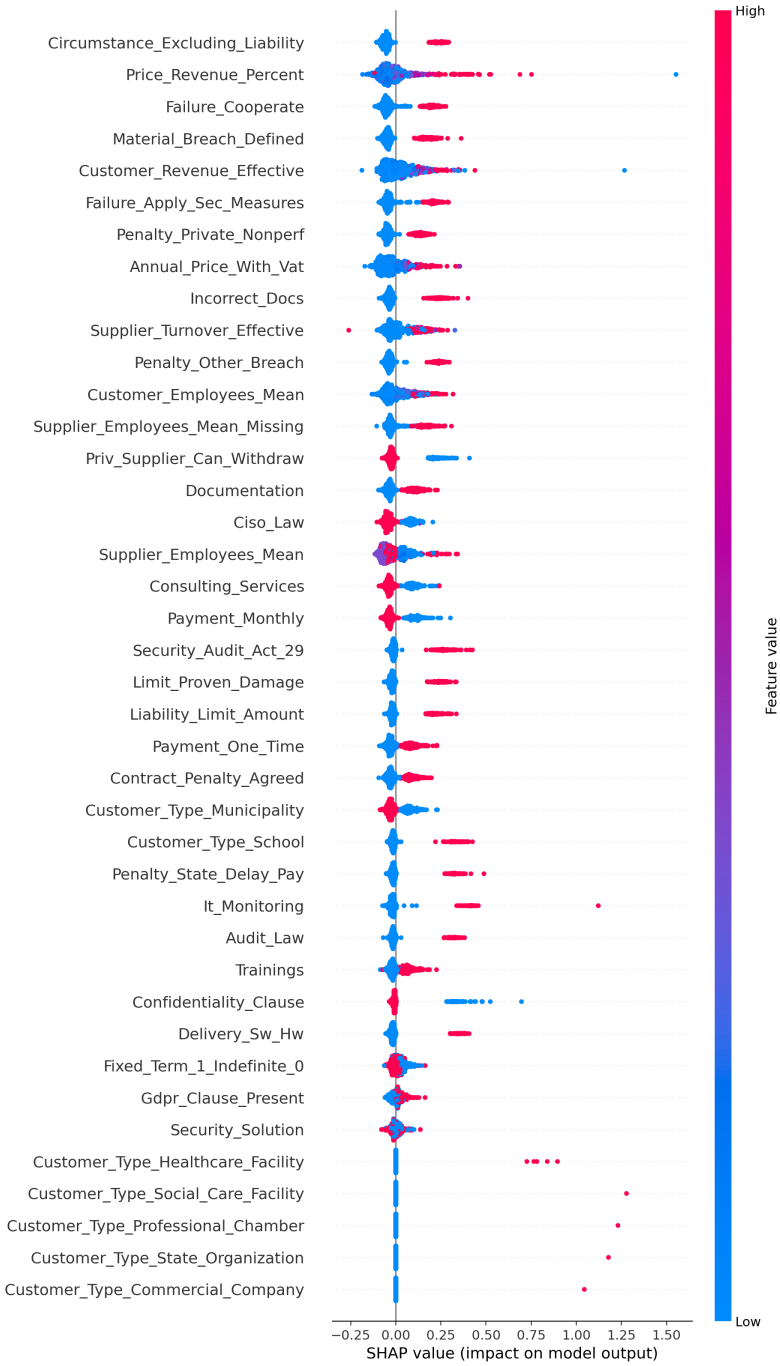


Fig. 3 SHAP beeswarm plot for all attributes. Each dot represents one feature value, with the x-axis showing its contribution to the LSCP score (red = high value, blue = low value)

presence or absence in the contracts contributed to the determination of contract anomality. In this regard, three categories of attributes could be distinguished, namely:

1. attributes that contribute to the contract normality;
2. attributes that contribute to the contract anomality; and
3. attributes without a significant effect on contract classification.

The attributes commonly contained in contracts, the presence of which therefore adds to contract normality, include the following attributes:

- 'Customer_Type_Municipality', that confirms that subjects that regularly outsource the provision of cybersecurity services are different municipalities that lack the necessary personnel that could ensure the provision of such services within their organisation;
- 'Payment_Monthly' that corresponds to the obligation of the service recipient to provide remuneration for services provided in regular monthly instalments;
- 'Ciso_Law' and 'Consulting_Services', which specify the nature of services that regularly form the subject-matter of contracts concluded in the field of cybersecurity, which include the provision of consulting services and the service of cybersecurity manager;
- 'Priv_Supplier_Can_Withdraw' that confirms the established protection of private service providers through their right to withdraw from the contract in case of its non-performance; and
- 'Confidentiality_Clause', reflecting the inherent need to protect the confidentiality of information provided between contractual parties within the scope of cybersecurity contracts.

Correspondingly, the absence of the above-defined attributes adds to contract anomality.

On the other hand, the attributes, the presence of which aids to contract anomality, include attributes that relate to:

- the nature of service recipients, where the attribute 'Customer_Type_School' identifies the least common recipient entering into the examined contracts;
- the periodicity of payment, where the attribute 'Payment_One_Time' allows for the identification of the least-common form of payment for services rendered;
- the specification of cybersecurity services provided, where the provision of information and cybersecurity documentation ('Documentation'), employee training ('Trainings'), monitoring of IT environment ('IT_Monitoring'), delivery of software or hardware ('Delivery_Sw_Hw'), security or external audit pursuant to applicable cybersecurity regulation ('Security_Audit_Act_29') adds to contract anomality;
- the limitation of liability, where the inclusion of any contractual provisions that restrict liability, e.g. to the specific pecuniary amount ('Liability_Limit_Amount') or to the amount of demonstrably caused damage ('Limit_Proven_Damage'), or exclude liability in defined circumstances ('Circumstance_Excluding_Liability') or due to the contractual party's failure to cooperate ('Failure_Cooperate'), provide correct documentation ('Incorrect_Docs') or act in accordance with the recommendations of the supplier and adopt the necessary security measures ('Failure_Apply_Sec_Measures'),

can be evaluated as contributing to the classification of contract in question as anomalous;

- sanctions that may be imposed in case of contractual breach, where, similarly, most attributes in this category including the stipulation of a contractual penalty in the contract as such ('Contract_penalty_agreed'), or inclusion of contractual penalties for the supplier's failure to perform the contract ('Penalty_private_non performance'), delay in remuneration payment ('Penalty_state_delay_pay') or for the infringement of other contractual obligations ('Penalty_other_breach') all contribute to contract anomality, and
- the protection of personal data ('GDPR_Clause_Present'), where the inclusion of such provisions in the contracts similarly adds to contract anomality. Given the nature of cybersecurity services which often focus on systems containing personal data, the absence of personal data protection provisions in contracts is surprising.

The last category of attributes with no discernible impact on contract classification include attributes such as the provision of the service of security solution, if included in the contractual subject-matter. The dataset also includes continuous attributes (e. g. 'Supplier_Turnover_Effective', 'Customer_Revenue_Effective'), the objective of which is to provide contextual information on the contractual parties that are not included in the provisions of the contracts examined.

4.2 Results for the most anomalous contracts

In this section, we analyse the most anomalous contracts,i.e. the set of 29 contracts that, based on the analysis, exhibited the highest degree of anomaly is characterised by several common features (top 5% contracts).

A visual representation of all attributes in the most anomalous contracts is shown in Figure 4 in the form of a beeswarm plot. In this visualization, each dot corresponds to one (row, feature) pair. The x-axis represents the SHAP value $\phi_j(x)$, indicating how much a given feature shifts the LSCP score either upward (to the right, more anomalous) or downward (to the left, more normal). The color encodes the raw feature value for that instance (red = high value, blue = low value), while vertical jitter reflects the density of points.

For the sake of clarity, these factors may be categorised into three groups. The first group concerns the nature of the contracting party on the side of the entities to which the services were provided. Upon examination of the content of the contracts, it is apparent that, within this sample of 29 contracts, healthcare institutions (primarily hospitals) are particularly significantly represented, as are territorial self-government units (regions), larger cities, ministries, and large state-owned enterprises. With regard to the entities included in this group within the analysed sample of the most anomalous contracts, it is evident that these are economically significant and strategic entities.

The second group of common features of these contracts is represented by the fact that the overall wording of the contract is complex and includes detailed regulation of the contractual relationship between the service provider and the entity to which the service is provided. In almost all cases, these contracts contain highly detailed provisions governing mutual rights and obligations, issues of liability, warranties, support

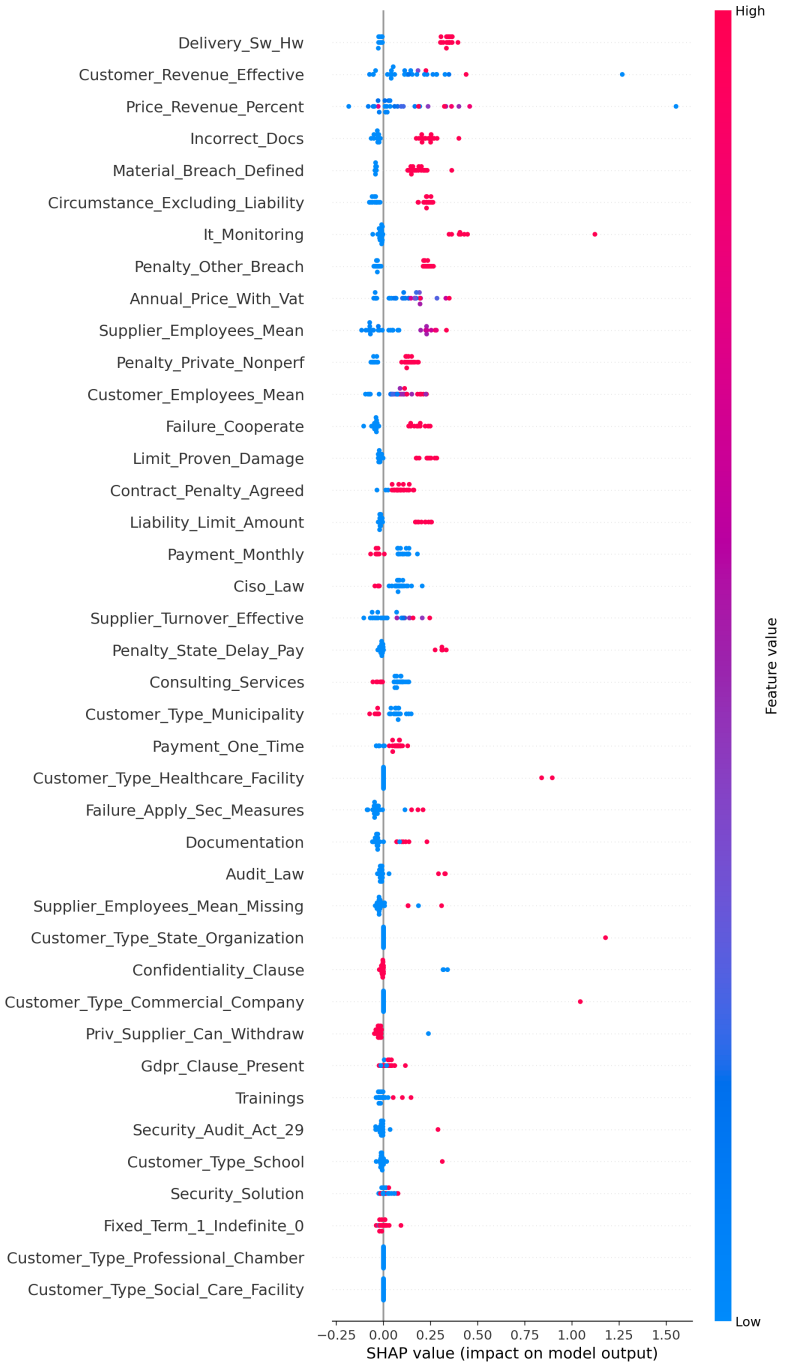


Fig. 4 SHAP beeswarm plot for all attributes in the most anomalous (top 5%) contracts. Each dot represents one feature value, with the x-axis showing its contribution to the LSCP score (red = high value, blue = low value)

and servicing, sanctions, as well as a detailed description of the technical parameters of the service provided.

The third category relates to the fact that the same observed attributes did not emerge as anomalous across these contracts; rather, each individual contract contains a specific combination of anomalies that renders it anomalous. It therefore cannot be stated with certainty that, within this particular sample, these specific attributes will, under all circumstances, contribute to the anomalous nature of a contract.

The analysed sample included one contract which, by its nature, did not appear to exhibit the characteristics of a large and significant entity on the recipient's side, nor the characteristic of complexity in the regulation of the contractual relationship. A relatively interesting finding is that the model identified the ratio of the price/value of the contract to the revenue of the entity to which the services were provided as the most anomalous attribute (+1.55). However, the value of the service cannot be regarded as exceeding a reasonable economic framework, as it amounts to EUR 35 per month excluding VAT. This finding may be explained by the fact that the municipality concerned recorded a negative financial result in the relevant year, which affected the assessment of the value of the service and produced the second most anomalous attribute for this particular contract, namely the negative value of the revenue of the entity to which the services were provided in the year in which the contract became effective (+1.27).

4.3 Results for the most normal contracts

This section examines the contracts with the lowest anomaly scores. In presenting the sample of the 29 most normal contracts (bottom 5% contracts), we likewise apply the categorisation into three groups used for the most anomalous contracts – namely, the nature of the entities on the side of the service recipient, the nature of the contract, and the common attributes exhibiting normality.

A visual representation of all attributes in the most normal contracts is shown in Figure 5 in the form of a beeswarm plot. In this visualization, each dot corresponds to one (row, feature) pair. The x-axis represents the SHAP value $\phi_j(x)$, indicating how much a given feature shifts the LSCP score either upward (to the right, more anomalous) or downward (to the left, more normal). The color encodes the raw feature value for that instance (red = high value, blue = low value), while vertical jitter reflects the density of points.

As regards the nature of the entities, these are primarily smaller municipalities, smaller social service facilities, and small towns. From an economic perspective, these are therefore less prominent or significant entities than those represented in the sample of the most anomalous contracts.

With regard to the nature of the individual contracts, the situation is precisely the opposite of that observed in the case of the most anomalous contracts. As a rule, these are simple contracts regulating only the essential elements, while the elaboration of mutual rights and obligations, sanctions, or liability provisions tends to be relatively limited and concise, with the contractual relationship being governed primarily by the applicable statutory framework.

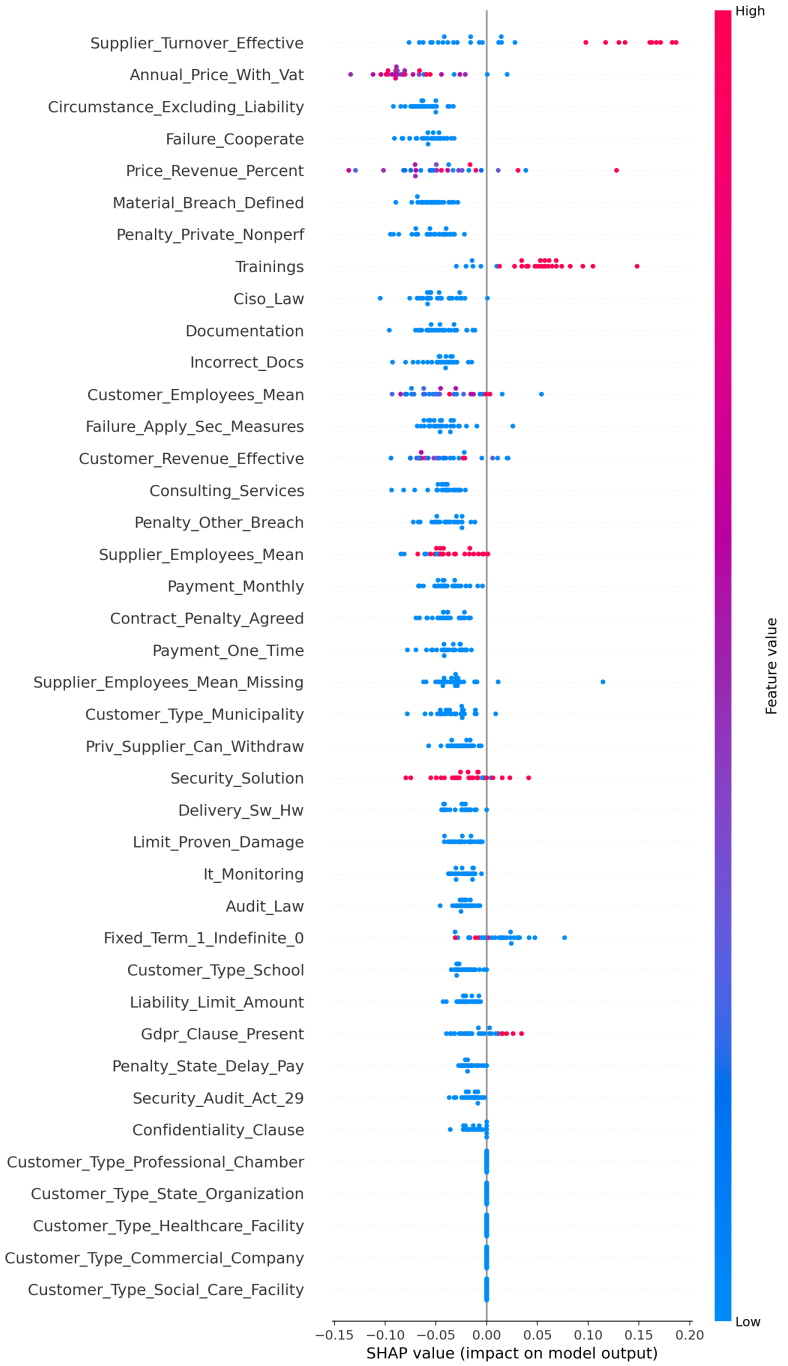


Fig. 5 SHAP beeswarm plot for all attributes in the most normal (bottom 5%) contracts. Each dot represents one feature value, with the x-axis showing its contribution to the LSCP score (red = high value, blue = low value)

Likewise, it is not possible to determine unequivocally which specific attributes within the analysed sample guarantee the normality of a contract, since these attributes must be assessed individually in relation to each particular contract, as they are mutually interconnected and may, and generally will, influence one another.

4.4 Results of regarding selected contracts

As part of the discussion, we focused on specific contracts that exhibit the highest anomaly. These include the contract with ID 10597488 (Figure 6) and the contract with ID 10080768 (Figure 7). On the other hand, we also examined contracts that can be considered standard (their anomaly level is the lowest within the analyzed set of contracts). These are the contracts with ID 7233005 (Figure 8) and the contract with ID 6332291 (Figure 9). The individual figures show the SHAP explanation for selected contracts. The plots illustrate the contribution of individual attributes to the model output. The baseline prediction corresponds to the expected value of the model $\mathbb{E}[f(X)] = -0.239$. Positive values (red colour) increase the anomaly score, while negative values (blue colour) decrease it. In the following text, we provide a more detailed analysis of these contracts.

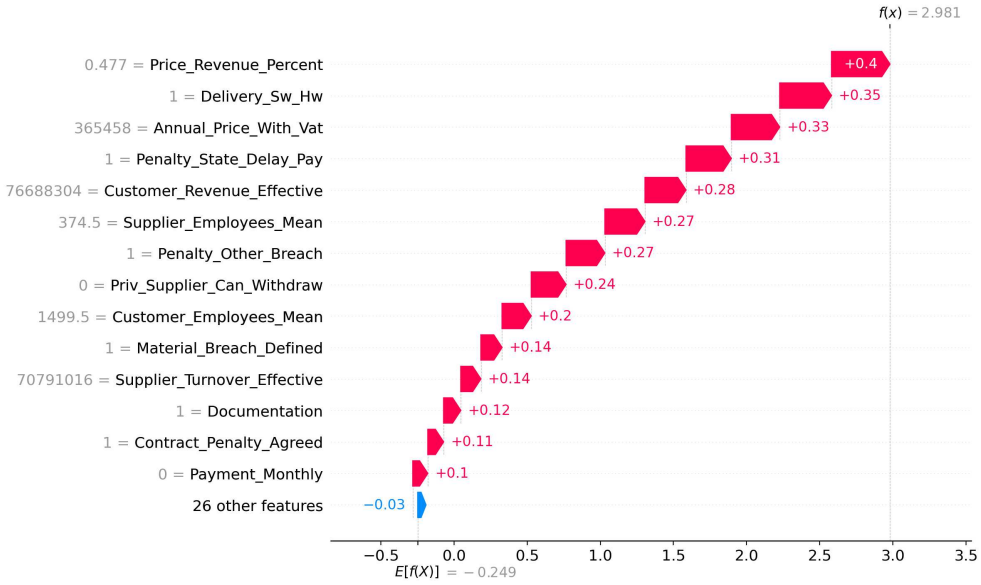


Fig. 6 SHAP explanation for contract ID 10597488. The plot illustrates the contribution of individual attributes to the model output

The contract identified in the dataset as ID 10597488 was, based on the results of the analysis, evaluated as the most anomalous contract in the analyzed set. It was concluded between a city and a commercial company providing cybersecurity services. The contract was concluded pursuant to Section 536 of the Commercial Code, i.e.,

as a contract for work. The work was defined in the contract as "Raising the level of cyber and information security of the City of XYZ."

The mean value, indicated in the lower center part of the chart as E[f(X)] = –0.249, represents the baseline value from which the contribution of the individual attributes to the resulting degree of the contract's anomalousness is derived. Attributes shown in red increase the resulting value and thus contribute to the contract's anomalousness, whereas attributes shown in blue act in the opposite direction and contribute to its normality. As this is the most anomalous contract in the dataset, the visualization is dominated almost exclusively by attributes with a positive contribution value, shown in red, which raise the overall result to f(x) = 2.981, shown at the top right of the chart.

In the case analyzed, the most significant positive influence came from the high share of the contract price in the supplier's turnover (Price_Revenue_Percent, +0.4), the high annual value of the contract including VAT (Annual_Price_With_VAT, +0.33), and the fact that the subject matter of the contract was the delivery of software or hardware (Delivery_Sw_Hw, +0.35). The degree of anomalousness was further increased by the existence of a contractual penalty against the public-administration entity for delay in payment for the subject matter (Penalty_State_Delay_Pay, +0.30), the high average number of the supplier's employees (Supplier_Employees_Mean, +0.27), the existence of a contractual penalty for breach of other contractual obligations (Penalty_Other_Breach, +0.26), and the fact that the supplier had no contractually established right to withdraw from the contract (Priv_Supplier_Can_Withdraw, +0.22). The result was also influenced by the high revenue of the customer (Customer_Revenue_Effective, +0.20), the absence of monthly payments (Payment_Monthly, +0.16), the definition of material breach of contract (Material_Breach_Defined, +0.14), the existence of documentation obligations (Documentation, +0.14), the high number of the customer's employees (Customer_Employees_Mean, +0.14), the absence of consulting services (Consulting_Services, +0.14), and the supplier's turnover (Supplier_Turnover_Effective, +0.13). Conversely, the remaining 26 variables had only a negligible effect on the resulting prediction value (–0.03)—that is, on what might contribute to the contract's normality.

The classification of the contract as the most anomalous follows primarily from the fact that it is an economically significant contract with a high value of the subject matter, which at the same time represents a substantial share of the supplier's turnover. The contract also differs from the contracts commonly occurring in the analyzed dataset in that it contains a more extensive regulation of contractual rights and obligations, including several sanction mechanisms, documentation obligations, and a definition of material breach, while the subject of performance was the delivery of a software or hardware solution rather than the mere provision of consulting services. These attributes can be explained by the fact that, in financially significant contracts with a more complex subject of performance, the parties naturally tend to regulate their mutual rights and obligations—as well as security and sanction mechanisms—in greater detail, which the model, given their lower incidence in the analyzed dataset, evaluated as anomalous characteristics of the contract.

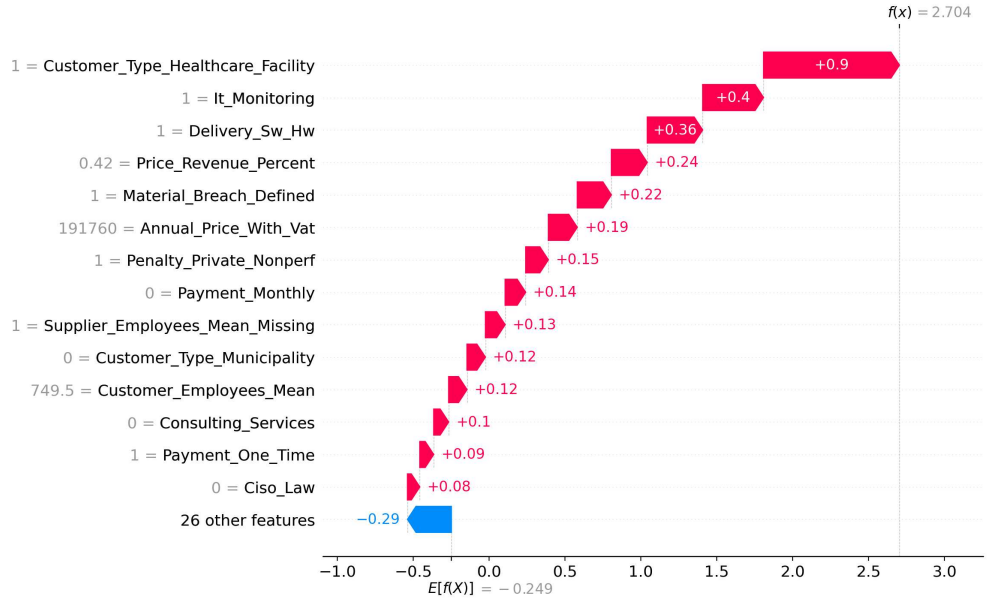


Fig. 7 SHAP explanation for contract ID 10080768. The plot illustrates the contribution of individual attributes to the model output

The contract identified in the dataset as ID 10080768 was, based on the results of the analysis, evaluated as the second most anomalous contract in the analyzed set. It was concluded between a hospital and a commercial company providing cybersecurity services. The legal basis for concluding the contract was Section 269(2) of the Commercial Code, which refers to the legal framework governing innominate contracts. The contract was designated as a Service Agreement, specifically for the "Deployment of security information and event management."

The mean value, indicated in the lower center part of the chart as E[f(X)] = –0.249, represents the baseline value from which the contribution of the individual attributes to the resulting degree of the contract's anomalousness is derived. Attributes shown in red increase the overall result and thus contribute to the contract's anomalousness, whereas attributes shown in blue act in the opposite direction and contribute to its normality. Unlike the most anomalous contract, the visual representation of the results also includes an attribute contributing to the contract's normality, shown in blue (Supplier_Employees_Mean), in addition to the remaining 26 attributes contributing to normality, albeit in a rather negligible manner. As this is the second most anomalous contract in the dataset, the chart is dominated by attributes with a positive contribution, which raise the overall result to f(x) = 2.704, shown at the top right of the chart.

In the case analyzed, the most significant positive influence came from the facts that the customer was a healthcare facility (Customer_Type_Healthcare_Facility, +0.9), that the subject matter of the contract was IT monitoring (It_Monitoring,

+0.42), and that the contract included the delivery of software or hardware (Delivery_Sw_Hw, +0.36). The degree of anomalousness was further increased by the existence of a contractual penalty for the supplier's non-performance (Penalty_Private_Nonperf), the definition of material breach of contract (Material_Breach_Defined), the one-time form of payment (Payment_One_Time), and the overall value of the subject matter (Annual_Price_With_VAT). Conversely, the average number of the supplier's employees (Supplier_Employees_Mean, –0.08), together with the combined effect of the remaining 26 variables, slightly lowered the resulting value.

The classification of the contract as the second most anomalous follows primarily from a specific combination of attributes that occurred less frequently in the analyzed dataset. Particularly significant were the facts that the customer was a healthcare facility, that the subject of performance was the deployment of an IT-monitoring system (SIEM), and that the contract included the delivery of a software or hardware solution—that is, a technically and organizationally more demanding subject of performance than was typical for most of the analyzed contracts. These circumstances also justify the more detailed regulation of contractual rights and obligations, including security mechanisms such as contractual penalties or the definition of material breach. Although the model evaluated the presence of these attributes as anomalous characteristics owing to their lower incidence in the dataset, from a substantive standpoint they can be regarded as a natural consequence of the greater complexity of the contract and the heightened cybersecurity requirements in the healthcare sector.

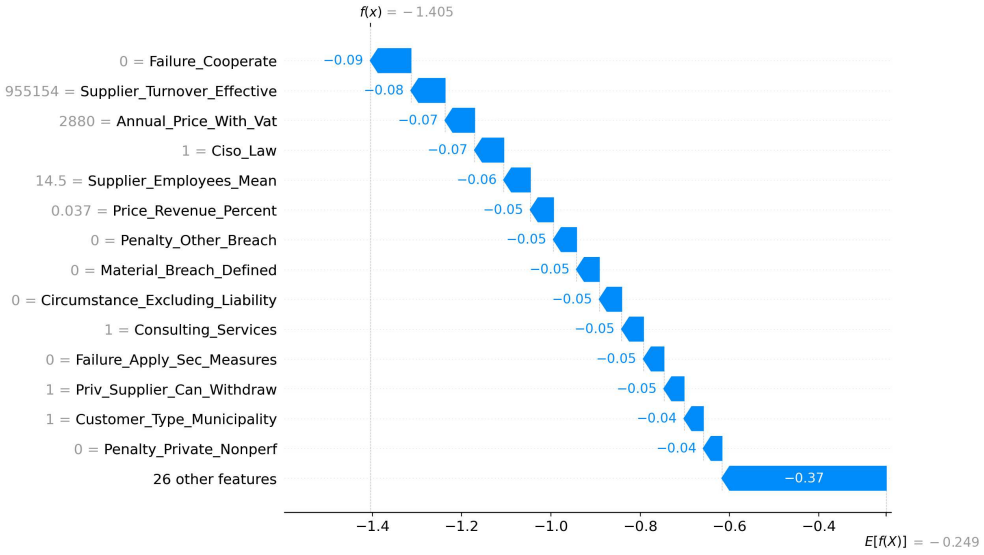


Fig. 8 SHAP explanation for contract ID 7233005. The plot illustrates the contribution of individual attributes to the model output

The contract identified in the dataset as ID 7233005 was, based on the results of the analysis, evaluated as the most normal contract in the analyzed set. It was concluded between a city and a commercial company operating in the field of cybersecurity services. The parties entered into an innominate contract pursuant to Section 269(1) of the Commercial Code, in conjunction with the provisions of Act No. 69/2018 Coll. on Cybersecurity and Decree No. 362/2018 Coll., which establishes the content and scope of security measures. The subject matter of the contract was the provision of consulting services in the implementation of cybersecurity.

The mean value, indicated in the lower right part of the chart as $E[f(X)] = -0.249$, represents the baseline value from which the contribution of the individual attributes to the resulting degree of the contract's anomalousness—or, conversely, its normality—is derived. Attributes shown in blue lower the resulting anomaly value and thus contribute to the contract's normality, whereas attributes shown in red act in the opposite direction and contribute to its anomalousness. As this is the most normal contract in the analyzed dataset, the visualization is dominated by attributes with negative values, which reduce the overall result to $f(x) = -1.405$, shown at the top center of the chart.

In the case analyzed, the attributes with the most significant influence on classifying the contract as the most normal were the low annual value of the subject matter including VAT (Annual_Price_With_VAT, –0.07), the low share of the contract price in the supplier's turnover (Price_Revenue_Percent, –0.07), the lower average number of the supplier's employees (Supplier_Employees_Mean, –0.07), the absence of a definition of material breach of contract (Material_Breach_Defined, –0.07), and the absence of a provision governing breach of the duty to cooperate in the performance of the contract (Failure_Cooperate, –0.07). The degree of anomalousness was further reduced by the provision of consulting services (Consulting_Services, –0.06), the absence of a provision governing breach of duties in the application of security measures (Failure_Apply_Sec_Measures, –0.05), the absence of a contractual penalty for breach of other contractual obligations (Penalty_Other_Breach, –0.05), the presence of a reference to the legal framework governing the status of the cybersecurity manager (Ciso_Law, –0.05), the absence of provisions on circumstances excluding liability (Circumstance_Excluding_Liability, –0.05), the existence of the supplier's right to withdraw from the contract (Priv_Supplier_Can_Withdraw, –0.05), the absence of provisions governing incorrect documentation (Incorrect_Docs, –0.05), the absence of a contractual penalty for the supplier's non-performance (Penalty_Private_Nonperf, –0.04), and the absence of specific documentation obligations (Documentation, –0.04). The remaining 26 variables likewise had a negative effect (–0.39), further reducing the resulting value and thereby contributing to the contract's classification as the most normal.

The classification of the contract as the most normal rests on the fact that the payment for the services to be delivered by the supplier is low and, at the same time, does not constitute a large share of the supplier's revenue, while the supplier also does not have many employees. Given that this is an ordinary contract of no great financial significance in terms of the supplier's overall revenue, the absence of a definition of material breach was assessed as contributing to normality. This can

be justified by the argument that the larger the contract, the more naturally the parties will tend toward a more precise regulation of the possibility of terminating the contractual relationship—and thus toward defining what they mutually regard as a material breach of contract.

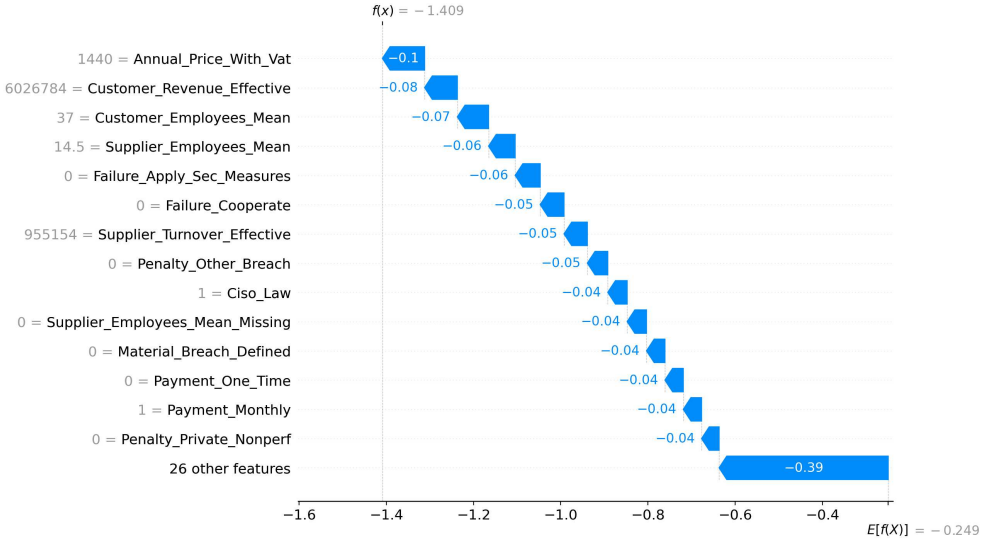


Fig. 9 SHAP explanation for contract ID 6332291. The plot illustrates the contribution of individual attributes to the model output

The contract identified in the dataset as ID 6332291 was, based on the model's results, evaluated as the second most normal contract in the analyzed sample. It was concluded between a municipality and a commercial company providing cybersecurity services. From a commercial-law perspective, it was an innominate contract concluded pursuant to Section 269(1) of the Commercial Code, in conjunction with the provisions of Act No. 69/2018 Coll. on Cybersecurity and Decree No. 362/2018 Coll., which establishes the content and scope of security measures. The subject matter of the contract was the provision of consulting services in the implementation of cybersecurity.

The mean value, indicated in the lower right part of the chart as E[f(X)] = –0.249, represents the baseline value from which the contribution of the individual attributes to the resulting degree of the contract's normality is derived. Attributes shown in blue lower the resulting value and thus contribute to the contract's normality, whereas attributes shown in red act in the opposite direction and contribute to its anomalousness. In this case, a single attribute is shown in red which, although not significantly, partially contributed to a slight anomaly in the contract. As this is the second most normal contract in the analyzed dataset, the visualization is strongly dominated by attributes that reduce the overall result to f(x) = –1.409, shown at the top center of the chart.

In the case analyzed, the attributes with the most significant influence on the overall result were the lower revenue of the customer (Customer_Revenue_Effective, –0.11) and the low annual value of the subject matter including VAT (Annual_Price_With_VAT, –0.11). The degree of anomalousness was further reduced by the absence of a one-time form of payment (Payment_One_Time, –0.06), the absence of a provision governing breach of the duty to cooperate (Failure_Cooperate, –0.06), and the absence of a provision governing breach of duties in the application of security measures (Failure_Apply_Sec_Measures, –0.06). The overall result was further lowered by the lower average number of the supplier's employees (Supplier_Employees_Mean, –0.05), the lower number of the customer's employees (Customer_Employees_Mean, –0.05), the absence of a contractual penalty for breach of other contractual obligations (Penalty_Other_Breach, –0.05), the low share of the contract price in the supplier's turnover (Price_Revenue_Percent, –0.05), the absence of IT monitoring (It_Monitoring, –0.05), the monthly form of payment (Payment_Monthly, –0.05), and the absence of a definition of material breach of contract (Material_Breach_Defined, –0.05). Conversely, the only attribute that slightly increased the resulting value was the supplier's higher effective turnover (Supplier_Turnover_Effective, +0.05). The remaining 26 variables likewise had a negative effect (–0.41), further lowering the overall result and contributing to the contract's normality.

The classification of the contract as the second most normal follows primarily from the fact that it is a financially less significant contract whose value is low not only in absolute terms but also represents a negligible share of the supplier's economic strength. At the same time, the contract does not contain several of the more detailed provisions governing individual instances of breach of contractual obligations, nor any special sanction mechanisms—a feature assessed as consistent with the contracts commonly occurring in the analyzed dataset. The monthly payments for performance of the subject matter were likewise assessed as standard. Only the supplier's higher effective turnover had a slightly opposite effect; however, in light of the low value of the subject matter and the overall predominance of attributes contributing to normality, it was unable to significantly affect the contract's final classification.

5 Discussion

In this section, we present several reflections, whether on the SCA method itself or on the subject of the article and the search for anomalies in cybersecurity service contracts.

5.1 The concept of anomaly and its practical application

In general terms, the concept of an anomaly evokes the notion of a variable that deviates from the norm, typically carrying a negative connotation. However, for the purposes of the present research conducted on the specific dataset, it is important to emphasise that the term anomaly should not be interpreted as having either an inherently positive or negative meaning. Rather, it refers to a contractual provision or

another characteristic arising from the contractual relationship that statistically deviates from the norm, without such deviation, in itself, implying either a beneficial or detrimental effect on the contractual relationship as a whole. Accordingly, an attribute that exhibits anomalous behaviour in one contract (i.e., contributes to the contract's anomalous classification) may, in another contract, be statistically evaluated as contributing to contractual normality, since the overall assessment of a contract takes into account the interaction of all other relevant attributes contained within the dataset.

The results of the analysis indicate, inter alia, that contracts drafted in a relatively simple manner were classified by the model as normal, whereas more complex and sophisticated contractual arrangements were identified as anomalous. This finding raises an important question: should a simply drafted contract be regarded as the norm, and is it genuinely anomalous for a contract to regulate a broader range of situations that may arise during the contractual relationship, for example, in the event of a breach of contractual obligations, by anticipating and addressing a variety of potential contingencies?

In our view, it is entirely natural that where the legal relationship between the contracting parties is more complex—particularly in light of the value of the contract or the size of the entity to which the services are to be provided—the parties will seek to regulate that relationship in greater detail. At the same time, the statutory legal framework should adequately reflect practical contractual needs so that contracting parties can reasonably rely on it without the necessity of regulating every conceivable situation in their agreements.

Another question arising from these findings is whether an anomaly reflects merely the complexity of the contractual arrangement or whether it instead signals a departure from the principle of fair commercial dealing. As discussed elsewhere in this paper, certain aspects of contract law cannot be adequately captured through a binary assessment. An illustrative example is a contractual penalty imposed on a party exercising its contractual right of withdrawal, requiring that party to pay the full contract price despite the fact that the other party is no longer obliged to perform the agreed service.

A final consideration concerns the principle of freedom of contract, which is fundamental to contractual relationships. Where a contractual provision has been accepted by both parties, it may reasonably be assumed that they intended to be bound by it. Nevertheless, it is conceivable that certain contractual practices become so widespread over time that, as a result of the evolution of market practice, a machine learning model ultimately classifies such provisions as normal or standard, even though they may be questionable from the perspective of fairness or contractual balance.

Although the research was primarily conceived as a scientific analysis of contractual practice, its results also open up scope for practical applications. In identifying anomalies, it is possible not only to determine the attributes that contribute most to anomality, but also those that contribute most to normality. On this basis, it would be possible to identify standard (normal) and non-standard contractual terms (anomalous) for cybersecurity contracts. It would subsequently be possible to develop a methodological checklist setting out the essential contractual elements that should form part of these types of contracts. Such a document could serve as a practical tool,

particularly for public-sector entities, in the drafting or review of contracts, while also drawing attention to provisions that occur only sporadically in practice or are frequently overlooked, even though they may have a significant impact on the protection of the rights and obligations of the contracting parties.

5.2 Limitations of binary coding

The coding of contractual provisions, primarily into binary variables (0/1), entails certain methodological limitations inherent to this type of research. The coding does not capture all possible eventualities and in practice represents a certain form of abstraction. Covering a larger number of eventualities would require a greater number of attributes, which may adversely affect the performance of various machine learning models.

One example concerns the contractual right of withdrawal. Under Slovak law, a contracting party may withdraw from a contract even where such right has not been expressly stipulated in the contract itself, as the conditions governing the exercise of this right are regulated by Act No. 513/1991 Coll., the Commercial Code. This code distinguishes between withdrawal in cases of material and non-material breach of contract. For the purposes of the contractual analysis and the coding of contractual provisions into the dataset, however, it was necessary to examine whether a particular contract granted the supplier an express contractual right to withdraw, extending beyond the rights already provided by the Commercial Code. Where the contract did not expressly provide for the supplier's right to withdraw in the event of a specified contractual breach, the dataset recorded the absence of such a contractual provision (0). Strictly speaking, however, virtually every supplier may withdraw from a contract under the conditions prescribed by the Commercial Code. The recorded absence therefore does not indicate the non-existence of the legal right to withdraw, but rather the absence of an express contractual regulation supplementing the statutory framework for the specific contractual relationship.

A solution to this shortcoming in coding could be the use of fuzzy set theory for the coding of legal texts. Perez (2015) argues that legal normativity is a matter of degree and that fuzzy sets are a legitimate tool for capturing it, developing a degree-theoretic model of legal normativity that he terms "fuzzy law" and drawing on three sources: the theory of complementary pairs, fuzzy set theory, and defeasible reasoning.

5.3 Contract Withdrawal in Relation to Contractual Penalties

Similar methodological limitations may also be observed with respect to the contractual penalty clause. Under Slovak private law, a contractual penalty constitutes a form of liquidated damages agreed upon by the contracting parties, typically to secure compliance with a specific contractual obligation. During the analysis, however, a limited number of contracts contained provisions stipulating a contractual penalty in the event of the premature termination of the contractual relationship, particularly where the contract had originally been concluded for a fixed term. For coding purposes, such provisions were recorded as the presence of a contractual penalty for the breach of another contractual obligation (1). From a legal perspective, however, such clauses are

incompatible with Slovak law. The exercise of the contractual right of withdrawal constitutes the exercise of a legal entitlement vested in a contracting party rather than a breach of a contractual obligation. Consequently, where withdrawal is exercised under circumstances permitted by either the contract or the applicable legislation, it cannot lawfully be sanctioned by means of a contractual penalty. In this regard, the Supreme Court of the Slovak Republic has expressly held that withdrawal from a contract does not constitute a breach of a contractual obligation but rather the exercise of a contractual right and, therefore, cannot be sanctioned by a contractual penalty (Supreme Court of the Slovak Republic 2001).

5.4 Information asymmetry in cybersecurity service contracts

One of the fundamental premises of contract law is the assumption that contracting parties possess a sufficient capacity to assess the subject matter of the contractual relationship and the consequences of their conduct (Miller and Oberdiek 2025). In the case of the contracts subject to this analysis, this assumption is significantly weakened or, in certain respects, entirely absent. The reason for this lies in the highly specialised nature of the activity that constitutes the subject matter of the analysed contracts.

In the examined sample of contracts, similarly to other contracts involving the provision of professional services, the client relies to a significant extent on the professional judgment of the expert. This circumstance modifies the traditional understanding of contractual equivalence, as the contractual relationship is characterised by a significant information asymmetry between the contracting parties.

In the case of the analysed sample of contracts, it may further be observed that a significant factor contributing to the emergence of the identified anomalies is precisely the information asymmetry between the contracting parties. While one party (the proposer/seller) possesses specific professional knowledge and experience, the other party (the acceptor/buyer) is generally constrained by limited expertise and is dependent on the professional explanations provided by the expert.

6 Conclusion

In this paper, we demonstrated a possible approach to the examination and identification of anomalies within the scope of systematic content analysis of contracts, focusing on contracts concluded in the field of cybersecurity as our case study. The proposed approach may find its application in relation to other legal documentation (e. g. judicial decisions, terms of services), with the objective of identifying the established legal practice and possible deviations from it. Crucial in this regard is the definition, extraction and analysis of selected attributes, whose presence or absence in contracts may determine its anomalous nature. The attributes examined in this paper focused on standard aspects of contracts regarding their subject matter, liability, confidentiality, etc. In this regard, we calculated the impact of individual attributes on the classification of contracts as anomalous and provided their categorization into different categories, reflecting their relevance for anomaly detection. As explained above, the presence of certain attributes in contracts aided in its classification as anomalous,

or, correspondingly, determined the 'standard' formulation of contractual provisions in the examined case study.

Future research in this area could also take into account existing methodological guidance and model contractual documentation published by competent public authorities, including the model cybersecurity contracts issued by the National Security Authority. A comparative analysis of these model documents and contracts could provide valuable insights into the extent to which recommended contractual provisions are implemented in practice and identify discrepancies between recommended standards and actual contractual arrangements.

A further avenue for research concerns a more detailed examination of contractual penalty mechanisms in relation to the sanctioning framework established under cybersecurity legislation. Such an analysis would make it possible to assess the extent to which contractual provisions reflect statutory requirements or supplement them through private-law arrangements governing the rights and obligations of the contracting parties.

References

Angiulli F, Pizzuti C (2002) Fast outlier detection in high dimensional spaces. In: Elomaa T, Mannila H, Toivonen H (eds) Principles of Data Mining and Knowledge Discovery. 6th European Conference, PKDD 2002, Helsinki, Finland, 19–23 August 2002, Proceedings, Lecture Notes in Computer Science, vol 2431. Springer, Berlin, pp 15–26, https://doi.org/10.1007/3-540-45681-3_2

Baldwin J, Davis G (2003) Empirical research in law. In: Cane P, Tushnet M (eds) The Oxford Handbook of Legal Studies. Oxford University Press, Oxford, p 880–900

Burns K, Hutchinson T (2009) The impact of 'empirical facts' on legal scholarship and legal research training. The Law Teacher 43(2):153–178. https://doi.org/10.1080/03069400903004236

Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: a survey. ACM Computing Surveys 41(3):1–58. https://doi.org/10.1145/1541880.1541882

Dechtiar M, Katz DM, Wang H (2025) Software engineering meets legal texts: LLMs for auto detection of contract smells. Machine Learning with Applications 20:100639. https://doi.org/10.1016/j.mlwa.2025.100639

Esposito A, Di Martino B, Ammendolia R, et al (2023) Time anomaly detection in the duration of civil trials in Italian justice. Connection Science 35(1):2283394. https://doi.org/10.1080/09540091.2023.2283394

Evans M, McIntosh W, Lin J, et al (2007) Recounting the courts? Applying automated content analysis to enhance empirical legal research. Journal of Empirical Legal Studies 4(4):1007–1039. https://doi.org/10.1111/j.1740-1461.2007.00113.x

Gerards J, Erken E, Loven C (2023) The expanding methodological toolbox of the ECHR scholar. Law and Method https://doi.org/10.5553/REM/.000072

Goldstein M, Dengel A (2012) Histogram-based outlier score (HBOS): a fast unsupervised anomaly detection algorithm. In: KI-2012: Poster and Demo Track, Saarbrücken, Germany, pp 59–63

Hall MA, Wright RF (2008) Systematic content analysis of judicial opinions. California Law Review 96(1):63–122. URL https://lawcat.berkeley.edu/record/1121706/files/fulltext.pdf

He Z, Xu X, Deng S (2003) Discovering cluster-based local outliers. Pattern Recognition Letters 24(9–10):1641–1650. https://doi.org/10.1016/S0167-8655(03)00003-5

Jaccard P (1901) Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin de la Société Vaudoise des Sciences Naturelles 37:547–579

Joanes DN, Gill CA (1998) Comparing measures of sample skewness and kurtosis. Journal of the Royal Statistical Society: Series D (The Statistician) 47(1):183–189. https://doi.org/10.1111/1467-9884.00122

Kaltenhäuser K, Slaats T, Muller M, et al (2025) Beyond accuracy: rethinking outlier detection in asylum data. ACM Journal on Responsible Computing 2(1):1–28. https://doi.org/10.1145/3649355

Lagioia F, Ruggeri F, Drazewski K, et al (2019) Deep learning for detecting and explaining unfairness in consumer contracts. In: Araszkiewicz M, Rodríguez-Doncel V (eds) Legal Knowledge and Information Systems: JURIX 2019. The Thirty-second Annual Conference, Madrid, Spain, 11–13 December 2019, Frontiers in Artificial Intelligence and Applications, vol 322. IOS Press, Amsterdam, pp 43–52, https://doi.org/10.3233/FAIA190305

Leeuw FL (2015) Empirical legal research: the gap between facts and values and legal academic training. Utrecht Law Review 11(2):19–33. https://doi.org/10.18352/ulr.315

Li Z, Zhao Y, Botta N, et al (2020) COPOD: copula-based outlier detection. In: 2020 IEEE International Conference on Data Mining (ICDM). IEEE, Sorrento, Italy, pp 1118–1123, https://doi.org/10.1109/ICDM50108.2020.00135

Li Z, Zhao Y, Hu X, et al (2022) ECOD: unsupervised outlier detection using empirical cumulative distribution functions. IEEE Transactions on Knowledge and Data Engineering 35(12):12181–12193. https://doi.org/10.1109/TKDE.2022.3159580

Lippi M, Palka P, Contissa G, et al (2019) CLAUDETTE: an automated detector of potentially unfair clauses in online terms of service. Artificial Intelligence and Law 27(2):117–139. https://doi.org/10.1007/s10506-019-09243-2

Liu FT, Ting KM, Zhou ZH (2012) Isolation-based anomaly detection. ACM Transactions on Knowledge Discovery from Data 6(1):3:1–3:39. https://doi.org/10.1145/2133360.2133363

Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems 30 (NIPS 2017). Curran Associates, Red Hook, NY, pp 4765–4774

Miller PB, Oberdiek J (eds) (2025) Oxford Studies in Private Law Theory: Volume III. Oxford University Press, Oxford, https://doi.org/10.1093/9780198961208.001.0001

Monahan J, Walker L (2011) Twenty-five years of social science in law. Law and Human Behavior 35(1):72–82. https://doi.org/10.1007/s10979-009-9214-8

Mukhsimbayev B, Kuralbayev A, Bekbaganbetov A, et al (2020) Anomaly detection between judicial text-based documents. In: 2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT). IEEE, Tashkent, Uzbekistan, pp 1–5

Oldfather CM, Bockhorst JP, Dimmer BP (2012) Triangulating judicial responsiveness: automated content analysis, judicial opinions, and the methodology of legal scholarship. Florida Law Review 64(5):1189–1249. URL https://scholarship.law.ufl.edu/flr/vol64/iss5/2/

Panarelli M, Galassi A, Lagioia F, et al (2025) Is it worth using LLMs for unfair clause detection in terms of service? In: Maranhão J (ed) Proceedings of the Twentieth International Conference on Artificial Intelligence and Law (ICAIL 2025), Chicago, IL, USA, 16–20 June 2025. ACM, New York, pp 139–149, https://doi.org/10.1145/3769126.3769218

Pedregosa F, Varoquaux G, Gramfort A, et al (2011) Scikit-learn: machine learning in Python. Journal of Machine Learning Research 12:2825–2830

Perez O (2015) Fuzzy law: a theory of quasi-legal systems. Canadian Journal of Law and Jurisprudence 28(2):343–370. https://doi.org/10.1017/cjlj.2015.19

Pevný T (2016) LODA: lightweight on-line detector of anomalies. Machine Learning 102(2):275–304. https://doi.org/10.1007/s10994-015-5521-0

Ruggeri F, Lagioia F, Lippi M, et al (2022) Detecting and explaining unfairness in consumer contracts through memory networks. Artificial Intelligence and Law 30(1):59–92. https://doi.org/10.1007/s10506-021-09288-2

Salehijam M (2018) The value of systematic content analysis in legal research. Tilburg Law Review 23(1–2):34–42. https://doi.org/10.5334/tilr.5

Schölkopf B, Platt JC, Shawe-Taylor J, et al (2001) Estimating the support of a high-dimensional distribution. Neural Computation 13(7):1443–1471. https://doi.org/10.1162/089976601750264965

Schrepel T (2024) A systematic content analysis of innovation in European competition law. European Journal of Law and Economics 58(2):355–395. https://doi.org/10.1007/s10657-024-09817-9

Shyu ML, Chen SC, Sarinnapakorn K, et al (2006) Principal component-based anomaly detection scheme. In: Lin TY, Ohsuga S, Liau CJ, et al (eds) Foundations and Novel Approaches in Data Mining, Studies in Computational Intelligence, vol 9. Springer, Berlin, p 311–329, https://doi.org/10.1007/11539827_18

Smit NM, Morgan RM, Lagnado DA (2018) A systematic analysis of misleading evidence in unsafe rulings in England and Wales. Science and Justice 58(2):128–137. https://doi.org/10.1016/j.scijus.2017.09.005

Spreeuwenberg S, van Engers TM, Gerrits R (2001) The role of verification in improving the quality of legal decision-making. In: Verheij B, Lodder AR, Loui RP, et al (eds) Legal Knowledge and Information Systems: JURIX 2001. The Fourteenth Annual Conference, Frontiers in Artificial Intelligence and Applications, vol 70. IOS Press, Amsterdam, p 1–15

Supreme Court of the Slovak Republic (2001) Judgment of the Supreme Court of the Slovak Republic, case no. 2 cdo 141/2000. Published as No. 58/2001 in the Collection of Judicial Decisions and Opinions (ZSP)

Weber RP (1990) Basic Content Analysis, 2nd edn. Quantitative Applications in the Social Sciences, SAGE Publications, Newbury Park, CA, https://doi.org/10.4135/9781412983488

Zhao Y, Nasrullah Z, Hryniewicki MK, et al (2019) LSCP: locally selective combination in parallel outlier ensembles. In: Proceedings of the 2019 SIAM International Conference on Data Mining (SDM). SIAM, Philadelphia, pp 585–593, https://doi.org/10.1137/1.9781611975673.66