Differentiation of Legal Rules and Individualization of Court Decisions in Criminal, Administrative and Civil Cases: Identification and Assessment Methods

Egor Trofimov
The All-Russian State University of Justice, Moscow, Russia
trofimovev.rpa@gmail.com
Oleg Metsker
The All-Russian State University of Justice, Moscow, Russia
olegmetsker@gmail.com
Georgy Kopanitsa
Georgy Kopanitsa ITMO University, Saint Petersburg, Russia
georgy.kopanitsa@gmail.com

Summary
The diversity and complexity of criminal, administrative and civil cases resolved by the courts makes it difficult to develop universal automated tools for the analysis and evaluation of justice. However, big data generated in the scope of justice gives hope that this problem will be resolved as soon as possible. The big data applying makes it possible to identify typical options for resolving cases, form detailed rules for the individualization of a court decision, and correlate these rules with an abstract provisions of law. This approach allows us to somewhat overcome the contradiction between the abstract and the concrete in law, to automate the analysis of justice and to model e-justice for scientific and practical purposes. The article presents the results of using dimension reduction, SHAP value, and p-value to identify, analyze and evaluate the individualization of justice and the differentiation of legal regulation. Processing and analysis of arrays of court decisions by computational methods make it possible to identify the typical views of courts on questions of fact and questions of law. This knowledge, obtained automatically, is promising for the scientific study of justice issues, the improvement of the prescriptions of the law and the probabilistic prediction of a court decision with a known set of facts.

Key words:
artificial intelligence, digital state, indicator, law, machine learning, data mining, lawmaking, e-justice, big data

1. Introduction

The diversity and complexity of criminal, administrative and civil cases resolved by the courts makes it difficult to develop universal automated tools for the analysis and evaluation of justice. However, big data generated in the scope of justice gives hope that this problem will be resolved as soon as possible. Modern methods of big data processing, intellectual analysis of data and machine learning are suitable for processing large amounts of judicial data that go far beyond the usual statistics. Big data in the scope of justice and computational methods from processing and analysis provide the opportunity to extract knowledge on a wide range of issues. In addition, big data is inaccessible for manual processing, but makes it possible to identify complex (implicit) connections between many phenomena and processes reflected in court decisions using computational methods.

In the context of this general problem, there is an important legislative and judicial problem concerning the conflict between the differentiation of legislative regulation and the individualization of judicial decision. Legislator is guided by the formulation of general rules, but court must, on the basis of these general rules, consider individual facts and make individual decision. To overcome this difficulty, courts create precedents or generate a standard practice for resolving similar cases. This allows the courts to reproduce differentiated provisions of law or further differentiate legal rules for resolving cases if the provisions of the law do not allow reaching adequate individualization of court decisions. The introduction of computational methods in the study of the problem of individualization of court decisions and their compliance with the provisions of the law makes it possible to create high-tech tools for the consistent elimination of the contradiction between lawmaking and justice.

2. Related works

The problem of identifying ways to individualize a single court decision based on an abstract law has been considered for a long time. The architecture of the expert system PROLEXS for a heterogeneous domain, included a database of precedents in the domain area of Danish law on real estate lease. To select typical use cases, neural

Manuscript received December 5, 2022
Manuscript revised December 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.12.16

networks and a hybrid architecture were used, which allowed the entire system to function and match rules and cases with expert knowledge in a small domain. [1]
James Popple proposed a SHYSTER system. The system was based on cases, contained tokenization and parsing modules, and calculated the proximity of the case under consideration of known precedents by the weight of the attributes. [2] However, to match the rules, it was necessary to develop a module in the law domain.
The LexrideLaw system made it possible to extract arguments from judicial decisions and provide access to them through nodes in the ontology of litigation or through keyword search in relational questions. [3]

Kevin D. Ashley and Stefanie Brüninghaus created the SMILE+IBP (Smart Index Learner Plus Issue-Based Prediction) program, which for a trade secret domain that could produce a case solution prediction by extracting facts from an array of cases and factoring them based on classification concepts and stereotypical patterns. [4]
Those approaches were mainly focused on achieving high accuracy in predicting a court decision based on given conditions, facts, and features. The approach reflects the ideas of individualization of court decision but conflicts with abstract law that was mentioned above. The difficulty lies in that an actual court decision is not the result of applying an algorithm. It is important that it is impossible to achieve 100% accuracy of the predictive function of a computer system. On the other hand, it is possible to solve the problem of comparing an abstract legal prescription of law with a court decision as the results of legal qualification of some facts. The application of big data makes it possible to identify typical options for resolving cases, form detailed rules for the individualization of a court decision, and correlate these rules to abstract provisions of law. This approach allows to somewhat overcome the contradiction between the abstract and the concrete in law, to automate the analysis of justice and to model e-justice for scientific and practical purposes.

3. Methods and materials

The study of the individualization of court decisions was carried out on a database collected by the authors of more than 28.3 million court decisions (including more than 8.8 million decisions in cases of administrative offenses, more than 8.9 million decisions in criminal cases, more than 10.6 million decisions on civil and administrative cases) from open public sources. Based on this information array, datasets were formed for certain categories of criminal, administrative and civil cases. This article presents the results of calculations for such categories of cases as disorderly conducts, migration offenses, pilferages, and breaches of Motor Third Party Liability Insurance (MTPL) conditions.

The fields of the received documents were filtered, leaving only significant data, which reduced the volume of fields by an average of 2.5 times. The data was processed by Apache Spark, which is designed for big data processing and was chosen due to its user-friendly API, fast execution, and interactivity. Python was used to process all the data along with a Jupyter notebook. Data from JSON is converted to Parquet to increase the speed of processing and reading frames, factorization and model training. The transformation algorithm was in the MapReduce paradigm. [5]

When analyzing a process, the data must be highly consistent with the specified process. Improving the consistency of data regarding the real process is one of the most important tasks of data mining and text analysis. Data on verdicts consists of a set of different elements that contain structured and semi-structured information, as well as information presented in an unstructured form in natural language. This data brings value for the analysis and identification of process elements. In addition, extracting data from natural language texts is an important scientific task.

Natural language processing (NLP) included parsing, text segmentation, tokenization, classification, mapping [6-9]. Tokenization was carried out both in an expert way (based on n-gram frequency dictionaries) and in an automated way, followed by an expert evaluation of the calculation results using previously developed and tested models. In both cases, the costs and benefits were comparable. Although with scaling of calculations and accumulation of experience, automated evaluation will certainly have an advantage.

The developed data structuring method is based on analytical queries and expert (legal) hypotheses. According to the task of identification and extraction, algorithms for learning rules (templates) and selecting the necessary data for structuring are formed. As a rule, templates for decisions of different types of legal proceedings and different instances differ. For this reason, it is required to separate these documents at the pre-processing stage.
The processed data is analyzed using a template. In case of an unsatisfactory result, the template was modified, and the data is processed again. The cycle of template modification and data processing continued until the required identification result was achieved. Then these rules were divided into induction and probabilistic methods. Successful patterns can be used to develop domain-specific data processing libraries. The identified entities are recorded as attributes in the analysis data sets. When applying this method, errors in data analysis often come down to the wrong choice of template or the analyzed part of the judicial document. [10]

Data cleaning was performed to obtain high-quality data suitable for data mining and machine learning and

meeting the requirements of validity, accuracy, completeness, consistency, and uniformity. [11]
As a result of preprocessing, including structuring, datasets were formed for computational experiments in a volume that ensures high reliability of the results (as a rule, 98–99%). Since law enforcement is a social activity, the assessment of reliability was carried out based on well-tested sociological methods in the subject area of law according to the rules for calculating the statistical error. The general population was determined using judicial statistics. The confidence probability was taken as 99% since judicial acts are accepted "as is". Empirically, it was found that with a random selection of court decisions, the quality of the results obtained can be considered unchanged starting from a sample size of 10 thousand court decisions in the dataset.

To analyze and evaluate the individualization of court decisions, various computational methods were used, described below. The application of methods to the above problem was based on the separation of questions of fact and questions of law. Questions of fact are the features of the circumstances of the case (violations of the law, procedural actions, etc.). Questions of law are judgments on issues that require resolution in accordance with the law (on the guilt of a person, on the type and amount of punishment, on satisfaction of a claim, etc.). Questions of fact act as analyzed data, questions of law act as targets.

4. Results

4.1 Dimension reduction

In the following experiment t-SNE method of dimension reduction was used from scikit-learn (see fig. 1). This method is good for non-linear dependencies identifications and their visualization.

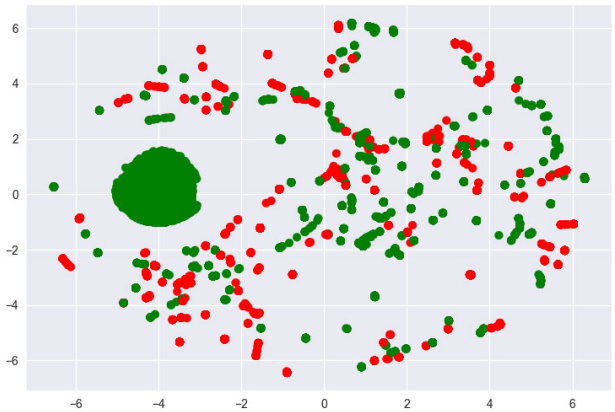

Fig. 1. Dimension reduction by the t-SNE method on the example of administrative cases of disorderly conduct

As features were used circumstances of administrative cases of disorderly conduct, as target was used kind of punishment. The figure demonstrates the alcohol-related violations (marked in red). The green color indicates opposite cases. The green cluster cases are characterized by the first-time-violation cases committed by sober persons. Although the administrative law does not provide for special provisions on administrative liability in cases of disorderly conduct for drunk and sober persons, the courts have formed a stable practice of different resolution of these two categories of cases with the appointment of more severe and more lenient punishment for violators, respectively.

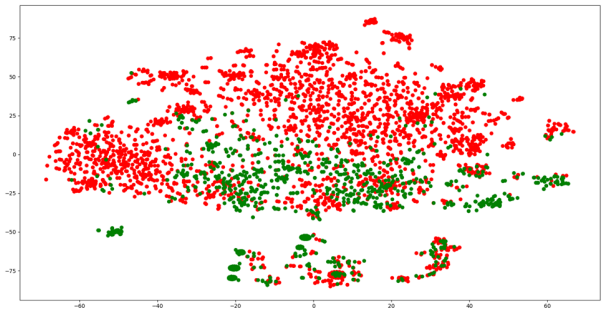

Fig. 2. Dimension reduction on the example of civil cases of MTPL insurance

In the previous example, the space of 11 features (presence or absence of typical mitigating and aggravating circumstances) was reduced. This example (see fig. 2) presents a cluster analysis based on machine learning on 2000 tokens identified automatically in civil cases of MTPL insurance. The cases, in which the court satisfied the claims of the victims, and the cases in which the insurance companies won (colored in different colors), have signs of clustering.

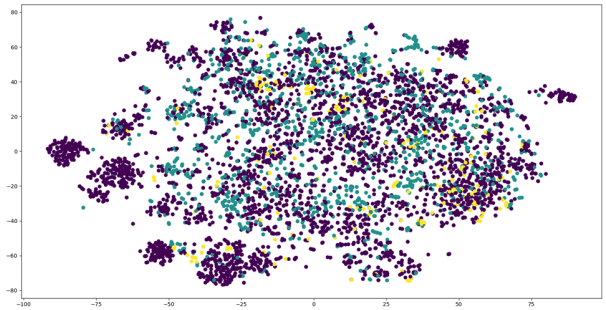

Fig. 3. Dimension reduction on the example of administrative cases on pilferage

This method shows good results not only for the analysis of a positive or negative solution to the main issue of law resolved by the court (satisfy the claim or refuse, convict or acquit), but also in the analysis of differentiated rules for sentencing. Dimension reduction allows you to

identify facts that entail the appointment of various punishments prescribed by law alternatively. For administrative cases of pilferage (see fig. 3), purple clusters are clearly visible, corresponding to cases for which compulsory work was assigned. Fine (yellow) and arrest (green) do not form distinct clusters.

This result shows that the courts impose compulsory work under certain sets of facts regarding an individual case of pilferage, while under other sets of facts, both fines and arrest can be imposed at the discretion of the court. For example, the courts impose compulsory work when the offender commits pilferage for the first time, steals alcoholic beverages and does not have a permanent job. Such reasoning was developed by the courts because of the individualization of justice. They have evolved into widespread unwritten rules that make the written prescriptions of the law even more differentiated.

4.2 SHAP value

With the use of SHAP value on significant data arrays of court decisions makes it possible to identify the influence of various tokens (facts) on the solution of a legal issue.

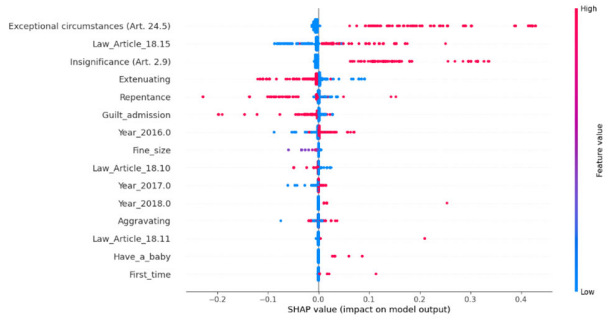

Fig. 4. Distribution of the significance of features for the outcome of appeals in cases of migration offenses

Figure 4 presents the example of migration offenses, which shows the positive and negative impact of the court's decision on the satisfaction of the offender's appeal. Features of the composition of the facts associated with a single case act as regulators of the individual resolution of the case by the court. At the same time, they show what facts force the courts to differentiate the prescriptions of the law and create detailed rules for the administration of justice.

In this example, the establishment of any extenuating circumstances ("Extenuating") by the court leads to the refusal to satisfy the offender's appeal (the red dots are shifted to negative values). Satisfaction of the appeal is associated with the absence of extenuating circumstances (blue points are shifted to positive values). This paradox is related to the fact that mitigating circumstances are established by the court to mitigate the punishment in a situation where the offense was actually committed, and the court sees no grounds for satisfying the appeal. In particular, the meaning of this practice of justice is well explained by such extenuating circumstances as repentance ("Repentance") and admission of guilt ("Guilt admission"), in which the offender pleads guilty or repents of the offense.

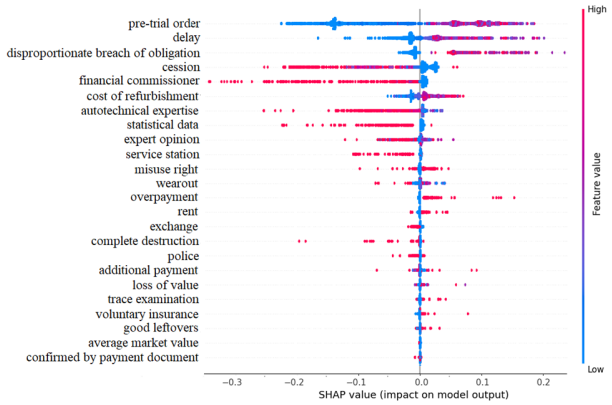

Fig. 5. Using SHAP value on the array of decisions of courts of general jurisdiction

Therefore, the court dismisses the appeal and upholds the judgment of the lower court and its decision on the imposed sentence. In contrary, the establishment of circumstances excluding the proceedings ("Exceptional circumstances (Art. 24.5)") shows their high significance for the cancellation of the earlier decision. The same result is shown by the establishment of the insignificance of the offense ("Insignificance (Art. 2.9)"), which entails exemption from administrative responsibility.

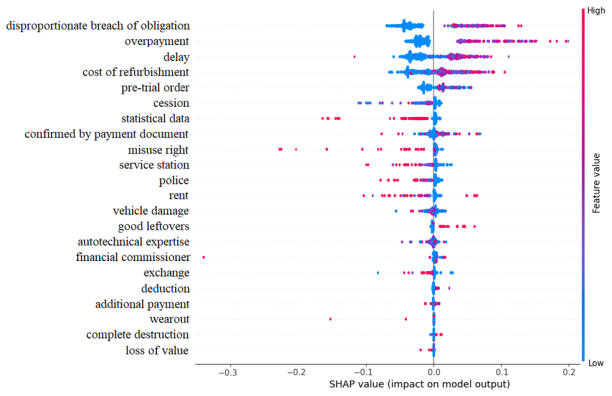

Fig. 6. Using SHAP value on the array of decisions of commercial courts

Another option for using SHAP value is to compare the significance of facts (tokens) for resolving similar cases by the court. Figures 5 and 6 show the facts that have the greatest impact on the satisfaction of civil claims or the refusal of a claim in civil cases of MTPL insurance. Figure

5 shows the results of SHAP value for the array of decisions of courts of general jurisdiction (claims by injured citizens against insurance companies). Figure 6 shows the results obtained for the array of decisions of commercial courts (claims of injured legal entities against insurance companies).

For a court of general jurisdiction and a commercial court, some facts have an opposite meaning for the satisfaction of a claim or the dismissal of a claim. For example, a court of general jurisdiction considers the fact of renting a vehicle in favor of an injured citizen, but commercial courts consider it a basis for dismissing the claim of an injured legal entity. The situation becomes directly opposite in the presence of the fact of cession, which in a dispute with the insurance company is considered in favor of the injured legal entity but plays against the injured citizen.

4.3 P-value

P-value allows assessing the validity of the court's choice of alternative solutions. Sometimes the law allows alternative court decisions but does not prescribe the facts that the court must establish when making a particular choice. In such cases, the court must individualize justice. Analyzing the facts that are used in the reasoning of the court, one can determine how differentiated justice is.

	fine / compulsory works	fine / arrest	compulsory work / arrest
guilty plea	0,043	0,428	0,023
repentance	<0,001	<0,001	0,033
have a child	0,252	<0,001	<0,001
health status	0,883	<0,001	0,143
first-time	<0,001	<0,001	0,176
repentance	<0,001	<0,001	<0,001
reimbursement	0,002	0,127	<0,001
drunkenness	<0,001	<0,001	0,620

Fig. 7. Using p-value on the administrative cases in pilferage array

On fig. 7 and fig. 8 shows the p-values when assessing the significance of the same fact for imposing two different punishments in similar cases. On fig. 7 shows the results obtained on the administrative cases on pilferage array, while fig. 7 shows the results on an array of similar criminal cases on pilferage, when the violation is repeated. Purple cells showed a significant degree of differentiation of justice and the high significance of the analyzed facts for the individualization of court decision (p-value<0.05).

5. Discussion

Dimension reduction allows you to determine the heterogeneity in the array of cases and their clustering. If the clusters are formed by significant facts and correlate with differentiated court decisions, they are an indicator of the functioning of a certain rule for the administration of justice, which can be directly provided for by law or can be formed by the courts themselves during the individualization of justice.

	fine / compulsory works	fine / correctional labor	fine / restriction of liberty	fine / imprisonment	compulsory works / correctional labor	compulsory works / restriction of liberty	compulsory works / imprisonment	correctional labor / restriction of liberty	correctional labor / imprisonment	restriction of liberty / imprisonment
drunkenness	0,001	0,001	0,146	0,001	0,838	0,007	0,027	0,010	0,073	0,087
pregnancy	0,001	0,002	0,597	0,001	0,525	0,001	0,191	0,001	0,150	0,001
reimbursement	0,001	0,785	0,014	0,001	0,001	0,672	0,979	0,030	0,001	0,626
first-time	0,457	0,409	0,322	0,001	0,105	0,097	0,001	0,797	0,001	0,001
guilty plea	0,075	0,963	0,053	0,440	0,080	0,541	0,001	0,054	0,502	0,001
apology	0,568	0,979	0,600	0,792	0,562	0,271	0,232	0,626	0,828	0,671
repentance	0,204	0,251	0,068	0,305	0,336	0,999	0,022	0,456	0,632	0,167
have a child	0,083	0,918	0,037	0,804	0,123	0,001	0,022	0,033	0,918	0,006
recidivism	0,001	0,002	0,930	0,001	0,024	0,001	0,001	0,008	0,001	0,001
facilitating an investigation	0,303	0,082	0,003	0,001	0,304	0,013	0,001	0,168	0,088	0,785
difficult life circumstances	0,077	0,020	0,068	0,005	0,361	0,758	0,066	0,688	0,707	0,438
surrender	0,073	0,687	0,118	0,001	0,029	0,808	0,099	0,060	0,001	0,476

Fig. 8. Using p-value on the criminal cases in pilferage array

The variety of facts that influence the decision of the court is difficult to comprehend without a certain level of simplification or abstraction. A person (judge, lawyer, plaintiff, offender, etc.) needs to have a visual model or scheme that transforms algorithms into simple steps. Dimension reduction methods make it possible to solve this problem. As a result, dimensionality reduction facilitates classification, visualization, and compression of high-dimensional data. [12] It is advisable to reduce the data to the internal dimension of the data. The intrinsic dimensionality of data is the minimum number of parameters needed to account for the observed properties of the data. [13] When the differentiation of the law and the individualization of the judicial decision are investigated, a two-dimensional or three-dimensional space is sufficient. This problem is fundamentally solved, in particular, by the t-SNE method. [14] [15]

The highest accuracy for large modern datasets is often achieved by complex models that even experts struggle to interpret. SHAP helps users interpret the predictions of

complex models and assigns each feature an importance value for a particular prediction. [16]

The p-value is a number, calculated from a statistical test, that describes how likely you are to have found a particular set of observations if the null hypothesis were true. P-values are used in hypothesis testing to help decide whether to reject the null hypothesis. Although p-values are helpful in assessing how incompatible the data are with a specified statistical model, contextual factors must also be considered, such as the design of a study, the quality of the measurements, the external evidence for the phenomenon under study, and the validity of assumptions that underlie the data analysis. [17]

The use of computational method for big legal data processing must consider limitations and doubts about the scope of applicability of algorithms and the reliability of the resulting calculations. However, in the scope of law, the quality of analysis and evaluation performed with the help of algorithms should be controlled by experts at all stages of implementation and applying of algorithms, until sufficient evidence has been accumulated that computational methods provide the solution of research and practical problems at a level no worse than that of an expert.

6. Conclusion

Processing and analysis of arrays of court decisions by computational methods make it possible to identify the typical views of courts on questions of fact and questions of law. This knowledge, obtained automatically, is promising for the scientific study of justice issues, the improvement of the prescriptions of the law and the probabilistic prediction of a court decision with a known set of facts. Knowing how the court uses the law to justify its decisions creates the conditions for the legislator to correctly express his will and can help to avoid defects in legislative regulation. Predicting the outcome of a trial will help avoid unnecessary disputes, reduce attorney fees, and reduce the burden on the judiciary.

Acknowledgments

The reported study was funded by RFBR, project number 20-011-00837.

References

[1] Walker, R.F., Oskamp, A., Schrickx, J.A., Opdorp, G.J., van den Berg, P.H.: *PROLEXS: creating law and order in a heterogeneous domain*. International Journal of Man-Machine Studies. 35 (1), 35–68 (1991)

[2] Popple, J.: *A pragmatic legal expert system*. Aldershot: Dartmouth (1996)

[3] Gifford, M.: *LexrideLaw: an argument based legal search engine*. In: Proceedings of the 16th International Conference on Artificial Intelligence and Law, 271–272 (2017)

[4] Ashley, K.D., Brüninghaus, S.: *Automatically classifying case texts and predicting outcomes*. Artificial Intelligence and Law, 17, 125–165 (2009)

[5] Metsker, O., Trofimov, E., Petrov, M., Butakov, N.: *Russian court decisions data analysis using distributed computing and machine learning to improve lawmaking and law enforcement*. Procedia Computer Science, 156, 264–273 (2019)

[6] Aletras, N., Tsarapatsanis, D., Preoţiuc-Pietro, D., Lampos, V.: *Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective*. PeerJ Computer Science, 2, e93 (2016)

[7] Backus, J.W.: *The syntax and semantics of the proposed international algebraic language of the Zurich ACM-GAMM Conference*. In: Proceedings of the International Conference on Information Processing, 125–131 Paris: UNESCO (1960)

[8] Doan, S., Conway, M., Phuong, T.M., Ohno-Machado L.: *Natural language processing in biomedicine: a unified system architecture overview*. Methods in Molecular Biology, 1168, 275–294 (2014)

[9] Friedman, C., Rindflesch, T.C., Corn, M.: *Natural language processing: state of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine*. Journal of Biomedical Informatics, 46 (5), 765–773 (2013)

[10] Metsker, O., Trofimov, E., Sikorsky, S., Kovalchuk, S.: *Text and data mining techniques in judgment open data analysis for administrative practice control*. Communications in Computer and Information Science, 947, 169–180 (2019)

[11] Osborne J.W.: *Best practices in data cleaning: a complete guide to everything you need to do before and after collecting your data*. Thousand Oaks, CA: Sage (2012)

[12] van der Maaten, L., Postma, E., van den Herik, J.: *Dimensionality reduction: a comparative review*. Journal of Machine Learning Research, 10, 66–71 (2009)

[13] Jimenez, L.O., Landgrebe, D.A.: *Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data*. IEEE Transactions on Systems, Man and Cybernetics, 28 (1), 39–54 (1997)

[14] Roweis, S., Hinton, G.: *Stochastic neighbor embedding*. In: Advances in Neural Information Processing Systems 15, 857–864 (2002)

[15] van der Maaten, L.J.P., Hinton, G.E.: *Visualizing data using t-SNE*. Journal of Machine Learning Research, 9, 2579–2605 (2008)

[16] Lundberg, S.M., Lee, S.-I.: *A unified approach to interpreting model predictions*. In: Advances in Neural Information Processing Systems 30, 4765–4774 (2017)

[17] Wasserstein, R.L, Lazar, N.A.: *The ASA's statement on p-values: context, process, and purpose*. American Statistician, 70 (2), 129–133 (2016)

Egor Trofimov received the M.E. degree, from Khabarovsk State Univ. of Econ. and Law in 2003. He received the Dr. Juris. degree from RUDN Unive. in 2013. After working as a Head of the Dept., ARSUJ, he has been a deputy director for science at ARSUJ since 2014. His research interest includes anti-corruption, and AI&law. He is a member of RAL.

Oleg Metsker received the M.E. degree, from ARSUJ in 2018. He received the Cand. Sc. Tech. degree from ITMO Univ. in 2019. He has been a researcher at ARSUJ since 2020. His research interest includes machine learning. He is a member of RAL.

Georgy Kopanitsa received the M.E. degree, from Tomsk Polytech. Univ. in 2007. He received the Cand. Sc. Tech. degree from Tomsk Polytech. Univ. in 2011. He has been a researcher at ITMO Univ. since 2018. His research interest includes machine learning. He is a member of RAL.