International Labour Organization

Funded by the European Union

Data mining and machine learning: Supporting labour inspectorates to address undeclared work

Ada Huibregtse, International Labour Organization

Eleni Alogogianni, Hellenic Labour Inspectorate, Greece

Data mining and machine learning: Supporting labour inspectorates to address undeclared work

ADA HUIBREGTSE, International Labour Organization

ELENI ALOGOGIANNI, Hellenic Labour Inspectorate, Greece

Email corresponding author

huibregtse@ilo.org

alogogianni.eleni@hli.gov.gr

This paper was presented at the

“8th Conference of the Regulating for Decent Work Network”

on Ensuring decent work in times of uncertainty

at the International Labour Office Geneva, Switzerland

10–12 July 2023

ISBN 9789220402719 (web PDF)

This publication was produced with the financial support of the European Union through Employment and Social Affairs Platform (ESAP) 2 project. Its contents are the sole responsibility of the author and do not necessarily reflect the views of the European Union.

Abstract

Does adopting Data Mining and Machine Learning (DM&ML) methods to identify labour law violations likely support labour inspectorates to ensure the enforcement of labour law and keep pace with rapidly changing labour relations? Inspection authorities have taken advantage of technological developments to collect more and better-quality data. A wealth of data is stored on the servers of numerous labour law enforcement institutions with the potential for a better understanding of who violates labour law, how they do it and why. Yet that data rarely is used for improving their inspection rationale and strategies.

This paper focuses on revealing the superior predictive power of DM&ML approaches – compared to manually configured red-flag approaches – in targeting businesses for inspections and in increasing labour inspection's efficiency in uncovering undeclared work. Unlike manually set red-flag methods, DM&ML tools increase the predictive accuracy by, first, automatically considering variables that traditional paradigms will omit about who, how and why one engages in undeclared work or other labour law violations; and second, by identifying changes in behavioural patterns significantly fasterw than experts or practitioners.

In this study we demonstrate the application of a sophisticated and interpretable machine learning method, the Associative Classification, in the process of planning actions to face undeclared work and other labour law violations of Albania's State Inspectorate of Labour and Social Services. Interpretable machine learning produces "white-box" classifiers that present their results in explainable terms to humans, improving users' domain knowledge and their acceptability and trust in the models' outputs. In our research application, we use actual data of around 12,600 onsite inspection visits performed across Albania between 2021 and 2022. We build data mining models used in two ways: first, as an effective prediction tool for classifying risky employers, hence contributing to scheduling targeted onsite visits to deal with specific labour law violations; second, as a knowledge provision tool that explains to users how the predictions are made and reveals the most dominating employers' attribute patterns associated with various labour law violations, thus enhancing the ability of inspectors identifying these violations. We present the models' classification outputs, their prediction assessment metrics and paradigms of extracted knowledge. We prove that the proposed methodology using DM&ML approaches is much more effective than the current inspection selection methods using red-flag indicators employed by the authority.

Keywords: data mining, labour inspection, machine learning, undeclared and grey work, labour inspection

▶ 1. Introduction

Labour inspectorates are the most prominent public institutions that have a mandate to enforce compliance with labour legislations and/or Occupational Safety and Health (OSH). However, some employers do not comply with these laws and regulations, putting workers and/or other employers in disadvantageous positions or at risk. Labour violations resulting from undeclared work[1] negatively affect, in most cases, the size of health and pensions budgets, while OSH violations are associated with loss of life, illnesses and, as a result, increased costs in health and other safety net state budgets. Additionally, labour inspectorates bear the cost of preventing and detecting violations, transforming undeclared into declared work and ensuring occupational safety and health standards. In this study, we seek to identify a more effective way (using DM&ML for predictive modelling) for labour inspectorates to uncover occurrences of undeclared work, thereby creating more public value.

Undeclared work, defined by the European Commission (EC) in its Communication 98/2019 (1998) as "paid activities that are lawful as regards their nature but not declared to the public authorities", and by the ILO as "all economic activities by workers and economic units that are – in law or practice – not covered or insufficiently covered by formal arrangements" (ILO 2002), is a complex phenomenon and easily affected by economic trends, technological innovations and specific legal and economic policies. In tight economic times or in the presence of tax hikes, both employers and workers may be more motivated to engage in undeclared work. Increased occurrences of undeclared work could also be expected when trust in government is low or decreases due to corruption, heavy bureaucratic procedures or dissatisfaction with how taxpayers' money is spent. Data from special Eurobarometer surveys in 2013 and 2019 indicate that the EU's proportion of workers "engaging in undeclared work because they believe they receive nothing back from the state, so it makes no sense to pay taxes" has doubled from 5 per cent in 2007 to 11 per cent in 2019.[2] Developing countries are fertile ground for undeclared work because the rule of law is relatively lax, corruption is high, trust in government or any public institutions is very low, law enforcement and public service delivery capabilities are low, and labour markets and industrial relations are disrupted easily by technological change.

To achieve their goal of addressing undeclared work, inspectorates in developing countries need to increase their institutional effectiveness and efficiency and improve the quality of services to employers and workers. Successful inspection planning, resulting in targeting employers that violate labour law rather than the ones that do not, increases institutional effectiveness and contributes to reaching institutional strategic and policy goals. Additionally, effective inspection planning contributes to increased inspection yields. In other words, the amount of resources and costs required to perform a successful inspection decreases as the share of successful inspections increases.

With digitalization and optimization of work processes and documentation, inspectorates in developed economies have collected a vast amount of information. Generally, the processing, storage and analysis of this information feed organizational decision-making processes and affect their performance effectiveness and efficiency. However, traditional methods of planning inspections to uncover labour law violations are based more frequently on educated guesses by inspectors and administrators planning inspections and less frequently on data analysis, cross-checks, correlations or inferential statistics. One of the tools employed by these inspectorates for targeted inspections is rule-based or red-flags risk assessments, where a set of rules is developed and consistently used based upon inspectors' current understanding of reality.

With the aim of catching up with developed countries, public institutions in emerging and developing economies falling in the upper middle-income category are racing to digitize and optimize their work

1 This study is a product of the Employment and Social Affairs Platform 2, a project funded by the European Commission and implemented by the International Labour Organization. An ILO team worked in close collaboration with Albania's State Inspectorate of Labour and Social Services (ASLISS) to develop a risk assessment system that will increase this inspectorate's effectiveness and efficiency in uncovering undeclared work as well as other labour law violations. We would like to thank the ASILSS team (Eljo Muçaj, Chief Labour Inspector; Irida Qosja, Deputy Chief Inspector; Albana Kuka, Head of the Analytics Unit; and Eda Beqiri, Head of the OSH Unit) for their dedication to build this system and for their transparency and openness during the entire course of the project. Without their help, this and the future studies planned under this project would not be possible.

2 Special Eurobarometer No. 498 conducted in 2019 (26,514 respondents), Special Eurobarometer No. 402 conducted in 2013 (26,257 respondents) and Special Eurobarometer No. 284 conducted in 2007 (25,346 respondents). Sample size after excluding the United Kingdom. See Collins and Horodnic (2020).

processes. Although in the first stages of e-governance and digitalization of processes and services, some labour inspectorates are increasing the size of their databases. Yet most of these databases remain either underused or inefficiently used due to a lack of knowledge about how data can be managed, scarce human capacities and limited financial resources. Few inspectorates in the upper-middle income category use risk assessment systems of any kind for inspection planning and targeting.

Inspectorates in developed, emerging and upper-middle-income categories can benefit from powerful analytics. While rule-based risk assessment tools might be superior to the absence of such tools in inspection planning, their effectiveness and efficiency could improve significantly by using innovative data-driven approaches, such as data mining and machine learning. A few pioneering studies utilizing data from the Hellenic Labour Inspectorate have explored and found that the adoption of DM&ML by inspectorates can increase significantly the accuracy of inspection planning and, as a result, contributes to reducing undeclared work and efficient use of taxpayers' money (Alogogianni and Virvou 2021, 2022 and 2023).

This paper applies an innovative and interpretable machine learning method, Associative Classification, to data from 12,600 onsite inspection visits performed by Albania's labour inspectors between the years 2021 and 2022 and employer characteristics data acquired from the General Tax Administration to determine, with a higher degree of accuracy than rule-based risk assessment tools, the profile of employers more likely to engage in undeclared work. Using "white-box" models that are inherently transparent, we generate human-readable explanations, thus improving the users' domain knowledge and their acceptability and trust in the models' outputs. Albanian State Inspectorate of Labour and Social Services and those in other countries that have collected a sizable dataset can use this method for an effective prediction of risky employers, hence increasing the likelihood that scheduled onsite visits identify undeclared work or other labour law violations. A tool based on this method can build knowledge about the patterns of behaviours and characteristics of risky employers as well as identify emerging patterns that are not immediately perceivable to human observers. In this study, we present the models' classification outputs, their predictive assessment metrics and paradigms of extracted knowledge. Results indicate that the proposed methodology using DM&ML approaches is more effective in several ways than the current inspection visits' selection methods using red-flag indicators employed by the authority. The structure of this study consists of a literature review to nest our argument in the existing body of knowledge, an outline of its theory and hypothesis, then covers our methodology, data analysis, findings, conclusions and ideas for future research.

▶ 2. Literature review

Artificial intelligence technologies have penetrated almost all organizational processes and, to a more varying degree, decision-making. According to the McKinsey Global Institute (2018), by 2030 some 70 per cent of companies in developed economies may have adopted at least one type of artificial intelligence technology, and less than half may have fully absorbed the five specified categories. The International Data Corporation estimated that 40 per cent of digital transformation initiatives in 2019 will use artificial intelligence services and that 75 per cent of business applications will use artificial intelligence by 2021 (Crews 2019). Private sector enterprises, their leaders and managers are leveraging digitalization, big data and artificial intelligence faster than public institutions, and they are at the forefront of innovation.

Many large companies around the globe, especially financial services, as well as some small and medium-sized enterprises in the developed world, are taking advantage of big data and powerful analytics techniques such as data mining and machine learning to create powerful advantages over their competitors through artificial-intelligence-augmented decision-making (Bean 2018, Miller 2018). Davenport (2019, 11) notes that from 250 cognitive-aware managers whose organizations already were pursuing artificial intelligence, 58 per cent responded that they were employing machine learning in their businesses. Nasdaq's SMARTS implemented machine learning within its surveillance technology software to analyse abnormal market events and inform their subsequent categorization by surveillance analysts to predict analysts' decision-making and detect market abuse across the Nordic markets (*NASDAQ OMX* 2017). Amazon, Google and the Chinese internet and e-commerce giants Alibaba, Baidu and Tencent have put economies of scalable learning into widespread practice across the end-to-end service realization and delivery life cycle (Miller 2018, 18).

Data mining and machine learning (DM&ML) provides decision-makers with support to disentangling complex, uncertain and equivocal situations compared to the use of heuristics and intuition only. These are tools to address their limitations and biases exacerbated by an increased amount and complexity of information that they are expected to account for in their decisions. DM&ML reduces bias in decision-making resulting from selective perceptions whether by overplaying information consistent with or downplaying information contradicting one's own views, or filtering information to reflect a person's experience. Statistically-based machine learning algorithms provide cognitive insights to decision-makers by detecting and interpreting patterns in data and ultimately reduce the cost of making predictions (Remus and Kotteman 1986; Agrawal, Gans and Goldfarb 2018; Miller 2018; Davenport 2019; Duan et al. 2019; Shrestha, Raj, Krishna and von Krogh 2021).

DM&ML can be applied to improve effectiveness and efficiency across a variety of fields. It can reinvent business models and ecosystems, improve safety protocols on the work floor, remake the customer experience and target the most appropriate consumer groups. With the application of machine learning to inform patient-provider decision-making, Brnabic and Hess (2021) identified 34 publications using a wide variety of approaches, algorithms and validation strategies. Tudoran (2022) analysed clickstream data using machine learning methods to help e-commerce better understand differences in customer decision-making and support these businesses' strategic decisions. Qian et al. (2020) has found that using machine learning could improve the efficiency of taxi dispatch at Chengdu Shuangliu airport. Rodríguez-Padial, Marín and Domingo (2017) argued and found support that a combination of Principal Component Analysis algorithms and Machine Learning using Artificial Neural Network algorithms improves control over the maintenance function of an industrial plant, provides information on strategic productive areas and allows for the discovery of hidden behaviour patterns in work orders. Nallathambi et al. (2023) used machine learning to identify the correlation between unsafe behaviours and influential factors in hazardous chemical warehouse accidents, estimate the impact of human factors that contribute to human errors that caused firework industry disasters, explosions and incidents, and ultimately proposed an expert system to address occupational hazards.

Applications of DM&ML in the area of tax evasion provide some of the most relevant information for our study. They inform policymakers, managers and researchers on the effective functioning of a public institution with law enforcement authorities similar to labour inspectorates. A growing body of knowledge has emerged around tax evasion primarily due to tax administrations' authority to fill government coffers. In 2019, worldwide, 102 tax administrations used data science and analytics tools to understand historical data and build predictive models. Of those, 58 use some form of artificial intelligence facilitated through Cognos, Julia, Neo4j, Python, R, SAS, SPSS or SQL. Tax authorities in Hungary, Latvia, the Netherlands and the United States, among others, use DM&ML in the areas of value-added tax, personal and corporate taxes, customs and automatic exchange of information (Tamás Czinege 2019; HAI 2020; NTCA 2020; NÁV 2022; PWC 2022). The U.S. Internal Revenue Service, in cooperation with Stanford University, implements an “active learning system” allowing for iterative updates of models to identify tax audits likely to reveal miscalculations or fraud (HAI 2020).

Findings from DM&ML applications either incorporated into tax authorities’ systems or implemented by researchers across the world indicate that supervised and unsupervised machine learning employing associations rules, logistic models, artificial neural networks, the Bayesian network, decision trees and so forth are very effective in identifying patterns and rules, and produce classifications and predictions. They can also be used to identify discrepancies in VAT applications (Wu et al. 2012; Matos et al. 2014; Fox et al. 2014; González and Velásquez 2013; Perez et al. 2019).

Both the EU and the ILO identify DM&ML as an important method to be incorporated into labour inspection systems – with the aim of addressing undeclared work and other labour law violations. However, labour inspectorates in Europe are only in the first stages of employing these methods. A promising study using DM&ML that utilizes labour inspection data about undeclared work come from the Hellenic Labour Inspectorate (Alogogianni and Virvou 2021, 2022, 2023). These research studies prove that the benefits of such innovative data-driven approaches are multiple, both in knowledge and efficiency. Most importantly, following the appropriate data-engineering techniques and interpretable machine learning methods, the prediction of undeclared work increases significantly from 6 per cent to 70 per cent, while also improving the overall accuracy of the inspections to 70 per cent. In parallel, it is shown that the extraction of comprehensible insights regarding patterns of undeclared and grey work should not be neglected, since they can enhance the labour inspectors’ knowledge and ability to identify more successfully these violations when being in the field. Lastly, these studies discussed the feasibility of adopting such techniques into the business processes of the labour inspectorates for planning, proving that they offer flexibility and do not require advanced machine learning knowledge from the systems’ users.

▶ 3. Theory and hypothesis

Labour inspectorates have the authority to address undeclared work, but in upper-middle-income countries they often lack financial and human resources, tools and sophisticated procedures to detect and prevent undeclared work. To fulfil this obligation, it is not sufficient to address the hindrances listed earlier. Applying “smarter” and more powerful approaches and tools to validate current knowledge and create a new, more accurate understanding about which employers engage in undeclared work – and why – can transform labour inspection into a more effective tool for addressing undeclared work. In this study, we seek to answer the question: can data mining and machine learning augmented decision-making improve the effectiveness of labour inspectorates in addressing undeclared work?

The adoption of DM&ML has the capacity to classify and analyse large amounts of information about complex issues such as undeclared work and to identify patterns of compliance/non-compliance behaviour with little induced human bias. With reduced human bias, the accuracy of data processing and analysis conducted by any labour inspectorate can increase the accuracy of predictions of who, how frequently and when an employer is engaged in undeclared work, thus generating valuable knowledge for labour inspectorate officials. When these predictions are incorporated into inspection planning with the aim of targeting risky employers, we expect that more inspection visits will identify cases of undeclared work than when they are not incorporated. An increased proportion of inspection visits identifying undeclared work will increase the overall institutional effectiveness. Increased effectiveness results in increased inspection efficiency since the same input in terms of human, operation and financial resources is translated into the discovery of more cases of undeclared work.

This line of reasoning rests upon the assumption that labour inspection management are interested in improving their institution's performance. If this assumption holds true, labour inspectorate officials (decision-makers, managers and implementers) will be willing to adopt DM&ML systems and use information and knowledge generated through these systems to identify the actors, behaviours and actions that challenge or support labour inspectorates' probability of achieving their objectives and fulfilling their mission. Equipped with knowledge, they can create strategies, policies and action plans that seek to minimize challenges and maximize actions conducive to success.

In this paper, the public institution taken as a case study is the Albanian State Inspectorate of Labour and Social Services (ASILSS) which is in the process of adopting a machine learning risk analysis system to improve the accuracy of inspections in identifying labour law violations and undeclared work. If ASILSS replaces the current risk analysis tool, in which variables are assigned weights based on ASILSS officials' understanding of undeclared work, with a data mining and machine learning risk assessment system (RAS) where human judgements and biases are kept to a minimum, RAS will generate more accurate predictions about employers likely to engage in undeclared work.

If ASILSS decision-making officials base their planning decisions on machine learning/data mining findings, it is expected to increase the success rate of inspections that identify labour law violations and undeclared work. Any identification of labour law violations is followed by an administrative measure that aims to rectify violations and improve the implementation of labour standards.

The increased success rate in identifying labour law violations will increase ASILSS's effectiveness in achieving its mission of compliance with labour law standards and, more specifically, undeclared work. It identifies more labour violations for the same number of inspections. While the cost of an inspection is unchanged, the number of inspections that have little to no effect on improving compliance with labour standards decrease. Fewer complying companies are subjected to unnecessary inspections.

Limitations of data mining and machine learning

So far there are no comprehensive cost-benefit studies determining the value (increased effectiveness) of adopting or scaling big data and machine learning methods vis-à-vis financial investment in technology/ capital, human resources and capacity building.

Possible selection bias in the criteria (political, electoral support) used to select businesses for inspection in the past can be reinforced by machine learning findings. Thus, a few sub-groups of companies will be more likely to be selected for inspection leaving potential violators uninspected. ASILSS inspection history shows that in 2019 a large number of inspections were conducted randomly (although in specific sectors). Hence, random inspections must continue in parallel with planned audits based on machine learning risk analysis, the aim being to feed the prediction models with new unbiased inspection data and detect recent irregularities in the labour market that remain hidden.

▶ 4. Data and methodology

As discussed, the present study leans on the business environment of the Albanian State Inspectorate of Labour and Social Services (ASILSS) and uses data from past inspections integrated with company data declared by the businesses to the Tax Authorities. More specifically, ASILSS owns a case management system called "The Matrix of Penalties" (MOP), which was installed in 2019 and has been used since then by the labour inspectors to enter data and manage their inspection cases. On a regular basis, the ASILSS receives an updated file with companies' data from the Tax Authority and uploads it to the MOP. Inspection cases in the MOP are linked with companies based on their Tax ID, and, upon their completion, the inspector fills in several details related to the audit outcome such as the number of employees found working; unregistered and partially registered workers; those formalized after the audit; all the detected violations related to labour relations and occupational safety and health; and the suggested measures to be taken by the company to be regulated as per the labour law.

For the purposes of this study and the design and implementation of a new Risk Analysis System using DM&ML being developed for ASILSS under the ESAP 2 project, the researchers formally requested and received from ASILSS all the inspection data in the form of statistical reports drawn from the MOP as well as the last updated file of company data coming from the Tax Authority. The researchers also ensured the ASILSS that the use and management of this data would comply with GDPR provisions and be used only for research.

a. Data preparation

Several data pre-processing steps were necessary to result in a dataset appropriate for machine learning. This phase includes integration of the inspection data coming from the MOP and the company data of the Tax Authority, understanding of data elements and selection of data attributes as per the study goals, data selection based on their appropriateness and completeness, and discretization and categorization of data attribute values for enhanced understandability of the machine learning models outcome.

For this machine learning application, only the inspection cases of 2021 and 2022 are to be used. Data from 2019 and 2020 require further investigation to ensure the same quality of data as those from 2021 and 2022. ASILSS's inspection activity in 2020 was mainly focused on advising companies regarding compliance with the labour law provisions related to the Covid-19 pandemic, and they were not performing all their regular inspections as per their organizational targets. In addition, 2019 was the first year of the MOP productive use, thus the inspection data are less reliable regarding their accuracy, completeness and links to company data from the Tax Authority.

Our focus on undeclared and grey work determined the selection of the attributes. Notably, the inspection case reports drawn from the MOP included several inspection findings, many of which relate to occupational safety and health. These are irrelevant to the current application, hence they were omitted. Additionally, the study does not aim to predict any business reactions after the audit since these rely entirely on human decisions (for example, if the undeclared employees are registered or any undeclared overtime is later paid), thus such attributes are also excluded from the dataset. Lastly, all data elements identifying companies, inspectors and employees were also excluded for anonymization purposes. In total, the MOP reports included 64 attributes, out of which only the inspection type, the region of the inspected business and the findings related to undeclared and grey work were kept for further processing.

The *inspection type* is related to the inspection trigger and is included in our study to investigate its correlation to the inspection result. It takes eleven different values, yet the most dominant is the "planned based on risk analysis tool" reaching more than 80 per cent. All the rest of the inspection types are grouped into a second category: "other inspection type". The *region* indicates the area where the inspected company branch is established and takes values from among the thirteen main regions of Albania, which are grouped into the four districts of the country as illustrated in table 1. The last dataset attribute created based on the inspection findings is the *result* of the inspection, which can be undeclared work detection (*UDW*); grey work detection (*GREY*); a combination of these two (*UDW & GREY*); detection of other violations, e.g., OSH (*OTHER_VIOL*); and no violation detection (*NO_VIOL*).

The rest of the dataset attributes come from the Tax Authority file of companies and include the *business sector* taking values out of the five main business sector categories as per the Tax Authority's categorization: *employees*, whose values are discretized in five ranges, as shown in table 1, and illustrates the size of the businesses as per their number of employees; *company registration*, an attribute constructed from the registration date and taking four categorical values indicating the seniority of the company; and the *legal form* and *company type* of the business, as indicated in the Tax Authority file.

In total, eight categorical attributes are included in the dataset taking the values summarized in table 1. After data integration and cleaning, for example, deleting inspection cases that contained null values in several attributes and those that could not be linked with a company existing in the Tax Authority file, the resulting total number of 2021–2022 inspection cases is 12,660, whose ratio per different attribute values is illustrated in the last column of table 1.

Table 1. All inspections performed between 2021 and 2022

Table 1. All inspections performed between 2021 and 2022			
Total number of cases: 12,660			
Attributes	Values	Range/Description	Ratio %
INSPECTION TYPE	RA_PLANNED	Planned inspections using the current Risk Analysis tool	80.94
	OTHER_INSP_TYPE	All other types of inspections including accidents, complaints, re-inspections, randoms, campaigns, sub-contractors, authorizations, etc.	19.06
BUSINESS SECTOR	PRODUCTION	Production businesses	27.64
	SERVICES	Businesses offering services	29.84
	TRADE	Businesses doing trade	24.05
	CONSTRUCTION	Construction businesses	16.86
	TRANSPORT	Transport businesses	1.61
EMPLOYEES	EMPL_1_10	Businesses with 1-10 employees	45.24
	EMPL_11_50	Businesses with 11-50 employees	33.67
	EMPL_51_200	Businesses with 51-200 employees	14.53
	EMPL_OVER_200	Business more than 200 employees	5.47
	EMPL_NOT_DEC	Businesses that have not declared any employees	1.10
COMPANY REGISTRATION	REG_LESS_5	New businesses registered less than 5 years ago	16.03
	REG_5_10	Businesses registered between 5 and 10 years ago	30.97
	REG_10_20	Businesses registered between 10 and 20 years ago	35.49
	REG_MORE_20	Businesses registered more than 20 years ago	17.51
LEGAL FORM	LIM_LIAB	Businesses of Limited Liability legal form	66.86
	PHYS_PER	Physical Person businesses	29.03
	OTHER_LF	Other legal form businesses, such as joint stock, nonprofit organizations, public entities and so forth	4.12
COMPANY TYPE	LARGE	Businesses registered as “large” in Tax Authorities	81.41
	SMALL_NO_VAT	Small businesses without VAT	14.45
	SMALL_VAT_OTHER	Small businesses with VAT or other types	4.14
REGION	SOUTHWEST_REGION	Businesses established in Berat/Fier/Vlorë	18.30
	CENTRAL_REGION	Businesses established in Tiranë/Durrës	51.55
	SOUTHEAST_REGION	Businesses established in Elbasan/Korcë /Gjirokastër/ Sarandë	16.57
	NORTH_REGION	Businesses established in Shkodër/Lezhë /Dibër/Kukës	13.58
RESULT	UDW	Inspection cases detecting undeclared work	5.81
	GREY	Inspection cases detecting grey work or undeclared overtime	15.46
	UDW & GREY	Inspection cases detecting undeclared AND grey work/ undeclared overtime	2.54
	OTHER_VIOL	Inspection cases detecting other violations related to other labour relation provisions or Occupational Health and Safety (OSH)	63.23
	NO_VIOL	Inspection cases that resulted in no violations	12.95

b. Machine learning datasets creation

Several interpretations can be drawn from the ratios of table 1, yet, most importantly, one can notice the high percentage of "other violations" resulting from inspection visits, which can be labour relations violations other than undeclared or grey work, or OSH violations. To target the different types of violations through machine learning predictive modelling, the models should "learn" from past experiences related to those specific types of violations so that they are able to distinguish between "right" and "wrong". Hence, if we aim at targeting undeclared work, we need to use a dataset to train our models that include both the inspection cases that detected undeclared work and those that found employers with no violations. The rest of the inspection cases deviating from our target violations are considered "noise" to our machine learning application and were omitted from the dataset. Correspondingly, if we aimed at targeting OSH violations, the models should be trained with a dataset containing both the relative OSH violations and the cases with no violations to help the models understand the differences between those two classes.

Under this reasoning and considering the scope of our research objectives, we created three different datasets to use for machine learning, building classifiers and evaluating their results: one for targeting undeclared work, one for grey work and one for both of these violations. Each one of these datasets is created by adding the inspection cases that detected at least one of the target violations to those of no violations. Hence, three ML datasets were created as given in table 2 below.

Table 2. Machine learning datasets

Table 2. Machine learning datasets						
	UDW		**GREY**		**UDW-GREY**	
	YES	NO	YES	NO	YES	NO
Total #	1,058	1,640	2,279	1,640	3,015	1,640
Ratio %	39.21	60.79	58.15	41.85	64.77	35.23

Table 3 illustrates the distribution per the attribute values of the inspection cases that revealed (a) undeclared work, (b) grey work and (c) no violations and reveals several interesting insights, for example, undeclared work is found mostly in the *services* business sector, whereas grey work is usually detected in the *production* sector. This practically enforces the reasoning for why we need to create different training datasets for machine learning and thus prediction models per target violation, since the violations in the labour market display differentiations in their attributes.

Table 3. Datasets to target Undeclared Work (UDW) and Grey Work (GREY)

Table 3. Datasets to target Undeclared Work (UDW) and Grey Work (GREY)				
Attributes	**Values**	**UDW**	**GREY**	**NO VIOLATIONS**
INSPECTION TYPE	**RA_PLANNED**	75.80	75.12	82.01
	OTHER_INSP_TYPE	24.20	24.88	17.99
BUSINESS SECTOR	**PRODUCTION**	25.90	34.53	28.29
	SERVICES	45.27	30.14	31.71
	TRADE	14.93	21.19	23.72
	CONSTRUCTION	13.04	12.51	15.06
	TRANSPORT	0.85	1.62	1.22
EMPLOYEES	**EMPL_1_10**	69.09	43.66	52.01
	EMPL_11_50	22.21	33.13	22.80
	EMPL_51_200	6.05	16.24	14.57
	EMPL_OVER_200	2.08	1.23	2.26
	EMPL_NOT_DEC	0.57	5.75	8.35
COMPANY REGISTRATION	**REG_LESS_5**	26.47	17.73	21.83
	REG_5_10	38.75	32.25	28.41
	REG_10_20	26.75	34.58	34.45
	REG_MORE_20	8.03	15.45	15.30
LEGAL FORM	**LIM_LIAB**	41.21	63.10	66.95
	PHYS_PER	58.03	33.30	26.77
	OTHER_LF	0.76	3.60	6.28
COMPANY TYPE	**LARGE**	53.97	81.00	75.18
	SMALL_NO_VAT	37.05	14.92	18.35
	SMALL_VAT_OTHER	8.98	4.08	6.46
REGION	**SOUTHWEST_REGION**	15.88	17.60	31.59
	CENTRAL_REGION	30.34	51.78	46.52
	SOUTHEAST_REGION	32.04	25.41	11.89
	NORTH_REGION	21.74	5.22	10.00
RESULT	**Cases with violation (number)**	1,058	2,279	1,640

c. Predictive modelling

As introduced, in this ML application we employ Associative Classification (AC) to construct the predictive models. AC is an advanced ML technique; it proves efficient with increased predictive accuracy compared to other DM and ML methods and offers an interpretability of results. Outputs interpretability is essential in this application domain for two main reasons. First, the models can offer understandable reasoning to the users of why a company should be inspected, thus improving their trust in the models' suggestions. Second, they provide a general knowledge of the prevalence and patterns of the targeted violations, hence developing labour inspectors' ability to easily identify such patterns.

Notably, the AC classifiers are comprised of a set of Class Association Rules (CARs), that is, readable, and understandable rules of the if-then form that are produced by a three-phase process: rule discovery; rule sorting; pruning to create the classifier; and testing for evaluation of the classification results. Several AC algorithms differentiating the three main phases have been proposed and used by the machine learning community. In this study, we use the *Classification Based on Associations* (*CBA*) (Liu et al. 1998), which has demonstrated high prediction accuracy and builds robust and compact classifiers.

The training and testing of the predictive models are based on the *Stratified 10-fold Cross Validation* method, which divides the dataset into ten stratified parts, each maintaining the same class distribution of the dataset. The nine parts are used for training the model, and the one part is used for testing. The process iterates along the ten parts of the dataset, using a different fold for testing each time and the remaining parts for training. At the end of the process, classification results for all the data instances of the dataset are collected and placed in the Confusion Matrix, illustrated in table 4.

Table 4. Confusion Matrix – classification results

Table 4. Confusion Matrix – classification results			
		PREDICTED	
		YES	NO
ACTUAL	YES	TP	FN
	NO	FP	TN

True Positives (TP) represent the positive cases (that is, the inspection cases that detected the target violation) correctly predicted by the classifier; whereas False Negatives (FN) are the positives falsely predicted as negatives. The negatives, on the other hand, (that is, the cases that detected no violations), when wrongly predicted as positives are named False Positives (FP), and when correctly classified, they are the True Negatives (TN).

d. Evaluation metrics

The values in the Confusion Matrix can be used to compute various prediction evaluation metrics well identified by the machine learning community, thereby enabling comparisons between the produced models. In this application, we employ the following common metrics.

Accuracy (*Acc*) represents the total prediction accuracy of the model – both of the positive and the negative data instances. It is computed as the ratio of the total correct classifications to all the classifications.

$$\mathrm{Acc} = (\mathrm{TP} + \mathrm{TN}) / (\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN})$$

Error rate (*Err*) is the complementary value of Accuracy, and it represents the ratio of the misclassifications.

$$\mathrm{Err} = (\mathrm{FP} + \mathrm{FN}) / (\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN})$$

Precision (*p*), or else *Positive Predictive Value* (*PPV*), focuses on evaluating the prediction performance of the positives since in most application domains, as also in the present study, the positive cases are of most interest to the researchers. Precision represents the model's *exactness* and is calculated as the ratio of positives correctly classified by the model to all the predicted positives.

$$\mathrm{p} = \mathrm{TP} / (\mathrm{TP} + \mathrm{FP})$$

Similarly, *Recall* (*r*), or else *Sensitivity* or *True Positive Rate* (*TPR*), corresponds to the ratio of positives correctly predicted by the model to all the actual positives and represents the model's *completeness*.

$$r = TP / (TP + FN)$$

Precision and Recall are complementary values; therefore, we also employ the F_1-score which is the harmonic mean of these two.

$$F_1\text{-}score = (2 * p * r) / (p + r)$$

Specificity (s), or the True Negative Rate (TNR), is also a significant evaluation measure in this study, corresponding to the ratio of true negatives to all actually negatives.

$$s = TN / (TN + FP)$$

e. Data engineering for enhanced positives prediction

In application domains where the correct prediction of positives is of more significance than the general prediction accuracy and the cost of misclassifying the positive values is considerably higher than the cost of negatives misclassification, data engineering techniques can be applied to the training datasets throughout the stratified 10-fold cross-validation process to enhance the prediction of positives at the cost, though, of increased false positives. Users should examine such approaches when there is a high difference in the misclassification costs of positives and negatives and investigate the outcomes of the different models produced by applying or not applying data engineering methods. Deciding which model fits best in their application domain also should take into consideration other details such as the human and financial resources available for proceeding with actions based on the models' results.

In the current application domain, the produced predictive models are used for planning onsite inspections. For example, the businesses classified as positives shall be suggested for inspection, whereas the ones predicted as negatives shall not be prioritized for an inspection visit. Consequently, misclassifying businesses involved in undeclared work to the negative class results in missed detection of undeclared work, which results in state revenue losses and missed social security contributions as well as disadvantageous positions for undeclared employees. On the other hand, compliant businesses misclassified to the positive class shall go through an unnecessary onsite audit, the cost of which can be counted in human and financial resources for an authority. From the view of the state, the misclassification cost of the positives, especially in the case of undeclared work prediction, is considered significantly higher compared to the cost of misclassifying the negatives; thus, in this study, we implement a simple yet effective data engineering approach that has proved to boost the prediction of the positives when undeclared work is involved (Alogogianni and Virvou 2022).

This approach refers to eliminating the negatives that overlap with positives in the training data space by removing from the training data instances belonging to the negative class that have the same values in all the attributes with a data instance belonging to the positive class and exists in the training dataset. In this way, we create well-identified class clusters in the training data that assist the generation of unambiguous CARs that promote the classification to the positive class more than to the negative. In the UDW dataset of table 2, the ratio of negative data instances overlapping with positives reaches more than 30 per cent, thereby significantly affecting the models' performance in effectively identifying undeclared work, as also illustrated in the following section that presents the models' results.

Results

Prediction performance

Using the datasets of table 2 plus the variation of overlaps elimination data engineering in the UDW dataset, we produced four different prediction models to target the UDW violations. Table 5 presents the confusion matrix results of the UDW Classifier produced by the UDW dataset, and table 6 exhibits the prediction outcomes of the UDW Classifier produced by the same dataset but with applying overlaps elimination in the training data. By examining the classification results, one may notice that the prediction of the actual positives is increased in the second classifier, yet this model also produces more false positives. These differences are illustrated in the evaluation metrics presented below in table 9.

Table 5. Confusion Matrix of the UDW Classifier

Table 5. Confusion Matrix of the UDW Classifier					
				PREDICTED	
	Total	**Ratio %**	**RESULT**	**YES**	**NO**
ACTUAL	1,058	39.21	YES	621	437
	1,640	60.79	NO	355	1285

Table 6. Confusion Matrix of the UDW Classifier (With no overlaps)

Table 6. Confusion Matrix of the UDW Classifier (With no overlaps)					
				PREDICTED	
	Total #	**Ratio %**	**RESULT**	**YES**	**NO**
ACTUAL	1,058	39.21	YES	872	186
	1,640	60.79	NO	834	806

Table 7 displays the confusion matrix of the classifier generated to target grey work, and table 8 illustrates the results of the model targeting both undeclared and grey work. Overlaps elimination in the training data is not applied in these cases because the ratio of the negatives in the datasets is already limited compared to the positives. For that reason, the practice of omitting further negative data instances from the training data would result in models classifying almost all the cases to the positive class. Indeed, this imbalance towards the positives affects the classifiers' predictions, which seem to favour the positive class producing several false positives. This effect is also depicted in the prediction evaluation metrics in table 9.

Table 7. Confusion Matrix of the GREY Classifier

Table 7. Confusion Matrix of the GREY Classifier					
				PREDICTED	
	Total #	**Ratio %**	**RESULT**	**YES**	**NO**
ACTUAL	2,279	58.15	YES	1822	457
	1,640	41.85	NO	954	686

Table 8. Confusion Matrix of the UDW-GREY Classifier

Table 8. Confusion Matrix of the UDW-GREY Classifier					
				PREDICTED	
	Total #	Ratio %	RESULT	YES	NO
ACTUAL	3,015	64.77	YES	2,590	425
	1,640	35.23	NO	1,147	493

Using the values in tables 5, 6, 7 and 8 and the metrics described in paragraph 4.d., the prediction performance metrics for all the models are calculated and summarized in table 9. Given that the models' positive predictions shall trigger onsite inspections, an initial evaluation of the models' successfulness should be evaluated based on the metric *Precision (p)* because this metric illustrates the success of the positive predictions. For example, the precision of the UDW classifier reaching 63.63 per cent translates to this success rate in undeclared work prediction if this model had been used for targeting undeclared work. Compared to the actual success ratio of 39.21 per cent of the UDW dataset, this classifier proves to offer a significantly improved performance rate. The precision of the UDW classifier built with no overlaps in the training data reaches 51.11 per cent, which proves less accurate in the positive predictions than the previous model, yet its success rate is still more than that of the actual dataset (39.21 per cent). Most importantly, though, it manages to identify 82.42 per cent of the actual cases of undeclared work (*Recall (r)*), which should not be neglected in situations when the authority aims at identifying as many of the actual undeclared work cases as possible in a targeted area. So, based on these outcomes, decision-makers should be able to identify which of the two models fits best their needs on each occasion, and correspondingly allocate their human and financial resources to targeted onsite inspections.

Table 9. Prediction performance metrics of the classifiers

Table 9. Prediction performance metrics of the classifiers						
Classifier	Acc	Err	Precision	Recall	F1-score	Specificity
UDW	70.64	29.36	63.63	58.7	61.07	78.35
UDW (no overlaps)	62.19	37.81	51.11	82.42	63.09	49.15
GREY	64	36	65.63	79.95	72.09	41.83
UDW_GREY	66.23	33.77	69.31	85.9	76.72	30.06

Similarly, the GREY classifier displays an improved positive prediction performance reaching 65.63 per cent compared to the actual rate of 58.15 per cent, and the last model targeting both undeclared and grey work proves efficient reaching almost 70 per cent compared to the actual of 64.77 per cent. The *Recall* rate of these models is also considerably high, reaching 80 per cent and 86 per cent correspondingly, yet one can notice that these models have not learned well to identify the negative cases and may trigger several unnecessary inspections. This can be seen as a deficiency of the current models due to the lack of a large number of inspection cases resulting in no violations, which can be improved in the future if more such cases are included in the training datasets. This performance could also be improved if we apply data engineering techniques towards decreasing the positive data instances in the training data space, such as random undersampling, and achieve a class balance between positives and negatives, an approach that shall be tested in our future research.

b. Interpretability

Apart from the increased prediction performance in detecting labour law violations, these models offer understandable outputs to the users: increasing their confidence in the classifiers' suggestions, enabling them to be involved in the inspection planning process and enhancing their knowledge regarding violation patterns in the labour market.

Table 10 exhibits some of the Class Association Rules generated by the UDW classifier, where one may identify the most prevalent patterns of attributes associated with undeclared work and those linked with no violations. For instance, a *small with no VAT* company offering *services*, employing *1 to 10 employees*, being *registered between 5 to 10 years*, and being established in the *southeast part* of the country is highly likely to be involved in undeclared work (first rule). However, a company offering services but employing more than 200 employees is less probably to be found with undeclared workers.

Table 10. Class association rules of the UDW Classifier

Table 10. Class association rules of the UDW Classifier	
Attribute values	**UDW**
SERVICES, EMPL_1_10, REG_5_10, SMALL_NO_VAT, SOUTHEAST_REGION	YES
OTHER_INSP_TYPE, SERVICES, EMPL_1_10, SOUTHEAST_REGION	YES
OTHER_INSP_TYPE, PHYS_PER, SMALL_NO_VAT, SOUTHEAST_REGION	YES
RA_PLANNED, SERVICES, REG_5_10, PHYS_PER, NORTH_REGION	YES
SERVICES, EMPL_1_10, PHYS_PER, SMALL_NO_VAT, SOUTHEAST_REGION	YES
OTHER_INSP_TYPE, SMALL_NO_VAT, SOUTHEAST_REGION	YES
SERVICES, EMPL_OVER_200	NO
RA_PLANNED, EMPL_OVER_200, REG_10_20	NO
TRADE, REG_MORE_20, CENTRAL_REGION	NO
OTHER_LF, SOUTHWEST_REGION	NO
EMPL_OVER_200, SOUTHWEST_REGION	NO
EMPL_OVER_200, CENTRAL_REGION	NO
PRODUCTION, OTHER_LF	NO
TRADE, EMPL_1_10, REG_10_20, LARGE, SOUTHWEST_REGION	NO
SERVICES, EMPL_51_200, LIM_LIAB	NO
TRADE, LIM_LIAB, LARGE, CENTRAL_REGION	NO
RA_PLANNED, REG_10_20, LIM_LIAB, LARGE, SOUTHWEST_REGION	NO

Lastly, table 11 illustrates some of the rules generated by the GREY classifier, where one may realize some differences in the patterns of undeclared and grey work. For instance, undeclared work is more prevalent in the services sector, whereas grey work can be found more often in the trade or the production business sectors. ASILSS planning based on the current risk assessment conducts most inspections in these three sectors. However, a higher degree of granularity of sectors might enhance our understanding about which subsectors are more prone to undeclared or grey work, thus supporting better inspection planning.

Table 11. Class association rules of the GREY Classifier

Table 11. Class association rules of the GREY Classifier	
Attribute values	**GREY**
TRADE, EMPL_1_10, REG_10_20, SOUTHEAST_REGION	YES
REG_10_20, PHYS_PER, LARGE, SOUTHEAST_REGION	YES
OTHER_INSP_TYPE, PRODUCTION, EMPL_11_50, LIM_LIAB, CENTRAL_REGION	YES
RA_PLANNED, TRADE, REG_10_20, SOUTHEAST_REGION	YES
OTHER_INSP_TYPE, PRODUCTION, EMPL_11_50, CENTRAL_REGION	YES
OTHER_INSP_TYPE, PRODUCTION, EMPL_51_200, LIM_LIAB, CENTRAL_REGION	YES
OTHER_INSP_TYPE, PRODUCTION, REG_5_10, LIM_LIAB, LARGE, CENTRAL_REGION	YES
SERVICES, REG_5_10, SOUTHEAST_REGION	YES
EMPL_11_50, REG_10_20, SOUTHEAST_REGION	YES
OTHER_INSP_TYPE, PRODUCTION, LARGE, SOUTHEAST_REGION	YES
OTHER_INSP_TYPE, PRODUCTION, EMPL_11_50, LIM_LIAB	YES
PRODUCTION, EMPL_11_50, REG_10_20, LIM_LIAB, CENTRAL_REGION	YES
OTHER_LF, SOUTHWEST_REGION	NO
OTHER_INSP_TYPE, PRODUCTION, LARGE	NO
PRODUCTION, EMPL_1_10, REG_LESS_5, LIM_LIAB	NO
OTHER_INSP_TYPE, EMPL_1_10, SOUTHWEST_REGION	NO
RA_PLANNED, SERVICES, EMPL_OVER_200	NO
RA_PLANNED, REG_10_20, LIM_LIAB, NORTH_REGION	NO
EMPL_1_10, REG_LESS_5, LIM_LIAB, SOUTHWEST_REGION	NO
EMPL_1_10, REG_LESS_5, SMALL_VAT_OTHER	NO
TRADE, EMPL_1_10, PHYS_PER, SOUTHWEST_REGION	NO
RA_PLANNED, EMPL_1_10, LIM_LIAB, LARGE, NORTH_REGION	NO
EMPL_1_10, LIM_LIAB, LARGE, NORTH_REGION	NO
TRADE, EMPL_1_10, SMALL_NO_VAT	NO

Planned inspections informed by the current risk assessment tool seem to be helpful with two enterprise profiles. One is related to undeclared work and includes physical person enterprises that have been operating between five and ten years in the service sector in the northern region. The other is related to grey work and includes enterprises that have been in operation between 10 and 20 years in the trade sector in the southeast region. Our modelling suggests that inspections triggered as result of any other reason except planning are better predictors of incidences of undeclared or grey work. Unplanned inspections may be better predictors because enterprises that have recorded previous violations (re-inspection), occurrences of accidents at work or employ sub-contractors (who are responsible for their own labour law compliance) might tend to violate the law more broadly, including not declaring or under-declaring their employees. Yet, 80.94 per cent of inspections are planned and only 23.81 per cent of them yield undeclared or grey work violations. These findings call for improvement of the current risk assessment tool and better understanding of unplanned inspections so that the knowledge generated from their understanding can be integrated more effectively into the RAS and planning process.

According to our findings, micro to small enterprises with up to ten workers have the highest probability of engaging in undeclared work and small enterprises have the highest probability of engaging in grey work compared to larger enterprises. Although, they comprise over 93 per cent (INSTAT 2021) of all businesses in Albania, only 45.24 per cent of ASILSS's total inspections targets micro enterprises, as shown in table 1. Small enterprises, which according to our model are more likely to engage in grey work, consist of 5.39 per cent of all enterprises, yet 34.33 per cent of inspected enterprises are from this category. Following the same pattern, 5.47 per cent of total inspections target enterprises with over 200 employees although they comprise less than 0.5 per cent of total enterprises (*Ibid.*). These findings suggest that a lot of ASILSS resources are going to inspections that most likely will yield no violations. This leads us to conclude that if the ASILSS were to switch its target from medium and large to smaller and micro-size enterprises, its inspections will uncover more cases of undeclared and grey work.

Another notable finding is that undeclared and grey work is more frequently uncovered in the southeast region. Its interpretation, however, is not straightforward. Interviews and data collected for evaluations of several development projects conducted by one of the authors of this study between 2010 and 2020 show that the southeast region populations have a stronger culture of compliance with the law than other regions.[3] It seems counterintuitive that a region with a higher compliance with the law will have more cases of non-compliance with UDW. The higher incidence of cases with undeclared or grey work might be also a result of labour inspectors in this region reflecting their region's law compliance culture through a more careful audit process. The available data, however, does not provide us with an explanation for this finding, and examination of other data is necessary to provide a logical reason for it. This is because if the finding that undeclared and grey work occurs more often in the southeast is not true, and if ASILSS puts more resources to conduct inspections in this region, then ASILSS's effectiveness and efficiency will be reduced. The fact that the central region is predicted with a higher propensity of undeclared and grey work based on data coming from unplanned inspection data is another reason for further exploration in this direction.

3 A culture of compliance can be measured as the rate of payment collection by the energy and water companies with the law or crime rate. Between 2010 and 2020, electricity and water collection rate in the southeast region is significantly higher than in other regions. There is no reason to believe that this has changed in the past two years. Furthermore, the crime rate in the southeast is lower than in all other regions (INSTAT 2021; authors calculations).

6. Conclusions

Data mining and machine learning can improve the effectiveness and efficiency of labour inspectorates by facilitating their leaders and managers to make smarter decisions faster regarding resource allocation and contributing toward achievement of their institutional objectives: reducing undeclared and grey work. DM&ML is also a useful tool for the accumulation and update of knowledge which can be used by senior management for strategic planning or by labour inspectors to do their jobs more effectively.

For its application, ASILSS should use its complete dataset (2019–2022) and continue to update it with more recent inspection data and other attributes from the Tax Authority aiming for improved prediction performance of the generated models. When faced with class imbalance between positives and negatives, such as in the UDW-GREY dataset, increased observations and application of data engineering methods in the training set can address it.

An evaluation of the results of the DM&ML approach to validate the outcomes and to understand why UDW and grey work occur more often in a given set of conditions than in others is in order. This information can be used to design and implement more effective approaches to address either issue.

Bibliography

Agrawal, Ajay, Joshua Gans and Avi Goldfarb. 2018. "Prediction Machines: The Simple Economics of Artificial Intelligence." *Harvard Business Press,* September.

Alogogianni, Eleni, and Maria Virvou. 2021a. "Data Mining for Targeted Inspections against Undeclared Work by Applying the CRISP-DM Methodology." In *2021 12th International Conference on Information, Intelligence, Systems & Applications (IISA)*. IEEE.

____. 2021b. "Addressing the Issue of Undeclared Work – Part I: Applying Associative Clasification per the CRISP_DM Methodology." *Intelligent Decision Technologies* 15 (4): 721–47. Available online: https://doi.org/103233/IDT-210213.

____. 2022. "Undeclared Work Prediction Using Machine Learning: Dealing with the Class Imbalance and Class Overlap Problems." In *2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA)*. IEEE.

____. 2023. "Handling Class Imbalance and Class Overlap in Machine Learning Applications for Undeclared Work Prediction." *Electronics* 12 (4): 913. Available online: https://doi.org/10.3390/electronics12040913.

Brnabic, Alan and Lisa M. Hess. 2021. "Systematic Literature Review of Machine Learning Methods Used in the Analysis of Real-World Data for Patient-Provider Decision Making." *BMC Medical Informatics and Decision Making* 21: 1–19. Available online: https://doi.org/10.1186/s12911-021-01403-2.

Bean, Randy. 2018. *How Big Data and AI Are Driving Business Innovation in 2018*. Cambridge: Massachusetts Institute of Technology. Available online: https://www.proquest.com/blogs-podcasts-websites/how-big-data-ai-are-driving-business-innovation/docview/1994096490/se-2.

Collins and Horodnic. 2020. "Trends in the Undeclared Economy and Policy Approaches. Evidence from the 2007, 2013 and 2019 Eurobarometer Surveys." Available online: https://www.ela.europa.eu/sites/default/files/2022-06/KE-02-20-972-EN-N_0.pdf.

Crews, C. J. 2019. "What Machine Learning Can Learn from Foresight: A Human-Centered Approach for Machine Learning-based Forecast Efforts to Succeed, They Must Embrace Lessons from Corporate Foresight to Address Human and Organizational Challenges." *Research-Technology Management* 62 (1): 30–33.

Davenport, Thomas H. 2018. *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*. Cambridge: MIT Press.

Duan, Yanqing, John S. Edwards and Yogesh K. Dwivedi. 2019. "Artificial Intelligence for Decision Making in the Era of Big Data – Evolution, Challenges and Research Agenda: SSIS." *International Journal of Information Management* 48 (10): 63. Available online: https://doi.org/10.1016/j.ijinfomgt.2019.01.021.

EC. 2019. Communication from the Commission (98) 219 final, 7 April 1998.

HAI. 2020. "How an Active Learning System Can Help Close the U.S. Tax Gap. Stanford: Stanford University Institute for Human-Centered Artificial Intelligence." Available online: https://hai.stanford.edu/news/how-active-learning-system-can-help-close-us-tax-gap.

ILO. 2002. Decent Work and the Informal Economy. Geneva: ILO. Available online: http://www.ilo.org/public/english/standards/relm/ilc/ilc90/pdf/rep-vi.pdf.

INSTAT. 2021. "Statistikat strukturore të ndërmarrjeve ekonomike." Available online: https://www.instat.gov.al/al/temat/industria-tregtia-dhe-sh%C3%ABrbimet/statistikat-strukturore-t%C3%AB-nd%C3%ABrmarrjeve-ekonomike/#tab2.

Kottemann, Jeffrey E., Fred D. Davis and William E. Remus. 1994. "Computer-Assisted Decision Making: Performance, Beliefs, and the Illusion of Control." *Organizational Behavior and Human Decision Processes* 57 (1): 26. Available online: https://doi.org/10.1006/obhd.1994.1002.

Liu, Bing, Yiming Ma, and Ching-Kian Wong. 2001. "Classification Using Association Rules: Weaknesses and Enhancements." In *Data Mining for Scientific and Engineering Applications; Massive Computing*, edited by Robert. L Grossman, Chandrika Kamath, Philip Kegelmeyer, Vipin Kumar, Raju R. Namburu, 591–605. Boston: Springer.

Lokanan, Mark, Vincent Tran, and Nam Hoai Vuong. 2019. "Detecting Anomalies in Financial Statements using Machine Learning Algorithm: The Case of Vietnamese Listed Firms." *Asian Journal of Accounting Research* 4 (2): 181–201. Available online: https://doi.org/10.1108/AJAR-09-2018-0032.

McKinsey Global Institute. 2018. "Notes from the AI Frontier: Modeling the Global Economic Impact of AI on the World Economy." Discussion Paper, September. Available online: https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/notes%20from%20the%20frontier%20modeling%20the%20impact%20of%20ai%20on%20the%20world%20economy/mgi-notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy-september-2018.pdf.

Miller, Steven. 2018. "AI: Augmentation, More So Than Automation." *Asian Management Insights* 5 (1): 1–20. Available online: https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1081&context=ami.

Nallathambi, Indumathi, Padmaja Savaram, Sudhakar Sengan, Meshal Alharbi, Samah Alshathri, Mohit Bajaj, Moustafa H. Aly and Walid El-Shafai. 2023. "Impact of Fireworks Industry Safety Measures and Prevention Management System on Human Error Mitigation using a Machine Learning Approach." *Sensors* 23 (9): 4365. Available online: https://doi.org/10.3390/s23094365.

NASDAQ OMX. 2017. "Nasdaq Launches New Machine Learning Technology for Surveillance Efforts on Nordic Markets: Nasdaq's SMARTS, in Collaboration with Nasdaq Nordic, Has Deployed Machine Learning Capabilities Which Analyze Abnormal Market Events and Improves the Alerting Processes Surrounding Market Abuse Detection." *NASDAQ OMX's News Release Distribution Channel*, 6 September.

NÁV. 2022. Online Invoice System. Budapest: National Tax and Customs Administration of Hungary. Available online: https://onlineszamla.nav.gov.hu/home.

NTCA. 2020. BR Roadmap 2020–2025. Amsterdam: Netherlands Tax and Customs Administration. Available online: https://www.sbr-nl.nl/english/what-is-sbr/sbr-roadmap-2020-2025.

PWC. 2022. PWC 2022 AI Business Survey. Available online: https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-business-survey.html.

Qian, Yingmiao, Shuhang Chen, Jianchang Li, Qinxin Ren, Jinfu Zhu, Ruijia Yuan and Hao Su. 2020.

"A Decision-Making Model Using Machine Learning for Improving Dispatching Efficiency in Chengdu Shuangliu Airport." *Complexity* 2020: 1–16. Available online: https://doi.org/10.1155/2020/6626937.

Rademaker, Thomas J., Emmanuel Bengio and Paul François. 2019. "Attack and Defense in Cellular Decision-Making: Lessons from Machine Learning." *Physical Review X* 9 (3) (July). Available online: https://doi.org/10.1103/PhysRevX.9.031012.

Remus, William E., and Jeffrey E. Kottemann. 1986. "Toward Intelligent Decision Support Systems: An Artificially Intelligent Statistician." *MIS* 10 (4): 403–18. Available online: http://www.jstor.org/stable/249197.

Rodríguez-Padial, Néstor, Marta Marín and Rosario Domingo. 2017. "An Approach to Integrating Tactical Decision-Making in Industrial Maintenance Balance Scorecards Using Principal Components Analysis and Machine Learning." *Complexity*: 15. Available online: https://doi.org/10.1155/2017/3759514.

Ruzgas, Tomas, Laura Kižauskienė, Mantas Lukauskas, Egidijus Sinkevičius, Melita Frolovaitė and Jurgita Arnastauskaitė. 2023. "Tax Fraud Reduction Using Analytics in an East European Country." *Axioms* 12 (3): 288. Available online: https://doi.org/10.3390/axioms12030288.

Shrestha, Yash Raj., Vaibhav Krishna and Georg von Krogh. 2021. "Augmenting Organizational Decision-making with Deep Learning Algorithms: Principles, Promises, and Challenges." *Journal of Business Research* 123: 588–603. Available online: https://doi.org/10.1016/j.jbusres.2020.09.068.

Tamás Czinege, Csilla. 2019. "Risk Management in order to Enhance Compliance of Taxpayers in Hungary." Budapest: IOTA. Available online: https://www.iota-tax.org/sites/default/files/documents/publications/IOTA_Papers/iota_paper_risk_analysis_in_hungary.pdf.

Tudoran, Ana A. 2022. "A Machine Learning Approach to Identifying Decision-making Styles for Managing Customer Relationships." *Electron Markets 32*: 351–374. Available online: https://doi.org/10.1007/s12525-021-00515-x.

List of acronyms and abbreviations

AC	Associative Classification
ASILSS	Albanian State Inspectorate of Labour and Social Services
CARs	Class Association Rules
CBA	Classification Based on Associations
DM&ML	Data mining and machine learning
ESAP 2	Employment and Social Affairs Platform 2
FN	False negatives
FP	False positives
GDPR	General Data Protection Regulation
MOP	Matrix of Penalties
OSH	Occupational Safety and Health
RAS	Risk assessment system
TN	True negatives
TP	True positives
UDW	Undeclared work

ilo.org

International Labour Organization
Route des Morillons 4
1211 Geneva 22
Switzerland