Engineering Applications of Artificial Intelligence 133 (2024) 108417

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Application of artificial intelligence as a knowledge creation instrument in tax procedures

Karri Koivula [a], Ahm Shamsuzzoha [b,c,*], Mohammad Shamsuzzaman [d]

[a] School of Technology and Innovations, University of Vaasa, PO Box 700, FI-65101, Vaasa, Finland
[b] School of Technology and Innovations, Digital Economy Research Platform, University of Vaasa, PO Box 700, FI-65101, Vaasa, Finland
[c] Daffodil International University, Daffodil Smart City, Birulia, 1216, Bangladesh
[d] Department of Industrial Engineering and Engineering Management, University of Sharjah, P. O. Box 27272, Sharjah, United Arab Emirates

ARTICLE INFO

Keywords:
Artificial intelligence
Machine learning
Neural networks
Self-organizing maps
Tax procedures

ABSTRACT

This study set out to find whether deep learning algorithms neural networks and self-organizing maps could be utilized in a value-adding way in the Finnish Tax Administration in the handling of income tax related claims by limited liability companies. According to research positive outcomes in artificial intelligence (AI) utilization have been attained outside Finland. The research was carried out according to the action design research method in which the focus of the research is concurrently building a suitable artifact for the organization and learning (design principles) from the creation and intervention itself. Research began with problem formulation followed by building, intervention, and evaluation. As a result, the project team consisting of three members created two functional artifacts: one based on neural networks, and another based on self-organizing maps. Creation of the artifacts was done in cycles as alpha, beta and gamma where alpha and beta were a neural network and gamma a self-organizing map. Alpha reached a macro average of 0.75–0.78 in classification and beta 0.77–0.79. Gamma gave a different point of view on the problem and was able to clearly identify the class's non-estimated customers in a topographical map. The artifacts were limited to function only as knowledge creation instruments due to legal and ethical limitations present in the context. Results suggest that it is recommendable to approach problems with more than one artifact. The preliminary results of this research were validated by applying the concept in a case organization, followed by an analysis of the results in an end-user setting.

1. Introduction

The Finnish Tax Administration's main task is to carry out the taxation related payments, tax control, and recovery of unpaid taxes. According to Finnish laws (Tax administration act (2010/502) 1 § and 2 §), this department is accountable to the ministry of finance. Currently, this department is suffering from long processing time to tax claims and making the necessary adjustment procedure. The adjustment procedure of claim refers to the processing of tax claims. Due to long processing times, the tax claims for adjustment are viewed as a bottleneck procedure. The long processing time is partially responsible for not filing the tax obligations within a set time limit by the company's taxpayers At the beginning of each year, a tax percentage is estimated which is deducted until a company returns a filled in tax return file and submit to the tax administration. The potential identification of such examples would be beneficial as knowledge to help pre-emptively tackle similar cases in the future.

Positive examples of various artificial neural network (ANN) solutions in tax administrations around the globe paved way for this research. Studies by Xiangyu et al. (2018), Pérez López et al., (2019); Chen et al., (2011) all were able to reach recognition rates between 80% and 89% of erroneous or fraudulent tax reports by using neural networks or other AI solutions. As it was with the study by Zhang et al. (2020) the impacts of utilizing methods like multi-modal deep neural networks can significantly increase the productivity of manual labour. Faúndez-Ugalde et al. (2020) deployed AI in Latin American countries, where the evidence shows that broad principles derived from each country's declaration of fundamental right allow taxpayers to protect their right to obtain tax related information. Rahayu (2021) studied the utilization of AI in tax audit in Indonesia and noticed that the benefits of deploying an AI system are unclear due to its novelty and lack of testing, making it difficult to forecast. Tax authorities in Australia, Canada, Norway, and the United Kingdom have created an AI-powered model to anticipate

* Corresponding author. School of Technology and Innovations, Digital Economy Research Platform, University of Vaasa, PO Box 700, FI-65101, Vaasa, Finland.
E-mail addresses: koivula.karri@gmail.com (K. Koivula), ahsh@uwasa.fi (A. Shamsuzzoha), mshamsuzzaman@sharjah.ac.ae (M. Shamsuzzaman).

https://doi.org/10.1016/j.engappai.2024.108417
Received 12 December 2023; Received in revised form 24 March 2024; Accepted 5 April 2024
Available online 25 April 2024

Abbreviations

ADR	Action Design Research
AI	Artificial Intelligence
ANN	Artificial Neural Network
DL	Deep Learning
ISOMAP	Isometric Mapping
LLC	Limited Liability Company
ML	Machine Learning
MLP	Multi-Layer Perceptron
PCS	Principal Component Analysis
SL	Supervised Learning
SOM	Self-Organizing Map
UL	Unsupervised Learning

significant-risk taxpayers for tax audits, where a person poses a significant risk regarding responding constructively to various tax-related operations to boost tax revenue (OECD, 2020).

The focus of this research was to create and test knowledge-creating artifacts in taxation environment. The secondary focus of this research was to do it so that laws, regulations, and ethics are considered from the beginning of the process. With the help of these artifacts the researcher aimed to create information that could help pre-emptively decrease the number of customers that are not fulfilling their tax reporting obligations. Lastly, this paper wants to encourage public and private organizations alike in Finland as well as elsewhere to bravely give a chance for AI solutions and see if your organization could benefit from it. This paper wanted to highlight that AI solutions are worthy options for analysing data without excluding ethics and legal perspectives.

Two different AI solutions based on ANN and self-organizing maps (SOM) referred to as IT artifacts, were built according to the action design research (ADR) research method. ANN and SOM were chosen due to their popularity in practical use cases and among researchers. Comparable peer algorithms were not created as the nature of the research was limited to testing these two algorithms with an open mind on the data. Encouraging previous studies regarding NN usage in different scenarios was a main drive for this research. The artifacts were created to function as pre-emptive knowledge-creating instruments as they were used to extract valuable information related to a bottleneck process in the tax administration (Finnish Tax Administration, 2021a). Additionally, the ADR research method generated design principles to make such artifacts in similar settings.

The research problem is whether AI could add value to taxation in the given problem domain. The research objective is to answer the research problem of how AI could be applicable in taxation in the limited liability companies' (LLC) claim for adjustment procedure. A suitable IT artifact would be created with the case organization to address the research problem. Performance and creation are afterwards analysed. Therefore, the task of this paper was to create an IT-artifact that would solve the research problem and meet the expectations of the research goal in accordance with the research objective.

Based on the above circumstances, this research study aimed to create a knowledge-creating artifact based on AI that could help an organization to decrease the number of estimated tax years pre-emptively and concurrently make a stronger sense of the potential usage possibility of such artifacts. Additionally, as the tax claim for adjustment procedure strictly follows by the predefined laws and regulations, the artifact and its creation focus on recognizing examples that go through the procedure and not participating in the decision-making of any sort. This study is interested in finding out if an AI solution could detect and recognize the underlying characteristics in corporate taxpayers that would explain the number of late filings.

Reducing the number of late filings would benefit the corporate taxpayers as they would receive their tax decisions more swiftly. The tax administration would like to improve the quality of its services and reduce the number of tax claims. This study investigates whether AI could add value to the taxation process and what challenges might it be faced when utilizing AI solutions in taxation procedure. This research focuses on a specific group of customers within the limited liability companies (LLC) claim for tax adjustment procedure. The creation of the artifact follows the ADR research method. The researcher formed research questions before the creation of the IT artifact began. Research questions are formulated as:

RQ 1: How can AI be deployed to the case organization in Finland to create value in its current taxation system?

RQ 2: How can AI be deployed in Finnish taxation system without violating its existing rules and regulations?

The first research question revolves around what kind of value and information AI creates in the case organization and how it can be achieved. Reasons and timing of the usage are considered and planned accordingly. The second research question focuses on challenges that AI usage should consider, such as trustworthiness, legality, and ethical restrictions, and how they are addressed. The rest of the article is organized as follows: Section 2 illustrates a theoretical framework that covers the basics of artificial intelligence, machine learning, deep learning, and artificial neural networks. Section 3 outlines the description of a limited liability company and its taxation procedure. The study methodology is described in Section 4, while study results are presented in Section 5. The overall study outcomes are concluded in Section 6 along with future research directions.

2. Theoretical framework

Nowadays, artificial intelligence, machine learning, and deep learning have become ubiquitous concepts in literature (Khandani et al., 2010; Kumar, 2017; de Laat et al., 2020; Bag et al., 2021). In the recent decades, the interest in their utilization opportunities in many sectors has significantly grown due to the exponential growth of computer power and the increased availability of data, allowing for more powerful and sophisticated information technology solutions. Moreover, technological maturity has also lowered the threshold, and various open-source libraries and active communities enable the utilization of various algorithms such as neural networks in practice (Pérez López et al., 2019; Zobeiry and Humfeld, 2021; Wang and Wang, 2022). Brief explanations of AI and its associated technologies and tools are explained in the following sub-sections.

2.1. Artificial intelligence

Artificial intelligence can be viewed as the replication of biological, analytical, and decision-making capabilities (Akerkar, 2019). It is often defined as the imitating of the science and engineering that augmenting human intelligence through artificial means and techniques to make intelligent machines (Alpaydin, 2014; Margaret et al., 2020). To be considered as intelligent, a system should be able to learn in a changing environment (Miller, 2019; Conati et al., 2021). Artificial intelligence-based artifacts are seen as value-adding, since the knowledge created by them can potentially save time, liberate resources, and expedite processes. Due to the legal and ethical limitations, the artifacts were limited to only knowledge creation instruments (Enholm et al., 2021; Muhlroth and Grottke, 2022). "Machine Learning is the systematic study of algorithms and systems that improve their knowledge or performance with experience" (Flach, 2012). According to Finlay (2018), the main ingredients that fuel most AI and ML applications include data input, data preprocessing, predictive models, decision rules, and output.

The issue of comprehending enormous data has resulted in the creation of new statistical methods, as well as the emergence of new fields such as data mining, machine learning, and bioinformatics. Many of these technologies share comparable foundations but are frequently expressed using distinct nomenclature (Hastie et al., 2009). LeCun et al. (2015) explained that DL techniques have made significant advancements in recognizing words, visual identification of objects, and recognizing objects. DL is a method of machine learning which depends extensively on the knowledge of human cognition, statistics, and practical mathematics as it has evolved over the past few years (Goodfellow et al., 2016). Due to exponential growth in computational power and the availability of a vast amount of data (big data), learnings methods such as ML and DL have become more attractive (Ryman-Tubb et al., 2018). In ML, we are interested in discovering patterns and useful approximations from data (Alpaydin, 2014). Data input can be almost anything from sensory inputs such as videos to filed online forms such as tax returns. Data preprocessing refers to turning data inputs into a computer-friendly format (Finlay, 2018).

Deep learning is viewed as a subfield of machine learning that utilizes multiple layered ANNs to solve problems (Mirjalili and Raschka, 2019; Fink et al., 2020). Instead of analysing data linearly, neural networks enable machines to process data nonlinearly (Alpaydin, 2014). At its core, DL divides the learning process into connected steps, also known as layers, that are assigned to different sections of the main problem available to the whole network (Rahman et al., 2020). According to Kelleher and Tierney (2018), the strength of DL models lies in their ability to utilize previously gathered knowledge from the previous layers to their advantage in the following layers, which is referred to as backpropagation. In backpropagation, previously accumulated feedback from events in the network is used in future calculations within the network (Rahman et al., 2020). The performance of different AI models such as ANN and self-organizing maps (SOM) requires tuning and defining the parameters of the algorithms (parameter tuning). This can be achieved by rigorously testing to find which parameters result in best results for the problem at hand. This was also the case in this paper.

Unsupervised learning (UL) includes only the input as output data is missing or excluded (Alpaydin, 2014, p.11). Unsupervised learning includes unlabelled or data of unknown structure. It is used to deduce important information from data to find patterns (Lee, 2019, p. 5). UL is suitable for finding regularities in the data and detecting naturally occurring groups, such as in the k-means clustering algorithm (Alpaydin, 2014, 11; Rahman et al., 2020, p. 21). An example of an UL method is SOM. SOM draws a topographical map of the data where similar observations are positioned closer, and an ordered representation of the data is created. As Kohonen (2013, p. 52–53) presented, Fig. 1 depicts the core idea of SOM: input data mapped out where Mc best represents X. Models (Mi) in the same circle are more like Mc than M to other observations on the map. Illustration of SOM is presented in Fig. 1.

2.2. Artificial neural networks

An ANN mimics the human brain and its functions; hence the neural in the neural network refers to biological neurons in the brain (Graupe, 2013). A neural network consists of layers that contain neurons that perform the required mathematical calculations (Rahman et al., 2020). The neurons in the layers together form a parallel and interconnected network as each of the layers and their neurons might connect (Rahman et al., 2020; Alpaydin, 2014). UL algorithm SOM is also considered a type of ANN (Kohonen, 2016). According to Mirjalili and Raschka (2019) feedforward neural network (FNN) is one of the prominent neural networks. Multilayer perceptron is a feedforward neural network.

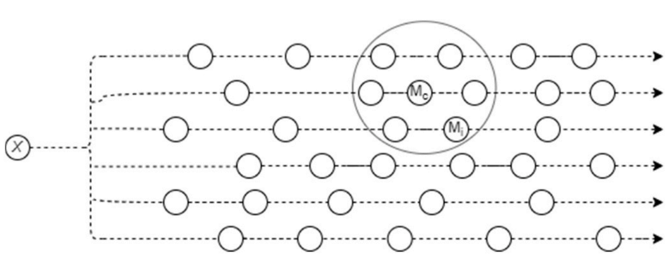

Fig. 1. Illustration of self-organizing maps (SOM).

2.2.1. Multilayer perceptron

The multilayer perceptron is a feedforward neural network model as all the connections move towards the output (Kelleher and Tierney, 2018). Perceptron refers to the ensemble of a neuron and its input connections and weights (Alpaydin, 2014). There are some constant parameters to consider when training a multilayer perceptron network, such as the number of hidden layers in the network as the increased number of hidden layers makes the network 'deep', activation function, or the calculation of a neuron's activation threshold, and the batch size or size of the data section is passed to the network in the training phase (Jung, 2018). Additionally, epochs or the number of passings of data passing through the network (Bayar et al., 2015) and learning rate or how quickly the network optimizes itself (Mirjalili and Raschka, 2019).

In Fig. 2 (adapted from Kelleher and Tierney, 2018), there are three layers of neurons: (1) input layer, A and B, (2) hidden layer C, D, and E, and (3) output layer F.

Neurons in a neural network are doing a set of operations.

1. Multiplying each input by a weight
2. Adding together the results of the multiplications
3. Pushing the result through an activation function

According to Kelleher and Tierney (2018), "all the connections between the neurons in a neural network are directed and have a weight associated with them." The weight applied to an input that a neuron receives is the weight on the connection coming to the neuron when the multi-input regression function over its inputs is calculated. As seen in Fig. 1, the flow of information in the network between the neurons is presented by arrows. The neural network in Fig. 2 is considered fully connected because each neuron is connected to all the neurons in the subsequent layer. The tags in the arrows reveal the weight that the neuron at the end of the arrow applies to the information passing through the connection. In Fig. 2, the calculation performed by neuron F of the network can be defined as:

$$Output = \varphi(\omega_{C,F}C + \omega_{D,F}D + \omega_{E,F}E)$$

* φ = activation function

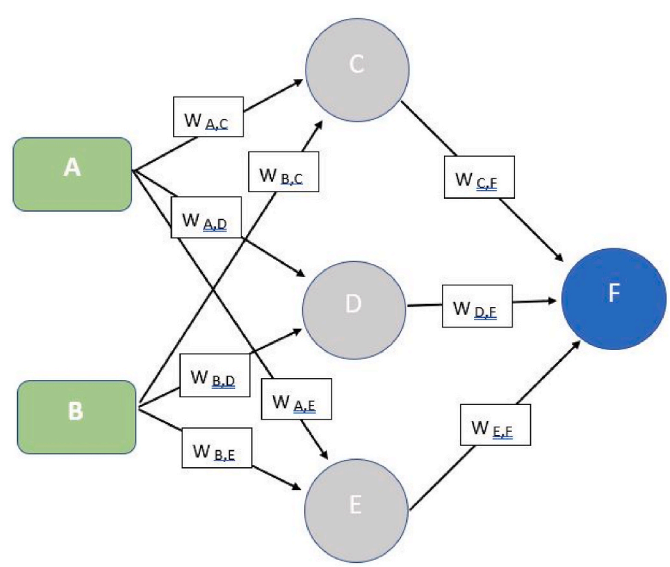

Fig. 2. Display of three layers of feedforward neural network (adapted from Kelleher and Tierney, 2018).

* *ω = weight applied to the neuron

2.2.2. Predicting future events in taxation using DL models

Tax officials in Finland have already utilized and expressed a growing interest in AI usage in taxation (Finnish Tax Administration, 2021b). Chen et al. (2011) developed an automatic detection model for discovering erroneous tax reports in their study. The study was motivated by the criticality of tax reporting and the large number of errors found in reports in recent years. Detecting erroneous tax reports is tedious and depends on experienced personnel. Therefore, the need for an automatic solution exists to reduce the workforce needed for the job. The model in the study by Chen et al. (2011) was carried out with various NN methods compared to each other. The different approaches were multi-layer perceptron's, learning vector quantization, decision tree, and hyper-rectangular composite neural networks methods." Data consisted of construction companies residing in Taiwan. No matter which NN approach was used, the correct recognition rate reached nearly 80 %. The best performing approach, hyper-rectangular composite neural network, was able to digest almost 250 valuable rules for identifying erroneous tax reports from the data.

Studies by Xiangyu et al. (2018) and Pérez López et al. (2019) focused on tax evasion. Xiangyu et al. (2018) developed a neural network model to tackle the issue of tax evasion in automobile sales enterprises in China. The NN-based recognition model's object was to determine behaviour related to tax evasion. Pérez López et al. (2019) utilized in their research an MLP neural network model to identify tax fraud concerning personal income tax returns in Spain. The result in both cases was a success. Xiangyu et al. (2018) reached a recognition accuracy of 89 %. The result was assessed with Receiver Operating Characteristic (ROC) curve, which showed that the classification effect was good. Pérez López et al. (2019) achieved an efficiency rate of 84.3%.

Moreover, the NN by Pérez López et al. (2019) offered information on the probability of each taxpayer's inclination to evade taxes. MLP is beneficial for classifying fraudulent/non-fraudulent taxpayers based on the results. The robustness of the model was confirmed with the ROC curve, which verified the NN's high predictive capacity.

A study by Rahimikia et al. (2017) focused on tax evasion with a more complex approach. In their study, Rahimikia et al. (2017) created a novel hybrid intelligent system to detect corporate tax evasion in Iran. Hybridity came from combining NN, SVM, and LR classification models with harmony search (HS) optimization algorithm, which is inspired by the improvisation process of musicians. The system was tested in the food and textile sectors. Researchers concluded that the system could accurately detect hidden patterns in tax returns that could point toward tax evasion. The results offer valuable, sector-wise information about the financial structure of tax evasion. The hybrid system is seen as a useful tool to detect tax evaders and an identifier of patterns suggesting tax evasion.

Additionally, tax officials have utilized NNs in social media. Zhang et al. (2020) developed a proof-of-concept NN to identify transaction-based tax-evading activities in the hidden economy of social media. Dataset consisted of 'Instagram posts about #lipstick and manually annotated sampled posts with multiple labels related to sales and tax evasion activities'. The purpose of the NN detection model was to identify suspicious social media posts. The posts deemed more suspicious by NN were analysed afterwards by tax officials. As the NN model identifies the suspicious posts, first, the productivity of manual work is improved from 22 percent to 72 percent. The NN model improves manual labour efficiency as the tax officers will not have to select the posts randomly (seeTable 1).

2.2.3. Performance evaluation metrics

Several ways exist to measure the performance of a neural network model. The performance of the neural network created is evaluated with the help of accuracy, precision, recall, and f1-score, all derived from the confusion matrix.

The confusion matrix (CM) presents the performance of a learning algorithm. CM is a square matrix that reports the count of true positive (TP), true negative (TN), false positive (FP), and false negative (FN), as presented in Fig. 3 (adapted from Rokach, 2009). Table 2 illustrates how they are calculated (Adapted from Giussani, 2020).

3. Taxation procedure of limited liability company (LLC) in Finland

This section describes the basic principles of an LLC in Finland and an overview of its taxation procedure It also focuses on automation, monitoring, and ethical principles for AI in the Finnish Tax Administration.

3.1. Principles of LLC's taxation

According to the Finnish limited liability companies act (2006/624), chapter 1, 1 §, subsection 1, LLC is a separate taxpayer. The tax rate for corporations is 20% (Income tax act 124 §, subsection 2). According to the act on assessment of assets in taxation ((2005/1142) 2 §, subsection 1), net worth (positive/negative) for a non-public LLC is calculated by subtracting the total amount of liabilities from the total amount of

Table 1
Various use cases of DL models in the taxation system.

Reference	Authors	Year	Problem	Method(s) used	Results obtained
Application of neural networks for detecting erroneous tax reports from construction companies	Chen et al.	2011	Decreasing the amount of erroneous tax reports with automation to free personnel for other tasks	Multilayer perceptron's, learning vector quantization, decision tree, hyper-rectangular composite NN	Almost 80% recognition rate achieved with all of them
Intelligent Identification of Corporate Tax Evasion Based on LM Neural Network	Xiangyu et al.	2018	Detection of tax evasion activities in automobile sales enterprises in China	LM neural network	Recognition rate of 89% achieved
Tax Fraud Detection through Neural Networks: An Application Using a Sample of Personal Income Taxpayers	Pérez López et al.	2019	Detection of tax evasion in personal income tax returns in Spain	MLP neural network	Recognition rate of 84% achieved
Detecting corporate tax evasion using a hybrid intelligent system: A case study of Iran	Rahimikia et al.	2017	Detection of corporate tax evasion in Iran	NN (multilayer perceptron), SVM, LR classification models with harmony search optimization algorithm (hybrid intelligent system)	MLP combined with HS optimization algorithm reached 90.07% and 82.45% accuracy, 85.48% and 84.85% sensitivity, and 90.34% and 82.26% specificity in the food and textile sectors
Detecting Transaction-based Tax Evasion Activities on Social Media Platforms Using Multi-modal Deep Neural Networks	Zhang et al.	2020	Identification and filtration of tax evasion related posts related in social media	Multi-modal deep neural network	NN filtrated posts and directed suspicious one to personnel for manual inspection. Productivity of manual work went from 22 % to 72 %

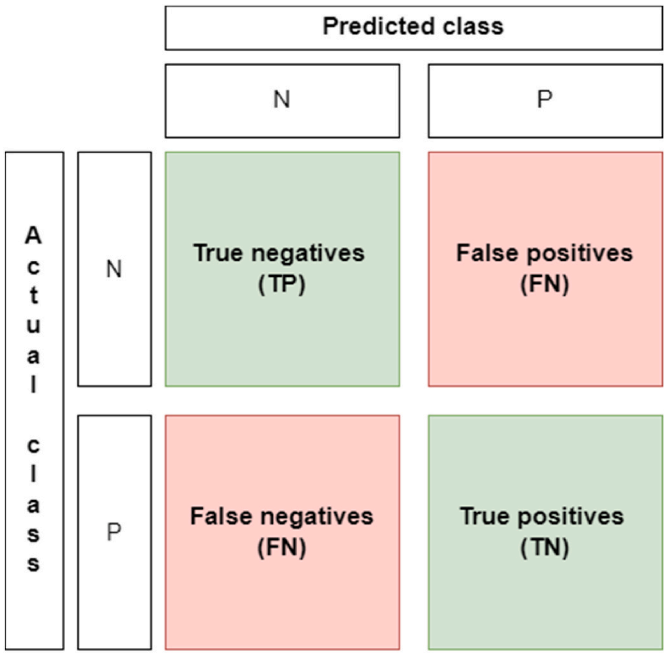

Fig. 3. Display of confusion matrix (adapted from Rokach, 2009).

Table 2
Display of calculation of evaluation metrics (Adapted from Giussani, 2020).

Metric	Calculation	Definition
Accuracy	$\frac{TP + TN}{FP + FN + TP + TN}$	Sum of correct predictions divided by all predictions.
Precision	$\frac{TP}{TP + FP}$	Correctness of the model per class. Useful in imbalanced class problems.
Recall	$\frac{TP}{FN + TP}$	The number of correctly evaluated instances per class is divided by all the correct examples in the class. Useful in imbalanced class problems.
F1-score	$\frac{2 \text{ x Precision x Recal}}{\text{Precision} + \text{Recall}}$	A balanced combination of precision and recall.

assets.

3.2. Tax assessment procedure for LLC

LLCs are obliged to fill out and return their income tax returns four months after their accounting period has ended. An LLC which is neglecting its obligation results in taxation being estimated by the tax administration. Estimated taxation is a result of an LLC not returning its tax return. Tax administration must send a hearing to the taxpayer of the estimation to do this. Taxation will be estimated if the reporting obligation is not fulfilled within the time reserved in the hearing. (Tax assessment act 7 §, 8 §, and 27 §; Tax administration's decision on reporting duties and notes (A123/200/2016)).

For corporations, the taxation period ends at the latest ten months after the end of their tax year closing month (= end of the accounting period). If the taxpayer has not filed their tax return within ten months, the tax decision is done based on the estimation that is made by the tax administration. To adjust a closed tax year, an adjustment claim is required. Processing time for a closed tax year is 12 months (Tax assessment act 49 § and 61 §; Finnish Tax Administration, 2021).

3.3. Automated taxation in Finland

The automated decision-making in Finnish taxation system is necessary due to the immense amount of tax-related work. Many tax-related procedures are carried out automatically without revising by a tax official (Finnish Tax Administration, 2020). Automation is directed to undisputed matters, which are not selected to manual control and could be solved without consideration. The tax administration does not utilize artificial intelligence in automation of tasks that require consideration and decision-making (Finnish Tax Administration, 2020).

3.4. Ethical principles for AI in the Finnish Tax Administration

In 2018, the Finnish Tax Administration released its ethical principles for AI. The Finnish Tax Administration joins in the pursuit of attaining ethical and trustworthy AI. According to the Finnish Tax Administration (Finnish Tax Administration, 2019), ethical principles is considered in all decision-making considering AI. The ethical principle of AI is consisted of four main principles: (1) reliable data, (2) a human is always responsible, (3) AI follows laws and regulations, and (4) tax administration takes part in public discussion on responsible and ethical AI applications (Finnish Tax Administration, 2019c).

4. Study methodology

This section presents justification of chosen methodology, data used, how ADR was utilized to form the IT artifact(s), how the creation proceeded, how the IT artifact performed, what was learned, and what kind of design principles came out of the process. This project aimed to create a suitable IT artifact for the case organization, the Finnish Tax Administration, according to the ADR research method.

4.1. Justification of methodology

This study aimed to create a suitable IT artifact for the Finnish Tax Administration as the case organization. To fulfil the aim, this study is chosen action design research (ADR) as a research method. This research method was chosen due to its flexibility, authenticity, and organization centricity (Sein et al., 2011). As a result, this study created two functional artifacts: one based on neural networks, and another based on self-organizing maps. Neural networks were chosen based on results from promising previous studies and self-organizing maps was chosen during the process of creating the artifacts as just one type of artifact was not seen enough. Project team decided that the research would only focus on neural networks and self-organizing maps other than PCA, kernel PCA, IOMAP, etc., which was enough to get the study results successfully. However, other classics methods such as PCA, kernel PCA, Laplacian Eigenmaps, ISOMAP, and Diffusion maps, the author could enhance the analysis by generating additional numerical results and comparing them with a broader range of methods, rather than solely relying on self-organizing maps. No additional artifact types were considered or created.

In this study, the ADR method is used to focus on building innovative IT artifacts in the organizational settings, while simultaneously learning from the intervention and assessing it concurrently (Alter, 2015). This method is seen as applicable when the establishment of an "in-depth understanding of the artifact–context relationship is needed to develop a socio-technical design agenda for a specific class of problems" (Sein et al., 2011).

In this study, the ADR project team consisted of three professionals working on different tasks within the case organization: the researcher who also worked as a tax specialist focusing on limited liability companies and their income taxation, an analytics expert whose role within the organization focuses on guiding data analytics and claims of adjustment procedure representative who oversees the claims of adjustment procedure within the organization. The researcher was responsible for the research paper and creation, testing and presentations of the IT artifact to other team members. The analytics expert and claims of adjustment procedure representative had supportive roles by guiding and offering valuable comments during the creation phases of the artifacts from their perspective within the organization. They did not participate in the hands-on creation of the paper and the artifacts. Based on the comments, the researcher adjusted the artifacts and tried out different parameter settings. Project team members also helped the researcher to attain the data to be used in the creation of the artifacts.

Several meetings and altogether three development cycles occurred during the study period to create a suitable IT artifact that could assist the procedure by offering new insights. As a result of utilizing the ADR in the case organization's restricted and authentic setting, the study concluded by creating two different IT artifact solution concepts namely: NN as an SL type of solution and SOM-based algorithm as an UL type of solution. Moreover, preliminary design principles also emerged from the ADR process.

4.2. Data collection

The study data collected consisted of private limited liability companies in Finland. The data included the tax years 2017–2019 of the companies conducting solely business activities under the act on business income only. The companies with personal and agricultural income sources were excluded. Data for tax years 2017, 2018, and 2019 consisted of information from 94 889 companies in the original dataset and 203 617 companies in the final dataset including missing data that was not originally available. Data gathered from the tax system included specific information about the companies per tax year, as presented in Table 3.

4.3. Problem formulation

As one of the researchers was an employee processing the claims in the LLC's claims of adjustment procedure, he had a piece of firsthand knowledge of the problem in its context. The binary nature of the problem (has been estimated or not) and the actual need to decrease the amount of processing time in the claim for adjustment procedure resulted in the researcher presenting an idea to create an IT artifact based on AI algorithms to tackle this issue. Long-term commitment from both the researcher and the organization was secured with a contract. Roles and responsibilities in the ADR team were set at the beginning of the project.

To create such an artifact, the researcher was handed anonymous data on LLCs from tax years 2017–2019. Data selection was performed intuitively by professionals working in the problem area who have developed a thorough understanding of the problem in its context. Intuitively selected means that it was decided beforehand for example to include only data of companies in the business income source since they are most likely more comparable to one another. Fig. 4 depicts the different competences of the team members in the project.

The IT artifact was not intended as a decision-making instrument but as a knowledge creation instrument due to legal and ethical limitations to what an automated solution is allowed to do and what is expected from an AI solution within the problem context. The information provided by the artifact would only be used proactively.

The researchers and the organization were interested in how well the artifact could perform as a knowledge-creating pre-emptive instrument. The IT artifact manufactured was not meant to be handed over to the organization. The artifact was tested on the data received from the organization, and its performance was analysed with the ADR team. Fig. 5 presents the blueprint for the BIE plan in the project.

5. Results analysis

After being handed the data, the authors began creating the artifact. The artifact aimed to recognize the estimated tax year of different companies from non-estimated ones as efficiently as possible. The initial knowledge creation target was to determine if such technologies could potentially be used for taxation benefits. The performance was monitored with precision, recall, and f1-score. Running times were not monitored as the focus was solely on the other metrics.

IT artifact in this project was a piece of code analysing data with DL libraries to deduce a correct outcome (supervised learning). The creation of the IT artifact started from scratch and was mainly based on trial and error with Keras (2021). The characteristics of the NN were modified and tested in cycles. The data available was trimmed down to find out the variables that had the most substantial impact on the result. Trimming refers to finding out which information that influences the end results. By taking out variables such as main line of business the researcher aimed to find which information influences the end results. The results from alpha, beta and gamma were attained with the following variables: total taxable business income, total tax-deductible business costs, and net worth. Alpha, beta, and gamma were results of a process of trial and error in which each of them was afterwards presented to other team members for comments and insight. Based on the comments the artifacts were modified and tested and subsequently presented again for team members.

Keras (2021) is a DL application programming interface written in Python on the ML platform TensorFlow. The code was written with Spyder IDE (Spyder, 2021), a scientific Python development environment. The first functional version of the IT artifact, the alpha version, and its results were analysed with the project team. The results are presented in Tables 4–6.

5.1. Alpha

Table 4, Table 5, and Table 6 present the results for the alpha performance with the tax years 2017, 2018, and 2019 test sets respectively.

Table 3
Description of specific information of the collected dataset.

1. The main line of business is a 2-digit code expressing to which specific business sector the company belongs
2. Starting date in taxation
3. Starting date in the trade register
4. Closing date in taxation [a]
5. Closing date in trade register [a]
6. Reason for closing the trade register [a]
7. Home municipality
8. Tax year (2017, 2018, or 2019)
9. Net sales per year (€)
10. The total taxable business income per year (€)
11. Purchases, variation in stocks and inventory per year (€)
12. Total tax-deductible business costs per year (€)
13. Assets total per year (€)
14. Liabilities total per year (€)
15. Taxation estimated at some point (yes/no)
16. Taxation still estimated (yes/no)

[a] = information expressed if available.

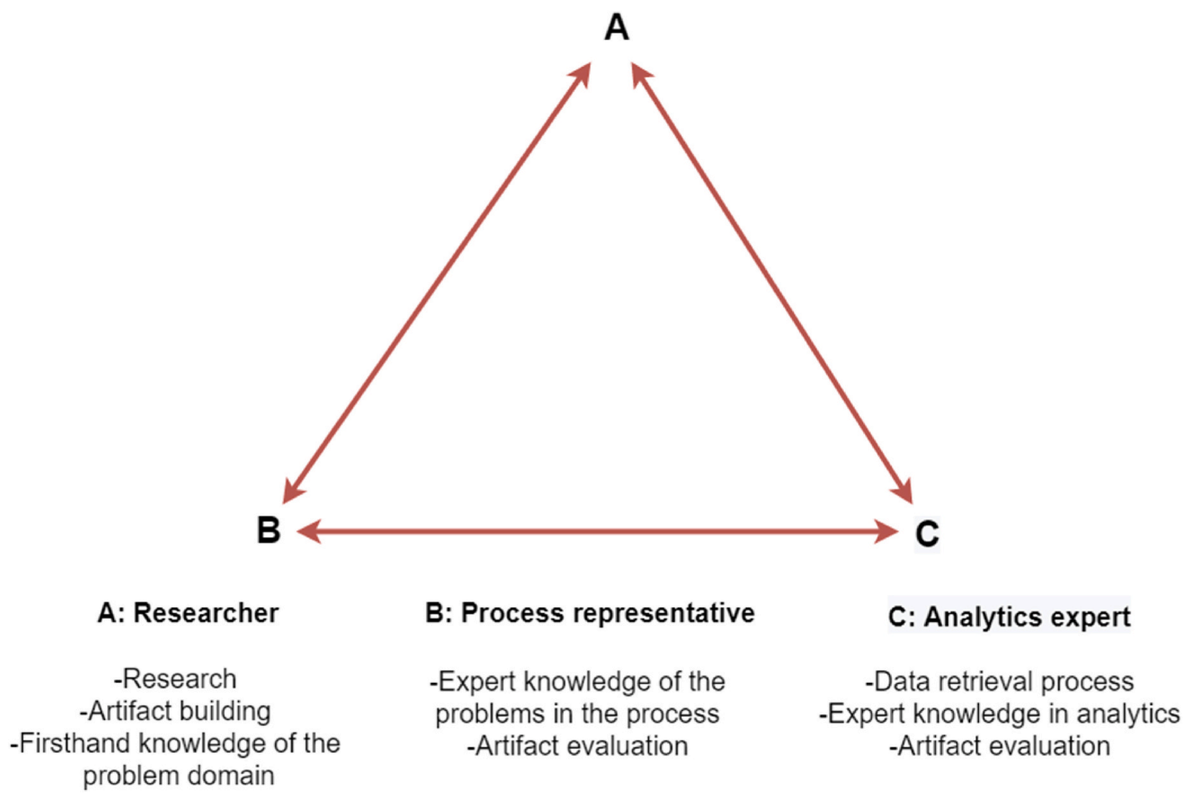

Fig. 4. BIE viewpoints in the project.

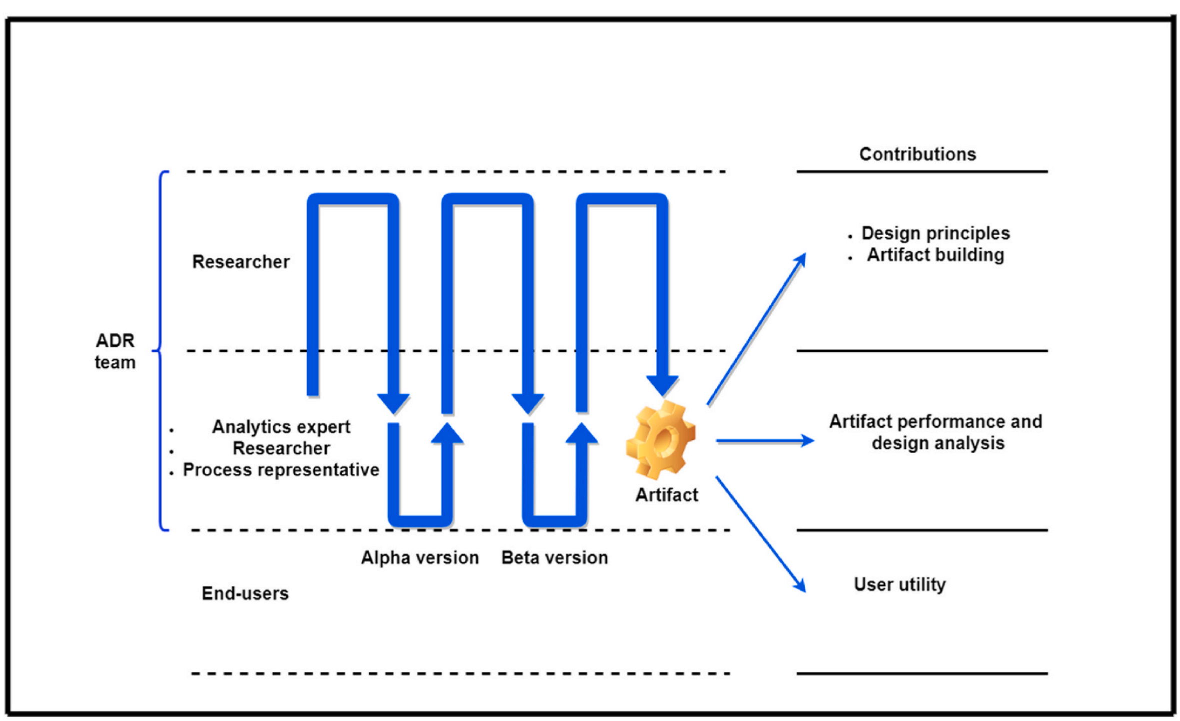

Fig. 5. Original IT-dominant BIE plan in the tax administration ADR project.

Precision, recall, and f1-score are calculated for the non-estimated companies (0) and estimated companies (1). The number of companies in each group is presented under "Number of companies." The macro average refers to the average of the NN's performance in both classes. Tables 5 and 6 are structured the same way as Table 4. They present the performance results attained with the artifact created with 2017 test data.

As the data was highly imbalanced, it proved to be an arduous task to get the neural network to recognize the estimated companies' class (1) as well as possible. During the alpha phase different balancing methods for the dataset were tested but they did not influence the end results. The precision, recall, and f1-score for the 0-class were excellent in the alpha-

Table 4
Alpha, performance, 2017 test set.

Estimation status for the tax year	Precision	Recall	F1-score	Number of companies
0	0.97	0.98	0.98	26822
1	0.61	0.55	0.58	1645
macro average	0.79	0.76	0.78	28467

Table 5
Alpha, performance, 2018

Estimation status for the tax year	Precision	Recall	F1-score	Number of companies
0	0.98	0.98	0.98	93062
1	0.58	0.57	0.58	4920
macro average	0.78	0.77	0.78	97982

Table 6
Alpha, performance, 2019

Estimation status for the tax year	Precision	Recall	F1-score	Number of companies
0	0.98	0.98	0.98	96388
1	0.48	0.56	0.52	3677
macro average	0.73	0.77	0.75	100065

version. However, the estimated class (1) results were not on an acceptable level. The neural network's f1-score for the estimated class achieved 0.52–0.58, which was not on par with the f1-score of the 0-class's 0.98. Reaching at least a moderate f1-score was set as a target for the beta version.

The project team analysed the performance of the alpha version and concluded that further refinements were required, and new approaches were suggested as well as different testing scenarios.

5.2. Beta

The researcher selected three different approaches to be tested in creating the beta version. Alpha-version worked as a starting point for the beta. The test plan for the beta version can be seen in Table 7.

Minor improvements to the model's performance were achieved compared to alpha. Tables 8–10 present the result for 2017, 2018, and 2019 respectively in the same way as the results for the alpha version were previously presented.

Table 7
Project team's test plan for the beta.

Test approach	Results
Creation of new artificial variables from data	Variable 1 Is the company passive and empty (no assets, liabilities, sales, purchases, or other taxable activity)? Yes (1) or no (0). Using only variable 1, the NN achieved similar results as the alpha version. Variable 2 Is the company passive but has 10 000 or more in assets? Yes (1) or no (0). This variable did not affect the performance, and acceptable results were not achieved using only this variable or having it as an additional variable.
Creating the artifact from the tax year 2018's data and tested on 2019 (leaving 2017 out altogether)	By leaving out data from the tax year 2017, no improvements were achieved nor significant drops in performance.
Eliminating unnecessary variables from the data	By eliminating variables, a slightly better performance was achieved. The remaining variables: •total taxable business income, •total tax-deductible business costs, and •net worth

Table 8
Beta, performance, 2017 test set.

Estimation status for the tax year	Precision	Recall	F1-score	Number of companies
0	0.98	0.98	0.98	17909
1	0.62	0.59	0.60	1069
macro average	0.80	0.78	0.79	18978

Table 9
Beta, performance, 2018

Estimation status for the tax year	Precision	Recall	F1-score	Number of companies
0	0.98	0.98	0.98	93062
1	0.58	0.60	0.59	4920
macro average	0.78	0.79	0.78	97982

Table 10
Beta, performance, 2019

Estimation status for the tax year	Precision	Recall	F1-score	Number of companies
0	0.99	0.97	0.98	96388
1	0.49	0.64	0.55	3677
macro average	0.74	0.81	0.77	100065

The increased performance in recognizing the estimated companies (class 1) can be seen in the increased f1-score from alpha to beta in all the tax years from 2017 to 2019. Table 11 presents how f1-score increased during the artifact development from alpha to beta. F1-score %-increase refers to the relative increase in performance from alpha to beta.

The most encouraging finding was that the model's performance was maintained and slightly increased by eliminating variables. Throughout the testing phases in alpha and beta versions, the NN recognized the companies that were not estimated (returned their tax returns) with an f1-score of 98%. It was expected because the data was strongly imbalanced towards the non-estimated class. Moreover, as pointed out by the claim of the adjustment procedure representative, it has always been a challenge to recognize the estimated companies from others in practice.

During the testing phase of the beta version, it was discovered that a problem had occurred in the data retrieval process. As a result, fifty percent of the data had been missing. It was decided that no additional development cycles would take place, but instead, the finalized artifact would be tested with the whole dataset including the missing data as well as the original data which now doubled the amount of data. The dataset was still largely imbalanced between the classes.

Tables 12–14 present the results with the beta version of the artifact (no parameters changed) in 2017, 2018, and 2019 respectively, but utilizing the final dataset. Results are presented in the same way as in alpha and beta.

No significant improvements were achieved even though the amount of data was doubled, as seen in Table 15. As an additional point of view

Table 11
Artifact's performance increased from alpha to beta.

Version	Tax year	Precision	Recall	f1-score	f1-score %-increase
Alpha	2017	0.61	0.55	0.58	–
Beta	2017	0.62	0.59	0.60	3,5%
Alpha	2018	0.58	0.57	0.58	–
Beta	2018	0.58	0.60	0.59	1,7%
Alpha	2019	0.48	0.56	0.52	–
Beta	2019	0.49	0.64	0.55	5,77%

Table 12
Beta with additional data, performance, 2017 test set.

Estimation status for the tax year	Precision	Recall	F1-score	Number of companies
0	0.97	0.98	0.98	37320
1	0.64	0.57	0.60	2208
macro average	0.81	0.77	0.79	39528

Table 13
Beta with additional data, performance, 2018

Estimation status for the tax year	Precision	Recall	F1-score	Number of companies
0	0.98	0.98	0.98	193477
1	0.59	0.59	0.59	10140
macro average	0.78	0.79	0.79	203617

Table 14
Beta with additional data, performance, 2019

Estimation status for the tax year	Precision	Recall	F1-score	Number of companies
0	0.99	0.98	0.98	200284
1	0.51	0.63	0.56	7580
macro average	0.75	0.80	0.77	207864

Table 15
Artifact performance comparison.

Version	Tax year	Precision	Recall	f1-score	f1-score %-increase
Beta	2017	0.62	0.59	0.60	–
Beta (with all the data)	2017	0.64	0.57	0.60	0%
Beta	2018	0.58	0.60	0.59	–
Beta (with all the data)	2018	0.59	0.59	0.59	0%
Beta	2019	0.49	0.64	0.55	–
Beta (with all the data)	2019	0.51	0.63	0.56	1,8%

for analysing the problem, the analytics expert suggested that analysing the data with a self-organized maps algorithm could provide valuable insight into the problem domain from UL's perspective. It was decided that another development cycle (gamma) would be conducted by testing the performance of a SOM algorithm on the data. Therefore, changes to the original BIE plan were made. Changes and actualized BIE cycles are presented in Fig. 9.

5.3. Gamma

Based on the characteristics related to the beta version, an additional cycle was conducted that would utilize a UL method to tackle the issue of recognizing estimated companies. The artifact was built with a SOM library (Minisom, 2022) created by Vettigli (2018).

In Figs. 6–8, five points representing the green and red squares and the white/green points were selected. Green squares represent points where there are no estimated tax years. A white point indicates that the data differentiates from the rest. Red circles represent companies whose tax year was estimated. A point including red and green indicates that a clear distinction between the two was not attained. The income, expenses, and net worth values represent the median of the values in the specific points selected. Numbers in the X and Y axes represent the coordinates of each point. For example, in Fig. 6, the white point with a green circle has coordinates x = 11 and y = 3. The mentioned points specific company-related information in median values is presented in

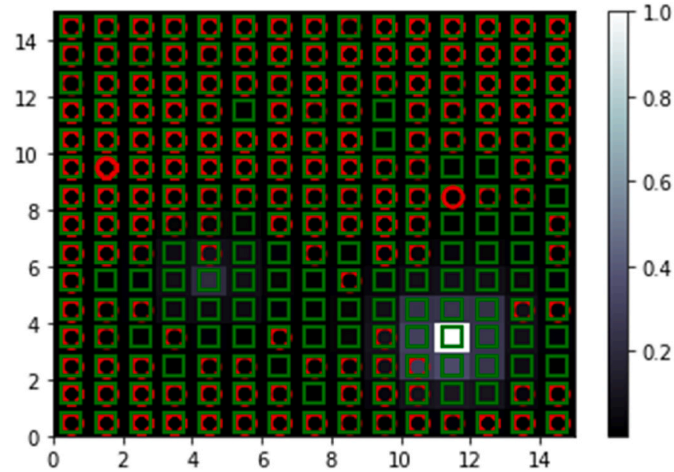

Fig. 6. Self-organizing map (SOM, 2017).

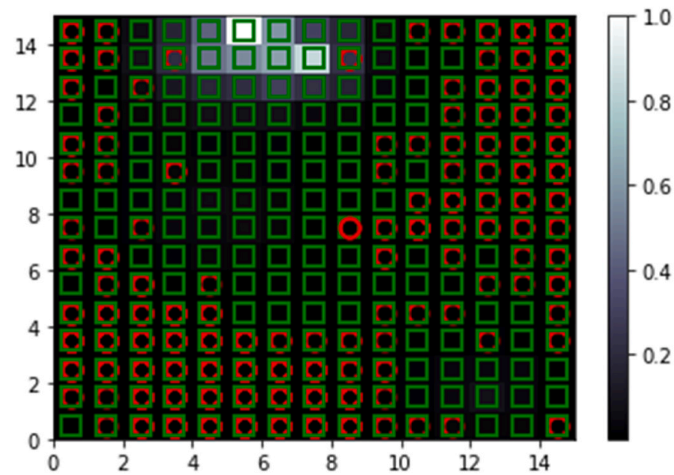

Fig. 7. Self-organizing map (SOM, 2018).

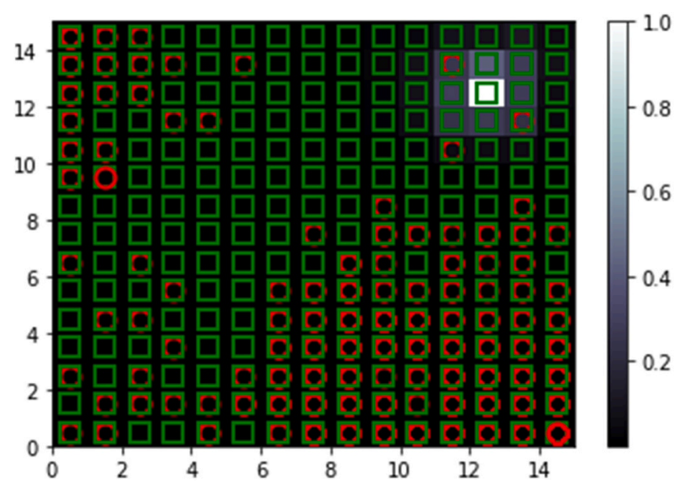

Fig. 8. Self-organizing map (SOM, 2019).

Table 15.

Based on the SOM map of the tax year 2017 (Fig. 6, Table 16) and the points selected, it is evident that the SOM algorithm does perform well at recognizing the non-estimated class. Income, expenses, and net worth were the variables that had the most effect on the result and therefore

Table 16
Datapoints, 2017.

Point	Coord.	Income	Expenses	Net worth	Firms
G	x=2, y=5	75 379 €	64 361 €	20 005 409 €	77
G	x=5, y=5	641 946 €	2 117 371 €	142 718 888 €	19
G	x=12, y=6	12 855 435 €	11 769 776 €	9 159 407 €	86
G	x=5, y=11	2 205 864 €	1 563 443 €	1 661 337 €	178
G	x=9, y=11	1 751 034 €	1 571 442 €	674 004 €	505
R	x=1, y=9	171 €	974 €	24 944 €	5072
R	x=11, y=8	3 137 743 €	2 800 569 €	1 534 862 €	169
R	x=14, y=14	- €	- €	116 €	30737
R	x=13, y=12	52 574 €	51 841 €	4 936 €	6915
R	x=4, y=9	223 652 €	212 214 €	29 866 €	5123
W	x=11, y=3	825 686 950 €	748 791 489 €	163 640 054 €	76

the others were discarded. Companies with high income and expenses and significant net worth fall in the green area. The area containing white represents part of the data that differs from the rest. The estimated class is not distinguishable. However, the areas that include red are less in income and expenses and more minor in net worth. The largest number of firms fall in points (14, 14) where the median income, expenses, and net worth are low.

The SOM map performs similarly with tax years 2018 and 2019 data (Figs. 7 and 8 respectively). More active and wealthy companies fall into the non-estimated class, and smaller and non-active tend to fall into the red/green areas (Tables 17 and 18). The white areas are prone to represent the largest and most active companies. Areas including red are lesser in income and expenses and more minor in net worth. The largest number of firms for the tax year 2018 fall in point (5, 0), where the median values of income, expenses, and net worth are low. For the tax year 2019, this is point (14, 0). A clear distinction between whether a company with lesser income, expenses net worth type falls into an estimated class is not attained.

SOM algorithm performed similarly to the beta version. The non-estimated class was detected more efficiently than the estimated class. The SOM algorithm provided information that narrows down where the estimated companies are more likely to occur.

The BIE cycle was concluded with the gamma version (Fig. 9). It is beneficial to develop and test more than one potentially suitable artifact. Results from beta and gamma provide information that an AI artifact has potential as a pre-emptive, knowledge-creating analysis instrument. Moreover, information created by two different algorithms strengthens and ratifies one another. End-user participation and testing in practice were absent from the project and they are left for future research projects. This project's AI-empowered analysis algorithms, the beta and gamma versions, are seen as potential tools for decreasing irregular taxation-related behaviour (Sein et al., 2011).

5.4. *Reflection and learning*

The artifact creation followed the principles laid out by ADR, considering project-related restrictions in time and scope. The researcher's dual roles (the research itself and the artifact's creation) slowed down the process. Having another person responsible for the actual artifact creation would have benefited the project. The project would have required more focus on setting up roles and responsibilities. However, the communication within the project team functioned well. As a result of the reassessment conducted after the beta version evaluation, a modification to the original BIE plan was added during the

Table 17
Datapoints, 2018.

Point	Coord.	Income	Expenses	Net worth	Firms
G	x=5, y=11	49 867 261 €	47 189 423 €	12 304 960 €	24
G	x=9, y=12	12 249 065 €	11 009 212 €	11 809 686 €	28
G	x=12, y=2	15 426 474 €	10 993 777 €	53 671 393 €	34
G	x=4, y=8	65 621 127 €	64 488 262 €	8 151 510 €	161
G	x=10, y=6	1 701 876 €	1 520 295 €	613 394 €	468
R	x=6, y=1	36 290 €	35 013 €	6 693 €	12708
R	x=3, y=0	1 639 €	3 061 €	167 178 €	3356
R	x=14, y=14	296 326 €	353 774 €	- 164 577 €	1027
R	x=8, y=7	5 060 955 €	4 592 170 €	2 161 655 €	120
R	x=5, y=0	- €	514 €	900 €	47328
W	x=5, y=14	1 453 362 342 €	1 442 106 548 €	246 206 271 €	42
W	x=7, y=13	20 411 718 €	5 592 390 €	933 404 590 €	36

Table 18
Datapoints, 2019.

Point	Coord.	Income	Expenses	Net worth	Firms
G	x=12, y=9	20 802 743 €	20 321 578 €	2 493 762 €	176
G	x=11, y=8	13 804 352 €	13 442 689 €	1 511 917 €	236
G	x=14, y=11	59 812 933 €	59 114 367 €	8 017 102 €	154
G	x=14, y=8	6 515 351 €	6 395 230 €	1 859 024 €	38
G	x=5, y=12	4 409 458 €	3 795 789 €	43 377 505 €	34
R	x=1, y=9	1 527 105 €	1 314 091 €	2 234 482 €	191
R	x=8, y=4	1 060 912 €	1 047 197 €	68 098 €	1391
R	x=13, y=1	44 207 €	41 660 €	9 563 €	12120
R	x=14, y=1	16 894 €	16 563 €	9 210 €	15569
R	x=14, y=0	- €	264 €	1 276 €	48569
W	x=12, y=12	1 465 951 470 €	1 435 013 955 €	309 162 510 €	41

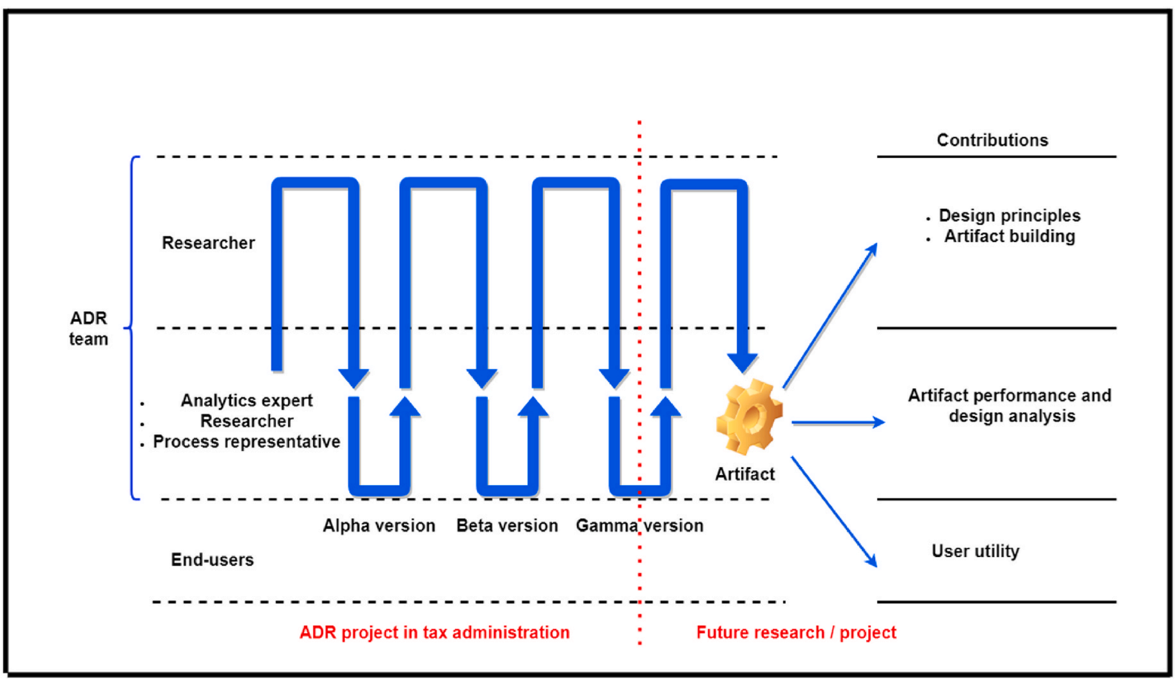

Fig. 9. Actualized IT-dominant BIE cycles in the project.

project.

The concept for the artifact stemmed from theory (theory-based artifact), and its creation was motivated by practice, thus creating a link between the two. Concurrent and authentic analysis of the artifact's performance was vital for the process. Everybody involved in the project could suggest and affect the direction of the artifact development.

The objective of the project was to create an IT artifact based on AI capable of identifying companies not returning their tax returns. After being handed the data, the researcher started creating the code in Python, following Keras guidelines. It was decided to analyse the model's performance with precision, recall, and F1-score. The initial artifacts creation ended with the beta version.

The development process from alpha to beta and gamma benefited from interventions as insightful ideas to test out helped shape the form of the artifact. As a result, from an intervention, the idea of creating an entirely different artifact for the same problem was initialized, resulting in the artifact's gamma version.

Gamma achieved similar results to beta by recognizing non-estimated companies and having issues drawing a distinct line between estimated and some non-estimated companies. The creation of gamma pinpoints the need to approach problems with more than one AI algorithm allowing users to attain knowledge from several different algorithms that build on one another. End-user testing was left out since the project was concluded with a gamma version. More development would be required to reach a finalized artifact that could be deployed to end-users.

5.5. Formalization of learning

The fourth stage in the ADR project required a change from specifics to generalization, divided into three levels according to the research method. The problem that ignited the project is that of a classification utilizing historical data (a generalization of the problem instance). In the context of this research, it refers to the artifacts ability to distinguish the customers that do not return their tax return from others.

The presented solution to the generalized problem is an AI-empowered instrument that is fed historical data and based on that, attempts to classify the customers into those that do return and those that do not return their tax returns. This insight provided by the artifact is used in decision-making. The knowledge-creating instrument offers a basis for pre-emptive actions as suggested in this project (a generalization of the solution instance). (Sein et al., 2011). The design principles were formed based on the answers to the research questions and how the artifact performed in practice. The research questions answered.

RQ1. How can AI be deployed to the case organization to create value in its current taxation system?

Value is created in the form of time, speed, and liberated resources. AI-empowered instruments can analyse and form inferences significantly faster than a human could. Therefore, such a solution would free time for human(s) to concentrate on more urgent matters. It is left for human(s) to analyse the results provided by AI to see if it is applicable. The ADR project in this study proved that a NN and a SOM algorithm could detect estimated companies even though room for improvements was left.

As the SOM analysis proved, AI is also better at detecting patterns and segmenting customers into corresponding sectors. Added value can be created by utilizing AI to help decision-making if boundaries on where and how to use the information are well examined. Moreover, it is required that AI does not function as a decision-maker, and a human is always behind every tax-related decision.

The information provided by AI in the problem context either provides specific information on companies within a particular sector as in the SOM analysis (descriptive) or information on whether the company will be estimated (imperative). To gain the most out of AI usage, it is suggested to approach a problem with more than one artifact. Doing this makes it possible to attain a more comprehensive view of the problem and its potential solution.

RQ2. How can AI be deployed so that it does not violate rules and regulations?

Challenges in using AI in taxation can be divided into three levels (preliminary set of design principles).

- Trustworthiness through accuracy

To be utilized as a knowledge creation instrument, adequate performance and accuracy are required and expected. Thorough testing, analysis, and continuous refinement are mandatory. Unless an organization-defined acceptable performance for an AI solution is not achieved, the solution should be discarded, and a new approach should be taken.

- Legal and ethical restrictions and limitations of use

Legal and ethical perspectives should always be considered when developing an AI solution. According to the Finnish Tax Administration's ethical principles for AI: AI should only use reliable data, follow laws and legislation, and constantly be monitored and managed by a human. Legal and ethical matters need to be addressed and considered when building AI solutions. They set restrictions, expectations, and requirements for an AI solution. A transparent and regulated AI solution is expected.

- Justification of usage

A preliminary inspection on where to use AI solutions should be conducted to deduce if significant improvements are achievable. Restrictions and limitations of use should always be considered. Problems with a lot of data available might make a desirable use case for AI. Organizations should prepare a few different AI approaches to tackle an issue that could potentially be solved entirely or partially with AI. Additionally, as the cost of using AI has decreased, organizations ought to have a low threshold for experimenting with them.

Lastly, seven meetings were held concerning the ADR project. Presentation and evaluation of alpha and beta versions formed the project's core. Outcomes achieved in the project were shared with the organization, including the gamma version and its findings. Encouraging results pave the way for future projects and the development of similar solutions. Dissemination of results was left out as the ADR process was not finished, and a finalized product was not created.

A summary of the ADR process focusing on creating a pre-emptive artifact is shown in Table 19. Table 19 is an adapted Table based on Sein et al. (2011).

Table 19
Summary of the ADR process (adapted from Sein et al., 2011).

Summary of the ADR Process in the pre-emptive artifact for tax project		
Stages and Principles		Artifact
Stage 1: Problem Formulation		
Principle 1: Practice-Inspired Research	Practical challenges in the case organization and the willingness to explore novel solutions worked as a launching point.	*Recognition*: Interest in utilizing AI solutions in taxation has grown. A transparent and holistic approach is vital. Simple problems with plenty of data are potential use cases such as recognizing estimated companies.
Principle 2: Theory-Ingrained Artifact	The overall theory related to ML and DL and their successful implementations in business created the foundation for the project.	
Stage 2: BIE		
Principle 3: Reciprocal Shaping	The team reciprocally solved problems and created development ideas. The data selected was mutually accepted. Members of the team could suggest development ideas.	***Alpha Version***: Initial version based on NN's created. ***Beta Version***: A modified second version of the NN artifact. No significant performance improvements.
Principle 4: Mutually Influential Roles	The ADR team consisted of researchers and practitioners to include theoretical, technical, and practical viewpoints. The researcher and artifact creator worked as an employee in the case organization.	***Gamma Version***: Third cycle with a new artifact utilizing UL algorithm SOM provided encouraging results.
Principle 5: Authentic and Concurrent Evaluation	Alpha & beta were evaluated with the project team but not released to a broader user setting. Development stopped with gamma.	
Stage 3: Reflection and Learning		
Principle 6: Guided Emergence	Authentic performance evaluation resulted in improvements (alpha to beta) and a new approach to the problem (gamma).	***Current Version and Realization***: A more comprehensive approach to solving AI problems is required. Further development is expected to arrive at testing in a wider setting within the organization.
Stage 4: Formalization of Learning		
Principle 7: Generalized Outcomes	Preliminary design principles were created.	***Ensemble Version***: Artifact and its creation are affected by the team through its observations and organizational practices, legal and ethical considerations, limitations, and restrictions.

6. Conclusions and future study recommendations

This research approached the subject of AI utilization in the Finnish Tax Administration from a holistic point of view and provided answers to the two identified research questions by implementing the ADR method. Research began as the needs of the case organization's interest in AI applications and their usage potential in taxation. Consequently, the study presented the idea of utilizing AI as an analysis tool in the LLC. As a result, two artifacts and a set of design principles emerged. The created artifacts functioned well enough to be considered helpful as knowledge-creating instruments. By limiting the role AI to knowledge-creation the risks related to AI usage are significantly reduced as the responsibility of decision-making is left to people. It is worth noting that to interpret the newly produced knowledge the personnel should have experience of AI applications as well as taxation procedures.

The artifacts can create value in the form of knowledge, time, speed, and liberated resources. The instruments analyse data faster than a human being, allowing humans to focus on more urgent matters. AI-based artifacts are investigative objects in the case organization and its activities.

Due to a holistic viewpoint, a better perception of the problem's context was reached while creating the artifacts with the case organization. According to the research method, even unexpected signals requiring drastic changes to the artifact were followed, resulting in an early version of a different kind of artifact instead of the original one. The emerging design principles emphasize the requirements to be filled to create useful and value-creating artifacts. In the case organization's context, AI usage was limited to knowledge creation with substantial human supervision requirements due to limitations. Based on the results of this research study, it is strongly recommended to test out different AI solutions in organizations to understand whether they can offer value of any kind.

The task was to create an IT artifact based on AI capable of identifying companies not returning their tax returns. The researcher created the code in Python, following Keras guidelines. The performance of alpha and beta was analysed with precision, recall, and F1-score. The performance of alpha and beta was seen as sufficient: alpha reached a macro average of 0.75–0.78 in classification and beta 0.77–0.79. Concurrent and authentic analysis and the involvement of all the projects members in the artifact creation process was vital. Gamma achieved similar results to beta by recognizing non-estimated companies and having issues drawing a distinct line between estimated and some of the non-estimated companies.

The created artifacts that worked on top of AI libraries were intuitively selected. To find more suitable libraries, a more profound comparison between different coding libraries would be required. The research process concluded after the third cycle (gamma), and preliminary design principles were created based on the results attained until that point. Moreover, the research would have benefited from testing scenarios in end-user settings and further development cycles for the artifacts to gain more organization-specific information.

It is also observed that the challenges associated with AI in taxation in legal and ethical boundaries need to be addressed when creating and using the artifacts. Challenges can be divided into three levels that also form the design principles derived from the ADR method: (1) trustworthiness through accuracy, (2) consideration of legal and ethical restrictions and limitations of use, and (3) justification of use. Different regulative demands most likely occur in organizations. Therefore, a proper preliminary inspection in the problem area is essential after considering the restrictions and limitations. Suggestions for future studies include but are not limited to profound research on AI implementation possibilities in taxation, securement of ethical principles in taxation while utilizing AI, and legal perspective on AI usage in taxation. In addition, a comparison of how different AI solutions could enhance automated taxation in Finland could act as a starting point for a future study.

It should also be noted that the dataset consisted of various companies in different sectors of business which could have a substantial effect on the results. Different companies operating in different markets might not be entirely comparable. Testing artifacts separately within similar or same main lines of businesses might result in more accurate results as opposed to testing artifacts on various types of companies in different markets and comparing them to one another.

Suggestions for future studies include but are not limited to profound research on AI implementation possibilities in taxation, securement of ethical principles in taxation while utilizing AI, legal perspective on AI usage in taxation and comparison of broader range of DL models and their performances to one another such as principal component analysis (PCA), kernel PCA etc. In addition, a comparison of how different AI solutions could enhance automated taxation in Finland as well as abroad could act as a starting point for a future study. As a recommendation the amount and quality of data of limited liability companies is vital. The information (data parameters to feed for NN for example) to choose as the basis require expertise and insight. Experts working in the field of taxation might have a strong understanding of which characteristics might have a strong impact on the end results. Moreover, additional methodology to deployment of AI in taxation can be deployed in USA, Europe, and Asia though qualitative method through data collection techniques with guided interviews with tax auditors, tax information system experts and documents review of specific tax office in a region or country.

CRediT authorship contribution statement

Karri Koivula: Writing – original draft, Project administration, Methodology, Data curation, Conceptualization. **Ahm Shamsuzzoha:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Conceptualization. **Mohammad Shamsuzzaman:** Writing – review & editing, Visualization, Formal analysis.

Declaration of competing interest

There is no conflict of interest to prepare this article.

Data availability

Data will be made available on request.

References

Akerkar, R., 2019. Artificial Intelligence for Business. Springer, Cham.
Alpaydin, E., 2014. Introduction to Machine Learning, third ed. The MIT Press, Cambridge, Massachusetts.
Alter, S., 2015. The concept of 'IT artifact' has outlived its usefulness and should be retired now. Inf. Syst. J. 25, 47–60.
Bag, S., Gupta, S., Kumar, A., Sivarajah, U., 2021. An integrated artificial intelligence framework for knowledge creation and B2B marketing rational decision making for improving firm performance. Ind. Market. Manag. 92, 178–189.
Bayar, N., Darmoul, S., Hajri-Gabouj, S., Pierreval, H., 2015. Fault detection, diagnosis and recovery using Artificial Immune Systems: a review. Eng. Appl. Artif. Intell. 46, 43–57.
Chen, J., Su, M., Chen, C., Hsu, F., Wu, C., 2011. Application of neural networks for detecting erroneous tax reports from construction companies. Autom. ConStruct. 20, 935–939.
Conati, C., Barral, O., Putnam, V., Rieger, L., 2021. Toward personalized XAI: a case study in intelligent tutoring systems. Artif. Intell. 298, 103503.
de Laat, M., Joksimovic, S., Ifenthaler, D., 2020. Artificial intelligence, real-time feedback and workplace learning analytics to support in situ complex problem-solving: a commentary. Int. J. Inform. Learn. Tech. 37, 267–277.
Enholm, I.M., Papagiannidis, E., Mikalef, P., Krogstie, J., 2021. Artificial intelligence and business value: a literature review. Inf. Syst. Front. https://doi.org/10.1007/s10796-021-10186-w.
Faúndez-Ugalde, A., Mellado-Silva, R., Aldunate-Lizana, E., 2020. Use of artificial intelligence by tax administrations: an analysis regarding taxpayers' rights in Latin American countries. Comput. Law Secur. Rep. 38, 105441.
Fink, O., Wang, Q., Svensen, M., Dersin, P., Lee, W.J., Ducoffe, M., 2020. Potential, challenges and future directions for deep learning in prognostics and health management applications. Eng. Appl. Artif. Intell. 92, 103678.

Finnish Tax Administration, 2019. Ethical Principles of AI. Retrieved 20.2.2022 from. https://www.vero.fi/en/About-us/finnish-tax-administration/operations/responsibility/finnish-tax-administrations-ethical-principles-for-ai/.
Finnish Tax Administration, 2020. Automaattinen Päätöksenteko Verotuksessa. Retrieved 26.3.2022 from. https://www.vero.fi/tietoa-verohallinnosta/verohallinnon_esittely/tietosuoja-ja-julkisuus/tietojen_kasittely_ja_tietosuoj/automaattinen-paatoksenteko/.
Finnish Tax Administration, 2021. Processing Times for Returns and Applications. Retrieved 26.3.2021 from. https://www.vero.fi/en/About-us/contact-us/processing-times-for-returns-and-applications/.
Flach, P., 2012. Machine Learning: the Art and Science of Algorithms that Make Sense of Data. Cambridge University Press, Cambridge.
Giussani, A., 2020. Applied Machine Learning with Python. In: ProQuest Ebook Central. EGEA Spa - Bocconi University Press. Retrieved 26.8.2021 from. https://ebookcentral-proquest-com.proxy.uwasa.fi/lib/tritonia-ebooks/detail.action?docID=6110795.
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT press.
Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2. Springer, New York, pp. 1–758.
Jung, A., 2018. Machine Learning: Basic Principles. https://doi.org/10.48550/arXiv.1805.05052.
Kelleher, J.D., Tierney, B., 2018. Data Science. The MIT Press, Cambridge, Massachusetts.
Keras, 2021. About Keras. Retrieved 15.10.2021 from. https://keras.io/about/.
Khandani, A.E., Kim, A.J., Lo, A.W., 2010. Consumer credit-risk models via machine-learning algorithms. J. Bank. Finance 34, 2767–2787.
Kohonen, T., 2013. Essentials of the self-organizing map. Neural Network. 37, 52–65.
Kumar, S.L., 2017. State of the art-intense review on artificial intelligence systems application in process planning and manufacturing. Eng. Appl. Artif. Intell. 65, 294–329.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Lee, Wei-Meng, 2019. Python Machine Learning. In: ProQuest Ebook Central. Retrieved 19.1.2022 from. https://ebookcentral-proquest-com.proxy.uwasa.fi/lib/tritoniaebooks/detail.action?docID=5747364.
Miller, T., 2019. Explanation in artificial intelligence: insights from the social sciences. Art Int. 267, 1–38.
Minisom, 2022. Minisom 2.3.0. Retrieved 15.06.2022 from. https://pypi.org/project/MiniSom/#files.
Mirjalili, V., Raschka, S., 2019. Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-Learn, and TensorFlow 2, third ed. Packt Publishing. Retrieved 20.10.2021 from. https://ebookcentral-proquestcom.proxy.uwasa.fi/lib/tritonia-ebooks/detail.action?docID=6005547.
OECD, 2020. The Impact of Big Data and Artificial Intelligence (AI) in the Insurance Sector. Retrieved from. www.oecd.org/finance/Impat-BigData-AI-in-theInsurance-sector.htm.
Pérez López, C., Delgado Rodríguez, M., de Lucas Santos, S., 2019. Tax fraud detection through neural networks: an application using a sample of personal income taxpayers. Future Internet 11, 86. MDPI AG.
Rahayu, S.K., 2021. Utilization of artificial intelligence in tax audit in Indonesia. Manage. Account. Rev. 20, 135–157.
Rahimikia, E., Mohammadi, S., Rahmani, T., Ghazanfari, M., 2017. Detecting corporate tax evasion using a hybrid intelligent system: a case study of Iran. Int. J. Account. Inf. Syst. 25, 1–17.
Rahman, M., Islam, D., Mukti, R.J., Saha, I., 2020. A deep learning approach based on convolutional LSTM for detecting diabetes. Comput. Biol. Chem. 88, 107329.
Ryman-Tubb, N.F., Krause, P., Garn, W., 2018. How Artificial Intelligence and machine learning research impacts payment card fraud detection: a survey and industry benchmark. Eng. Appl. Artif. Intell. 76, 130–157.
Sein, M., Ola, H., Sandeep, P., Matti, R., Rikard, L., 2011. Action design research. MIS Q. 35, 37–56.
Spyder, 2021. Overview of Spyder. Retrieved 20.2.2021 from. https://www.spyder-ide.org/.
Vettigli, G., 2018. MiniSom: minimalistic and NumPy-based implementation of the self organizing map. GitHub. Retrieved 15.12.2021 from. https://github.com/JustGlowing/minisom/.
Wang, Y.-Y., Wang, Y.-S., 2022. Development and validation of an artificial intelligence anxiety scale: an initial application in predicting motivated learning behavior. Interact. Learn. Environ. 304, 619–634.
Xiangyu, X., Youlin, Y., Qicheng, X., 2018. Intelligent identification of corporate tax evasion based on LM neural network. https://doi.org/10.23919/ChiCC.2018.8483951.
Zhang, L., Nan, X., Huang, E., Liu, S., 2020. Detecting Transaction-Based Tax Evasion Activities on Social Media Platforms Using Multi-Modal Deep Neural Networks. https://doi.org/10.1145/1122445.1122456.
Zobeiry, N., Humfeld, K.D., 2021. A physics-informed machine learning approach for solving heat transfer equation in advanced manufacturing and engineering applications. Eng. Appl. Artif. Intell. 101, 104232.