Scientax
Jurnal Kajian Ilmiah Perpajakan Indonesia

djp

VOLUME 4 NO 2 | APRIL 2023 Journal homepage: ejurnal.pajak.go.id ISSN 2686-5718

Artificial Neural Networks for Predicting Taxpaying Behaviour of Indonesian Firms

Arifin Rosid
Directorate General of Taxes, Jakarta, Indonesia. Email: arifin@pajak.go.id
University of Indonesia, Jakarta, Indonesia. Email: arifin.rosid@ui.ac.id

*Corresponding Author: arifin@pajak.go.id

ABSTRACT

Big data and sophisticated analytics might help tax authorities extract actionable data insights. In response, this paper employs an Artificial Neural Networks (ANN) model to predict and discover the determinants of firms' taxpaying behaviour. Examining 538,254 firm-level administrative data across fiscal years 2014 and 2019, this study is the first to apply ANN to exploit the taxpaying behaviour of Indonesian firms. Multi-Layer Perceptron Neural Network-based models were trained to predict three categories of taxpaying measurement—i.e., Corporate Tax Turnover Ratio (CTTOR)—across varying magnitudes of annual turnover. The models predicted the firms' taxpaying behaviour with an average accuracy rate above 92%. This study also reveals heterogeneous channels responsible for firms' taxpaying behaviour across groups. The findings demonstrate other business income and positive fiscal adjustment to be significant predictors of taxpaying behaviour for small and medium firms. In contrast, operating profit margin, other business expenses, and negative fiscal adjustment are prominent predictors for large corporations. The findings of this study can provide valuable assistance to decision-makers and relevant stakeholders in tax administrations by identifying potential areas of misreporting in annual tax returns. This evidence-based approach could enable tax administrations to develop more effective policies while potentially reducing the need for extensive monitoring and associated costs.

Keywords: corporate taxpayers, tax compliance, taxpaying behaviour, Artificial Neural Networks (ANN)

1. INTRODUCTION

Taxation is essential for a nation to meet its objectives for sustainable development (Bird, 2010). Infrastructure, maintaining economic growth, and eradicating poverty are just a few development initiatives that employ tax revenue. However, in many countries, tax revenue mobilisation remains significantly below the levels required to support sustainable development objectives—i.e., 15% of GDP is an often-quoted yearly target (Prichard et al., 2019).

Due to the complexity of tax and the time and resources needed to monitor and examine the tax returns of both individuals and firms, tax non-compliance is challenging to detect. As a result, one of the most important goals of tax authorities worldwide is to quantify and identify taxpaying behaviour (Pérez López et al., 2019). In this sense, the way tax authorities handle and evaluate the data at their disposal may be improved by big data and advanced analytics (Brondolo et al., 2022). Big data and sophisticated analytics might help tax authorities extract actionable insights from the

doi: 10.52869/st.v4i2.526
Received: August 15, 2022; Accepted: April 3, 2023; Published: April 27, 2023

Scientax: Jurnal Kajian Ilmiah Perpajakan Indonesia is Sinta 4 Journal (https://sinta.kemdikbud.go.id/journals/profile/9121)
How to Cite:
Rosid, A. Artificial Neural Networks for predicting taxpaying behaviour of Indonesian firms. *Scientax: Jurnal Kajian Ilmiah Perpajakan Indonesia*, *4*(2), 174–204. https://doi.org/10.52869/st.v4i2.526

information they already have while also supplying them with new tools to strengthen enforcement and discover tax fraud, evasion, and avoidance (Dom et al., 2022). Revenue bodies must be aware of the potential for a fresh wave of 'innovative technologies' to reshape tax administration (Asian Development Bank [ADB], 2022). These technologies are gaining popularity in an increasing number of revenue bodies worldwide. Among these is artificial intelligence (AI) (ADB, 2022).[1]

In recent years, AI has evolved into a technology that enables the administration of massive datasets and the application of algorithms that, despite their complicated structure, provide results that can be comprehended (Pérez López et al., 2019). Two tasks performed by the tax authorities that may be accomplished using AI are tax audits and tax collecting (Huang, 2018). This can be accomplished by the utilisation of work automation using computer-controlled tools and the use of AI. Several approaches may adopt AI to enhance services for tech-savvy taxpayers. Additionally, applying AI in taxes might help tax authorities analyse risks and spot unusual commercial practises (Wang & Wang, 2020). Thus, the development of new analytical tools has dramatically improved the efficiency and efficacy of tax administration, such as the enhanced identification of potential non-compliance via better risk assessment modeling and employing advanced analytical approaches (Organisation for Economic Co-operation and Development [OECD], 2020a).

Big data and neural networks are often linked. This is due to the common association of neural networks with challenging or impractical modelling tasks that cannot be accomplished by other kinds of models (Cook, 2020). For this reason, this paper aims to contribute through research on applying neural network models to administrative tax data to facilitate the identification of tax evasion by quantifying taxpayers' propensity to commit underreporting behaviour. With this objective in mind, one of the machine learning prediction techniques for supervised learning—the Artificial Neural Networks (ANN) model—is used. ANN method is excellent at solving random, ill-defined problems, highly non-linear, with many distinct and complex variables (Graupe, 2013).

In the last ten years, ANN has become a powerful and essential class of machine learning technologies (Cook, 2020). It is a non-parametric modeling tool that can perform the mapping of complex functions with sufficient accuracy (Zhang et al., 1999). ANN is widely used in cross-disciplinary research and tasks. For example, in finance, Sánchez-Serrano et al. (2020) used the ANN approach to create a predictive model of special audit opinion for consolidated financial statements. This study can predict audit opinion with an accuracy rate of 83%. AI approaches in finance are starting to be widely used, considering that behaviour is often non-linear and full of uncertainty (Bahrammirzaee, 2010). In the field of education, for example, Aryadoust and Baghaei (2016) examine the relationship between reading ability, lexical knowledge, and grammar of a group of students who use English as a foreign language. In this study, ANN accurately classified approximately 78% of students.

Several studies built on AI in the form of harmony search optimisation algorithms, support vector machines, genetic algorithms, decision trees, logistic regressions, and neural networks to discover tax evasion behaviour (González & Velásquez, 2013; Goumagias et al., 2012; Lin et al., 2012; Rahimikia et al. 2017; Warner et al., 2015). Specifically, the ANN approach has been applied in various tax studies across jurisdictions (Chen et al., 2011; Jang, 2019; Jupri & Sarno, 2018; Lin et al.,

[1] It is worth noting that while both AI and machine learning are related concepts, they are not interchangeable terms. AI refers to the broader concept of creating machines or systems that can perform tasks that would typically require human intelligence, whereas machine learning refers to a subset of AI that involves training a machine to recognize patterns in data and make predictions based on that data (Girasa, 2020).

2012; Pérez López et al., 2019; Rahimikia et al., 2017). For instance, Lin et al. (2012) used an AI approach to detect tax evasion in Taiwan. In Chile, González and Velásquez (2013) used decision trees, neural networks, and Bayesian networks to spot fraud tendencies for audited taxpayers.

Rahimikia et al. (2017) combined multilayer perceptron neural networks with various classification algorithms to identify corporate tax evasion in Iran. Pérez López et al. (2019) contribute to identifying tax fraud for personal income tax returns in Spain by using ANN as an advanced prediction technique with an efficiency rate of 84.3%. However, to the best of the author's knowledge, no empirical studies have used the ANN approach to predict the taxpaying behaviour of firms in Indonesia. As Saragih et al. (2022) argue, the potential advantage of an AI application for modernising Indonesia's tax administration system is that it would facilitate enforcement by including AI in tax audits to monitor taxpayers and in tax services to enhance the effectiveness of the tax authority. Subsequently, this paper is the first to exploit how neural network algorithms can be adopted to predict Indonesian firms' taxpaying behaviour and identify their significant predictors. In doing so, this study compared four different groups based on the size of firms' annual turnover.

The results are very encouraging. This study supports the notion that AI approaches are superior to conventional statistical methods for addressing various issues, particularly those involving non-linear patterns (see, for example, Aryadoust & Baghaei, 2016; Bahrammirzaee, 2010; Chen et al., 2011; Lin et al., 2012; Sánchez-Serrano et al., 2020). This study concludes that the ANN approach accurately predicts the taxpaying behaviour of Indonesian firms with different annual turnovers, with an accuracy rate above 92%.

The implementation of AI also allows this study to identify heterogeneous channels responsible for firms' taxpaying behaviour across groups. As Brondolo et al. (2022) posit, particularly in establishing a compliance risk management strategy, with supervised machine learning, the algorithm may reveal complex data patterns associated with successful case outcomes while deemphasising those that were not. The findings would be of benefit since the Directorate General of Taxes (DGT) has not yet incorporated AI in its operations (ADB, 2022). The present study raises the possibility of generating actionable data insights and identifying areas of misreporting at strategic levels.

This article contributes at two levels. First, at the taxation literature level, this paper adds empirical knowledge about how AI approaches can be applied in taxation studies. The practical implications of this study are essential because it is the first study to try to predict the taxpaying behaviour of Indonesian firms using administrative data. Second, on a practical level, the findings of this paper help Indonesia's tax authorities identify several unique, influential factors and provide actionable data insights that lead to income taxpaying behaviour.

This paper consists of six main parts. Section 2 describes the conceptual framework of this study. Section 3 describes the data and the analytical approach, while Section 4 describes the results and discussions. Section 5 concludes, and Section 6 discusses the practical implications and limitations.

2. CONCEPTUAL FRAMEWORK

2.1 Institutional Settings

The Indonesian government launched and implemented major and radical tax reform in 1983 by introducing self-assessment to its tax system as a strategic response to the threat to the national budget. The primary goal of the reform was to increase non-oil-related tax revenue to decrease the government's dependency on oil money (Heij, 2001). At that time, Indonesia's tax system underwent significant modifications that brought it into compliance with worldwide best practices (Alm, 2019). Most of these structural elements are still in place today.

However, the tax system has shown that it cannot generate optimal revenue collection, partly because it has developed through time in a fragmented, ad hoc fashion with little apparent attention given to how the system elements need to fit together (Alm, 2019). Although these issues are not unique to Indonesia (see, for example, International Monetary Fund, 2014; World Bank, 2016), the lessons from these Indonesian tax reform initiatives are often vague, perhaps even unaddressed, and specific (Alm, 2019). For this reason, the Indonesian tax authority, DGT, is still dealing with increasingly challenging and complex issues in collecting tax revenue, and both the growth and ratio of tax collection showed a decreasing trend (Directorate general of Taxes [DGT], 2021). In the 2019 fiscal year, Indonesia's tax ratio was 11.6% (ADB, 2022). This figure is much lower than the average tax ratio for Organisation for Economic Co-operation and Development (OECD) countries, which is 33.8%, followed by Latin America and the Caribbean (23.0%), Asia-Pacific (18.7%), and Africa (16.8%) (ADB, 2022).

From 2017 to 2021, income taxes in Indonesia contribute more than 55% of national tax revenue. Income tax revenue is divided into (i) oil and gas and (ii) non-oil and gas income taxes. Most income tax revenue comes from non-oil and gas income taxes (92%), comprising several types of income tax based on related sections of income tax law. Figure 1 shows that corporations contribute the most considerable portion of income tax revenue from 2017 to 2021, ranging from 26.9% to 34.3%. The Covid-19 pandemic resulted in a decline in corporation tax income in 2020, followed by a rise in 2021. In 2020, as Figure 1 shows, the contribution of Article 25 of Corporations and Article 22 on Imports reached the lowest points due to the pandemic and tax incentives provided for impacted industries.

International experience indicates that a small number of large corporations (usually less than 1%) account for 60%–70% of domestic tax revenues, whereas a significant number of small enterprises contribute for less than 5%–10%. Between these two categories, medium-sized firms contribute about 20%–30% of domestic tax revenues (York, 2011). In most nations, this distribution of corporate taxpayers is identical (York, 2011).

Due to its continually low tax ratio, DGT (2015) has identified several difficulties and is primarily concerned with persistent compliance problems. While AI is currently not yet implemented in the DGT (ADB, 2022), the DGT is concerned with establishing a data-driven organisation to improve tax compliance (DGT, 2022). It is envisaged that by being a data-driven organisation, DGT's ambition to become a trusted partner for nation-building and to collect state revenues via effective tax administration may be accomplished (DGT, 2022). Indeed, successful digitalisation initiatives understand that technology

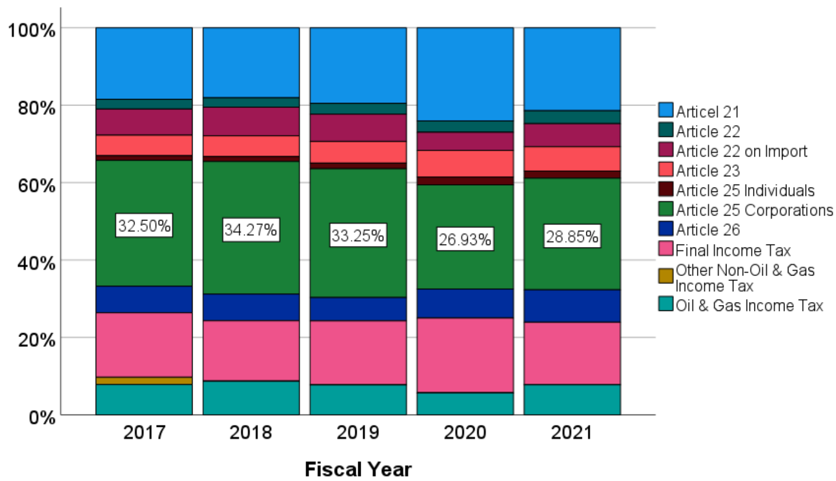


Figure 1 Type of Income Taxes in Indonesia and Their Contribution to Total
Source: Adapted from the DGT's internal data

is a tool for solving a specific issue rather than a goal in and of itself (Dom et al., 2022).

2.2 Tax Avoidance and Taxpaying Behaviour

Before attempting to analyse corporate taxpaying behaviour, one needs to comprehend tax compliance first. There are several ways to examine tax compliance, some of which can lead to more definitional concerns. The definitions include a variety of topics and are broad. Consequently, the definition of compliant behaviour is not universally agreed upon (Devos, 2014). Following Rosid et al. (2018), this study emphasises the benefit of adopting an operational approach instead of a conceptual approach. This paper defines tax compliance as 'taxpayers' willingness to correctly report tax liability in accordance with the prevailing tax law' (OECD, 2014). Thus, in contrast to forced willingness (*ex-post*), this concept is more concerned with the voluntary willingness (*ex-ante*) of taxpaying behaviour.[2]

Any arrangement or transaction that lowers a company's tax obligation is considered tax avoidance (Dyreng et al., 2008). While there are various ways to measure tax avoidance (Gebhart, 2017), Hanlon & Heitzman (2010) define tax avoidance as any tax reduction resulting from a transaction directly affecting the company's tax burden. Nevertheless, OECD finds it challenging to define the term 'tax avoidance'—and distinguish it from 'tax evasion.' However, it generally refers to tax planning undertaken by taxpayers with the intention of reducing their tax liabilities.[3] While such planning may be considered legal, it is typically contrary to the objectives of tax laws. Slemrod (2016) opines that the dividing line between illegal and legal tax evasion is blurry. For this reason, this study does not emphasise the legality of tax avoidance but rather how much the company reduces its tax payments.

In this sphere, effective tax rates (ETR) may be extensively used to gauge a firm's tax burden (Dyreng et al., 2008; Gebhart, 2017). ETR—calculated by using a measure of pre-tax earnings or cash flow as the denominator and specific estimates of the tax due as the nominator—is also one of the commonly used metrics of tax evasion in this area (Hanlon & Heitzman, 2010). This proxy represents the typical tax rate per revenue unit or cash flow. Thus, ETR is a potent predictor of the efficacy of a company's tax planning activities (Gebhart, 2017; Mills et al., 1998).

However, it is worth emphasising that ETR may only be able to capture non-conforming tax avoidance activities but ignore conforming tax avoidance activities (Badertscher et al., 2019). This is because conforming tax avoidance is more difficult to detect because it reduces book income and tax income at the same time.[4] When using a measurement such as ETR, this will change the numerator and denominator at the same time—preventing the company from being detected as tax evaders. In addition, conforming tax avoidance is more difficult to detect because there are not many measurements related to conforming tax avoidance (Badertscher et al., 2019; Hanlon & Heitzman, 2010). Regarding this, the noncompliance rate for firms concerning their size seems to be 'U-shaped, with medium-sized enterprises among the group of large corporations showing the lowest noncompliance rate (Slemrod, 2007).

For this reason—and for practicality—to gauge the level of tax avoidance, we instead chose

[2] It is important to note that this definition purposefully omits the registration, filing, and income reporting elements owing to the practical and legal challenges in acquiring the data.

[3] See the OECD Centre for Tax Policy and Administration's Glossary of Tax Terms at http://www.oecd.org/ctp/glossaryoftaxterms.htm, accessed 25 July 2022.

[4] Two strategies—non-conforming and conforming tax avoidance—can help businesses lower their income tax obligations. Companies that use non-conforming tax avoidance strategies decrease their income tax liability but not their book income. Companies that use conforming tax avoidance lower their income tax obligations by taking actions that lower their taxable and book earnings (Badertscher et al., 2019).

to utilise a different metric adopted by the Indonesian tax authority: the corporate tax to turnover ratio (CTTOR). CTTOR is a tax payment measure based on declared income tax payable scaled by annual turnover. This practical measure refers to the Director General of Taxes Circular Letter number SE-02/PJ/2016 concerning the Benchmark Behavioural Model (BBM).[5] By doing so, this paper provides actionable policy implications in addition to giving the Indonesian tax authorities an objective' empirical indicator' of the taxpaying behaviour of corporate income taxpayers.

Relating to this, the recently completed Voluntary Disclosure Program (VDP) provides some relevant indicators of persistent non-compliance behaviour.[6] As of the completion of the VDP's implementation on 30 June 2022, taxpayers had disclosed assets totalling IDR 594.82 trillion. They had paid a total of IDR 61.01 trillion in tax liabilities derived from those assets in the form of income tax.[7] Both under the tax amnesty and VDP schemes, the value of assets that were declared and the ransom money that was paid were the strong indications of underreporting behaviour among Indonesian taxpayers. In this sense, one justification for prevalent tax non-compliance in Indonesia might be the outdated information technology utilised in tax administration, which does not reflect the most recent computer technology used in other countries to track taxpayers and their incomes (Alm, 2019).

2.3 Artificial Neural Networks (ANN)

ANN is one of the most popular AI methods in various disciplines (Cook, 2020), including finance (Bahrammirzaee, 2010). It comprises mathematical approaches typically applied in prediction and classification studies (Aryadoust & Baghaei, 2016). ANN is a frequently utilised technique for predictive data mining analysis because of its accuracy, adaptability, and simplicity. Particularly in instances where the underlying processes are sophisticated (International Business Machines [IBM], 2021).8 ANN is grounded on the idea that the specific link between independent and dependent variables can be calculated using non-linear mathematical functions (Aryadoust & Baghaei, 2016). Thus, the real benefit of ANN is pattern identification and classification due to its non-linear, non-parametric nature of adaptive learning (Zhang et al., 1999).[9] Table 1 summarises the benefits and drawbacks of several classification algorithms.

[5] It is important to note that the policy outlined in SE-02/PJ/2016 has been revoked by SE-24/PJ/2019, which introduced a centralized framework for compliance risk management implementation within the DGT. The revocation of the use of BBM reflects this policy of centralization, but it does not imply the discontinuation of CTTOR usage. In practice, CTTOR remains essential in policymaking to increase taxpayer compliance at the DGT.

[6] Note that the Indonesian government recently introduced a tax amnesty program, i.e., in 2016. A tax amnesty is a 'forgiveness' period during which people are given a chance to pay back taxes that have gone unpaid without worrying about facing the financial penalties and/or criminal punishment that come with being found guilty of tax evasion. In the 2016 tax amnesty program, a total of IDR 4.87 trillion in unreported assets were reported to the tax authorities, with about three-quarters of these disclosed assets being domestic assets. See, https://setkab.go.id/realisasi-tax-amnesty-menkeu-tebusan-rp130-triliun-deklarasi-rp4-8134-triliun-dan-repatriasi-rp146-triliun/, accessed 24 July 2022.

[7] See, https://www.kemenkeu.go.id/publikasi/berita/pps-berakhir-menkeu-harta-yang-diungkap-rp594-82-triliun-dengan-pph-rp61-01-triliun/, accessed 26 July 2022.

[8] Data mining is the exploration and analysis of vast amounts of data using automated or semiautomatic methods to spot meaningful patterns and rules (Fayyad et al., 1996).

[9] The discussion of ANN involves a complex mathematical model. Readers can refer to Ripley (1996) and Haykin (1999) for a comprehensive discussion of theoretical and mathematical models. Readers interested in the practical aspects of ANNs might learn more in 'IBM SPSS Neural Networks 28' (IBM, 2021).

Table 1 Overview of Benefits and Drawbacks of The Classification Algorithms
Source: Adapted from Wendler & Gröttrup (2021, pp 761-762)

	Advantages	Disadvantages
Logistic regression	▪ Models are often very accurate ▪ Works well on small datasets ▪ Predicts probabilities ▪ Easy to interpret, in particular, the influence of each input variable	▪ It can only provide linear solutions ▪ Problems with high collinearity of the input variables
Linear discriminant analysis	▪ Typically, very fast building the model ▪ Works well on small datasets ▪ Optimal if data assumptions are fulfilled	▪ More restrictive assumptions than other methods (e.g., logistic regression) ▪ Usually, needs data preparation ▪ Sensitive to outliers ▪ Only applicable to linear problems
Decision trees	▪ Robust to outliers ▪ Model and decision rules are easy to understand ▪ Can handle different data types ▪ Fast in prediction and no assumptions on variable distributions needed. Thus, less effort for data preparation ▪ Can handle missing values	▪ Can be computationally expensive to train ▪ Large trees tend to overfitting ▪ Most of the time it does not find the optimal solution ▪ Prefers variables with many categories or numerical data
Neural networks	▪ Good performance on large datasets ▪ Very good at allowing nonlinear relations and can generate very complex decision boundaries ▪ Non-parametric ▪ No distribution assumptions needed ▪ Can handle noisy data ▪ Often outperforms other methods	▪ Training can be computationally expensive ▪ Results and effects of input variables are hard to interpret (black box algorithm) ▪ Tends to overfitting and does not always find the optimal solution ▪ Can only process continuous input variables

Generally, as Table 1 suggests, neural network models outperform other linear and non-linear models in terms of accuracy and predictive capability (Murorunkwere et al., 2022). From a quantitative standpoint, neural networks often consist of optimal combinations that enable more accurate predictions and estimates than other models (Pérez López et al., 2019). Without requiring the researcher to make prior assumptions about the connections between the dependent and independent variables, a neural network may mimic various statistical models. Instead, the learning process determines how the connections will take shape (IBM, 2021).

The findings of the neural network should generally resemble those of the linear regression model if there is a linear connection between the dependent and independent variables. The neural network will automatically get close to the 'correct' model structure if a non-linear connection is more suitable (IBM, 2021). In this sense, in addition to being crucial for traditional statistical decision theory, posterior probabilities are also crucial for many managerial decision-making scenarios (Zhang et al., 1999). Although there are other methods for estimating posterior probability, ANN is the only one currently known to directly estimate posterior probabilities without knowledge of the underlying group population distributions (Zhang et al., 1999).

The general ANN model is a three-layered structure of linked nodes: the input, the hidden, and the output units. The nodes between the input and output layers might create one or more hidden layers. As illustrated in Figure 2, every neuron in one layer has a connection to every other subsequent layer, but neurons corresponding to the same layer have no interactions between them (Cook, 2020). The input layer gets sensory input, the hidden layer executes the information processing, and the output layer

(a) ANN architecture (b) ANN activation node

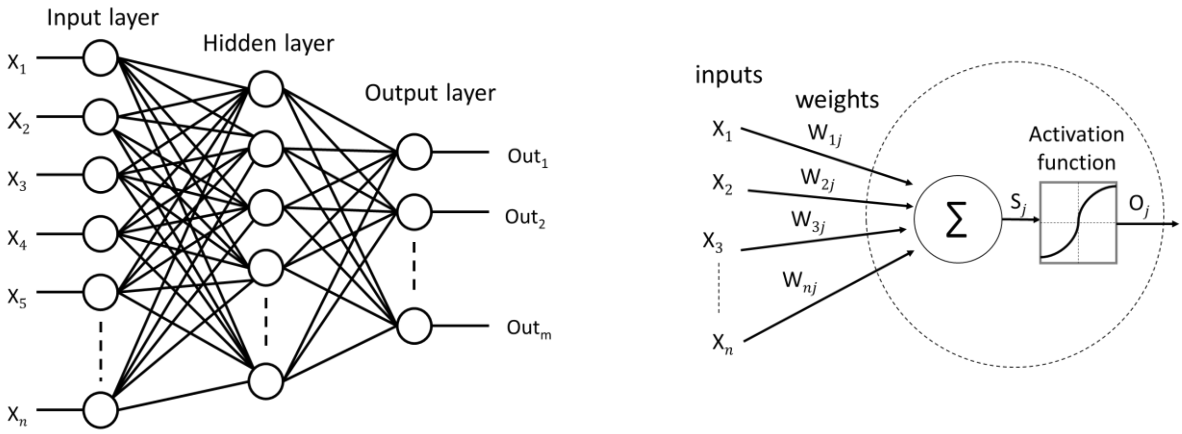


Figure 2 ANN Architecture and Neural Network Activation Node
Source: Adapted from Cook (2020)

generates the class label or anticipates continuous values.[10]

The primary purpose of the input layer is to disperse the data supplied to the neural network for further processing. The hidden and output layer nodes process the signals by adding synaptic weights or processing factors. Each layer has an extra node known as bias, which adds a new term to the output of all the nodes in the layer. All inputs to a node are weighted, integrated, and processed by a function known as a transfer function or activation function, which regulates the output flow from that node to facilitate communication with all nodes in the following layer (Pérez López et al., 2019).

The following equations represent the nett sum of the weighted inputs entering a node *j* and the output activation function that translates a neuron's weighted input to its output activation (often the sigmoid function):

$$S_j = \sum_{i=1}^{n} X_i \, W_{ij} \qquad (1)$$

and

$$O_j = \frac{1}{1+e^{S_j}} \qquad (2)$$

An activation function processes a neuron's inputs $x_0, x_1, \ldots, x_n$ with their associated weights $w_0, w_1, \ldots, w_n$ and the result is the neuron's output, which corresponds to:

$$\begin{aligned} Output &= f(w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + bias) \\ &= f\left(\sum_{i=1}^{n} x_i w_i + bias\right) \qquad (3) \end{aligned}$$

where w_i and x_i are the weight vector and input vector, respectively, *f* is the activation function used on the sum of products of each input and its corresponding weight and the bias (Murorunkwere et al., 2022).[11] Without an activation function, every

[10] This structure is also known as 'feedforward architecture' because the links in the network move forward from the input layer to the output layer without a feedback loop. The input layer consists of predictors, the hidden layer consists of units that cannot be observed (unobservable), and the output layer contains the response. The unit in the output is a collection of several functions from the unit in the hidden layer (IBM 2021).

[11] The neural network will dynamically select whether the model is linear or non-linear in this process. Because of this flexibility, the synaptic weights of the neural network are not intuitive. If the interpretation of the linear connection between the dependent and independent variables is the primary objective, a standard statistical model approach is preferable (IBM, 2021). The estimation of synaptic weights is based solely on training data and is therefore generally not used to interpret ANN test results. That is, although the data is partitioned into three

Table 2 Variables Under Study

	Variable name	Code	Description
1	Gross profit margin	GPM	Annual turnover minus cost of goods sold, scaled by annual turnover
2	Operating profit margin	OPM	Annual turnover minus cost of goods sold and operating expenses, scaled by annual turnover
3	Other income ratio	OIR	Other business income scaled by annual turnover
4	Other expense ratio	OER	Other business expense scaled by annual turnover
5	Positive fiscal adjustment ratio	PFAR	Positive fiscal adjustment scaled by annual turnover
6	Negative fiscal adjustment ratio	NFAR	Negative fiscal adjustment scaled by annual turnover
7	Corporate Tax Turnover Ratio	CTTOR	Income tax payable scaled by annual turnover

ANN is merely a basic linear function (Agostinelli et al., 2015).

Even though traditional linear equations are simple and quick to solve, they are restricted in their complexity and lack the capacity to learn and find intricate data mappings. In response, ANN makes it simple to manage large datasets and, despite the complexity of their algorithms, provide easily interpretable results—which is why they are widely used in the financial industry, marketing, forecasting, and increasingly in risk assessment and fraud detection (Murorunkwere et al., 2022). In addition to facilitating the categorisation of each group of taxpayers as low, moderate, and high tax payments, the neural network exposes firms' propensity for taxpaying behaviour based on the variables under study.

3. RESEARCH METHODS

3.1 Empirical Data

Taxable income and tax liability information is required to quantify company tax compliance behaviour (Salihu et al., 2013). A key strength of the present study was access to a large set of actual data. This study uses administrative data from fiscal years 2014 to 2019, consisting of 538,254 useable anonymous corporate tax records. This firm-level data offers a unique period and natural setting covering six fiscal years prior to the Covid-19 pandemic—i.e., when the economy plummed significantly (OECD, 2020b). As shown earlier in Figure 1, during the pandemic, the contribution of corporate income tax reached the lowest point. During these unprecedented times, the economy plummeted substantially, with more than three-quarters of tax administrations experiencing a decline (OECD, 2022). One of the key advantages of adopting six years period is that it may prevent year-to-year volatility in the variables under investigation. Table 2 details the variables.

Accordingly, this study relies heavily on Corporate Annual Income Tax Return Form 1771, 1771-I, and 1771-III. These data include detailed information on corporate income reporting, particularly from the perspective of the income statement—i.e., annual turnover (part 1a Form 1771-I), cost of goods sold (part 1b Form 1771-I), operating expenses (part 1c Form 1771-I), other business income (part 1e Form 1771-I), other business expense (part 1f Form 1771-I), total positive fiscal adjustment (part 5m Form 1771-I),[12] total negative fiscal adjustment (part 6e Form 1771-I), taxable income (part 8 Form 1771-I), and income tax payable (part 4 Form 1771).

Following Rosid and Ariyani (2022), this paper reduced the original data to obtain a usable sample by removing observations that fit these criteria. First, we do not include businesses with an

categories, training, testing, and holdout, the estimation of synaptic weights is based solely on training data (IBM, 2021).

[12] In this context, the terms ‘positive’ and ‘negative’ relate to the consequence of the adjustment on taxable income; hence, positive denotes an increase in taxable income and vice versa.

annual turnover of less than IDR 5 billion.[13] Second, this analysis excludes entities with a CTTOR of greater than one or negative.[14] Third, this study excludes firms that are liable to schedular-final incomes taxes. These include firms in the construction service sector, real estate businesses, representative offices, shipping and air transportation enterprises, and financial brokerage firms. Finally, it is essential to note that this usable sample does not consider the impact of either tax audit or tax oversight activities, as well as the type of tax offices where the taxpayers administered. This study also does not consider the impact of loss carryforwards (LCF) on the total taxable income. Only 2.9% of firms within the usable sample report LCF. Thus, in this scenario, the inclusion should be inconsequential.

This research divides the data into four size-based categories. The reasons are twofold. First, the way the Indonesian tax authorities administer its taxpayers are size based—i.e., Primary Tax Office (PTO), Medium Tax Office (MTO), and Large Tax Office (LTO). Second, businesses of various sizes exhibit various compliance habits (Slemrod, 2007). Following Indonesia Government Regulation number 7 of 2021, the categories are as follows: i) between IDR 5 billion to 15 billion—categorised as '*small firms*'; ii) between IDR 15 billion to 50 billion—categorised as '*medium firms*'; iii) between IDR 50 billion to 100 billion—categorised as '*medium-large firms*'; iv) more than IDR 100 billion—categorised as '*large firms*.'[15]

3.2 Analytical Approach

This study builds on an applied research approach. Applied research intends to address specific practical questions or provide answers to real issues (Neuman, 2011).[16] Using the perspective of the Cross-Industry Standard Process for Data Mining (CRISP-DM), this study is in the modelling phase.[17] In this phase, data mining algorithms are constructed to extract knowledge from the data. The modelling process generates a model or group of models accurately representing the learned information. In this sense, intuitively, taxpaying behaviour appears to have a great deal to do with non-linear variables. In non-linear areas, the degree of accuracy of contemporary approaches in AI tends to be superior to classic statistical methods (Bahrammirzaee, 2010).[18] Based

[13] This analysis substitutes IDR 5 billion thresholds for IDR 4.8 billion for the eligibility test out of practicality. Note that firms with an annual turnover of less than IDR 4.8 billion can benefit from a final income tax. As governed by Government Regulation Number 46 of 2013, beginning in January 2014, taxpayers with annual sales below IDR 4.8 billion are taxed at one percent of monthly turnover. The rate is then reduced by 50%—at a half percent rate from monthly turnover—starting from July 2018 based on Government Regulation Number 23 of 2018. Notably, the exclusion of businesses with an annual turnover of less than IDR 5 billion constitutes a significant proportion of the submitted annual income tax returns, accounting for approximately 40-45% of the total initial sample.

[14] The reason is that income tax liability and annual turnover values need to be positive to measure taxpaying behaviour. Consequently, a negative number would suggest an error. On the other hand, the CTTOR value surpassing one would be illogical given that the maximum tax rate (applied to taxable income) is 25%.

[15] Note that according to Government Regulation number 7 of 2021 regarding Ease, Protection, and Empowerment of Cooperatives and Micro, Small, and Medium Enterprises, businesses with annual turnover between IDR 2 billion to 15 billion are categorized as 'small' businesses, while businesses with turnover between IDR 15 billion to 50 billion are categorized as 'medium' businesses. As of 31 July 2022, IDR one billion equals roughly USD 67,417.

[16] The implication is that applied research rarely has a strong relationship with activities to build, test, or relate theory in depth (Neuman, 2011).

[17] See, https://www.ibm.com/docs/en/spss-modeler/18.2.0?topic=dm-crisp-help-overview, accessed 15 July 2022. Note that iteration and adjusting in reaction to real-world data are key components of the CRISP-DM strategy, which is being used in Canada and New Zealand (OECD, 2016).

[18] The linear regression model has a fixed model structure and a set of assumptions applied before learning from the data. On the other hand, neural networks may estimate numerous statistical models without needing us to postulate beforehand a specific link between the dependent and independent variables—because the form of the relationship is determined during the model's learning process (IBM, 2021).

on these considerations—and referring to the research aim, the author employs an Artificial Neural Network (ANN) model as a primary approach.

More specifically, this paper uses the Multilayer Perceptron (MLP) module from IBM SPSS.[19] The concept of MLP in ANN departs from backpropagation learning error—an algorithm that is most often used in ANN (Pérez López et al., 2019). The MLP approach is more popular in ANN than other approaches (Zhang et al., 1999). A study conducted by Sánchez-Serrano et al. (2020) also found that the MLP method produces a higher level of accuracy than the Radial Basis Function (RBF) approach.[20]

Following Bekesiene et al. (2021), the model used in this study randomly classifies the data into three groups: (i) 60% training data, (ii) 20% testing data, and (iii) 20% holdout data—or commonly called 60%-20%-20% format.[21] In this format, the training data is used to find the weights and then to build the model, while the testing data is used to find errors and prevent overtraining during training mode. Finally, the holdout data is used to validate the model (IBM, 2021).[22]

There are seven independent variables used to build the model: : (i) gross profit margin, (ii) operating profit margin, (iii) other business income ratio, (iv) other business expense ratio, (v) positive fiscal adjustment ratio, (vi) negative fiscal adjustment ratio, and (vii) types of annual income tax return. The dependent variable in the model is an ordinal variable representing the three levels of CTTOR: (i) low CTTOR (< 0.59%); (ii) moderate CTTOR (0.59% to 1.19%); (iii) high CTTOR (>1.19%).

4. RESULTS AND DISCUSSIONS

4.1 Descriptive Statistics

The descriptive statistics for the seven variables, which provide the means, medians, standard deviations, variance, minimum, and maximum values, form the basis of this study. Table 3 presents the results.

Pearson's correlation coefficient was further employed to assess the relationships between the variables under study. Panel B of Table 3 presents the tested variables' correlation coefficient (r). The strongest positive significant correlation was identified between the relationship between the variable *other income ratio* and *other expense ratio* (OIR, r = 0.698, ρ < 0.001). The strength of the relationship between these two variables is somewhat high because it has an r value of 0.69 (Schober et al., 2018). A significant negative correlation—although weak—was identified between the relationship between the variable *positive fiscal adjustment ratio* and CTTOR (PFAR, r = -0.120, ρ < 0.001). The correlational relationships among other variables have a coefficient that varies with the tendency of a weak relationship because it has a value of < 0.40.[23] Figure 3 facilitates the visualisation of the associations between the level of CTTOR and the business sector and firms' size.

---

[19] There are two ANN-based prediction application modules within IBM SPSS: MLP and Radial Basis Function (RBF) (IBM 2021). MLP and RBF are the most widely used neural network architectures (Ripley, 1996).

[20] It should be noted that the MLP method is not always more accurate. For example, Jupri and Sarno (2018) compared four classification algorithms C4.5, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and MLP, to classify the level of compliance of taxpayers and concluded that the C4.5 algorithm is a more accurate classification algorithm.

[21] The study by Bekesiene et al. (2021) found that the 60%-20%-20% partition format is more optimal than the 50%-30%-20% and the 70%-20%-10% format.

[22] The partition format that is quite common is 70% for training data and 30% for testing data (70%-30%). This formation is usually adopted for relatively few observations (e.g., less than 1,000) and therefore does not have the data allocated to validate the model—i.e., the holdout data.

[23] According to Schober et al. (2018), the correlation coefficient 0.40-0.69 indicates a moderate correlation, 0.70-0.89 indicates a strong correlation and 0.90-1.00 indicates a powerful correlation.

Table 3 Descriptive Statistics and Relationships between Tested Variables

Panel A. Descriptive statistics (n = 538,254)

	Mean	Median	SD	Variance	Min.	Max.
Gross profit margin (%)	20.83	15.17	20.04	401.50	-98.60	100.00
Operating profit margin (%)	4.49	3.02	11.57	133.86	-175.64	100.00
Other income ratio (%)	1.20	0.04	8.62	74.38	-92.59	4,308.28
Other expense ratio (%)	1.59	0.01	9.53	90.87	-70.42	4,332.31
Positive fiscal adjustment ratio (%)	7.24	0.22	21.78	474.50	-104.07	100.00
Negative fiscal adjustment ratio (%)	0.60	0.00	4.51	20.33	-189.45	100.00
CTTOR (%)	0.89	0.42	1.35	1.82	0.00	23.37

Panel B. Descriptive statistics (n = 538,226)

Types of tax return	CTTOR Category						Total	
	Low CTTOR		Moderate CTTOR		High CTTOR			
	Observ.	%	Observ.	%	Observ.	%	Observ.	%
Nil	105,606	33.9%	1,294	1.5%	1,868	1.3%	108,768	20.2%
Underpaid	182,612	58.7%	81,791	94.3%	131,708	93.9%	396,111	73.6%
Overpaid	22,978	7.4%	3,653	4.2%	6,716	4.8%	33,347	6.2%
	311,196	100.0%	86,738	100.0%	140,292	100.0%	538,226	100.0%

Panel C. Relationship between variables under study (n = 538,254)

	GPM	OPM	OIR	OER	PFAR	NFAR	CTTOR
GPM	1	.476**	.023**	.080**	.065**	.065**	.349**
OPM	.476**	1	-.095**	.018**	.072**	.045**	.410**
OIR	.023**	-.095**	1	.698**	.020**	.108**	.074**
OER	.080**	.018**	.698**	1	.038**	.073**	.004**
PFAR	.065**	.072**	.020**	.038**	1	.098**	-.120**
NFAR	.065**	.045**	.108**	.073**	.098**	1	.011**
CTTOR	.349**	.410**	.074**	.004**	-.120**	.011**	1

**. Correlation is significant at the 0.01 level (2-tailed).

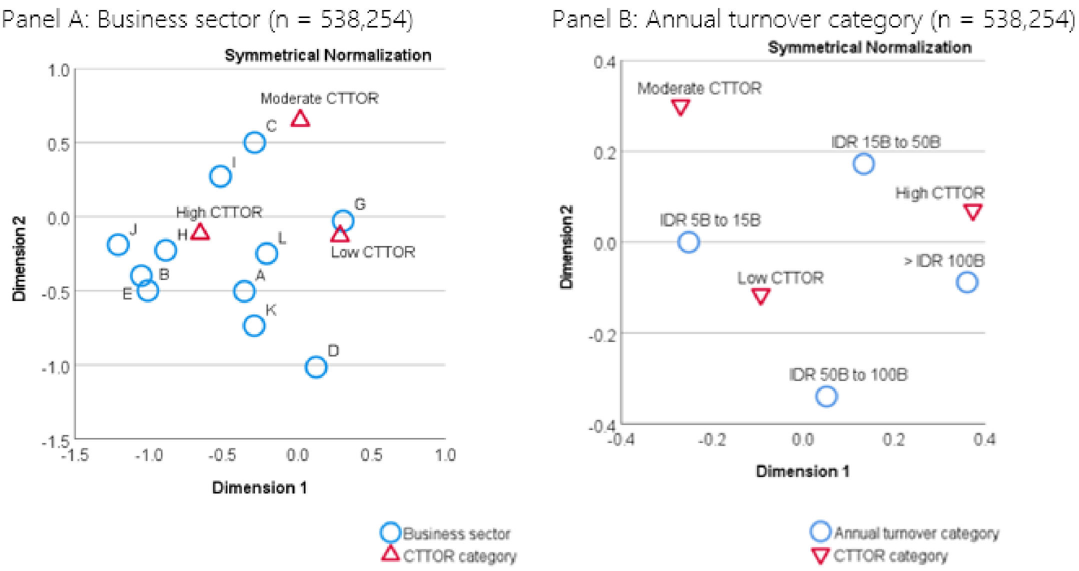
Note: GPM = gross profit margin; OPM = operating profit margin; OIR = other income ratio; OER = other expense ratio; PFAR = positive fiscal adjustment ratio; NFAR = negative fiscal adjustment ratio; CTTOR = corporate tax turnover ratio. For types of tax

For details on the number of observations and the percentage of firms in three categories of CTTOR for each business sector under study, see Appendix 1. For details on the share of CTTOR categories by business sectors and by size for 2014 to 2019, see Appendix 2.

4.2 Preliminary Analysis: ANOVA and Graphical Evidence

To see whether firms that vary in annual turnover categories also differ considerably in terms of variables under study—i.e., their gross profit margin, operating profit margin, other income ratios, other expense ratios, positive fiscal adjustment ratios, and negative fiscal adjustment ratios, this study employs preliminary analysis. This initial step builds on an analysis of variance (ANOVA) and provides visual evidence to see whether firms in four annual turnover categories also vary substantially in terms of the variables under investigation.

Table 4 describes that small firms have the highest GPM compared to other categories (M = 24, SD = 20.6). Large firms, on the other hand, have the lowest GPM (M = 16.23, SD = 17.62). However, regarding operational profitability, the OPM value for large firms is the highest (M = 5.03, SD = 11.91), followed by the OPM value for small firms (M = 4.77, SD = 11.69). Medium-large and large firms virtually have the same OPM values (M = 3.99, SD = 11.34 for medium firms; M = 3.96, SD = 11.15 for the rest). Additionally, large firms report the largest portion of the other income ratio (M = 1.75, SD = 6.44), whereas small firms show the lowest (M = 0.99, SD = 6.74). It is evident that the annual turnover and other income ratio are



A = Agriculture, B = Mining and quarrying, C = Manufacturing, D = Electricity and gas, E = Water supply, sewerage, etc., G = Wholesale and retail trade, H = Transportation and storage, I = Accommodation and food service, J = Information and communication, K = Financial and insurance, L = Others

Note : Panel A depicts a biplot displaying three levels of CTTOR and how they relate to the specific business sector. The proximity between the two legends suggests a commonality or trend. In panel A, firms in the wholesale and retail trade sector, for instance, tend to have a low CTTOR. In contrast, businesses in the transportation, storage, mining, and quarrying industries often have a high CTTOR. In Panel B, for instance, small firms have a stronger tendency to report a low CTTOR, but large corporations are inclined to have a high CTTOR.

Figure 3 Visual Associations: CTTOR, Business Sectors, and Annual Turnover

positively correlated—i.e., the higher the turnover, the greater the OIR.

Further, the value of other expense ratio reported by large businesses is likewise the largest (M = 2.54, SD = 7.00), while the ratio reported by small firms is the lowest (M = 1.13, SD = 8.75). It is evident that there is a positive correlation between annual turnover and other expense ratio, meaning

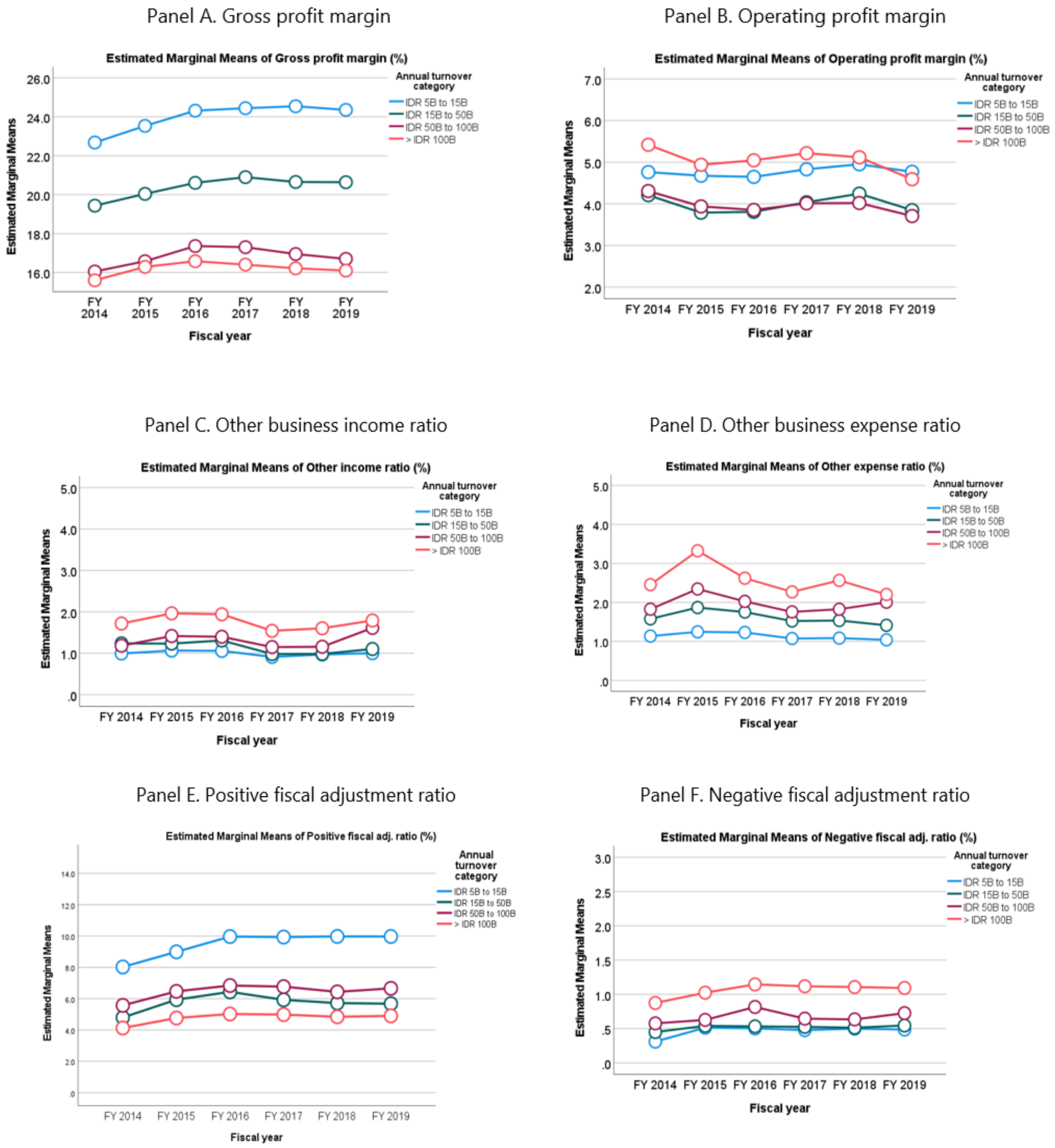
Table 4 ANOVA Results of Variables Under Study by Annual Turnover

		GPM (%)		OPM (%)	
Annual turnover	Observ.	Mean	SD	Mean	SD
Small firms (IDR 5B to 15B)	224,181	24.00	20.62	4.78	11.69
Medium firms (IDR 15B to 50B)	166,171	20.41	20.16	4.00	11.35
Medium-large firms (IDR 50B to 100B)	61,572	16.84	18.59	3.96	11.16
Large firms (> IDR 100B)	86,330	16.20	17.62	5.04	11.92
Total	538,254	20.83	20.04	4.49	11.57
		OIR (%)		OER (%)	
Annual turnover	Observ.	Mean	SD	Mean	SD
Small firms (IDR 5B to 15B)	224,181	1.00	6.74	1.13	8.75
Medium firms (IDR 15B to 50B)	166,171	1.13	5.94	1.60	6.14
Medium-large firms (IDR 50B to 100B)	61,572	1.32	18.18	1.95	18.52
Large firms (> IDR 100B)	86,330	1.75	6.44	2.54	7.00
Total	538,254	1.20	8.62	1.59	9.53
		PFAR (%)		NFAR (%)	
Annual turnover	Observ.	Mean	SD	Mean	SD
Small firms (IDR 5B to 15B)	224,181	9.50	25.78	0.46	4.34
Medium firms (IDR 15B to 50B)	166,171	5.74	18.70	0.52	4.20
Medium-large firms (IDR 50B to 100B)	61,572	6.48	20.30	0.67	4.85
Large firms (> IDR 100B)	86,330	4.79	15.55	1.07	5.17
Total	538,254	7.24	21.78	0.60	4.51

Note: GPM = gross profit margin; OPM = operating profit margin; OIR = other income ratio; OER = other expense ratio; PFAR = positive fiscal adjustment ratio; NFAR = negative fiscal adjustment ratio; CTTOR = corporate tax turnover ratio. All ANOVA results in this table are statistically significant at $\rho < 0.001$.

that the bigger the turnover, the higher the OER. In terms of positive fiscal adjustment ratio, small firms show the largest (M = 9.50, SD = 25.78), followed by medium-large firms (M = 6.48, SD = 20.30) and medium firms (M = 5.74, SD = 18.70). In contrast, large firms report the lowest (M = 4.79, SD = 15.55). Nevertheless, this group shows the largest negative fiscal adjustment ratio (M = 1.07, SD = 5.17). As Table 4 presents, annual turnover positively correlates with the negative fiscal adjustment ratio, with a larger turnover resulting in a higher NFAR.

Figure 4 plots the mean value of each variable evaluated throughout the fiscal year 2014 to 2019. It can be noticed that, in general, there are parallel trends in each panel. For instance, in terms of gross profit margin, it can be observed that the small firms continuously reported the most significant figure compared to other groups throughout the six-year observation period. Conversely, large firms consistently had the lowest



Note : These graphs show the results of a two-way ANOVA describing plots of the various means of the studied variables for groups with differing annual turnover for fiscal year 2014-2019. The vertical axis depicts the scale of estimated marginal mean scale. The horizontal axis represents fiscal years. All differences are statistically significant at the 0.01

Figure 4 The Difference of Variable Means Among Four Groups

GPM from 2014 to 2019. Further, as described earlier in Table 4, large firms have the lowest negative fiscal adjustment ratio figure. The advantage of providing the graphical illustration in Figure 4 is that the similar tendencies seem independent from the year of observation since the order is depicted consistently.

4.3 Artificial Neural Networks (ANN)

The primary objective of this study is to examine how accurately the multilayer perceptron (MLP) Neural Networks approach predicts the category of taxpaying behaviour of Indonesian firms by analyzing the tax record data of Indonesian firms for six fiscal years.

Table 6 shows the number of observations used to build the ANN model in four datasets. As indicated earlier, following Bekesiene et al. (2021), the administrative data in this study were divided into three groups with the following proportions: (i) training data 60% (n = 323,394); (ii) testing 20% (n = 107,297); and (iii) holdout data 20% (n =107,534). The application in the analysis excluded twenty-eight observations. The table suggests that the largest observation is for firms with annual turnover between IDR 5 billion to IDR 15 billion (42%), while the smallest proportion is firms with annual turnover between IDR 50 billion to IDR 100 billion (11%).

Table 5 describes the number of neurons in each layer and seven independent variables used in the analysis (input layer): (i) annual tax return type, (ii) gross profit margin, (iii) operating profit margin, (iv) other income ratio, (v) other expense ratio, (vi) positive fiscal adjustment ratio, (vii) negative fiscal adjustment ratio. In this analysis, one categorical variable is included in the factors group, and six scale variables are included in the covariates category, comprising seven inputs.

The application's automatic architecture feature shows two hidden layers in the model.[24] There are seven and five nodes in the first and second hidden layers, respectively. The output layer has three units that represent the category of taxpaying behaviour. The activation of the hidden layer function in this analysis uses a Sigmoid, while

Table 5 Network Information

Input Layer	Factors	1	Annual tax return type
	Covariates	1	Gross profit margin (%)
		2	Operating profit margin (%)
		3	Other income ratio (%)
		4	Other expense ratio (%)
		5	Positive fiscal adj. ratio (%)
		6	Negative fiscal adj. ratio (%)
	Number of Units[a]		9
	Rescaling Method for Covariates		Standardised
Hidden Layer(s)	Number of Hidden Layers		2
	Number of Units in Hidden Layer 1[a]		7
	Number of Units in Hidden Layer 2[a]		5
	Activation Function		Sigmoid
Output Layer	Dependent Variables	1	Three CTTOR Category
	Number of Units		3
	Activation Function		Softmax
	Error Function		Cross-entropy

Note: a. Excluding the bias unit. This network information summary is identical for all four grups in this study.

[24] Hidden layers allow the ANN to emulate non-linear patterns more accurately in the data. Without a hidden layer, the ANN will behave like an ordinary linear model which cannot detect non-linear patterns (Aryadoust & Baghaei, 2016).

Table 6 Case Processing Summary

	Small firms (IDR 5 billion to 15 billion)		Medium firms (IDR 15 billion to 50 billion)		Medium-large firms (IDR 50 billion to 100 billion)		Large firms (More than IDR 100 billion)	
Sample	N	Percent	N	Percent	N	Percent	N	Percent
Training	135,140	60.3%	99,779	60.0%	36,816	59.8%	51,659	59.8%
Testing	44,634	19.9%	33,167	20.0%	12,236	19.9%	17,260	20.0%
Holdout	44,393	19.8%	33,215	20.0%	12,517	20.3%	17,409	20.2%
Valid	224,167	100.0%	166,161	100.0%	61,569	100.0%	86,328	100.0%
Excluded	14		10		3		2	
Total	224,181		166,171		61,572		86,330	

Note: the 60%-20%-20% proportion is not precise due to the excluded observations and rounding. The excluded observations are due to missing data in the types of the annual tax return.

the output layer uses Softmax. Cross-entropy is used as an error function since the Softmax method is used as an activation function. Structure of Neural Networks for the Model Prediction in Appendix 3 depicts the network diagrams visualising the number of nodes in the input, hidden, and output layers.

The size of the input nodes represents the extent of the influence of the related independent variables on the dependent variable. Larger rectangles imply a more substantial influence of the related independent variable on the output. For example, the annual tax return type and operating profit margin nodes seem to have a weak influence on the CTTOR categories. These impacts will be discussed in detail later.

Next, Table 7 presents a summary of information related to the training (and testing) and validation results on the holdout sample. The value of the cross-entropy error is also presented for both the training sample and the testing sample—it shows the value of the error function minimised by the ANN model during the training phase. The reduced cross-entropy error values for the testing sample compared to the training samples for all groups indicate no overfitting of the training data in the network models.

These results justify the role of sample testing in preventing overtraining. Based on Table 7, the proportion of incorrect predictions from the training samples was low—i.e., 10.9%, 6.4%, 7.7%, and 6.8% for group small, medium, medium-large, and large firms, respectively. Likewise, the proportion of incorrect predictions from the testing samples was 10.8%, 6.4%, 7.6%, and 7.0% for small, medium, medium-large, and large firms, respectively. For the holdout sample, the proportion of incorrect predictions is 10.9%, 6.4%, 6.9%, and 7.1%, respectively, for each group. These results suggest that the models have performed well in predicting taxpaying behaviour.

Table 7 ANN Model Summary

		Small firms	Medium firms	Medium-large firms	Large firms
Training	Cross Entropy Error	42,037.16	22,814.93	9,211.94	10,952.95
	Percent Incorrect Predictions	10.9%	6.4%	7.7%	6.8%
	Stopping Rule Used	Max. number of epochs (100) exceeded	Max. number of epochs (100) exceeded	Max. number of epochs (100) exceeded	Max. number of epochs (100) exceeded
	Training Time	0:00:10,35	0:00:07,97	0:00:03,24	0:00:04,56
Testing	Cross Entropy Error	13,972.8	7,444.7	3,078.6	3,750.5
	Percent Incorrect Predictions	10.8%	6.4%	7.6%	7.0%
Holdout	Percent Incorrect Predictions	10.9%	6.4%	6.9%	7.1%

Dependent Variable: CTTOR Category

Table 8 provides information about the accuracy (i.e., confusion matrix) of the ANN model for training, testing, and holdout data samples. The formula for sensitivity (i.e., recall or true positive rate) is $\frac{TP}{TP+FN}$ x 100%. The formula for the specificity (i.e., true negative rate) is $\frac{TN}{TN+FP}$ x 100%, while the formula for the model's accuracy is $\frac{TN+TP}{TN+FP+TP+FN}$ x 100%. Each group has a relatively high accuracy rate of 89.1% to 93.6% in aggregate. For instance, the model accurately predicts 93.8% of firms have low CTTOR for all training, testing, and holdout data for small firms. For the moderate CTTOR category, the model predicts with an accuracy of 72.6%, 72.7%, and 72.2%, respectively, for the training, testing, and holdout data. For firms with high CTTOR, the model predicts with an accuracy of 89.3%, 89.7%, and 89.8%, respectively, for training, testing, and holdout data.

More specifically, in the training data of small firms, 75,006 firms are correctly classified as having low CTTOR, with 3,103 firms and 1,847 firms falsely predicted to have moderate and high CTTOR, respectively. In the moderate CTTOR group, the model correctly predicts 16,862 firms, with 3,472 firms and 2,876 firms falsely classified as low and high CTTOR, respectively. Furthermore, in the high CTTOR category, the model accurately predicts 28,566 firms, whereas 947 and 2,461 firms, respectively, are incorrectly labelled as low and

Table 8 Accuracy of classification

Panel A

		Predicted for small firms (IDR 5 billion to 15 billion)				Predicted for medium firms (IDR 15 billion to 50 billion)			
Sample	Observed	Low	Moderate	High	Percent Correct	Low	Moderate	High	Percent Correct
Training	Low	75,006	3,103	1,847	93.8%	53,201	1,211	1,398	95.3%
	Moderate	3,472	16,862	2,876	72.6%	1,300	13,690	1,319	83.9%
	High	947	2,461	28,566	89.3%	474	703	26,483	95.7%
	Percent	58.8%	16.6%	24.6%	89.1%	55.1%	15.6%	29.3%	93.6%
Testing	Low	24,762	1,024	604	93.8%	17,716	384	431	95.6%
	Moderate	1,139	5,586	954	72.7%	418	4,520	467	83.6%
	High	333	760	9,472	89.7%	183	230	8,818	95.5%
	Percent	58.8%	16.5%	24.7%	89.2%	55.2%	15.5%	29.3%	93.6%
Holdout	Low	24,675	1,027	599	93.8%	17,804	396	430	95.6%
	Moderate	1,174	5,546	961	72.2%	455	4,602	433	83.8%
	High	270	796	9,345	89.8%	156	258	8,681	95.4%
	Percent	58.8%	16.6%	24.6%	89.1%	55.4%	15.8%	28.7%	93.6%

Panel B

		Predicted for medium-large firms (IDR 50 billion to 100 billion)				Predicted for large firms (More than IDR 100 billion)			
Sample	Observed	Low	Moderate	High	Percent Correct	Low	Moderate	High	Percent Correct
Training	Low	21,191	335	472	96.3%	27,356	763	1,030	93.8%
	Moderate	947	3,819	499	72.5%	634	6,064	616	82.9%
	High	315	252	8,986	94.1%	275	182	14,739	97.0%
	Percent	61.0%	12.0%	27.0%	92.3%	54.7%	13.6%	31.7%	93.2%
Testing	Low	7,121	114	169	96.2%	9,162	299	316	93.7%
	Moderate	308	1,270	159	73.1%	206	2,000	232	82.0%
	High	104	70	2,921	94.4%	84	63	4,898	97.1%
	Percent	61.6%	11.9%	26.6%	92.4%	54.8%	13.7%	31.6%	93.0%
Holdout	Low	7,182	106	136	96.7%	9,177	296	352	93.4%
	Moderate	299	1,331	139	75.2%	240	1,995	206	81.7%
	High	104	81	3,139	94.4%	75	72	4,996	97.1%
	Percent	60.6%	12.1%	27.3%	93.1%	54.5%	13.6%	31.9%	92.9%

Dependent Variable: CTTOR Category (i.e., low, moderate, high).

moderate CTTOR categories. This model results in an overall accuracy rate of 89.1%, 89.2%, and 89.1% for training, testing, and holdout sample in the group A. For the medium firms, the model correctly predicts 95.3%, 95.6%, and 95.6% of firms with moderate CTTOR for training, testing, and holdout data, respectively. For the moderate CTTOR category, the group B model predicts with an accuracy of 83.9%, 83.6%, and 83.8% for training, testing, and holdout data, respectively.

For companies with a high CTTOR, the model predicts with an accuracy of 93.6% for all training, testing, and holdout data. In medium firms' category training data, 53,201 entities are accurately identified as having a low CTTOR, whereas 1,211 firms and 1,398 firms are incorrectly categorised as having a moderate and high CTTOR, respectively. In the moderate CTTOR category, the model accurately predicts 13,690 businesses, whereas 1,300 and 1,319 entities are incorrectly labelled as low and high CTTOR, respectively. In addition, the model successfully predicts 26,483 enterprises in the high CTTOR group, but 474 and 703 firms are misclassified in the low and moderate CTTOR categories, respectively. This model gives the medium firms' category a total accuracy of 93.6% for all training, testing, and holdout sample data.

Similarly, as Panel B of Table 8 shows, the models for medium-large and large firms also demonstrate high prediction accuracy. For instance, the model in the medium-large category results in an overall accuracy rate of 92.3%, 92.4%, and 93.1% for training, testing, and holdout sample, respectively. Similarly, the ANN model gives large firms' categories a total accuracy of 93.2%, 93%, and 92.9% for training, testing, and holdout sample data, respectively.

The dependent variable in the present study comprised three levels of CTTOR: low (1), moderate (3), and high (3). Specificity, also known as the real negative rate, is the percentage of high-CTTOR corporations that were accurately categorised into Group 3. Sensitivity is the fraction of actual low-CTTOR firms that were correctly placed into Group 1 (Fawcett, 2006). When a high-CTTOR is misclassified as 1 or 2, or a low-CTTOR is incorrectly classified as 2 or 3 (a false negative), an error has been made (a false positive; Swets, 1988). As Figure 5 depicts, these indices are visually shown in a Receiver Operating Characteristic (ROC) curve for the dependent variable.

With a diagonal or no-discrimination line connecting the bottom left corner of the graph to the diagonally opposite corner, ROC shows 1-specificity on the x-axis versus sensitivity on the y-axis. Poor classification is shown by points below the no-discrimination line, whereas successful classification is indicated by points above the no-discrimination line (Fawcett, 2006). The formula for sensitivity (i.e., recall or true positive rate) is $\frac{TP}{TP+FN}$ x 100%. The formula for the specificity (i.e., true negative rate) is $\frac{TN}{TN+FP}$ x 100%, while the formula for the model's accuracy is $\frac{TN+TP}{TN+FP+TP+FN}$ x 100%. Figure 5 describes the area under the curve (AUC) values range in this study range between 0.941 and 0.988, indicating the prediction results are very accurate (see Appendix 4 and Appendix 5 for the cumulative gain charts and the lift charts).

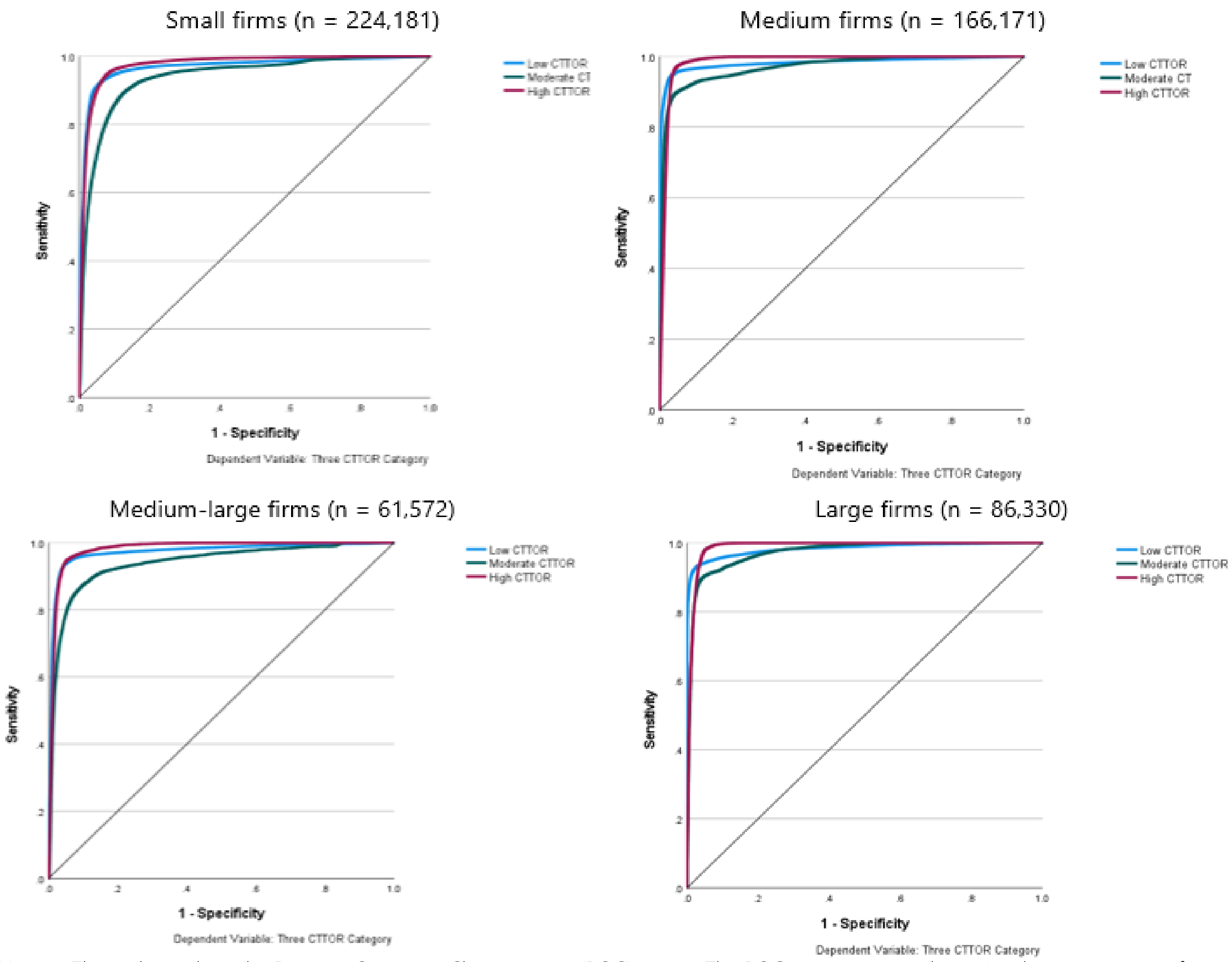
Further, Table 9 shows the assessment of the independent variables in the ANN models, which is measured by relative importance and normalized importance. The independent variable importance in the ANN models suggests that each group has heterogenous predictors. This study focuses on the three most important predictors.[25] For instance, as shown in Table 9, for small firms, the variable *positive fiscal adjustment ratio* has the highest score (0.259; normalized importance =

[25] As a predictive model, the ANN can be combined with the Decision Tree model for more complete analysis (IBM, 2021). One of the popular approaches in the Decision Tree model is Chi-Squared Automatic Interaction Detection (CHAID). In the CHAID approach, only three levels of the tree—which in this case refers to the independent variable—are generated (IBM, 2017). Based on these considerations, the author only focuses on the three most important variables.

100%), followed by *other income ratio* (0.197; normalized importance = 76.1%) and *operating profit margin* (0.138; normalized importance = 53.2%). Differently, for medium firms, the variable *negative fiscal adjustment ratio* has the highest score (0.179; normalized importance = 100%), followed closely by *positive fiscal adjustment ratio* (0.178; normalized importance = 99.4%) and *other income ratio* (0.178; normalized importance = 99.3%).

Meanwhile, for medium-large firms, the variable *other expense ratio* has the highest score (0.222; normalized importance = 100%), followed closely by *negative fiscal adjustment ratio* (0.207; normalized importance = 93.0%) and *operating profit margin* (0.201; normalized importance = 90.5%). Finally, for large firms, the variable *other*

Panel A: ROC Curves



Note: These charts show the Receiver Operating Characteristics (ROC) curves. The ROC curve is a two-dimensional representation of classification performance (Fawcett, 2006). It provides an overview of the sensitivity and specificity levels based on a combination of training and sampling data. A 45-degree diagonal line from the bottom left to the top right represents the no-discrimination line. A point below the no-discrimination line indicates an inaccurate classification and a point above the no-discrimination line indicates an effective classification result (Fawcett, 2006).

Panel B: Area Under the Curve

		Small firms	Medium firms	Medium-large firms	Large firms
CTTOR	Low CTTOR	0.970	0.974	0.981	0.983
Category	Moderate CTTOR	0.941	0.965	0.963	0.973
	High CTTOR	0.977	0.981	0.984	0.988

Note: The area under the curve (AUC) values range in this study range between 0.941 and 0.988. In general, the AUC values are classified into five categories: 0.50–0.60 (fail); 0.60–0.70 (less accurate); 0.70–0.80 (fairly accurate); 0.80–0.90 (accurate); and 0.90–1.00 (very accurate). Thus, the AUC value of the study shows that this prediction result is very accurate.

Figure 5 ROC Curves and Area Under the Curve *expense ratio* has the highest score (0.208; normalized importance = 100%), followed closely by *operating profit margin* (0.195; normalized importance = 93.9%) and *negative fiscal adjustment ratio* (0.191; normalized importance = 91.9%).

Interestingly, while the three most influential factors appear somewhat varied, the two variables with the lowest influence are similar—i.e., the *gross profit margin* and the *annual tax return type*. For instance, for small firms, *gross profit margin* has an importance score of 0.070 (normalized importance = 26.9%), and the *annual tax return type* has an importance score of 0.110 (normalized importance = 42.6%). For medium firms, *gross profit margin* has an importance score of 0.031 (normalized importance = 17.1%), and the *annual tax return type* has an importance score of 0.098 (normalized importance = 54.5%). The same patterns apply to both medium-large and large firms. These results imply that the prediction of the taxpaying behaviour of Indonesian firms does not depend on the gross profit margin and the type of annual tax return.

For easier comprehension, the results described in Table 9 are diagrammatically represented in Figure 6. As shown in the graph, for instance, the variable *positive fiscal adjustment ratio*, *other income ratio*, and *OPM* are the three most influential variables in predicting the CTTOR category for small firms. The variable *positive fiscal adjustment ratio*, *negative fiscal adjustment*, and *other income ratio* are the three most important variables in predicting the CTTOR category for medium firms. It is interesting to recall that medium-large and large firms have similar three most important factors: *other expense ratio*, *negative fiscal adjustment ratio*, and *operating profit margin*. Also, as **Error! Reference source not found.** shown, the two variables with the lowest influence for all four groups are the *gross profit margin* and annual *tax return type* variables.

The findings indicate that for small firms, *positive fiscal adjustment ratio*, *other income ratio*, and *OPM* have stronger relationships with taxpaying behaviour than *other income ratio* and *negative fiscal adjustment ratio*. Businesses in this category reported the highest mean value for

Table 9 Independent Variable Importance

Panel A

	Small firms (IDR 5 billion to 15 billion)		Medium firms (IDR 15 billion to 50 billion)	
	Importance	Normalized Importance	Importance	Normalized Importance
Annual tax return type	0.110	42.6%	0.098	54.5%
Gross profit margin (%)	0.070	26.9%	0.031	17.1%
Operating profit margin (%)	0.138	53.2%	0.172	95.8%
Other income ratio (%)	0.197	76.1%	0.178	99.3%
Other expense ratio (%)	0.114	43.9%	0.165	92.2%
Positive fiscal adj. ratio (%)	0.259	100.0%	0.178	99.4%
Negative fiscal adj. ratio (%)	0.113	43.7%	0.179	100.0%

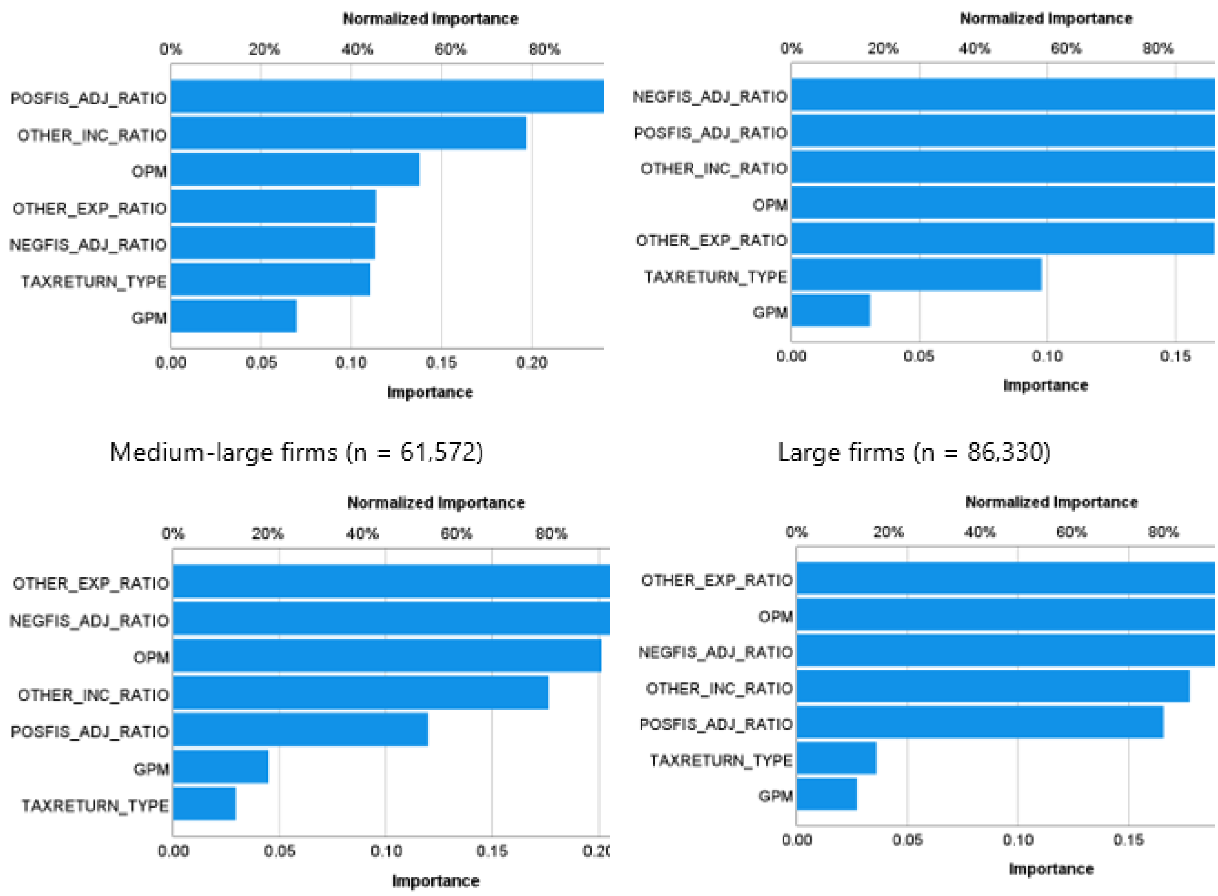
Panel B

	Medium-large firms (IDR 50 billion to 100 billion)		Large firms (More than IDR 100 billion)	
	Importance	Normalized Importance	Importance	Normalized Importance
Annual tax return type	0.029	13.2%	0.036	17.4%
Gross profit margin (%)	0.045	20.1%	0.027	13.1%
Operating profit margin (%)	0.201	90.5%	0.195	93.9%
Other income ratio (%)	0.176	79.3%	0.178	85.6%
Other expense ratio (%)	0.222	100.0%	0.208	100.0%
Positive fiscal adj. ratio (%)	0.120	53.8%	0.166	79.9%
Negative fiscal adj. ratio (%)	0.207	93.0%	0.191	91.9%

positive fiscal adjustment ratio—i.e., 9.5%, meaning that companies in this category will increase their taxable revenue by IDR 95.000 for every IDR one million annual turnovers. This number is higher than the population's average (7.24%). Firms in this category, in contrast, reported the lowest *other income ratio*, at only 1% of turnover, meaning that firms in this category will report other business income of IDR 1,000 for every IDR one million in annual turnover. This proportion is lower than the population average (1.2%), with firms with an annual turnover of more than IDR 100 billion reporting the largest other income ratio (1.7%).

Further, *operating profit margin* is the third most crucial predictor for firms in this category. This group's OPM is within the range of 4.78%—slightly higher than the population's mean of 4.49%—meaning that businesses in this category declared operating profits of IDR 47,800 for every IDR one million in sales. These findings suggest that part 5 Form 1771-I (positive fiscal adjustment), part 1e Form 1771-I (other business income), and part 1c Form 1771-I (operating expense) are potential areas of misreporting for small businesses when it comes to annual income tax returns.[26]

The best predictors of taxpaying behaviour for medium firms are *negative fiscal adjustment ratio*, *positive fiscal adjustment ratio*, and *other income ratio*. *Negative fiscal adjustment ratio* for firms in this category is 0.52%, which indicates that



Note: The graph displays the result of ANN's analysis separately for four groups of firms. The value of IVI indicates that the variable positive fiscal adjustment is of the most significant importance in the prediction of taxpaying behaviour for firms with annual turnover between IDR 5 billion and IDR 15 billion, while the negative fiscal adjustment is of the most significant importance for firms with annual turnover between IDR 55 billion and IDR 50 billion. For firms with an annual turnover of more than IDR 50 billion, the variable other business expense is the most important independent variable for predicting taxpaying behaviour.

Figure 6 Independent Variable Importance (IVI) by Annual Turnover Category

[26] Note that gross profit margin is the least important predictor for firms in this group. *Operating expense* is the single feature distinguishing operating profit margin from gross profit margin.

entities in this group reduced their taxable income of IDR 5,200 for every IDR one million sales. This percentage falls just short of the population mean of 0.60%. It should be noted that there is a positive correlation between *negative fiscal adjustment ratio* and turnover for businesses in this category—i.e., the negative fiscal adjustment increases with more turnover. Like for small firms, *positive fiscal adjustment ratio* is also a significant predictor of taxpaying behaviour for medium firms. Medium firms make average positive fiscal adjustments of IDR 57,400 for each IDR one million sales (5.74%). This number is slightly below the population average of 7.24%.

Further, the predictor comparable to that of small firms is *other income ratio*. Medium firms reported an OIR of 1.13%, indicating that medium firms report additional business income of IDR 11,300 for every IDR one million turnover. These findings imply that potential areas of misreporting for medium firms are part 6e Form 1771-I (negative fiscal adjustment items), part 5m Form 1771-I (positive fiscal adjustment items), and part 1e Form 1771-I (other business income).

The best predictors of taxpaying behaviour for medium-large firms are *other expense ratio*, *negative fiscal adjustment ratio*, and *operating profit margin*. *Other expense ratio* for medium-large firms is 1.95%, meaning that entities in this category claimed additional business expenses of IDR 19,500 for every IDR one million sales. This percentage is higher than the population means of 1.59%. It should be noted that there is a positive correlation between other expense ratio and turnover—i.e., the additional business expense increases with more turnover. Further, medium-large firms make average negative fiscal adjustments of IDR 6,700 for each IDR one million sales (0.60%). This number is slightly higher than the population average of 0.67%. The predictor comparable to that of small firms is operating profit margin. Medium-large firms reported an OPM of 3.96%, indicating that medium-large firms report operational profits of IDR 39,600 for every IDR one million turnover. These results suggest that part 1c Form 1771-I (operating expense), part 1e Form 1771-I (other business expenditure), and part 6e Form 1771-I (negative fiscal adjustment) are possible areas of misreporting for medium-large enterprises.

Interestingly, the most prominent predictors for large corporations appear to resemble those of medium-large firms: *other expense ratio*, *operating profit margin*, and *negative fiscal adjustment ratio*. Large firms reported the highest portion of other expense ratio among other categories. *Other expense ratio* for large firms is 2.54%, meaning that large firms claimed additional business expenses of IDR 25,400 for every IDR one million sales. This ratio is 60% higher than the population means of 1.59%. The predictor comparable to that of medium-large firms—and small firms—is the operating profit margin. Large corporations reported an OPM of 5.04%, indicating that large firms report operational profits of IDR 50,400 for every IDR one million turnover. Further, large firms make average negative fiscal adjustments of IDR 10,700 for each IDR one million sales (1.07%). This number is the highest among other categories, as the relationship between negative fiscal adjustment and turnover is positive.

These findings imply that potential areas of misreporting for large corporations are part 1e Form 1771-I (other business expense), part 1c Form 1771-I (operating expense), and part 6e Form 1771-I (negative fiscal adjustment). Interestingly, while other business income is a prominent predictor for small and medium firms, an additional business expense is crucial for medium-large and large firms. A possible explanation for this might be that large firms have more income visibility and therefore have less control over misreporting of income streams. For this reason, the potential areas of misreporting are expenses over which large firms have more control. In contrast, small and medium firms have less visibility in income and therefore have more control over misreporting concerning income streams.

5. CONCLUSIONS

This study utilised an ANN model to predict the taxpaying behaviour of Indonesian firms. Initially, this paper examined whether firms in four size-based categories—i.e., small, medium, medium-large, and large firms—also differ considerably in terms of their gross profit margin, operating profit margin, other income ratios, other expense ratios, positive fiscal adjustment ratios, and negative fiscal adjustment ratios. This study found that firms that vary in annual turnover categories also differ considerably in the variables under study. The results suggest that small firms have the highest gross profit margin compared to other groups. However, in terms of operational profitability, large firms reported the highest. Large firms also declared the most considerable portion of other business income and the largest share of other business expenses. Interestingly, small firms show the most significant positive fiscal adjustment ratio, while large corporations declare otherwise. In contrast, large firms report the largest negative fiscal adjustment ratio.

This study suggests that AI approaches may have advantages over conventional statistical methods in addressing various issues, particularly those involving nonlinear patterns (see, for example, Aryadoust and Baghaei, 2016; Bahrammirzaee, 2010; Chen et al., 2011; Lin et al., 2012; Sánchez-Serrano et al., 2020). However, further research is needed to determine the precise contexts in which AI approaches may be most effective and compare their performance with conventional statistical methods.

This paper demonstrated that the ANN models accurately predict the taxpaying behaviour of Indonesian firms across four groups. The classification accuracy rate was high, with an overall 92.2% accuracy in categorising the firms into low, moderate, and high CTTOR. This study used the MLP module from ANN and built on a model with a 60%–20%–20% formation. The results also showed that the channels responsible for taxpaying behaviour vary for different groups. For small firms, the three most influential factors in predicting taxpaying behaviour are the positive fiscal adjustment ratio, other income ratio, and OPM. This means that the extent to which firms positively adjust their commercial, the presence of other business income, and the share of the cost of goods sales are critical factors for small firms. The accuracy rate, significant predictors, and areas of concern for all groups can be summarised in Table 10 as follows.

These findings, while preliminary, suggest that it is possible to minimise the monitoring time and expense since the model is applied to a national-level big data set and directly exhibits several areas of concern in the annual tax returns. The findings would assist decision-makers in tax administrations about potential areas of misreporting, enabling them to develop evidence-based and effective policy actions.

6. IMPLICATIONS AND LIMITATIONS

6.1 Implications

Although further research is required to delve deeply into the effectiveness of machine learning in comprehending taxpaying behaviour, this work has revealed new insights into how potential areas of misreporting can be identified with minimal cost. Several practical implications for the Indonesian tax authorities have been shown by illustrating how various variables affect taxpaying behaviour across different groups, including their areas of concern in the annual income tax returns (as presented in Table 10).

This study provides findings relevant to making targeted strategic decisions—for example, the results of this analysis can be used to segment groups of firms deemed necessary to increase their level of tax payment, focusing on their variable interests. These may support the development of effective treatment strategies to improve compliance rates. Particularly in developing a compliance risk management plan, with supervised machine learning, the case selection algorithm may highlight complex data patterns linked to successful case outcomes while deemphasizing those that did not (Brondolo et al., 2022). It would be of assistance because the DGT has not yet used AI in its operations (ADB, 2022).

Table 10 Summary: Accuracy Rate, Predictors, and Areas of Concern

	Prominent predictors by firms' category			
	Small firms (IDR 5B to 15B)	Medium firms (IDR 15B to 50B)	Medium-large firms (IDR 50B to 100B)	Large firms (> IDR 100B)
Accuracy rate*	89.1%	93.6%	93.1%	92.9%
Annual return type	No	No	No	No
GPM	No	No	No	No
OPM	Yes	No	Yes	Yes
OIR	Yes	Yes	No	No
OER	No	No	Yes	Yes
PFAR	Yes	Yes	No	No
NFAR	No	Yes	Yes	Yes
Areas of concern within annual income tax return	Part 1c Form 1771-I, part 1e Form 1771-I, and part 5 Form 1771-I	Part 1e Form 1771-I, part 5m Form 1771-I, and 6e Form 1771-I	Part 1c Form 1771-I, part 1e Form 1771-I, and part 6e Form 1771-I	

Note: * accuracy rate based on holdout sample; GPM = gross profit margin; OPM = operating profit margin; OIR = other income ratio; OER = other expense ratio; PFAR = positive fiscal adjustment ratio; NFAR = negative fiscal adjustment ratio

As Dom et al. (2022) argue, getting insights into the possibilities of more advanced emerging technologies, like blockchain and AI, in lower-capacity settings when their usage becomes more widespread would be intriguing. In this sense, variations were identified in the channels responsible for taxpaying behaviour among different groups of corporate income taxpayers in Indonesia. These results can assist tax administrations in identifying potential areas of misreporting, reducing monitoring time and costs, and developing evidence-based and effective policy actions by identifying areas of concern in annual tax returns.

6.2 Limitations

Due to practical constraints, this study makes no distinction between legal tax avoidance and illegal tax evasion. In addition to using limited variables, this article does not address the impact of enforcement or monitoring activities. The findings also raise intriguing questions regarding the nature and the causal relationships between the predictors and the conforming and non-conforming tax avoidance among Indonesian firms. Unfortunately, this study cannot provide further detailed information on, for example, why—or how—factors such as positive and negative fiscal adjustment, operating profit margin, other business income, and other business expenses are the most critical factors in determining taxpaying behaviour, while in contrast, gross profit margin and the type of tax return are negligible. Studies with different empirical data and analytical approaches are needed to answer those questions. Therefore, a further study with more focus on the causal inference among the identified predictors and underreporting behaviour is suggested.

Acknowledgment

I thank the editor, the committee of Taxation Call for Paper 2022, and two anonymous referees for their valuable comments and suggestions. I would also like to express my appreciation to Yon Arsal, Assistant to the Minister of Finance for Tax Compliance; Ihsan Priyawibawa, Director of Tax Potential, Compliance, and Revenue DGT; Eureka Putra, Deputy Director of Policy Impact DGT; and all members of the Taxpayer Behaviour Studies team, including Amalia Indah Sujarwati, Bobby Indra Bachriansyah, Dike Danila, Eta Fasita, Irfan Yulianto, Jens Naki, Leonard Tantripal, Martina Merdekawati Putri, Novi Dewi Harini, Octa Naafi' Sutantri, and Tifara Ashari, for their valuable support and assistance. The views expressed here are those of the author and do not necessarily

reflect those of the individuals or organisations acknowledged here.

REFERENCES

[1] *Agostinelli*, F., Hoffman, M., Sadowski, P., & Baldi, P. (2015). *Learning activation functions to improve deep neural networks*. arXiv 2014. https://doi.org/10.48550/arXiv.1412.6830

[2] Alm, J. (2019). *Can Indonesia reform its tax system? Problems and options* (Tulane Economics Working Paper Series 1906). Tulane University, Department of Economics.

[3] Aryadoust, V., & Baghaei, P. (2016). Does EFL readers' lexical and grammatical knowledge predict their reading ability? Insights from a perceptron artificial neural network study. *Educational Assessment*, *21*(2), 135–156. https://doi.org/10.1080/10627197.2016.1166343

[4] Asian Development Bank. (2022). *A comparative analysis of tax administration in Asia and the Pacific* (5th ed.). Asian Development Bank. http://dx.doi.org/10.22617/TCS220183

[5] Badertscher, B. A., Katz, S. P., Rego, S. O., & Wilson, R. J. (2019). Conforming tax avoidance and capital market pressure. *The Accounting Review*, *94*(6), 1-30. https://doi.org/10.2308/accr-52359

[6] Bahrammirzaee, A. (2010). A comparative survey of artificial intelligence applications in finance: Artificial neural networks, expert systems and hybrid intelligent systems. *Neural Computing and Applications*, *19*(8), 1165-1195. https://doi.org/10.1007/s00521-010-0362-z

[7] Bekesiene, S., Smaliukiene, R., & Vaicaitiene, R. (2021). Using artificial neural networks in predicting the level of stress among military conscripts. *Mathematics*, *9*(6), 626. https://doi.org/10.3390/math9060626

[8] Bird, R. M. (2010). *Taxation and development*. Economic Premise 34, The World Bank. http://documents.worldbank.org/curated/en/757771468339579952/Taxation-and-development

[9] Brondolo, J. D., Chooi, A., Schloss, T., & Siouclis, A. (2022). *Compliance risk management: Developing compliance improvement plans*. Technical Notes and Manuals 2022/001, International Monetary Fund. https://doi.org/10.5089/9798400205910.005

[10] Chen, J. H., Su, M. C., Chen, C. Y., Hsu, F. H., & Wu, C. C. (2011). Application of neural networks for detecting erroneous tax reports from construction companies. *Automation in Construction*, *20*(7), 935–939. https://doi.org/10.1016/j.autcon.2011.03.011

[11] Cook, T. R. (2020). Neural networks. In P. Fuleky (Ed.), *Macroeconomic forecasting in the era of big data: Theory and practice* (pp. 161-189). Springer.

[12] Devos, K. (2014). *Factors influencing individual taxpayer compliance behaviour*. Springer. https://doi.org/10.1007/978-94-007-7476-6

[13] Directorate General of Taxes. (2015). *Keputusan Direktur Jenderal Pajak nomor KEP-95/PJ/2015 tentang rencana strategis Direktorat Jenderal Pajak tahun 2015-2019* [Strategic plan of the Directorate General of Taxes years 2015–2019, Decree of Director General of Taxes number KEP-95/PJ/2015]. Directorate General of Taxes.

[14] Directorate General of Taxes. (2021). *Annual Report 2020: Constantly optimizing opportunities amid challenging times*. Directorate General of Taxes.

[15] Directorate General of Taxes. (2022). *CRMBI langkah awal menuju data driven organization* [CRMBI First step towards data driven organization]. Directorate General of Taxes.

[16] Dom, R., Custers, A., Davenport, S., & Prichard, W. (2022). *Innovations in tax compliance: Building trust, navigating politics, and tailoring reform*. World Bank Group. https://doi.org/10.1596/978-1-4648-1755-7

[17] Dyreng, S. D., Hanlon, M., & Maydew, E. L. (2008). Long-run corporate tax avoidance. *The Accounting Review*, *83*(1), 61–82. https://www.jstor.org/stable/30243511

[18] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

[19] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). *From data mining to knowledge discovery in databases*. AI *Magazine*, *17*(3), 37-54. https://doi.org/10.1609/aimag.v17i3.1230

[20] Gebhart, M. S. (2017). Measuring corporate tax avoidance – An analysis of different measures. *Junior Management Science*, *2*(2), 43–60. https://doi.org/10.5282/jums/v2i2pp43-60

[21] Girasa, R. (2020). *Artificial intelligence as a disruptive technology: Economic transformation and government regulation*. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-35975-1

[22] González, P. C., & Velásquez, J. D. (2013). Characterization and detection of taxpayers with false invoices using data mining techniques. *Expert Systems with Applications*, *40*(5),1427–1436. https://doi.org/10.1016/j.eswa.2012.08.051

[23] Goumagias, N. D., Hristu-Varsakelis, D., & Saraidaris, A. (2012). A decision support model for

tax revenue collection in Greece. *Decision Support Systems*, *53*(1), 76–96. https://doi.org/10.1016/j.dss.2011.12.006

[24] Graupe, D. (2013). *Principles of artificial neural networks* (3rd ed.). World Scientific. https://doi.org/10.1142/8868

[25] Hanlon, M., & Heitzman, S. (2010). A review of tax research. *Journal of Accounting and Economics*, *50*(2-3), 127-178. https://doi.org/10.1016/j.jacceco.2010.09.002

[26] Haykin, S. S. (1999). *Neural networks: A comprehensive foundation* (2nd ed.). Macmillan College Publishing.

[27] Heij, G. (2001). The 1981-83 Indonesian income tax reform process: Who pulled the strings?. *Bulletin of Indonesian Economic Studies*, 37(2), 233–251. https://doi.org/10.1080/00074910152390900

[28] Huang, Z. (2018). Discussion on the development of artificial intelligence in taxation. *American Journal of Industrial and Business Management*, *8*(8), 1817-1824. https://doi.org/10.4236/ajibm.2018.88123

[29] International Business Machines. (2017). *IBM SPSS decision trees 25*. IBM. https://www.ibm.com/docs/en/SSLVMB_25.0.0/pdf/en/IBM_SPSS_Decision_Trees.pdf

[30] International Business Machines. (2021). *IBM SPSS neural networks 28*. IBM. https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/ IBM_SPSS_Neural_Network.pdf

[31] International Monetary Fund. (2014). *Indonesia: Tax policy and administration – Setting a strategy for the coming years*. Technical Assistance Report, IMF.

[32] Jang, S. B. (2019, January 28-30). *A design of a tax prediction system based on artificial neural network* [Conference paper]. 2019 International Conference on Platform Technology and Service (PlatCon), Jeju, South Korea. https://doi.org/10.1109/PlatCon.2019.8669416

[33] Jupri, M., & Sarno, R. (2018, March 6-7). *Taxpayer compliance classification using C4.5, SVM, KNN, I Bayes and MLP* [Conference paper]. 2018 International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia. https://doi.org/10.1109/ICOIACT.2018.8350710

[34] Lin, C. H., Lin, I. C., Wu, C. H., Yang, Y. C., & Roan, J. (2012). The application of decision tree and artificial neural network to income tax audit: The examples of profit-seeking enterprise income tax and individual income tax in Taiwan. *Journal of the Chinese Institute of Engineers*, *35*(4), 401–411. https://doi.org/10.1080/02533839.2012.655901

[35] Mills, L., Erickson, M. M., & Maydew, E. L. (1998). Investments in tax planning. *The Journal of the American Taxation Association*, *20*(1), 1-20.

[36] Murorunkwere, B. F., Tuyishimire, O., Haughton, D., & Nzabanita, J. (2022). Fraud detection using neural networks: A case study of income tax. *Future Internet*,*14*(6),168. https://doi.org/10.3390/fi14060168

[37] Neuman, W. L. (2011). *Social research methods: Qualitative and quantitative approaches* (7th ed.). Allyn & Bacon.

[38] Organisation for Economic Co-operation and Development. (2014). *Measures of tax compliance outcomes: A practical guide*. OECD. http://dx.doi.org/10.1787/9789264223233-en

[39] Organisation for Economic Co-operation and Development. (2016). *Advanced analytics for better tax administration: Putting data to work*. OECD. http://dx.doi.org/10.1787/9789264256453-en

[40] Organisation for Economic Co-operation and Development. (2020a). *Tax administration 3.0: The digital transformation of tax administration*. OECD. http://www.oecd.org/tax/forum-on-tax-administration/publications-and-products/tax-administration-3-0-the-digital-transformation-of-tax-administration.htm

[41] Organisation for Economic Co-operation and Development. (2020b). *Tax and fiscal policy in response to the coronavirus crisis: Strengthening confidence and resilience*. OECD.

[42] Organisation for Economic Co-operation and Development. (2022). *Tax administration 2022: Comparative information on OECD and other advanced and emerging economies*. OECD. https://doi.org/10.1787/1e797131-en

[43] Pérez López, C., Delgado Rodríguez, M. J., & de Lucas Santos, S. (2019). Tax fraud detection through neural networks: An application using a sample of personal income taxpayers. *Future Internet*, *11*(4), 86. https://doi.org/10.3390/fi11040086

[44] Prichard, W., Custers, A. L., Dom, R., Davenport, S.R., &Roscitt, M.A. (2019). *Innovations in tax compliance: Conceptual framework*. Policy Research Working Paper Series 9032, The World Bank. http://documents.worldbank.org/curated/en/816431569957130111/Innovations-in-Tax-Compliance-Conceptual-Framework

[45] Rahimikia, E., Mohammadi, S., Rahmani, T., & Ghazanfari, M. (2017). Detecting corporate tax

evasion using a hybrid intelligent system: A case study of Iran. *International Journal of Accounting Information Systems*, *25*, 1–17. https://doi.org/10.1016/j.accinf.2016.12.002

[46] Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge University Pres.

[47] Rosid, A., Evans, C., & Tran-Nam, B. (2018). Tax non-compliance and perceptions of corruption: Policy implications for developing countries. *Bulletin of Indonesian Economic Studies*, *54*(1), 25-60. https://doi.org/10.1080/00074918.2017.1364349

[48] Rosid, A., & Ariyani, F. (2022). *Does 'information reporting' really matter for tax compliance? The case of Indonesia*. Working Paper Series 22-05. http://dx.doi.org/10.2139/ssrn.4114000

[49] Salihu, I. A., Obid, S. N. S., & Annuar, H. A. (2013). Measures of corporate tax avoidance: Empirical evidence from an emerging economy. *International Journal of Business and Society*, *14*(3), 412-427.

[50] Sánchez-Serrano, J. R., Alaminos, D., Garcia-Lagos, F., & Callejón-Gil, A.M. (2020). Predicting audit opinion in consolidated financial statements with artificial neural networks. *Mathematics*, *8*(8), 1288. https://doi.org/10.3390/math8081288

[51] Saragih, A. H., Reyhani, Q., Setyowati, M. S., & Hendrawan, A. (2022). The potential of an artificial intelligence (AI) application for the tax administration system's modernization: The case of Indonesia. *Artificial Intelligence and Law*, *2022*. https://doi.org/10.1007/s10506-022-09321-y

[52] Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763-1768. https://doi.org/10.1213/ANE.0000000000002864

[53] Slemrod, J. (2007). Cheating ourselves: The economics of tax evasion. *The Journal of Economic Perspectives*, *21*(1), 25-48. https://doi.org/10.1257/jep.21.1.25

[54] Slemrod, J. (2016). *Tax compliance and tax evasion*. In The New Palgrave Dictionary of Economics (pp. 1-6). Palgrave Macmillan. https://doi.org/10.1057/978-1-349-95121-5_2771-1

[55] Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285-1293. https://doi.org/10.1126/science.3287615

[56] Wang, Y., & Wang, P. (2020). New personal tax collection management system based on artifcial intelligence and its application in the middle class. *Journal of Physics: Conference Series*, *1574*, 1–8. https://doi.org/10.1088/1742- 6596/1574/1/012105

[57] Warner, G., Wijesinghe, S., Marques, U., Badar, O., Rosen, J., Hemberg, E., & O'Reilly, U. M. (2015). Modeling tax evasion with genetic algorithms. *Economics of Governance*, *16*(2), 165-178. https://doi.org/10.1007/s10101-014-0152-7

[58] Wendler, T., & Gröttrup, S. (2021). *Data mining with SPSS modeler: Theory, exercises and solutions*. Springer. https://doi.org/10.1007/978-3-030-54338-9

[59] World Bank. (2016). *Indonesia economic quarterly: Resilience through reforms*. The World Bank. https://documents1.worldbank.org/curated/en/458091542207517162/pdf/Indonesia-economic-quarterly-resilience-through-reforms.pdf

[60] York, S. (2011). A risk-based approach to large businesses. In M. S. Khwaja, R. Awasthi, & J. Loeprick (Eds.), *Risk-based tax audits: Approaches and country experiences* (pp. 39-44). The World Bank.

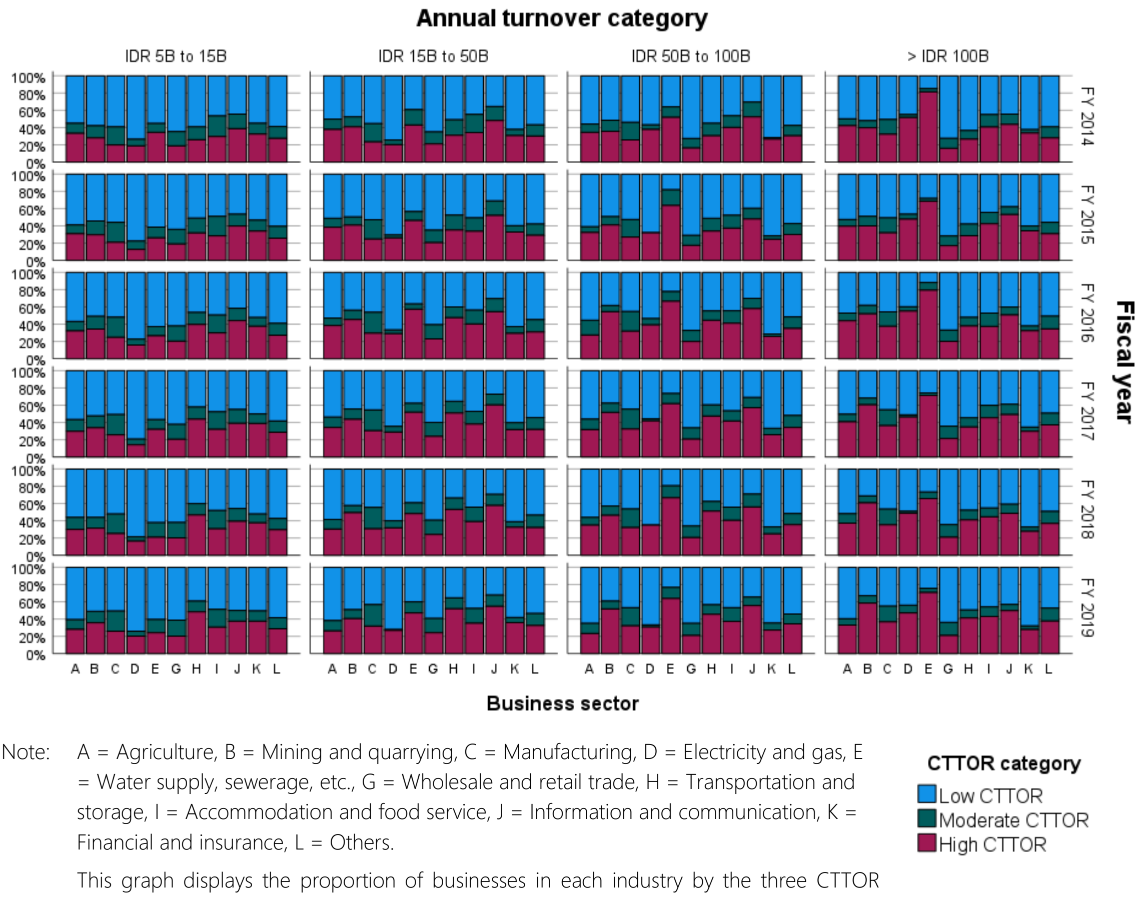
[61] Zhang, G., Hu, M.Y., Patuwo, B.E., & Indro, D.C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, *116*(1), 16–32. https://doi.org/10.1016/S0377-2217(98)00051-4

APPENDICES

Appendix 1 CTTOR Categories by Business Sectors

Business sector	CTTOR category						Total	
	Low		Moderate		High			
Agriculture	5,958	56%	1,148	11%	3,601	34%	10,707	100%
Mining and quarrying	3,223	45%	760	11%	3,222	45%	7,205	100%
Manufacturing	46,512	49%	20,068	21%	27,816	29%	94,396	100%
Electricity and gas	1,613	68%	132	6%	642	27%	2,387	100%
Water supply, sewerage, etc.	739	43%	198	12%	772	45%	1,709	100%
Wholesale and retail trade	192,364	63%	48,451	16%	63,520	21%	304,335	100%
Transportation and storage	11,257	45%	3,384	13%	10,538	42%	25,179	100%
Accomm. & food service	6,752	47%	2,692	19%	4,881	34%	14,325	100%
Information and communication	3,074	39%	1,080	14%	3,716	47%	7,870	100%
Financial and insurance	6,835	58%	972	8%	3,879	33%	11,686	100%
Others	32,869	56%	7,853	13%	17,705	30%	58,427	100%
	311,196	58%	86,738	16%	140,292	26%	538,226	100%

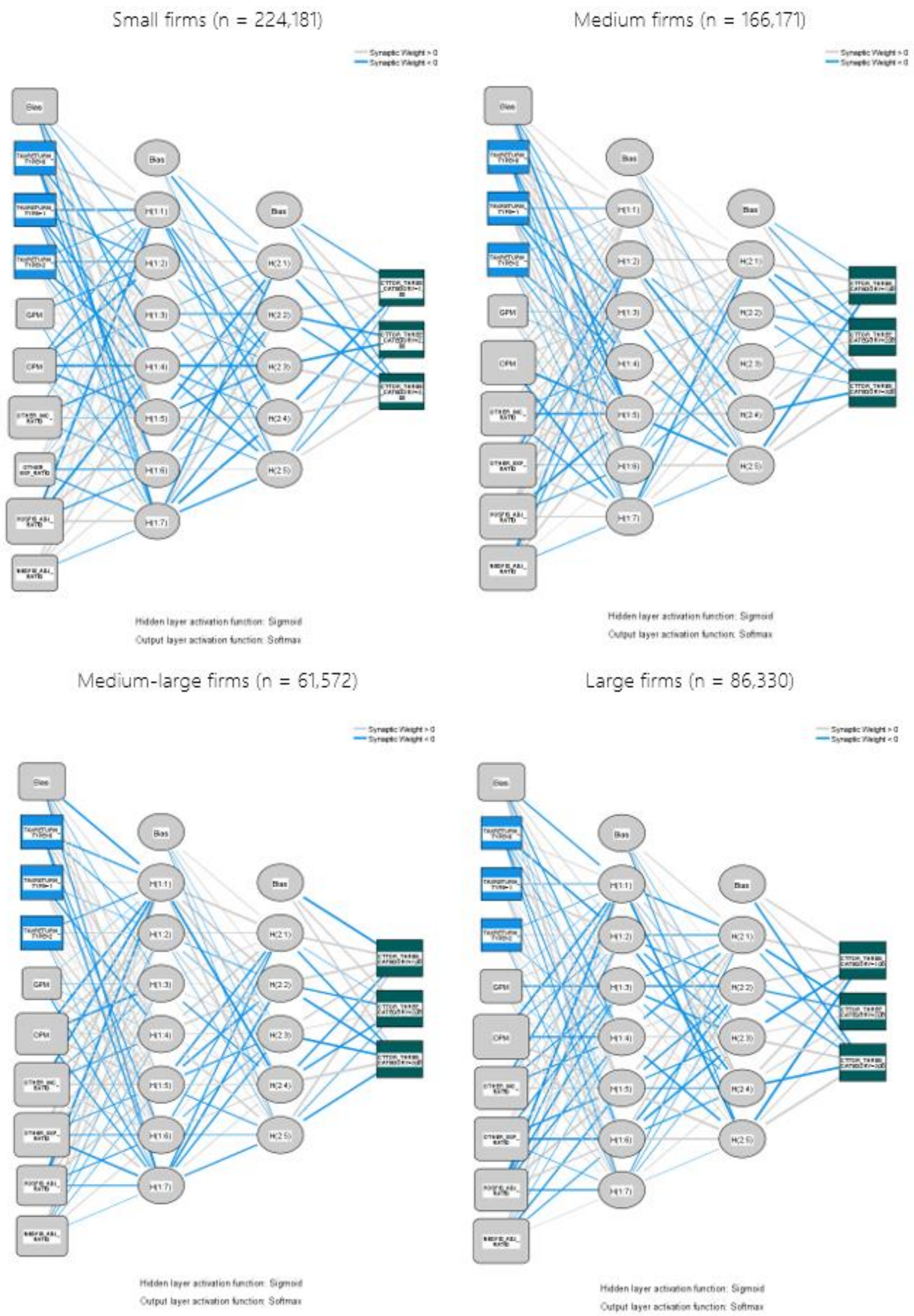
Appendix 2 CTTOR Categories by Business Sectors and by Size for 2014 to 2019



Note: A = Agriculture, B = Mining and quarrying, C = Manufacturing, D = Electricity and gas, E = Water supply, sewerage, etc., G = Wholesale and retail trade, H = Transportation and storage, I = Accommodation and food service, J = Information and communication, K = Financial and insurance, L = Others.

This graph displays the proportion of businesses in each industry by the three CTTOR categories: low (<0.59%), moderate(0.59% to 1.19%), and high (>1.19%). The graph is classified into four categories based on annual turnover size. For instance, the graphs show that firms in the wholesale and trade sector (G) have the lowest percentage of firms having high CTTOR—i.e., only 20 percent—across categories.

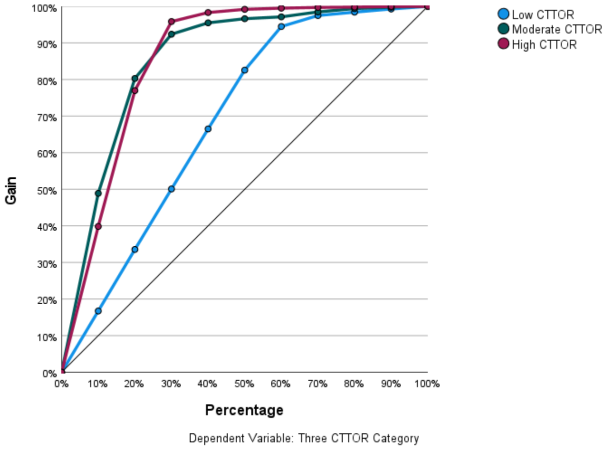
Appendix 3 Structure of Neural Networks for the Model Prediction



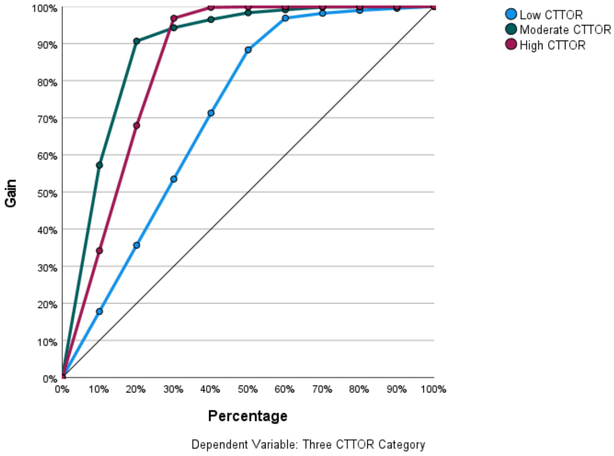
Note: These graphs show the ANN models in this study. There are seven independent variables used in the input layer. Each of the models has two hidden layers. There are seven and five nodes in the first and second hidden layers, respectively. The output layer has three units that represent the category of taxpaying behaviour. The activation of the hidden layer function in this analysis uses a Sigmoid, while the output layer uses Softmax.

Appendix 4 Cumulative Gain Chart

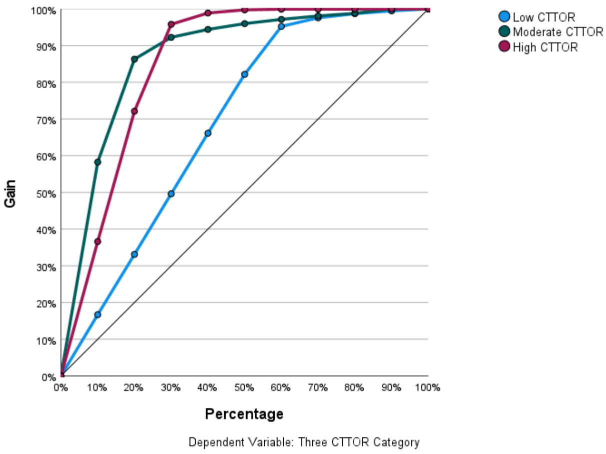
Small firms (n = 224,181)



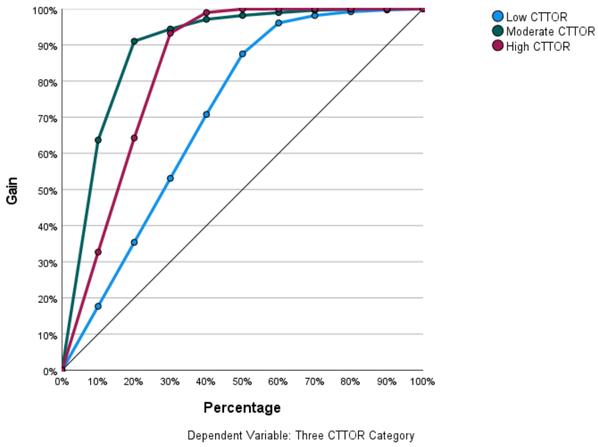
Medium firms (n = 166,171)



Medium-large firms (n = 61,572)



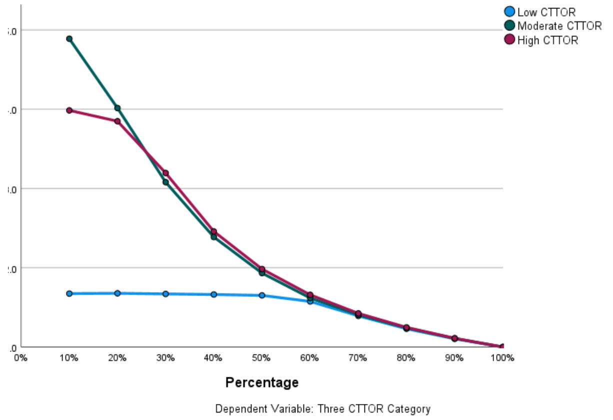
Large firms (n = 86,330)



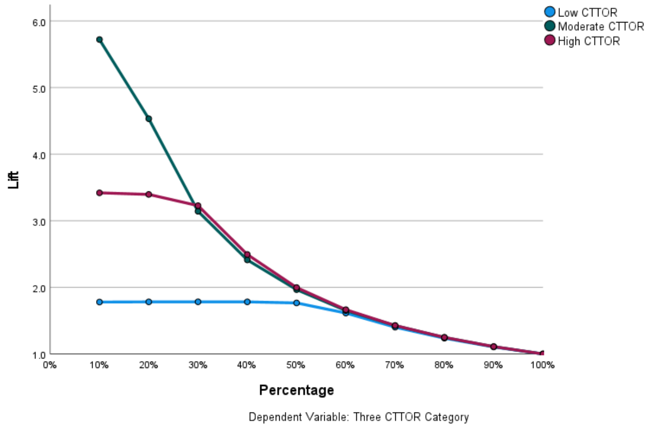
Note: The graph depicts the cumulative gains, which is the presentence of accurate classifications provided by the ANN model versus accurate classifications that may occur by chance (i.e., without using the model)). For example, the third point on the curve for the moderate CTTOR category in Group A is at (30%, 92%), meaning that if the network scores a dataset and sorts all of the cases by predicted pseudo-probability of failure, it would be expected that the top 30% would contain approximately 92% of all of the cases that fall into the moderate CTTOR category. Gain is a measure of the effectiveness of a classification model, calculated as the proportion of correct predictions made with the model versus the percentage of right predictions gained without a model (baseline). The more above the baseline a curve sits, the bigger the gain. A bigger overall gain signals better performance.

Appendix 5 Lift Chart

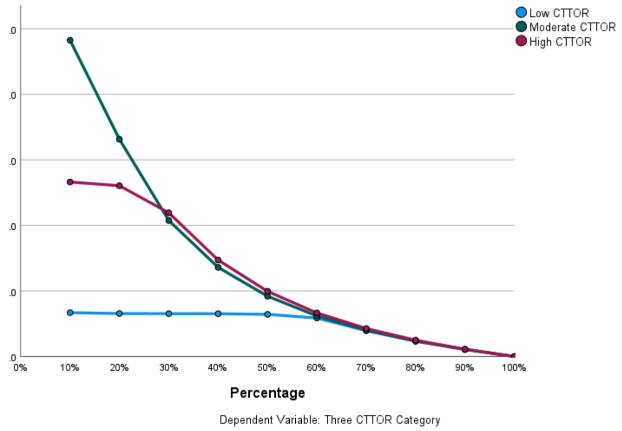
Small firms (n = 224,181)



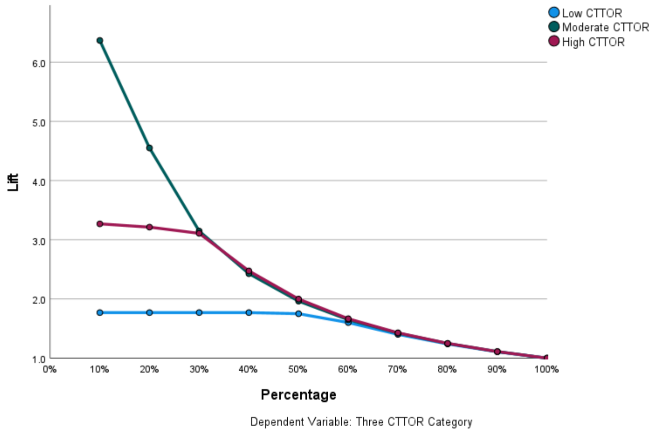
Medium firms (n = 166,171)



Medium-large firms (n = 61,572)



Large firms (n = 86,330)



Note: Lift charts, similar to gain charts, are visual tools for assessing the performance of classification models. However, in contrast to the confusion matrix that tests models for the entire data, the gain or lift chart analyses model performance in a data segment A lift chart leverages information to present a clear visual of the benefit of using a model in contrast to not using a model. For instance, in Group A, the factors from the gains diagram are utilised to compute the lift factor (i.e., the benefit): the lift at 92% for the moderate CTTOR category is 92%/30% = 3.1.