

Article

Application of Interpretable Artificial Intelligence for Sustainable Tax Management in the Manufacturing Industry

Ning Han ¹, Wen Xu ² , Qian Song ³, Kai Zhao ⁴ and Yao Xu ^{1,*} ¹ Business School, Qingdao University of Technology, Qingdao 266520, China; hanning@qut.edu.cn² School of Engineering, Cardiff University, Cardiff CF24 3AA, UK; XuW20@cardiff.ac.uk³ Business School, Loughborough University, Leicestershire LE11 3TU, UK; q.song@lboro.ac.uk⁴ School of Engineering and Architecture, University College Cork, T12 HW58 Cork, Ireland; kzhao@ucc.ie

* Correspondence: xuyao0607@126.com

Abstract: The long-term development of the manufacturing industry relies on sustainable tax management, which plays a key role in optimizing production costs. While artificial intelligence models have been applied to tax-related predictions, research on their application for predicting tax management levels is quite limited, with no studies focused on the manufacturing industry in China. To enhance digital innovation in corporate management, this study applies interpretable artificial intelligence models to predict the tax management level, which helps decision-makers maintain it within a sustainable range. The ratio of total tax expense to total profits (ETR) is used to represent the tax management level, which is predicted using decision trees, random forests, linear regression, support vector regression, and artificial neural networks with eight input features. Comparisons among the developed models indicate that the random forest model exhibits the best performance in terms of prediction accuracy and generalization capability. Additionally, the Shapley additive explanations (SHAP) technique is integrated with the developed model to enhance the interpretability of its predictions. The SHAP results reveal the importance of the input features and also highlight the dominance of certain features. The results show that the ETR from the previous year holds the greatest importance, being more than twice as significant as the second most important factor, whereas the effect of board size is negligible. Moreover, benefiting from the local interpretations using SHAP values, this approach aids managers in making rational tax management decisions.



Academic Editors: Alfonso Marino and Paolo Pariso

Received: 25 December 2024

Revised: 15 January 2025

Accepted: 27 January 2025

Published: 30 January 2025

Citation: Han, N.; Xu, W.; Song, Q.; Zhao, K.; Xu, Y. Application of Interpretable Artificial Intelligence for Sustainable Tax Management in the Manufacturing Industry. *Sustainability* **2025**, *17*, 1121. <https://doi.org/10.3390/su17031121>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: digital innovation; sustainable tax management; artificial intelligence; Shapley additive explanation; strategic management

1. Introduction

The implementation of legitimate and sustainable tax management effectively reduces costs, providing additional funds that can be allocated to operational activities, and thereby improving resource management [1,2]. The application of artificial intelligence in tax management enhances digital innovation in corporate management, contributing to the sustainable development of corporations. However, due to the significant variability in tax management strategies and the highly diverse behaviors of corporations, this real-world prediction problem is challenging [3].

For the prediction of complex real-world problems, artificial intelligence has been proven to be a powerful tool [4–11]. However, the application of artificial intelligence models for tax-related prediction is still in its early stages, with limited research conducted in this area. Xing et al. [12] developed a Levenberg–Marquardt neural network to detect

tax evasion behavior in the automobile sales industry in China, for which the license income, maintenance margin, agent insurance, as well as value-added tax burden are taken as indicators. Savić et al. [13] investigated the application of a hybrid unsupervised model, combining representational learning and clustering, to help tax authorities prevent and detect tax evasion cases, thereby improving their tax risk management. In the work of Rahimikia et al. [14], three artificial intelligence models, including logistic regression (LR), support vector machine (SVM), and multilayer perceptron (MLP) neural networks have been proposed for detecting corporate tax evasion in food and textile industries. The harmony search (HS) optimization approach was employed for determining the parameters of the artificial intelligence models, and the comparisons among the developed models demonstrate that the MLP neural network model achieves the best performance. In order to enhance the detection of transaction-based tax evasion in social e-commerce, Zhang et al. [15] developed a multimodal deep neural network based on 2041 labeled posts from Instagram. The input features of the developed model are the image attributes, comments, and tags.

Different from tax fraud detection, the prediction of the tax management level is more difficult. In the work of Lismont et al. [16], three prediction models, including LR, decision tree (DT), and random forest (RF), have been employed to predict the tax management level, considering a wider range of network features. Based on the results of a set of companies connected by shared board memberships, the RF model was proved to have the best prediction performance. Liu et al. [17] investigated the performance of the SVM model for predicting the tax gross, and comparisons were made between the SVM model and an artificial neural network (ANN). It was concluded that the SVM model exhibited superior prediction performance over the ANN model. Wahab and Bakar [18] evaluated the performance of K-nearest neighbor (KNN), classification and regression tree (CART), naive Bayes (NB), LR, SVM, ANN, and RF for predicting tax compliance of digital economic retailers. The comparisons indicated that the ensemble methods exhibited better classification performance than the single classifiers. Zheng and Li [19] developed a light gradient boosting machine (LightGBM) model to predict the potential tax arrears of corporations. The prediction performance of the developed model was enhanced by incorporating a knowledge graph, from which important features related to the relationships between the trust-breaking events, tax arrears events, and corporations were extracted and utilized. In the work of Guenther et al. [3], an extreme gradient boosting (XGBoost) model was developed to predict the effective tax rate (ETR), defined as the ratio of total tax expenses to total profits, for U.S. companies. Unlike classification predictions, regression predictions for social science problems are significantly more challenging. Despite the difficulties in regression predictions in the tax problem, the model developed in this study was able to achieve a coefficient of determination (R^2) greater than 0.5. The obtained results indicate that the artificial intelligence model outperforms humans in this prediction task.

Although the application of artificial intelligence provides a strong tool for the complex prediction problem, the black-box prediction process makes it unattractive for decision-makers who require transparency and interpretability in the prediction results to facilitate digital management. To overcome this drawback, interpretable techniques have been developed for artificial intelligence models, such as interpretable mimic learning (IML) [20] and Shapley additive explanations (SHAP) [21]. The SHAP technique serves as a useful tool to assist humans in understanding and evaluating the outcomes generated by artificial intelligence models [22–24]. The effectiveness of the SHAP technique has been demonstrated in various fields, including civil engineering [25–27], wastewater treatment [28], medicine [29], commerce [30], and cybersecurity [31].

Although several studies have been conducted to predict tax-related issues using artificial intelligence models, research on the prediction of tax management levels is still limited, with no research focused on the manufacturing industry in China, which has experienced remarkable developments in recent decades. In the present study, five artificial intelligence models are applied to predict the tax management levels of manufacturing corporations in China, aiming to guide them in maintaining their tax management level within a sustainable range for the following year. To enable relevant parties to better understand the underlying factors influencing the prediction results and also make the decision-making process more accountable, the SHAP technique is incorporated into the prediction process. In the next section, the database used for developing the models is introduced. Section 3 provides a detailed description of the development of the models. The results and discussions are provided in Section 4, followed by a brief conclusion in Section 5.

2. Data Description

The database contains 1436 data samples, which are collected from the Wind database. These samples are specifically from manufacturing corporations. Figure 1 shows the distribution of data samples. According to tax law, the fundamental corporate characteristics that are commonly recognized as most directly related to tax management behavior are selected as inputs for the models. These include the ratio of foreign turnover to sales (FT_Int), the ratio of fixed assets to total assets (PPE), the ratio of research and development expenditure to sales (R&D_Int), the logarithm of total assets (TA), the ratio of pre-tax profits to total assets (ROA), the ratio of intangible assets to total assets (IA_Int), and ETR. Other factors would need to influence tax management behavior through these key features. In addition, Minnick and Noga [2] pointed out that the board of directors also influences corporate tax management decisions. Given that board size (BSIZE) is a fundamental feature of the board of directors and data on this factor is available in the Wind database, it is also taken as an input for the models.

Based on these parameters of the corporations, this study aims to predict the ratio of total tax expense to total profits in the following year (ETR_f), which is used to represent the tax management level of the corporations in this study. The correlations between the outputs and inputs have been demonstrated in several studies. The costs associated with foreign tax management may discourage companies from actively managing taxes, resulting in a correlation between FT_Int and ETR_f [32]. Companies can depreciate eligible fixed assets over a specific period, which reduces taxable income and enhances tax management, indicating that PPE is also related to ETR_f [33]. According to tax law and related policies, eligible companies can claim additional tax deductions for their research and development expenditures, leading to a correlation between R&D_Int and ETR_f [34]. In addition, large companies attract more attention and are subject to stricter regulatory scrutiny, making TA, which represents firm size, correlated with ETR_f [35]. Corporate profitability also influences tax management decisions, as companies with higher profitability have greater incentives and the ability to manage taxes, revealing a strong correlation between ROA and ETR_f [36]. In addition, Markle and Shackelford [37] pointed out that intangible assets, represented by IA_Int, provide companies with tax management opportunities and are therefore associated with ETR_f. Since tax policies are generally long-term, current tax policies can influence those in subsequent periods, making the current ETR correlated with ETR_f [38]. Furthermore, Minnick and Noga [2] suggested that smaller boards are more likely to influence managers to engage in tax management, implying that BSIZE is also expected to be correlated with ETR_f.

It should be noted that the statutory corporate tax rate in China is 25% [39,40]. If the value of ETR_f exceeds 0.25, it implies that the corporation does not effectively manage its taxes. However, if the value of ETR_f is below 0.1, the corporation is likely to be accused of engaging in excessive tax management practices, which could be detrimental to its sustainable development. Therefore, the range of tax management levels between 0.1 and 0.25 is considered to represent sustainable tax management in this study. Before the training of the models, all the parameters were normalized between 0 and 1 to enhance the accuracy of the predictions. The entire database used in this study is included in the Supplementary Materials.

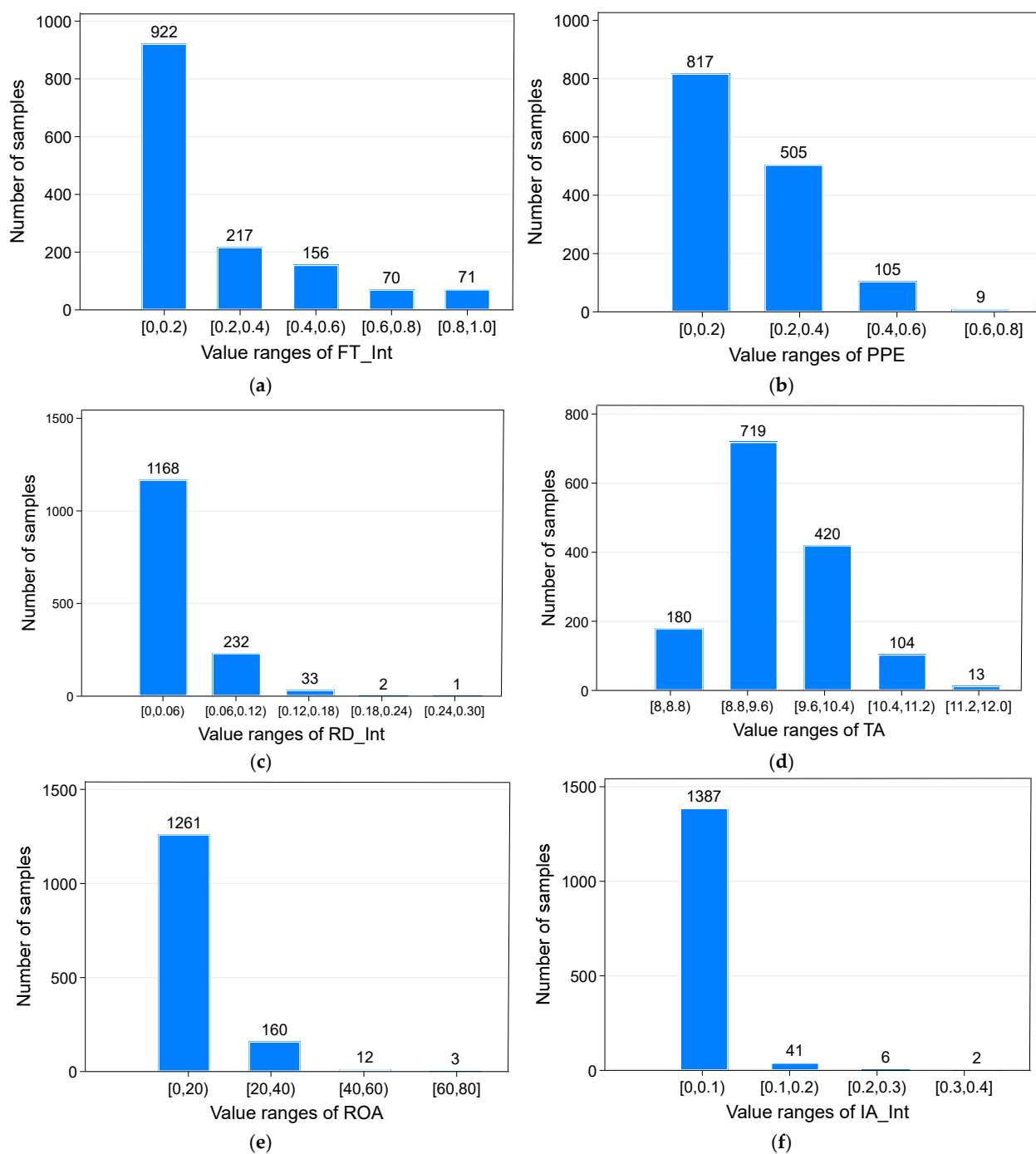


Figure 1. Cont.

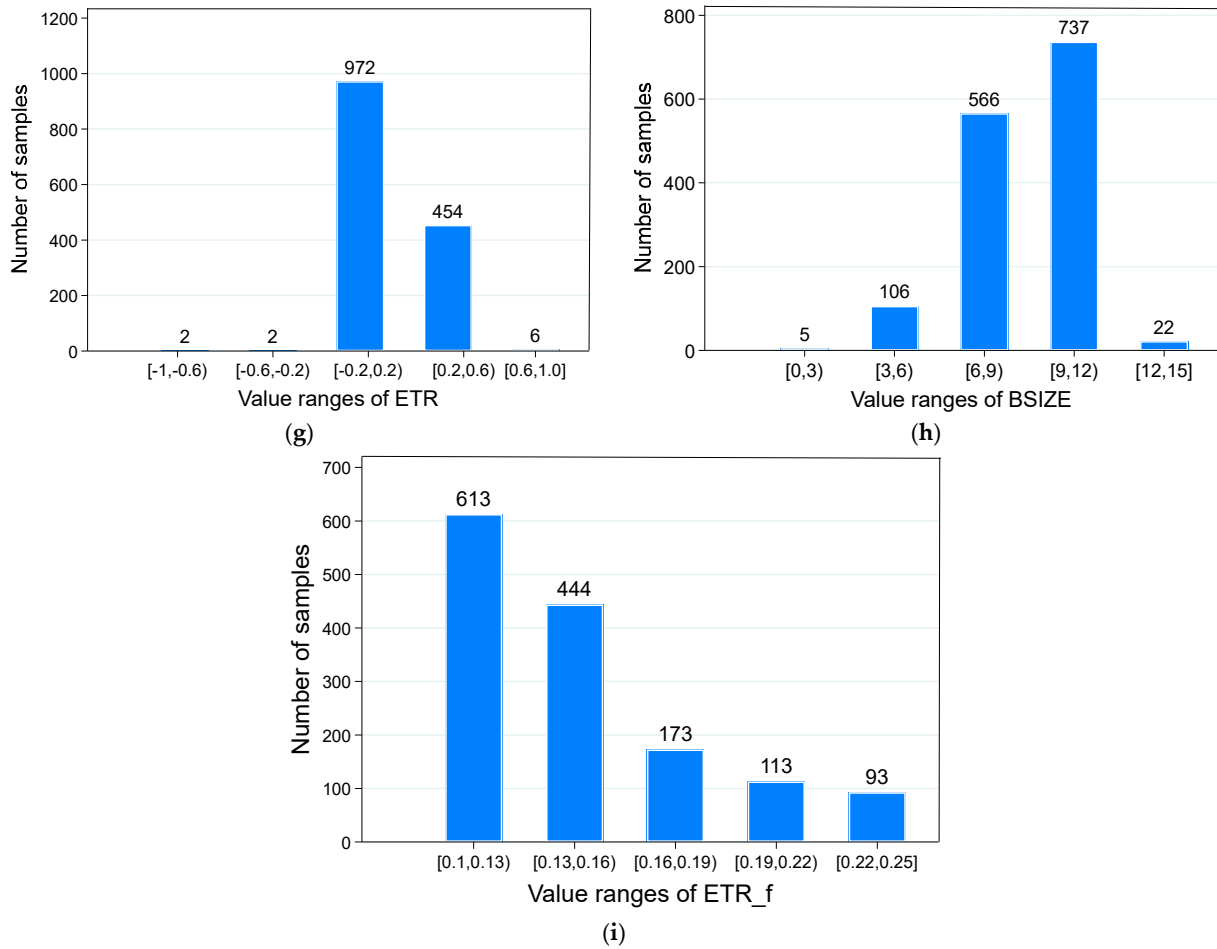


Figure 1. Distribution of data samples. (a) Input feature FT_Int, (b) Input feature PPE, (c) Input feature RD_Int, (d) Input feature TA, (e) Input feature ROA, (f) Input feature IA_Int, (g) Input feature ETR, (h) Input feature BSIZE, (i) Output feature ETR_f.

3. Model Development

Five artificial intelligence models are developed in this section for this prediction problem, including artificial neural networks, decision trees, random forests, support vector regression, and linear regression. This is the first study to investigate the application of these commonly used models to predict ETR values.

3.1. Artificial Neural Network Model

The ANN is recognized as one of the most powerful methods for the prediction of complex relationships [41,42]. As the name implies, the ANN is derived from the structure of the biological neuron system. Typically, an ANN model is composed of one input layer, one output layer, and several intermediate layers in between. The layers consist of neurons, and the neural network is constructed by connecting the neurons in the adjacent layers. The structure of an ANN model is shown in Figure 2.

During the training process, the inputs are fed forward from the input layers to the output layer. For the neurons in the intermediate layers, the received inputs from the preceding layer are weighted and summed, and then activated by an activation function, as shown below:

$$z = f\left(\sum_{i=1}^n w_i p_i + b_i\right) \quad (1)$$

where p_i is the input provided by the i th neuron in the preceding layer, w_i and b_i are the corresponding weight and bias, n is the number of neurons in the preceding layer, and f represents a activation function. In the present study, the Python-based open-source package, Keras [43], is applied to establishing the ANN model. Several activation functions are provided by this package, such as the sigmoid function (Equation (2)), tanh function (Equation (3)), and ReLU function (Equation (4)), as shown below:

$$\text{sigmoid} = \frac{1}{1 + e^{-z}} \quad (2)$$

$$\text{tan h} = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (3)$$

$$\text{ReLU} = \max(0, z) \quad (4)$$

Once the first prediction result is obtained, the error between the obtained result and the target result is calculated using a loss function. To minimize the loss function, a back propagation strategy is then conducted to optimize the weights and biases. The training of an ANN model requires undergoing this iterative optimization process. In addition, in this study, the mean squared error (MSE) is used as the loss function, as shown below:

$$L = \frac{1}{m} \sum_{i=1}^m (t_i - p_i)^2 \quad (5)$$

where m is the number of data samples, p_i is the prediction result, and t_i is the target value from the dataset. Apart from the MSE, R^2 is also used to measure the performance of the developed model; the expression of R^2 is shown as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^m (t_i - p_i)^2}{\sum_{i=1}^m (t_i - \bar{t})^2} \quad (6)$$

where \bar{t} is the average of the target values.

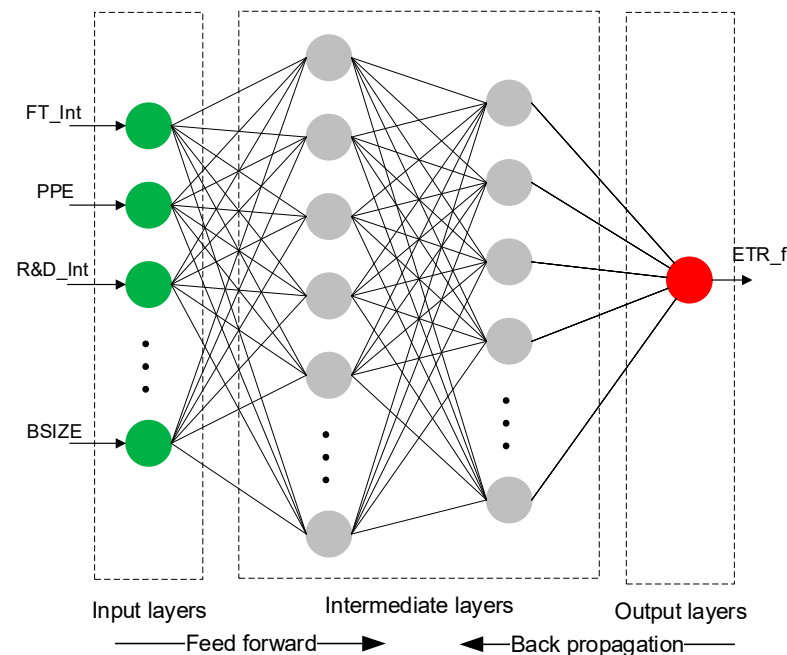


Figure 2. The structure of the artificial neural network.

3.2. Decision Tree Model

As shown in Figure 3, the DT model is derived from a tree structure [44]. The prediction process starts from the root node where an initial condition is applied to split the data. The split data are assigned to the lower nodes along the branches, where subsequent conditions are applied to further split the data. This process repeats until the data reaches the leaf nodes, which represent the final predicted results.

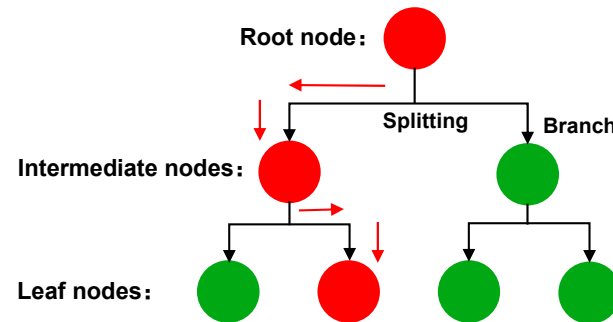


Figure 3. The tree structure of a decision tree model.

To ensure accurate predictions in regression problems, the feature and split point at each node must be carefully decided. For a DT regression model, each node typically splits the data into two subsets, with the aim of decreasing prediction errors at each step. The mean squared error criterion is commonly used to evaluate and select the suitable features and division points, as follows:

$$\text{MSE} = \frac{1}{n_L} \sum_{i=1}^{n_L} (y_i - \bar{y}_L)^2 + \frac{1}{n_R} \sum_{i=1}^{n_R} (y_i - \bar{y}_R)^2 \quad (7)$$

where \bar{y}_L and \bar{y}_R represent the average targets of the left split data and right split data, respectively, and n_L and n_R denote the number of data samples in the left split data and right split data, respectively. The process of generating new nodes continues until the maximum tree depth is reached.

3.3. Random Forest Model

The RF model is constructed by integrating several DT models, as shown in Figure 4, each of which is developed based on a randomly selected subset of the training dataset and a random subset of features [25]. These decision trees operate independently, producing individual predictions, and the mean of the predictions of all constituent tree models is taken as the final result of the RF model. The output of the RF model is obtained as shown below:

$$\text{Output} = \frac{1}{t} \sum_{i=1}^t f_i(\mathbf{x}) \quad (8)$$

where f_t represents the prediction result of the t th tree, and t denotes the number of decision trees in the RF model.

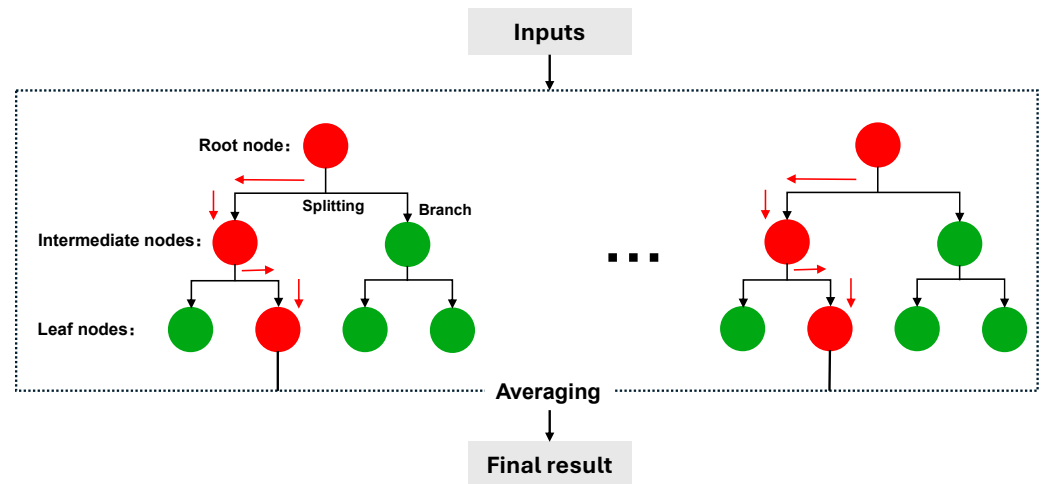


Figure 4. Illustration of the random forest model.

3.4. Support Vector Regression Model

As for the SVR model, the inputs are mapped into a higher dimensional space to predict highly non-linear relationships [45]. During the training process, the distance between the hyperplane and data points is gradually minimized by optimizing the risk function, which is expressed as follows:

$$\min : R = \frac{1}{2} \|w\|^2 + C \sum_i^m L_\varepsilon(y_i, h(x_i)) \quad (9)$$

where $h(x)$ represents the hyperplane, w denotes a weight parameter, C represents the regularization parameter, and L_ε is an insensitive loss function.

As shown in Figure 5, by introducing two slack variables, ζ_i and ζ_i^* , to denote the spacing from the outlying points to margin boundaries of the hyperplane, the function R can be converted into the following:

$$\min : R = \frac{1}{2} \|w\|^2 + C \sum_i^m L_\varepsilon(\zeta_i + \zeta_i^*) \quad (10)$$

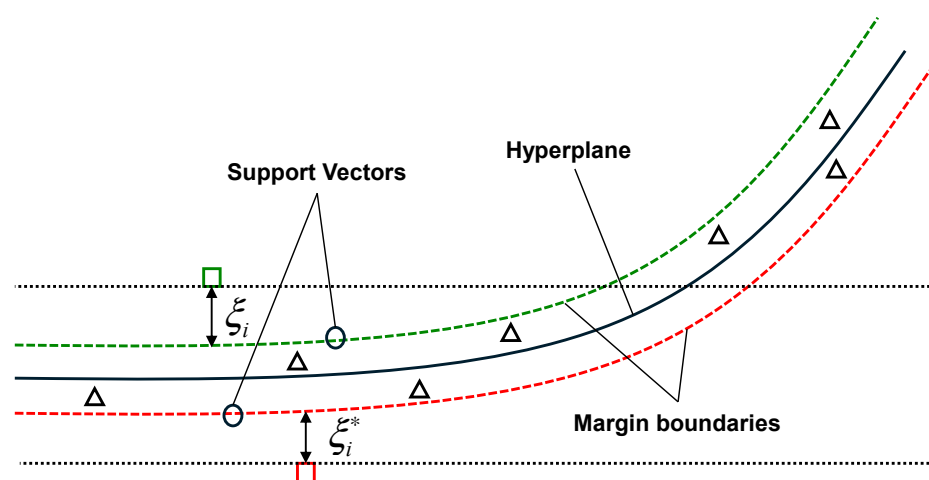


Figure 5. Illustration of the support vector regression model.

The minimization optimization problem can be easily solved using the Lagrange method [46].

3.5. Linear Regression Model

Although several advanced artificial intelligence models are proposed to solve this prediction problem, the linear regression (LR) model, which is simpler and more interpretable, is also employed in this study to assess whether the relationship between the input features and the output can be adequately captured by a linear model. Additionally, the results provided by the LR model are set as a baseline for comparison with the other models. The LR model can be expressed as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \quad (11)$$

where β_i is the regression coefficient, and ε is a residual error term.

3.6. Interpretation for Black-Box Prediction

In this paper, the SHAP method developed by Lundberg and Lee [21] is employed to interpret the black-box prediction of the artificial intelligence model. As an additive feature attribution method, SHAP is proposed based on game theory. The predictions can be explained for the overall dataset at a global level and also for every sample at a local level. The contribution of each feature to the prediction can be obtained and assigned with a SHAP value, based on which the decision-making process behind the prediction can be shown. The explanation model is defined as a linear summation of a based value and SHAP values as shown below:

$$E(\mathbf{x}') = s_0 + \sum_{i=1}^N s_i x'_i \quad (12)$$

where s_i represents the attribution of the i th feature (i.e., SHAP value), s_0 is a constant value when all input features are null, N represents the number of features, and \mathbf{x}' represents a vector of simplified inputs.

The SHAP value of the i th feature ϕ_i is defined as follows:

$$\phi_i(f, \mathbf{x}) = \sum_{\mathbf{z}' \subseteq \mathbf{x}'} \frac{|\mathbf{z}'|!(N - |\mathbf{z}'| - 1)!}{N!} [f(\mathbf{z}' \cup \{i\}) - f(\mathbf{z}')] \quad (13)$$

where $f(\mathbf{z}' \cup \{i\})$ and $f(\mathbf{z}')$ represent predictions with and without adding the i th feature, respectively, \mathbf{z}' represents a portion of \mathbf{x}' , and $|\mathbf{z}'|$ denotes the nonzero elements in \mathbf{z}' .

4. Results and Discussion

4.1. Predictions of the Developed Models

The prediction results of the developed models are shown in Figure 6. As expected, the LR model exhibits the poorest prediction performance with the lowest value of R^2 and the highest value of MSE, implying high non-linear relationships between the inputs and outputs. Compared to the SVR model, the DT model obtains better prediction results, as the values of R^2 and MSE of the SVR model are 0.39 and 0.035, respectively, while those for the DT model are 0.477 and 0.03, respectively. To make a reasonable comparison between the models, the grid search process is conducted to determine the hyperparameters of the models. The regularization parameter and the kernel coefficient of the SVR model are selected as 1 and 10, respectively. For the DT model, a maximum tree depth of 10 is selected, with the minimum number of samples needed at a leaf node set to 15. In addition, the minimum sample size needed to divide an internal node is assigned to three.

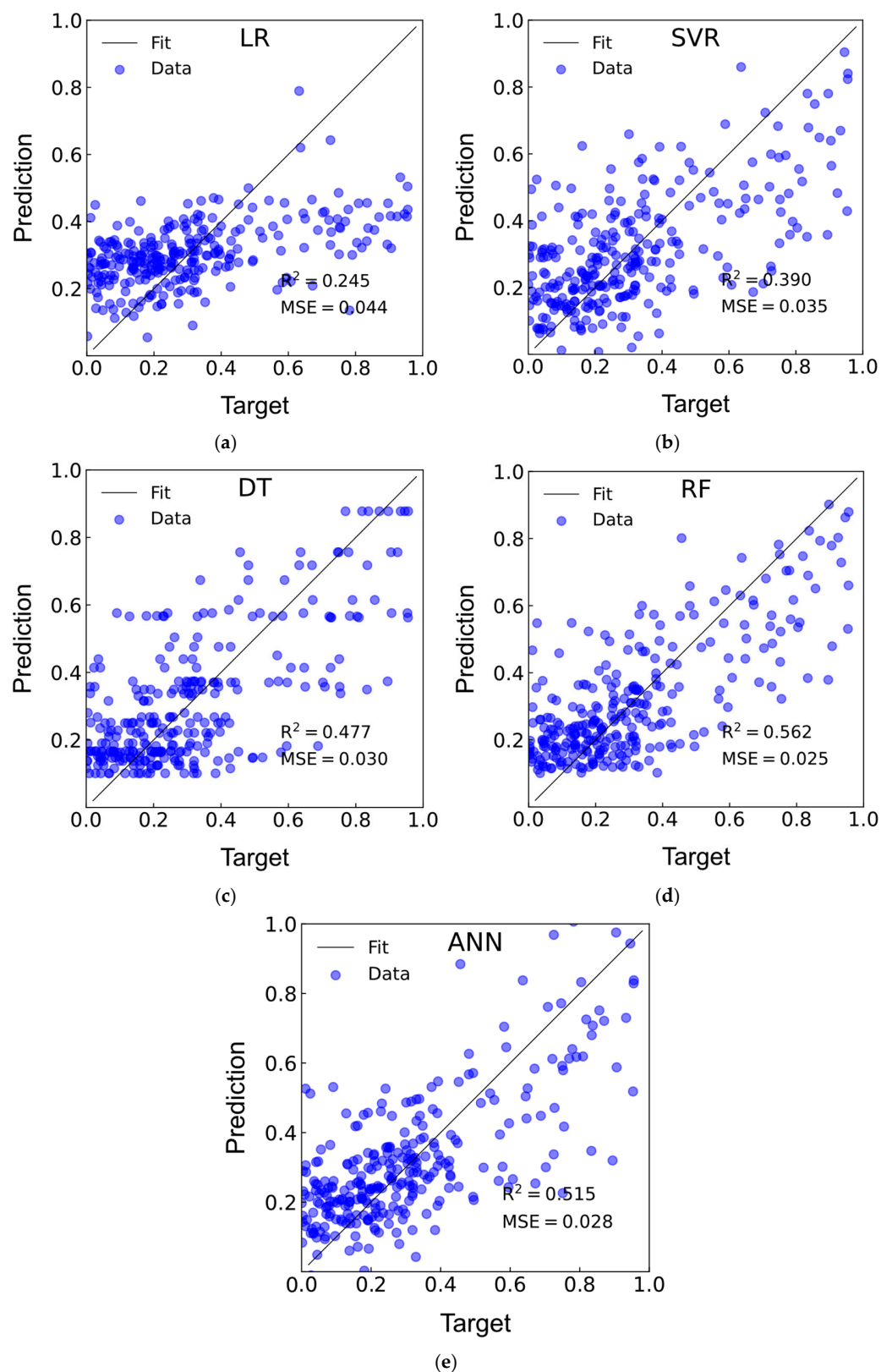


Figure 6. Predictions of the developed artificial intelligence models. (a) LR model, (b) SVR model, (c) DT model, (d) RF model, and (e) ANN model.

As can be seen from Figure 6, the RF model has the best prediction capability with the highest R^2 of 0.562 and the lowest MSE of 0.025. The prediction performance of the RF model is similar to that of the model developed in [3]. Compared to the DT model, the

prediction performance of the RF model is significantly improved due to the ensemble nature of multiple trees. According to the grid search results, the RF model consisting of 50 decision trees with a maximum depth of 20 is selected. The RF model is configured with a minimum of eight samples per leaf and ten samples needed to split an internal node. The ANN model shows only a slightly lower performance than the RF model, with the values of R^2 and MSE equal to 0.515 and 0.028, respectively. The ANN model has three intermediate layers; the number of neurons in each layer is 30, 50, and 70, respectively. The tanh function is used as the activation function, and the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm is selected as the optimization approach.

In this study, a 5-fold cross-validation test is performed to assess the generalization capability of the developed models in predicting tax management levels. Figure 7 shows the standard deviation of R^2 for the developed models across five trainings. As can be seen, the RF model achieves the lowest variability in R^2 with a standard deviation of 0.02, demonstrating that the RF model has the best generalization capability in this prediction task compared to the other models. The SVR and ANN models exhibit the same variability in R^2 , with a standard deviation of 0.05. The DT model has relatively poor generalization capability compared to the other models, and the comparisons between the RF and DT models indicate that the ensemble nature of the RF model contributes to more consistent prediction performance across various databases for predicting tax management levels.

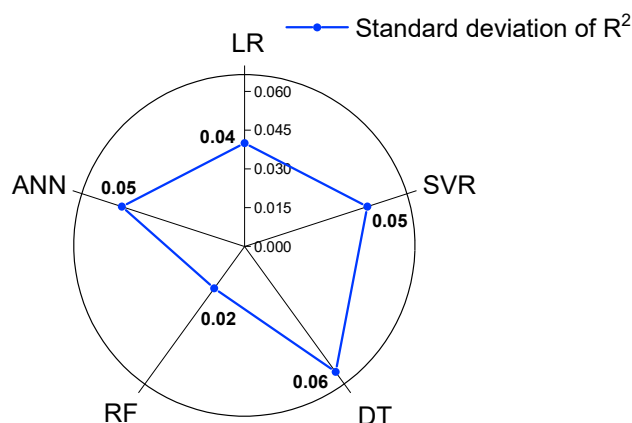


Figure 7. Standard deviation of R^2 for the 5-fold tests.

4.2. Interpretations for the Predictions

4.2.1. Interpretations at the Global Level

The SHAP values for the entire dataset are shown in Figure 8a, based on which the global interpretations of the predictions can be determined. Each data point is denoted by a dot, illustrating the relationship between the feature values and their corresponding SHAP values. Lower feature values are depicted in blue, whereas higher feature values are in red. As can be seen from Figure 8a, the SHAP value is plotted along the x-axis, where a positive SHAP value indicates a positive contribution to the ETR_f , whereas a negative SHAP value reflects a detrimental effect on the ETR_f . It can be observed that a higher ETR in the previous year generally leads to a higher ETR in the following year, with greater positive contributions from higher ETR values in the previous year and relatively smaller negative contributions from lower ETR values. The obtained results indicate a stronger positive effect of higher input ETR and a weaker negative effect of lower input ETR.

Although higher values of TA, FT_int, and IA_int generally lead to higher ETR values in the following year, these effects are less pronounced than those of the ETR value in the previous year. In contrast, higher values of R&D_int and ROA basically have negative SHAP values, resulting in lower ETR values in the following year. For both these input features, the positive effects from the lower feature values are more pronounced than

the negative effects from the higher feature values. In addition, the effect of PPE on the ETR values in the following year remains uncertain, and the effect is relatively small. Furthermore, the BSIZE has a negligible effect on the ETR values in the following year.

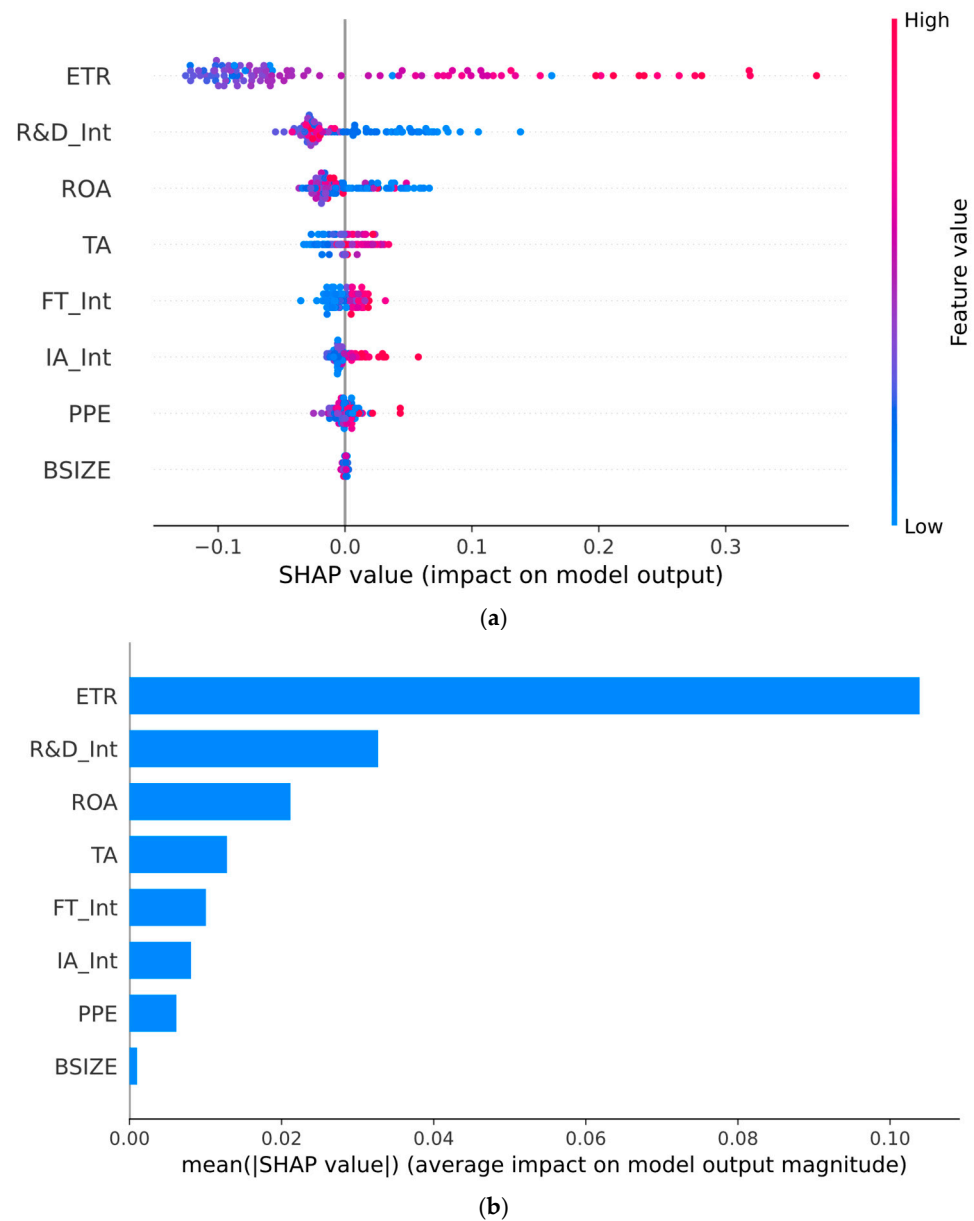


Figure 8. Global interpretation for the entire database. (a) A summary plot of SHAP values. (b) Average absolute SHAP values.

Figure 8b shows the mean absolute SHAP values that reflect the importance of the input features. It can be seen that the ETR in the previous year has the greatest importance for this prediction problem, which is even more than two times greater than the following R&D_int. The high importance of ETR is consistent with the results obtained in [3], despite the use of different databases, input features, and models in this study, further demonstrating its importance. For the following input features, the differences in importance are relatively small. As can be seen, the importance of BSIZE is the smallest, and the PPE is the second least important, which aligns with the results illustrated in Figure 8a.

4.2.2. Interpretations at the Local Level

To further illustrate the details of the black-box predictions in artificial intelligence models, specific interpretations for individual cases are presented in this section. Figure 9 shows the local interpretations of six randomly selected cases, where $E[f(x)]$ on the x-axis represents the baseline value for the local interpretation, and even contributions very close to zero are depicted with slight variations in arrow lengths to reflect their magnitude and direction. For the case shown in Figure 9a, the higher value of ETR in the previous year leads to an increase of 0.1 in ETR in the following year, from the base value of 0.298. The lower value of R&D_int leads to a slight increase in ETR in the following year, with a value of 0.01. The lower values of PPE, ROA, and IA_Int reduce the ETR in the following year. In addition, in this example, the contributions of the TA, FT_int, and BSIZE on the prediction are negligible. For the case presented in Figure 9b, the relatively higher R&D_int leads to a decrease of 0.03 in ETR in the following year.

Different from the two examples discussed above, the input features R&D_int and TA in the example shown in Figure 9c have relatively larger contributions to the prediction, as the R&D_int and TA lead to increases of 0.08 and 0.03 in ETR in the following year, respectively. As for the example shown in Figure 9d, the R&D_int has a positive contribution to the prediction with a value of 0.06, while the contributions of ROA, FT_int, PPE, and IA_Int are the same with a value of 0.01. As shown in Figure 9e, the contributions of PPE and FT_int are 0.03 and 0.02, respectively. The R&D_int and ROA have slight negative contributions with values of -0.01 in this case. As for the case presented in Figure 9f, the R&D_int, PPE, and ROA have positive contributions to the ETR in the following year, with the values of -0.03 , -0.02 , and -0.02 , respectively. The FT_int has a positive contribution of 0.01, whereas the IA_int has a negative contribution of -0.01 . Based on the local interpretations, the prediction process for each individual case is transformed into an open-box process, enabling the understanding and adjustment of the tax management level.

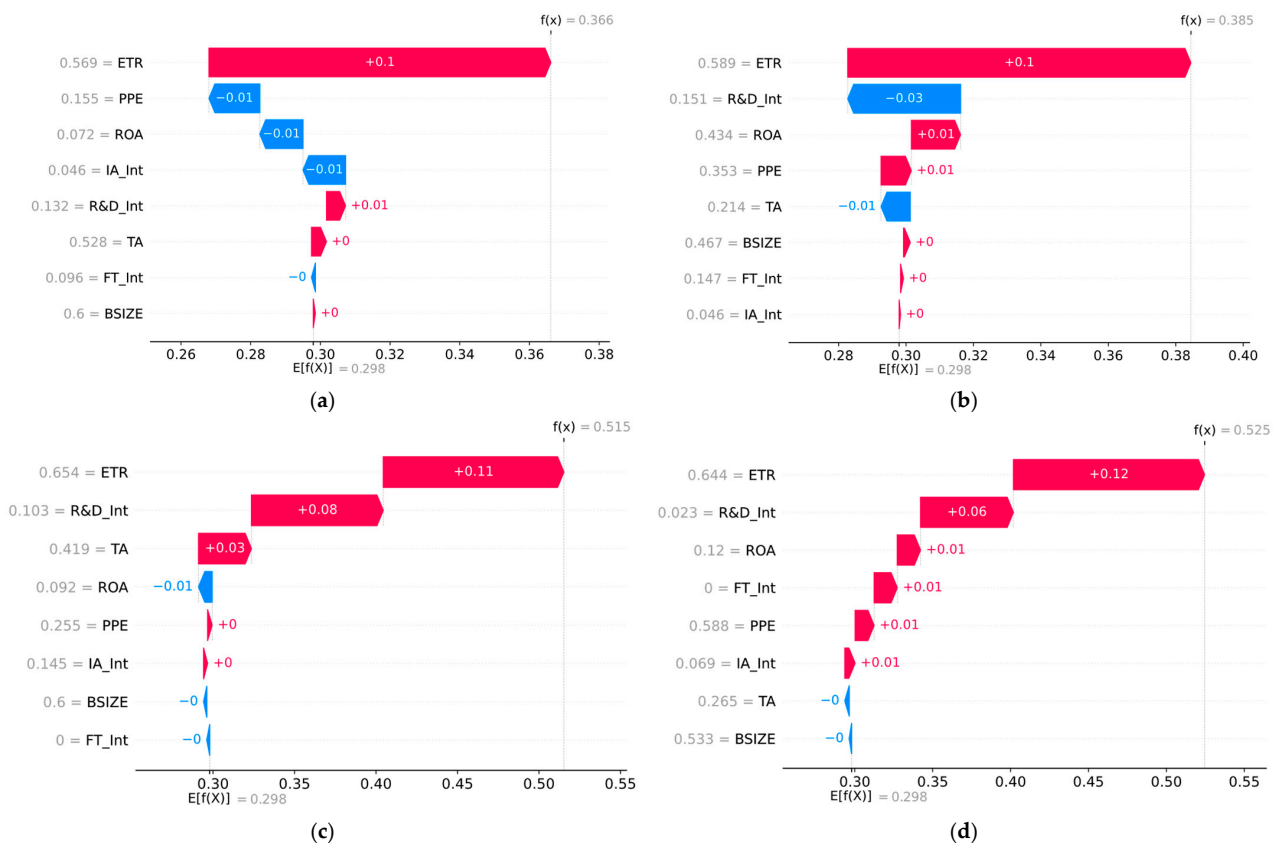


Figure 9. Cont.

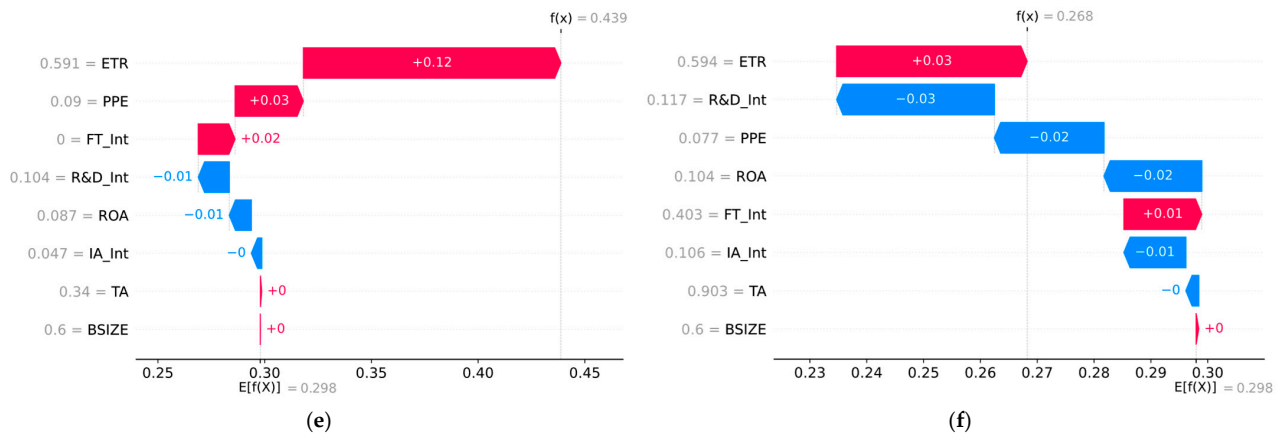


Figure 9. Local interpretation for individual examples. (a) SHAP interpretation for example 1, (b) SHAP interpretation for example 2, (c) SHAP interpretation for example 3, (d) SHAP interpretation for example 4, (e) SHAP interpretation for example 5, and (f) SHAP interpretation for example 6.

5. Conclusions

In this study, interpretable artificial intelligence models are developed to predict tax management levels, which provides an effective approach for corporations to dynamically adjust their tax management decisions to maintain the tax management levels within a sustainable range. In order to predict the ETR in the following year, eight input features are utilized in this study, and all the data samples are collected from the Wind database. A total of 1436 data samples are used for the training and testing of the models. The results show that the RF model attains the lowest value of MSE and the highest value of R^2 . The generalization capability of the developed models is assessed using a 5-fold cross-validation test, and the RF model exhibits more consistent prediction performance compared to the other models. Despite the stronger capability of artificial intelligence models for solving highly non-linear prediction problems compared to traditional models, their prediction process can be interpreted using interpretation approaches, assisting humans in understanding the prediction process. In this study, the SHAP approach is incorporated with the artificial intelligence model to make the prediction process more transparent and interpretable. The global SHAP results indicate that the input feature ETR in the previous year has the greatest effect on prediction, with a higher ETR in the previous year generally leading to a higher ETR in the following year. This effect is more than twice as significant as that of the input feature R&D int. On the other hand, the effect of the input feature BSIZE on the prediction is negligible. In this study, the practical value and advantage of the RF model are demonstrated through comparisons among the five models, as it exhibits better prediction accuracy and generalization capability. The obtained values of R^2 are not as high as those typically observed in engineering and natural science problems, where objective laws govern the underlying process, indicating the inherent difficulty in predicting tax management levels. However, benefiting from the local interpretations, the practical value of the predictions is enhanced, as decision-makers can still be supported in making informed decisions based on their judgment of the local interpretation. Nevertheless, the present work employs SHAP to achieve model interpretability, but it does not address other dimensions of interpretability, such as model fairness, rationality, and sensitivity to noise. Additionally, the consistency of SHAP values across different algorithms and datasets could be affected by variations in training data distributions or feature sets. Hence, future research is recommended to explore these aspects to enhance prediction interpretability. This study primarily uses fundamental corporate characteristics that are most directly related to tax management levels as inputs.

It is recommended to consider other input features, such as macroeconomic factors and industry-specific tax regulations, in future studies. Moreover, the use of more advanced optimization techniques for artificial intelligence models is also suggested for future work.

Supplementary Materials: The following supporting information can be downloaded at: <https://zenodo.org/records/14554569> (accessed on 25 December 2024), Table S1: Database for training and testing the developed models.

Author Contributions: Conceptualization, N.H.; methodology, W.X. and K.Z.; software, K.Z.; validation, N.H. and K.Z.; formal analysis, N.H., Y.X. and Q.S.; investigation, N.H., Y.X. and Q.S.; resources, N.H. and K.Z.; data curation, K.Z. and Q.S.; writing—original draft preparation, N.H.; writing—review and editing, Q.S., K.Z., W.X. and Y.X.; visualization, W.X., Y.X. and K.Z.; funding acquisition, N.H. and Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Shandong Province, China, grant number ZR2023QG139 and ZR2023QG037, the National Natural Science Foundation of China, grant number 72303124, and the Youth Innovation and Technology Team Project for Higher Education Institutions of Shandong Province, China, grant number 2024KJB003.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Materials. Further inquiries can be directed at the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Shin, Y.; Park, J.-M. The Effect of a Company's Sustainable Competitive Advantage on Their Tax Avoidance Strategy—Focusing on Market Competition in Korea. *Sustainability* **2023**, *15*, 7810. [CrossRef]
2. Minnick, K.; Noga, T. Do corporate governance characteristics influence tax management? *J. Corp. Financ.* **2010**, *16*, 703–718. [CrossRef]
3. Guenther, D.A.; Peterson, K.; Searcy, J.; Williams, B.M. How Useful Are Tax Disclosures in Predicting Effective Tax Rates? A Machine Learning Approach. *Account. Rev.* **2023**, *98*, 297–322. [CrossRef]
4. Salamian, F.; Paksaz, A.; Khalil Loo, B.; Mousapour Mamoudan, M.; Aghsami, M.; Aghsami, A. Supply Chains Problem During Crises: A Data-Driven Approach. *Modelling* **2024**, *5*, 2001–2039. [CrossRef]
5. Pereira, R.M.S.; Oliveira, F.; Romanyshyn, N.; Estevez, I.; Borges, J.; Clain, S.; Vasilevskiy, M.I. Classification of Real-World Objects Using Supervised ML-Assisted Polarimetry: Cost/Benefit Analysis. *Appl. Sci.* **2024**, *14*, 11059. [CrossRef]
6. Marino, A.; Pariso, P.; Picariello, M. Organizational and Energy Efficiency Analysis of Italian Hospitals and Identification of Improving AI Solutions. *Int. J. Energy Econ. Policy* **2024**, *14*, 628–640. [CrossRef]
7. Imran, M.; Qureshi, S.H.; Qureshi, A.H.; Almusharraf, N. Classification of English Words into Grammatical Notations Using Deep Learning Technique. *Information* **2024**, *15*, 801. [CrossRef]
8. Qin, J.; Hu, F.; Liu, Y.; Witherell, P.; Wang, C.C.L.; Rosen, D.W.; Simpson, T.W.; Lu, Y.; Tang, Q. Research and application of machine learning for additive manufacturing. *Addit. Manuf.* **2022**, *52*, 102691. [CrossRef]
9. Marino, A.; Pariso, P.; Picariello, M. Transition towards the artificial intelligence via re-engineering of digital platforms: Comparing European Member States. *Entrep. Sustain. Issues* **2022**, *9*, 350.
10. Alarfaj, F.K.; Malik, I.; Khan, H.U.; Almusallam, N.; Ramzan, M.; Ahmed, M. Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. *IEEE Access* **2022**, *10*, 39700–39715. [CrossRef]
11. Qin, J.; Liu, Y.; Grosvenor, R. Multi-source data analytics for AM energy consumption prediction. *Adv. Eng. Inform.* **2018**, *38*, 840–850. [CrossRef]
12. Xiangyu, X.; Youlin, Y.; Qicheng, X. Intelligent identification of corporate tax evasion based on LM neural network. In Proceedings of the 37th Chinese Control Conference (CCC), Wuhan, China, 25–27 July 2018; pp. 4507–4511.
13. Savić, M.; Atanasijević, J.; Jakovetić, D.; Krejić, N. Tax evasion risk management using a Hybrid Unsupervised Outlier Detection method. *Expert Syst. Appl.* **2022**, *193*, 116409. [CrossRef]
14. Rahimikia, E.; Mohammadi, S.; Rahmani, T.; Ghazanfari, M. Detecting corporate tax evasion using a hybrid intelligent system: A case study of Iran. *Int. J. Account. Inf. Syst.* **2017**, *25*, 1–17. [CrossRef]

15. Zhang, L.; Nan, X.; Huang, E.; Liu, S. Social E-commerce Tax Evasion Detection Using Multi-modal Deep Neural Networks. In Proceedings of the 2021 Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, Australia, 29 November–1 December 2021; pp. 1–6.
16. Lismont, J.; Cardinaels, E.; Bruynseels, L.; De Groote, S.; Baesens, B.; Lemahieu, W.; Vanthienen, J. Predicting tax avoidance by means of social network analytics. *Decis. Support Syst.* **2018**, *108*, 13–24. [[CrossRef](#)]
17. Liu, L.; Zhuang, Y.; Liu, X. Tax forecasting theory and model based on SVM optimized by PSO. *Expert Syst. Appl.* **2011**, *38*, 116–120. [[CrossRef](#)]
18. Raja Wahab, R.A.S.; Abu Bakar, A. Digital Economy Tax Compliance Model in Malaysia using Machine Learning Approach. *Sains Malays.* **2021**, *50*, 2059–2077. [[CrossRef](#)]
19. Zheng, J.; Li, Y. Machine learning model of tax arrears prediction based on knowledge graph. *Electron. Res. Arch.* **2023**, *31*, 4057–4076. [[CrossRef](#)]
20. Che, Z.; Purushotham, S.; Khemani, R.; Liu, Y. Interpretable deep models for ICU outcome prediction. In Proceedings of the AMIA Annual Symposium Proceedings, Chicago, USA, 12–16 November 2016; p. 371.
21. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, USA, 4–9 December 2017.
22. Wang, H.; Liang, Q.; Hancock, J.T.; Khoshgoftaar, T.M. Feature selection strategies: A comparative analysis of SHAP-value and importance-based methods. *J. Big Data* **2024**, *11*, 44. [[CrossRef](#)]
23. Salih, A.M.; Raisi-Estabragh, Z.; Galazzo, I.B.; Radeva, P.; Petersen, S.E.; Lekadir, K.; Menegaz, G. A perspective on explainable artificial intelligence methods: Shap and lime. *Adv. Intell. Syst.* **2024**, *7*, 2400304. [[CrossRef](#)]
24. Arenas, M.; Barceló, P.; Bertossi, L.; Monet, M. On the complexity of SHAP-score-based explanations: Tractability via knowledge compilation and non-approximability results. *J. Mach. Learn. Res.* **2023**, *24*, 1–58.
25. Liu, X.; Sun, G.; Ju, R.; Li, J.; Li, Z.; Jiang, Y.; Zhao, K.; Zhang, Y.; Jing, Y.; Yang, G. Prediction of load-bearing capacity of FRP-steel composite tubed concrete columns: Using explainable machine learning model with limited data. *Structures* **2025**, *71*, 107890. [[CrossRef](#)]
26. Zhou, C.; Wang, W.; Zheng, Y. Data-driven shear capacity analysis of headed stud in steel-UHPC composite structures. *Eng. Struct.* **2024**, *321*, 118946. [[CrossRef](#)]
27. Zhou, C.; Xie, Y.; Wang, W.; Zheng, Y. Machine learning driven post-impact damage state prediction for performance-based crashworthiness design of bridge piers. *Eng. Struct.* **2023**, *292*, 116539. [[CrossRef](#)]
28. Wang, D.; Thunell, S.; Lindberg, U.; Jiang, L.; Trygg, J.; Tysklind, M. Towards better process management in wastewater treatment plants: Process analytics based on SHAP values for tree-based machine learning methods. *J. Environ. Manag.* **2022**, *301*, 113941. [[CrossRef](#)] [[PubMed](#)]
29. Prendin, F.; Pavan, J.; Cappon, G.; Del Favero, S.; Sparacino, G.; Facchinetti, A. The importance of interpreting machine learning models for blood glucose prediction in diabetes: An analysis using SHAP. *Sci. Rep.* **2023**, *13*, 16865. [[CrossRef](#)] [[PubMed](#)]
30. Meng, Y.; Yang, N.; Qian, Z.; Zhang, G. What makes an online review more helpful: An interpretation framework using XGBoost and SHAP values. *J. Theor. Appl. Electron. Commer. Res.* **2020**, *16*, 466–490. [[CrossRef](#)]
31. AsSadhan, B.; Bashaiwath, A.; Binsalleeh, H. Enhancing Explanation of LSTM-Based DDoS Attack Classification Using SHAP With Pattern Dependency. *IEEE Access* **2024**, *12*, 90707–90725. [[CrossRef](#)]
32. Lee, N.; Swenson, C. Effects of overseas subsidiaries on worldwide corporate taxes. *J. Int. Account. Audit. Tax.* **2016**, *26*, 47–59. [[CrossRef](#)]
33. Gaertner, F.B. CEO after-tax compensation incentives and corporate tax avoidance. *Contemp. Account. Res.* **2014**, *31*, 1077–1102. [[CrossRef](#)]
34. Lanis, R.; Richardson, G. Is corporate social responsibility performance associated with tax avoidance? *J. Bus. Ethics* **2015**, *127*, 439–457. [[CrossRef](#)]
35. Watts, R.L.; Zimmerman, J.L. Towards a positive theory of the determination of accounting standards. *Account. Rev.* **1978**, *L111*, 112–134.
36. Rego, S.O. Tax-avoidance activities of US multinational corporations. *Contemp. Account. Res.* **2003**, *20*, 805–833. [[CrossRef](#)]
37. Markle, K.; Shackelford, D.A.J.T.L.R. Cross-country comparisons of the effects of leverage, intangible assets, and tax havens on corporate income taxes. *Tax Law Rev.* **2012**, *65*, 2013–2025.
38. Dyreng, S.D.; Hanlon, M.; Maydew, E.L. Long-run corporate tax avoidance. *Account. Rev.* **2008**, *83*, 61–82. [[CrossRef](#)]
39. Liang, Q.; Li, Q.; Lu, M.; Shan, Y. Industry and geographic peer effects on corporate tax avoidance: Evidence from China. *Pac.-Basin Financ. J.* **2021**, *67*, 101545. [[CrossRef](#)]
40. Chen, H.; Tang, S.; Wu, D.; Yang, D. The political dynamics of corporate tax avoidance: The Chinese experience. *Account. Rev.* **2021**, *96*, 157–180. [[CrossRef](#)]
41. Liu, X.; Qin, J.; Zhao, K.; Featherston, C.A.; Kennedy, D.; Jing, Y.; Yang, G. Design optimization of laminated composite structures using artificial neural network and genetic algorithm. *Compos. Struct.* **2023**, *305*, 116500. [[CrossRef](#)]

42. Abiodun, O.I.; Jantan, A.; Omolara, A.E.; Dada, K.V.; Umar, A.M.; Linus, O.U.; Arshad, H.; Kazaure, A.A.; Gana, U.; Kiru, M.U. Comprehensive review of artificial neural network applications to pattern recognition. *IEEE Access* **2019**, *7*, 158820–158846. [[CrossRef](#)]
43. Ketkar, N. Introduction to Keras. In *Deep Learning with Python: A Hands-On Introduction*; Apress: Berkeley, CA, USA, 2017.
44. Patalas-Maliszewska, J.; Łosyk, H.; Rehm, M. Decision-Tree Based Methodology Aid in Assessing the Sustainable Development of a Manufacturing Company. *Sustainability* **2022**, *14*, 6362. [[CrossRef](#)]
45. Abd El Aal, A.K.; GabAllah, H.M.; Megahed, H.A.; Selim, M.K.; Hegab, M.A.; Fadl, M.E.; Rebouh, N.Y.; El-Bagoury, H. Geo-Environmental Risk Assessment of Sand Dunes Encroachment Hazards in Arid Lands Using Machine Learning Techniques. *Sustainability* **2024**, *16*, 11139. [[CrossRef](#)]
46. Zhao, W.; Chen, P.; Liu, X.; Wang, L. Impact response prediction and optimization of SC walls using machine learning algorithms. *Structures* **2022**, *45*, 390–399. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.