iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.59282

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Tax Evasion Detection: A Fusion of AdaBoost Classifiers and Deep Learning Models

Ms. Manyala Naga Sailaja[1], Manoj Kumar[2], M Pooja[3], M Srivardhan[4]
[2, 3, 4]*UG Student, Department of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana*
[1]*Assistant Professor, Department of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana*

Abstract: ***Income Tax plays an important role in the developmental activities of a nation. However some companies try to avoid or minimize their tax by showing reduced profits. Such companies can be identified by studying and analyzing the balance sheet indicators like: Sale Volume Vs. Purchase Volume, Employee Growth Vs. Taxpaid, Total Sales Vs. Taxpaid, Claims Crediting, Net Income Vs. Tax Increase, Gross Loss Vs. Tax, Inventories Vs. Direct Income, Gross Revenue Vs. Taxpaid, Raw Material Purchase Cost (RMPC) Vs. Final Product Sales (FPS) and Fixed Assets Vs. Taxpaid. This project aims at Tax Evaders detection by using a dataset, containing the above mentioned identification factors in the balance sheet. The analysis is carried out by using AdaBoost classifier and also by using a Deep learning model.***
Keywords: Deep Learning, Machine learning, Tax Evasion Detection

I. INTRODUCTION

As there are millions and millions of companies in the blessed country, It's not even possibly conceivable to thoroughly and completely study the balance sheets of each and every single company to determine the tax evaders, right? Hence, In the existing system which currently exists, the Income Tax department randomly and arbitrarily selects the balance sheets for additional evaluation without any real method to it. The proposed system helps IT department for identifying and being able for identification of the companies that are most very likely to evade the tax by analyzing the data using Ada boost analysis and neural net analysis. The project, which is quite interesting, interestingly comprises of four modules. The primary module deals with building AdaBoost model and finding its Accuracy, which is a crucial aspect. The second module deals personally with the predictions of AdaBoost model, kind of fascinating. The third module comprises of building the deep neural net and finding its accuracy, which is an absolute key requirement. The final module deals with finding the predictions with the Deep Neural Net model which is an intricate process. AdaBoost is seriously an iterative classification method.

II. LITERATURE REVIEW

This segment audits past considers conducted on charge avoidance discovery. As said prior, machine learning calculations play an fundamental part in assess avoidance location, and most thinks about have utilized a combination of ml calculations. For case, a consider within the field displayed an architecture for the issue of money related extortion location by Chinese commercial companies, which included communication with the specialists within the field, utilize of information mining calculations, plan enlightening for information mining systems, and integration of information of specialists within the field.

Benkraiem, R., Uyar, A., Kilic, M., & Schneider, F. In this consider, we offer novel around the world to prove on the impacts of these two components on charge avoidance. In general, indeed in spite of the fact that solid examining benchmarks may moderate assess avoidance, the ethical behavior of firms encompasses a factually more vigorous impact in accomplishing this objective; these comes about hold after we control for endogeneity and diverse subperiods (some time recently, amid, and after the later worldwide budgetary emergency). More particularly, the moral behavior of firms is successful for moo- and middle-income nations with moo and tall levels of financial specialist assurance and low-efficacy corporate sheets; be that as it may, moral behavior and examining guidelines are commonly successful for high-income nations and nations with center level speculator security and center- and high-efficacy corporate sheets.

Uyar, A., Nimer, K., Kuzey, C., Shahbaz, M., & Schneider, F. Uyar, Ali, et al. "Can e-government initiatives alleviate tax evasion? The moderation effect of ICT." Technological Forecasting and Social Change 166 (2021): 120597.

Wu, Roung-Shiunn, et al. "Using data mining technique to enhance tax evasion detection performance." Expert Systems with Applications 39.10 (2012): 8769-8777.

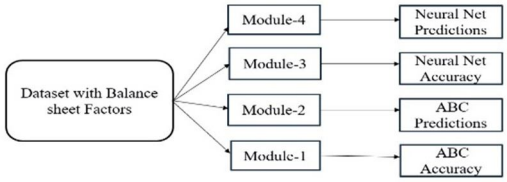
©IJRASET: All Rights are Reserved

Pappa, Evi, Rana Sajedi, and Eugenia Vella. "Fiscal consolidation with tax evasion and corruption." Journal of International Economics 96 (2015): S56-S75.
Di Gioacchino, Debora, and Domenico Fichera. "Tax evasion and tax morale: A social network analysis." European Journal of Political Economy 65 (2020): 101922.
El Bouchti, Abdelali, et al. "Fraud detection in banking using deep reinforcement learning." 2017 Seventh International Conference on Innovative Computing Technology (INTECH). IEEE, 2017.
Savić, Miloš, et al. "Tax evasion risk management using a Hybrid Unsupervised Outlier Detection method." Expert Systems with Applications 193 (2022): 116409.
Network, Tax Justice. "The State of Tax Justice 2020: Tax Justice in the time of COVID-19. 2020." URL: https://taxjustice.net/wpcontent/uploads/2020/11/The_St te_of_Tax_Justice_2020_ENGLISH. pdf (accessed 09.11.2022) (2021).
Zhang, Fa, et al. "TTED-PU: a transferable tax evasion detection method based on positive and unlabeled learning." 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC). IEEE, 2020.
El Bouchti, Abdelali, et al. "Fraud detection in banking using deep reinforcement learning." 2017 Seventh International Conference on Innovative Computing Technology (INTECH). IEEE, 2017.
Peng, Taoxin, and Alexander Telle. "A tool for generating synthetic data." Proceedings of the First International Conference on Data Science, E-learning and Information Systems. 2018.
Assefa, Samuel A., et al. "Generating synthetic data in finance: opportunities, challenges and pitfalls." Proceedings of the First ACM International Conference on AI in Finance. 2020.
Visitpanya, Narongchai, and Taweesak Samanchuen. "Synthesis of Tax Return Datasets for Development of Tax Evasion Detection." IEEE Access (2023).
Didimo, Walter, et al. "A visual analytics system to support tax evasion discovery." Decision Support Systems 110 (2018): 71-83.
Kovermann, Jost, and Patrick Velte. "The impact of corporate governance on corporate tax avoidance—A literature review." Journal of International Accounting, Auditing and Taxation 36 (2019): 100270.

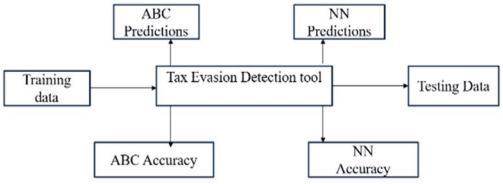
III. METHODOLOGIES

A. System Description:

There are four modules in the project. The AdaBoost model's construction and accuracy determination are the topics of the first module. The AdaBoost model's predictions are covered in the second module. The construction of the neural net and determining its correctness are covered in the third module. The fourth module focuses on using the neural net model to find the predictions.



Factors on the Balance Sheet Using the appropriate user interfaces, training and testing data must be supplied to the system as input. Next, the system computes Neural Net Accuracy & Predictions and AdaBoost Accuracy & Predictions.



B. Tools Used

The program used to design the project's graphical user interfaces was Qt Designer. The tool is located in the Anaconda tool's Library/Bin subdirectory.
After highlighting the main window, click the build button to build the main window. The properties window, located in the center section on the right hand side, can be used to set a title for the main window.

C. AdaBoost Classification Algorithm

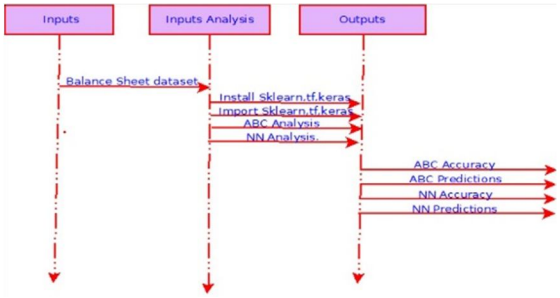
A machine learning approach for ensembles called adaptive boosting, or AdaBoost, can be used in a variety of classification and regression applications. In order to categorize data, this supervised learning strategy combines multiple weak or base learners (such as decision trees) to produce a strong learner. The training dataset's examples are weighted based on the effectiveness of previous classifications in order for AdaBoost to operate.

D. Neural Net Classification Algorithm

To find underlying correlations in a given set of data, a neural network is a collection of algorithms created to mimic the functions of the human brain. In this context, artificial or organic neural networks are referred to as neural networks. Neural networks can yield the best results without the need to modify the criteria because they are able to adapt to changing input.

1) Sequence Diagram:

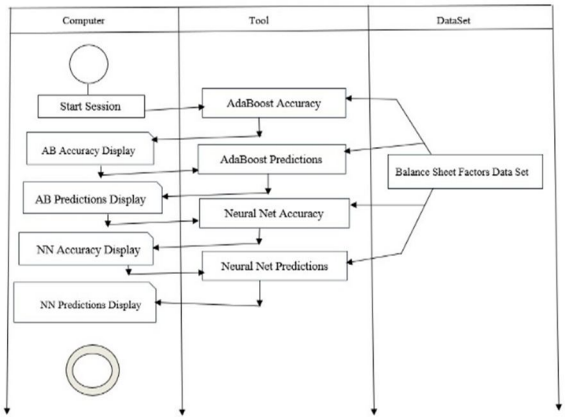
It is an interaction diagram that illustrates the relationships between objects and their respective functions. It is a message sequence chart construct. The interactions between objects are displayed chronologically in a sequence diagram.



From the sequence diagram above, we must proceed as follows: Enter the information requested as indicated by the above figure. Give the Balance Sheet Factors Dataset, and then compute the AdaBoost and Neural Net predictions' accuracy.

2) Activity Diagram

Activity diagram is basically a flow chart to represent the flow from one activity to another activity. The activity can be described as an operation of the system. So, the control flow is drawn from one operation to another. In activity diagram we can see that first provide the dataset comprising of Balance Sheet Factors, and then train the model, test the model and calculate the accuracies along with the predictions of AdaBoost & Neural Net.



3) Highlights and Philosophy

Python may be a programming dialect. Object-oriented programming and organized programming are completely bolstered, and numerous of its highlights bolster utilitarian programming and aspect-oriented programming (counting by meta programming and meta objects (enchantment strategies)). Numerous other standards are upheld using expansions, counting plan by contract and rationale programming.

Python employments energetic writing, and a combination of reference tallying and a cycle identifying trash collector for memory administration. It moreover highlights energetic title determination (late authoritative), which ties strategy and variable names amid program execution.

4) Modules Imported in Python Routines

a) OS Module: The OS module in Python gives a way of utilizing working framework subordinate usefulness. The capacities that the OS module gives permits you to interface with the basic working framework that Python is running on – be that Windows, Mac or Linux.

b) Framework Module: The sys module gives data approximately constants, capacities and strategies of the Python translator. dir(system) gives a rundown of the accessible constants, functions and strategies. Another plausibility is the assistance() work. Utilizing help(sys) gives important detail information. The module sys educates e.g. almost the maximal recursion profundity (sys.getrecursionlimit()) and gives the plausibility to alter (sys.setrecursionlimit()). The current adaptation number of Python can be gotten to as well by utilizing this module. The standard output (stdout) can be diverted e.g. into a record, so that we are able this record afterward with another program. The same is conceivable with the standard mistake stream, we can divert it into a record as well. We are able both stderr and stdout into the same record or into isolated files.

c) NumPy: NumPy's main question is the homogeneous multidimensional cluster. It could be a of components (as a rule numbers), all of the same sort, ordered by a tuple of positive integrability. In NumPy measurements are called axes. Python has a package called NumPy. It is an extension module for Python that is primarily written in C. This clearly shows that Numpy's precompiled scientific and numerical capabilities and features guarantee exceptionally fast execution. The foundation of NumPy is two older Python modules for cluster management and Numeric is one of them. Another precursor to NumPy is Numarray, which is also hypothesized and may be a modification of Numeric.

The more vital qualities of an ndarray protest are:

- ndarray.ndim

The number of tomahawks (measurements) of the array.

- ndarray.shape

The measurements of the cluster. This is often tuple of integrability demonstrating the estimate of the cluster in each measurement. For a framework with n columns and m columns, shape will be (n,m). The length shape tuple is hence the no of tomahawks, ndim

- ndarray.size

The Full number of components of the array. Typically break even with to the item of the components of shape.

- ndarray.dtype

An object describing the sort of the components within the cluster. One can make or indicate dtype's utilizing standard Python sorts. Also NumPy gives sorts of its claim. numpy.int32, numpy.int16, and numpy.float64 are a few examples.

- ndarray.itemsize

The estimate in bytes of each component of the cluster. For illustration, an cluster of components of sort float64 has itemsize 8 (=64/8), whereas one of sort complex32 has itemsize 4 (=32/8). It is comparable to ndarray.dtype.itemsize.

- ndarray.data

Regularly, we won't got to utilize this property since we are going the components in an cluster utilizing ordering offices.

IV. RESULTS AND DISCUSSION

We use three different types of files which are: (1) .py files (2) .ui file (3).txt files.
.ui files are the user interface files, created by using PyQt layout editor
.py files are python program files created either manually, or automatically. For instance, each .ui file has a corresponding .py file that is created automatically by using the PyUIC tool.

.txt files contains the generic useful information about the project
tax1.py, is the entry program for this project. Execution of this python program leads to the entry screen as follows:

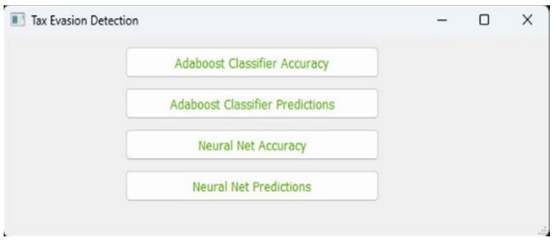


Fig:1

Fig:1 States the Graphical User Interface (GUI) of the project. Which contains four buttons, the first two buttons are used to calculate the accuracy & predictions of AdaBoost and the last two buttons are used for Neural Net classification.

If we select AdaBoost Classifier Accuracy Button it displays:

AdaBoost Classifier Accuracy:

```
(base) C:\8049>python tax1.py
The accuracy of Adaboost Classifier on testing data is: 78.80247763248451
```

Fig:2

The Adaboost Classifier accuracy on testing data is: 78.802%
If we select AdaBoost Classifier Predictions Button it displays:

AdaBoost Classifier Predictions:

```
(base) C:\8049>python tax1.py
The accuracy of Adaboost Classifier on testing data is: 78.80247763248451
The accuracy of Adaboost Classifier on testing data is: 78.80247763248451
ETC prediction on the first test set is: [1]
ETC prediction on the second test set is: [1]
ETC prediction on the third test set is: [0]
```

Fig:3

In this we consider three Test set where;
First test set is: [1] - Medium
Second test set is: [1] - Medium
Third test set is: [0] - Low
If we select Neural Net Accuracy Button it displays:

Neural Net Accuracy:

```
Models Loss and Accuracy are [0.2517312169075012, 0.8932244181632996]
$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
```

Fig:4

The Neural net accuracy of Models Loss and Accuracy [0.251, 0.893]
If we select Neural Net Predictions Button it displays:

Neural Net Predictions:

```
1/1 [==============================] - 0s 136ms/step
DL Model Prediction on the first test set is: [1]
1/1 [==============================] - 0s 31ms/step
DL Model Prediction on the second test set is: [2]
1/1 [==============================] - 0s 34ms/step
DL Model Prediction on the Third test set is: [0]
```

Fig:5

In this we consider three Test set where;
First test set is: [1] - Medium
Second test set is: [2] - High
Third test set is: [0] – Low
The results and discussion underscore the positive impact of Tax Evasion Detection using AdaBoost Classifiers and Deep Learning Models. It benefits national revenue by forcing evaders to pay tax.
Which brings the further development of nation at infrastructures like roads and irrigation projects.

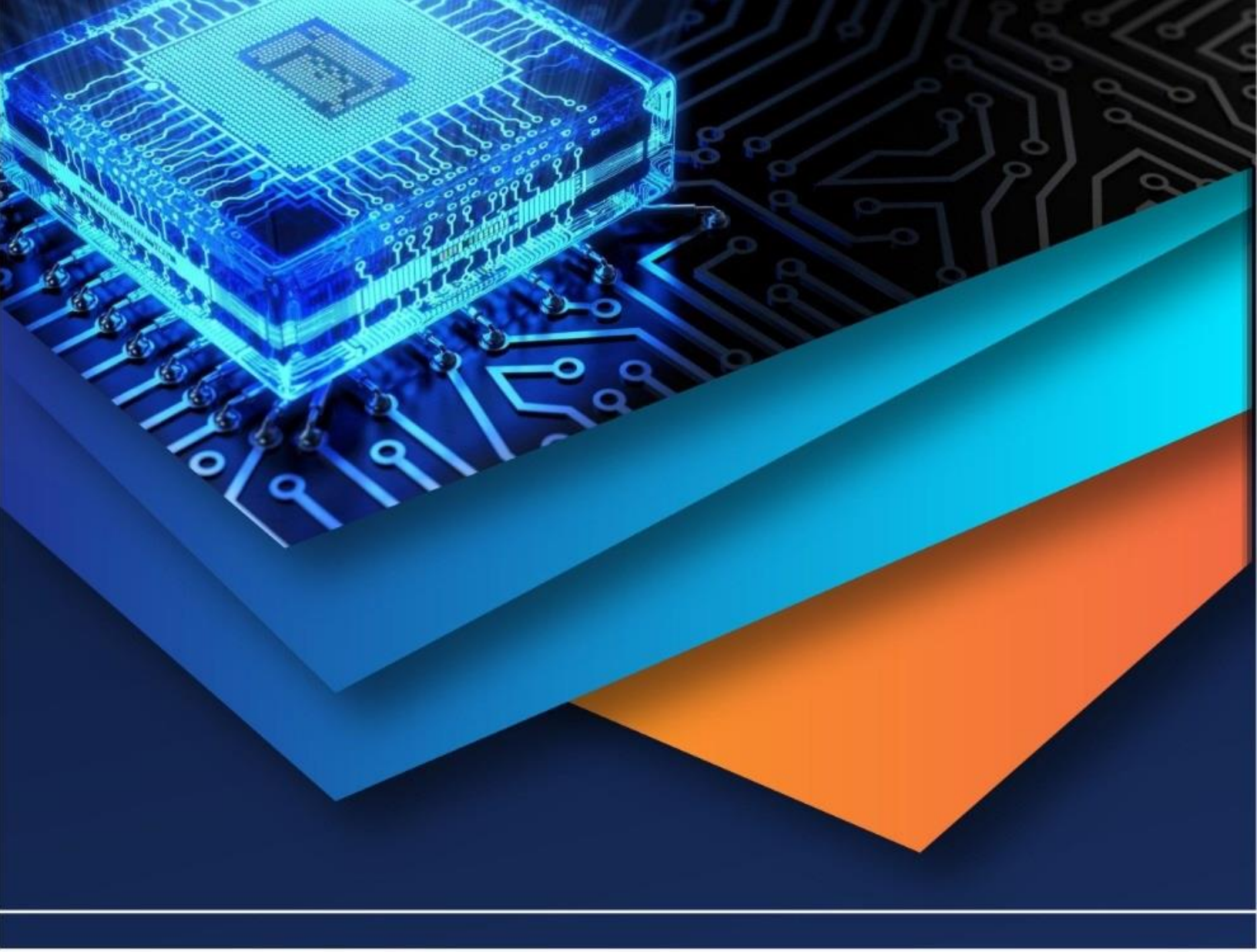
V. CONCLUSION

This project entitled “Tax Evasion Detection using Adaboost Classifier.” is useful to the income Tax department to determine the companies that are likely to evade the tax. This project is indirect useful to enhance the national revenue by forcing the tax evaders to pay the tax. Increased revenue leads to the further growth of the country as it can be applied to the advancement of nations infrastructure like Roads and Irrigation projects.

REFERENCES

[1] R. Benkraiem, A. Uyar, M. Kilic, and F. Schneider, ‘‘Ethical behavior, auditing strength, and tax evasion: A worldwide perspective,’’ J. Int. Accounting, Auditing Taxation, vol. 43, Jun. 2021, Art. no. 100380, doi: 10.1016/j.intaccaudtax.2021.100380.
[2] A. Uyar, K. Nimer, C. Kuzey, M. Shahbaz, and F. Schneider, ‘‘Can e-government initiatives alleviate tax evasion? The moderation effect of ICT,’’ Technol. Forecasting Social Change, vol. 166, May 2021, Art. no. 120597, doi: 10.1016/j.techfore.2021.120597.
[3] R.-S. Wu, C. S. Ou, H.-Y. Lin, S.-I. Chang, and D. C. Yen, ‘‘Using data mining technique to enhance tax evasion detection performance,’’ Expert Syst. Appl., vol. 3 9, no. 10, pp. 8769–8777, Aug. 2012, doi: 10.1016/j.eswa.2012.01.204.
[4] E. Pappa, R. Sajedi, and E. Vella, ‘‘Fiscal consolidation with tax evasion and corruption,’’ J. Int. Econ., vol. 96, pp. S56–S75, Jul. 2015, doi: 10.1016/j.jinteco.2014.12.004.
[5] D. Di Gioacchino and D. Fichera, ‘‘Tax evasion and tax morale: A social network analysis,’’ Eur. J. Political Economy, vol. 65, Dec. 2020, Art. no. 101922, doi: 10.1016/j.ejpoleco.2020.101922.
[6] M. Savić, J. Atanasijević, D. Jakovetić, and N. Krejić, ‘‘Tax evasion risk management using a hybrid unsupervised outlier detection method,’’ Expert Syst. Appl., vol. 193, May 2022, Art. no. 116409, doi: 10.1016/j.eswa.2021.116409.
[7] Tax Justice Network. The State of Tax Justice 2021. Accessed: Nov. 10, 2022. [Online]. Available: https://taxjustice.net/reports/the-stateof-tax-justice-2021/
[8] F. Zhang, B. Shi, B. Dong, Q. Zheng, and X. Ji, ‘‘TTED-PU: A transferable tax evasion detection method based on positive and unlabeled learning,’’ in Proc. IEEE 44th Annu. Comput., Softw., Appl. Conf. (COMPSAC), Madrid, Spain, Jul. 2020, pp. 207–216, doi: 10.1109/compsac48688.2020.00036.
[9] A. E. Bouchti, A. Chakroun, H. Abbar, and C. Okar, ‘‘Fraud detection in banking using deep reinforcement learning,’’ in Proc. 7th Int. Conf. Innov. Comput. Technol. (INTECH), Luton, U.K., Aug. 2017, pp. 58–63, doi: 10.1109/intech.2017.8102446.
[10] T. Peng and A. Telle, ‘‘A tool for generating synthetic data,’’ in Proc. 1st Int. Conf. Data Sci., E-Learn. Inf. Syst., Madrid, Spain, Oct. 2018, pp. 1–6, doi: 10.1145/3279996.3280018.
[11] S. A. Assefa, D. Dervovic, M. Mahfouz, R. E. Tillman, P. Reddy, and M. Veloso, ‘‘Generating synthetic data in finance: Opportunities, challenges and pitfalls,’’ in Proc. 1st ACM Int. Conf. AI Finance, New York, NY, USA, Oct. 2020, pp. 1–8, doi: 10.1145/3383455.3422554.
[12] P. Sophaphong, V. Rattanawiboonso, and C. Chanbanjong, ‘‘Factors and prevention guideline of tax evasion of listed companies in the stock exchange of Thailand,’’ J. Humanities Social Sci. Valaya Alongkorn, vol. 12, no. 3, pp. 47–57, 2017.
[13] W. Didimo, L. Giamminonni, G. Liotta, F. Montecchiani, and D. Pagliuca, ‘‘A visual analytics system to support tax evasion discovery,’’ Decis. Support Syst., vol. 110, pp. 71–83, Jun. 2018, doi: 10.1016/j.dss.2018.03.008.
[14] J. Kovermann and P. Velte, ‘‘The impact of corporate governance on corporate tax avoidance—A literature review,’’ J. Int. Accounting, Auditing Taxation, vol. 36, Sep. 2019, Art. no. 100270, doi: 10.1016/j.intaccaudtax.2019.100270.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)