Technology And Artificial Intelligence To Gain Insight Of Tax Data

[1]Detty Purnamasari, [2]Milda Safrila Oktiana, [3]Ulfa Hidayati, [4]Mario Mora Siregar, [5]Fanka Arie Reza

[1,3]Information Technology Universitas Gunadarma,[2,5]Magister of Information System Management Universitas Gunadarma, [4]Informatics Universitas Gunadarma
Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat
[1]detty@staff.gunadarma.ac.id, [2]mildaokt21@gmail.com, [3]ulfahidayati21@gmail.com, [4]mariomorasiregar@gmail.com, [5]fankawork@gmail.com

Abstract

Technology is used in various fields to effective and efficient tasks. Artificial intelligence methods are also used to analyze data to gain insights. The knowledge is an output from the process using artificial intelligence that useful in policy-making and decision-making within an organization. Tax management organizations in a country possess large and complex datasets, the implementation of artificial intelligence is necassary. This article reviews the concept of artificial intelligence and the technologies that can be used to process tax data and gain the insights. Linear regression and clustering techniques, which are part of artificial intelligence, can be used to gain insights into tax data. Clustering techniques can be implemented on taxpayer data to assign labels/classes to taxpayers based on the amount of tax paid.

Keywords : *artificial intelligence, clustering, insight using ai, tax, taxpayer*.
JEL Codes : H20, O33

INTRODUCTION

The Indonesian Directorate General of Taxes (in Indonesia called *Direktorat Jenderal Pajak Indonesia*) website states that taxes are mandatory contributions to the state owed by individuals or entities, which are compulsory under the law, without any direct compensation, and are used for state purposes for the greatest prosperity of the people. Taxes are also a primary source of revenue for governments in various countries. A country's economic life and the provision of infrastructure, goods, and services rely largely on tax revenues. (Nadiminti, V. et al., 2023) (Syahril, M.A.F., Hasan, H., 2024)

The State Budget (in Indonesia called *Anggaran Pendapatan dan Belanja Negara/APBN*) in Indonesia is largely sourced from taxes, which constitute the largest source of revenue. Taxes imposed on the public include income tax, land and building tax, motor vehicle tax (Mawarni, A. et al., 2024). The government continues to revise and refine tax regulations to optimize tax revenue (Nanda, A. W. M., Nafi'ah, B. A., 2025). The use of technology and artificial intelligence is one way to process tax data, generating insights that are then used in decision-making and policy-making.

The use of technological advances has transformed manual tasks into automated or digital ones. The e-filling application is one form of output that utilizes technology and replaces manual processes with automation. This application was launched to assist taxpayers in reporting or submitting tax returns, eliminating the need for transportation costs to go to the Tax Service Office for taxpayers. Tax returns are delivered to taxpayers through an application that can be accessed with internet access using a computer or mobile phone (Tarmidi, D. et.al., 2017).

The e-Invoice application is another example of a tax-related feature that utilizes technology. This application is provided by the Directorate General of Taxes (in Indonesia called *Direktorat Jenderal Pajak Indonesia*) to improve the existing tax collection and management system. (Mayori, E., Halimah, S. N., 2022)

Taxpayers can access all tax-related services through internet-based services offered by financial and tax management institutions, such as tax return reporting, which can be considered an implementation of the e-Tax concept (Mapesa et al., 2020). The use of digital (online) applications provides a large source of digital data. This data can become a source of information and knowledge when properly processed to support decision-making, one of which is through the use of artificial intelligence.

Research conducted (Darono, A., 2025) has demonstrated the existence of a centralized tax analytics system and application that generates data to support various tax administration functions, including taxpayer registration, compliance monitoring, dispute resolution, and law enforcement. A data mining approach has also been used (Ginanjar, A., Wibawa, A. S. 2025) using a linear regression method that can identify patterns in taxpayer financial reports.

This article conducts a literature review on the use of technology and artificial intelligence on tax data. Several implementations of artificial intelligence can be used to detect fraud in tax payments and group taxpayer data based on the amount of tax payments.

LITERATURE REVIEW

Artificial intelligence can be defined as a system implemented on a computer that can complete tasks that would normally require human labor or human intelligence. Implementing artificial intelligence on machines requires data as knowledge to enable learning and experience, allowing the machine to become as intelligent as a human. (Purnamasari, D., et.al., 2023)

Research reviewing the implementation of artificial intelligence in various fields has been conducted (Drus, Z., Khalid, H., 2019) on the topic of sentiment analysis. First, began by defining the research question. The steps taken were:

a. Determining the characteristics required for the study.
b. Finding and obtaining potentially relevant literature and selecting relevant literature.
c. Synthesizing relevant information from the literature.
d. Writing up the review results.

The methods that can be used to conduct a literature review are as follows: (Ochengo, J., Deng, H., 2023)

a. Systematic reviews, are arguably the most commonly used standard in literature review methods. Systematic reviews involve a comprehensive and structured approach to identifying, selecting, and critically appraising relevant studies.
b. Narrative reviews, also known as traditional or qualitative techniques, provide descriptive overviews of the literature on a specific topic.
c. Meta-analysis, is a quantitative review technique that involves statistical analysis to combine data from multiple studies.
d. Scope reviews, aim to map the existing literature on a broad topic or research question. These reviews are widely used to explore the diversity of research approaches and identify research gaps or emerging trends.

Research has been conducted using narrative review techniques (Putri, S. K. et.al., 2022) on the role of digital libraries in supporting learning and providing credible information to the public. The review involved five reference sources indexed in Google Scholar spanning the past ten years, analyzing research topics, methods, and results.

RESEARCH METHOD

This article is a preliminary study by conducting a literature review on the implementation of technology and artificial intelligence related to tax data, and also techniques in artificial intelligence in other fields but can be used to find knowledge in tax data.

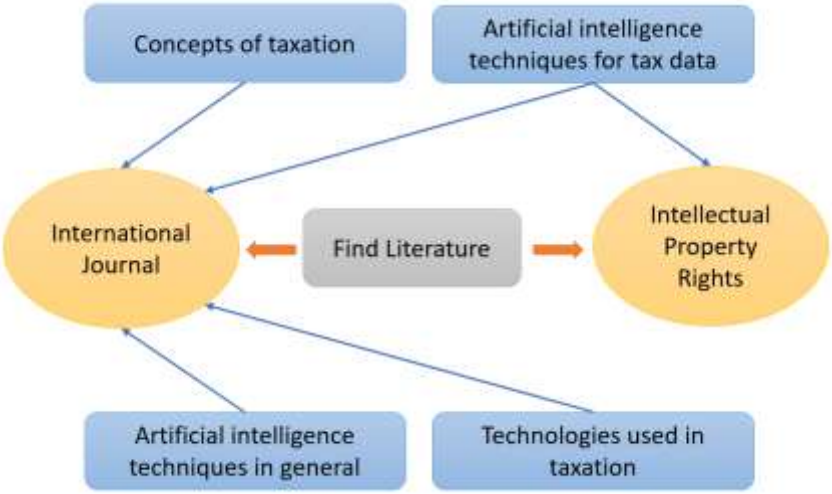

Figure 1. Research Methods in the Review Conducted

Figure 1 shows the techniques used in this article. The review of artificial intelligence research in this article uses a narrative review to obtain a descriptive overview of the literature used. The research or publications conducted, referenced, and reviewed include:

a. Research published in international journals
b. Applications related to tax data processing that have been registered as intellectual property rights

The research topics reviewed are as follows:

a. Concepts of taxation
b. Technology used in taxation
c. Artificial intelligence techniques in general
d. Artificial intelligence techniques for tax data

RESULTS AND DISCUSSION

The review in this article was conducted as a preliminary study to understand taxation and existing data related to taxes, as well as methods of artificial intelligence that can be used.

Figure 2 shows the flow for analyzing tax data. Data comes from internal and external sources, then stored in a data warehouse and processed using various techniques (one of which is the implementation of artificial intelligence or data science), resulting in information and knowledge (Darono, A., 2025). Tax data is processed and analyzed as information and knowledge used in decision-making or the development of new policies.

Figure 2. Tax data analytics workflow (adapted from Sakti 2021; DJP 2022)
Source: Darono, A. (2025)

Large amounts of data are used to analyze fundamental issues related to the implementation of tax collection and administration, as well as legal issues. Therefore, to assess the tax oversight system, more comprehensive algorithms are needed to reduce the number of taxpayers who do not pay taxes and increase the number of active taxpayers. (Liuhong, C., 2022)

One example of using clustering techniques (Leonardo, P., Tjen, C. 2020), with data analyzed using a content analysis approach. This study employed clustering techniques to gain insight into primary data obtained through interviews and questionnaires. This method focused on each word, focusing on the main topic being discussed.

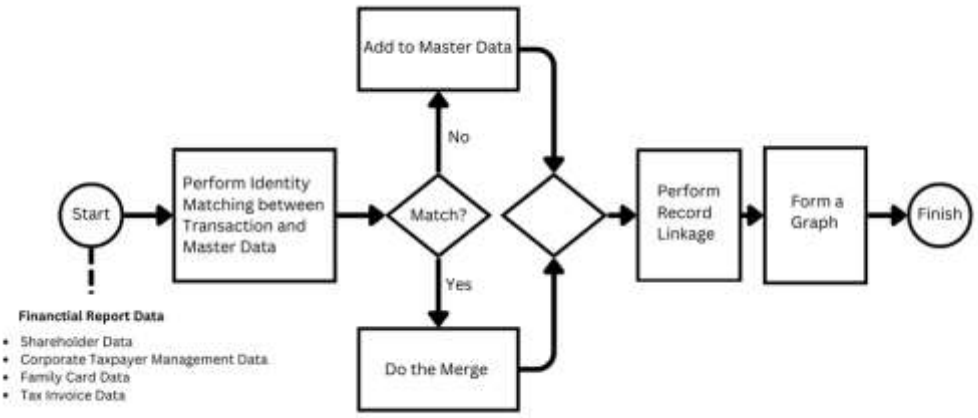

Figure 3. Graph Formation in Tax Data Fraud Detection
Source: Nuryani et.al. (2024)

Taxpayers are not only individuals, but also corporate taxpayers. Previous research (Nuryani et al., 2024) investigated the detection of tax evasion within complex corporate networks. Big data analytics techniques utilize an artificial intelligence approach. The approach involves graph formation and the use of Social Network Analysis (SNA) to detect tax evasion, including schemes such as issuing invoices without actual transactions and transfer pricing through special relationships. Figure 3 illustrates the graph formation associated with tax data used to detect fraud.

The linear regression method can also be used as a data mining approach to identify patterns in taxpayer financial reports (Ginanjar, A., Wibawa, A. S. 2025). The research also used t-SNE (t-distributed Stochastic Neighbor Embedding) to clearly visualize data groupings. The research demonstrates how to analyze the skewness of annual accounting reports.

One application of artificial intelligence-based tax data processing is the use of clustering techniques to assign labels or classes to tax data (Purnamasari, D. et. al., 2024). The stages for building a clustering model that then becomes a ready-to-use application using The Cross Industry Process – Data Mining (CRISP-DM), shown in Figure 4, consist of: (Purnamasari, D. et. al., 2023)

a. Business Understanding, defining and identifying the problem
b. Data Understanding, obtaining data and seeking deeper insights into the data available
c. Data Preparation, adjusting the data according to the required needs
d. Modeling, training the prepared data with a machine learning/deep learning algorithm
e. Evaluation, assessing the performance of each machine learning/deep learning algorithm trained in the modeling stage
f. Deployment, implementing the machine learning/deep learning model into a usable application

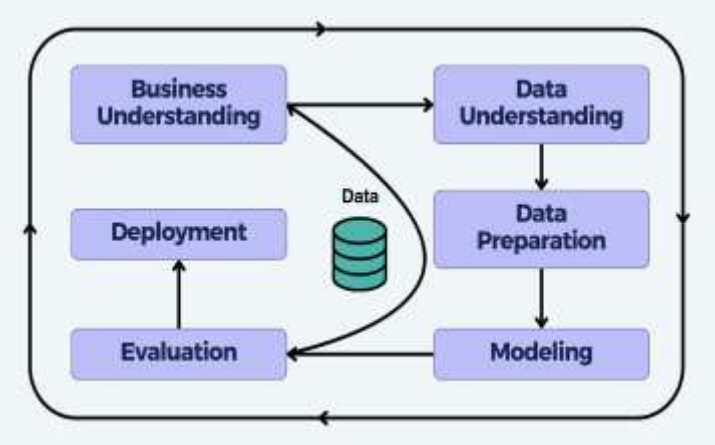

Figure 4. CRISP-DM Methodology

Figure 5(a) is an example of a visualization of scattered data with unknown groupings. Therefore, clustering techniques can be used to generate groups based on similar data. Figure 5(b) is an example of a visualization of clustering results with already grouped data. (Villanueva, O. I., et. al., 2022).

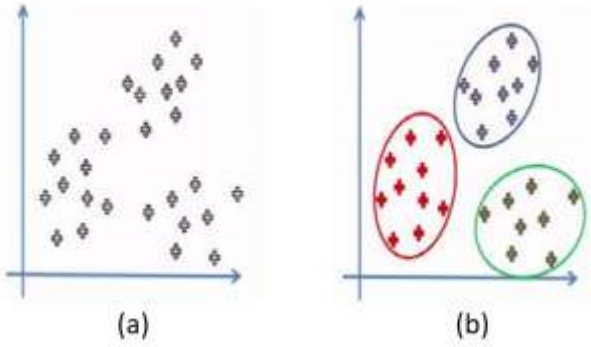

Figure 5. Visualization of Data Distribution and Cluster Results
Source: Villanueva, O. I., et. al., (2022)

An application for tax data processing using clustering techniques (Purnamasari, D. et. al., 2024) utilizes the K-Means algorithm. The steps in the K-Means algorithm are as follows: (Sucahyo, F. R., et. al., 2025)

a. Initialization of Centroids
b. Assigning Data Points to the Nearest Centroid
c. Updating Centroids
d. Reiteration of the Process

The software requirements for building models and creating artificial intelligence-based applications include:

a. Operating System: Windows 10
b. Programming Language and Libraries: Python 3.7+ with libraries such as Pandas, Scikit-learn, and Matplotlib.
c. Web Editor: Jupyter Notebook
d. Deployment Framework: Streamlit
e. Web Browser: Latest version of Google Chrome or Microsoft Edge to access the application dashboard

A clustering model is built to generate groupings in the data. Once the best number of clusters is found, the clusters are assigned labels or classes based on the data. Figure 6 shows an example of an application page that clusters taxpayer data based on tax payment amounts.

Figure 6. Taxpayer Cluster Application Home Page
Source: Purnamasari, D. et.al., (2024)

CONCLUSION AND SUGGESTION

The research in this article is the initial part of a literature review on the use of technology and artificial intelligence. Work related to artificial intelligence can be conducted using the CRISP-DM stages, and the initial step is business understanding. Therefore, this article also reviews taxation to gain an understanding of the tax data that will be used in its implementation with artificial intelligence. Techniques that can be used to discover knowledge from tax data include linear regression and clustering. Clustering techniques using the K-Means algorithm, for example, can be used to obtain groups of taxpayers based on taxpayer payments.

REFERENCES

Darono, A. (2025). Dealing with Last-Mile Analytics: evidence from Indonesian Tax Administration through Practice Research. Scientax: Jurnal Kajian Ilmiah Perpajakan Indonesia, Vol. 6 No. 2, ISSN 2286-5718

Direktorat Jenderal Pajak.Pengertian Pajak. https://pajak.go.id/id/pajak. Tanggal akses: 16 Agustus 2025

Drus, Z., Khalid, H. 2019. Sentiment Analysis in Social Media and Its Application: Systematic Literature Riview. Procedia Computer Science 161, halaman 707–714

Ginanjar, A., Wibawa, A. S. 2025. Data Mining to Detect Fraud Patterns in a taxpayer's Financial Statement. Scientax Jurnal Kajian Ilmiah Perpajakan Indonesia, vol. 6 no. 2, April, ISSN 2686-5718

Leonardo, P., Tjen, C. 2020. Penerapan Ketentuan Perpajakan pada Transaksi E-Commerce pada Platform Marketplace. Jurnal Pendidikan Akuntansi dan Keuangan, Vol. 8 No. 1, hal. 45-54

Liuhong, C. 2022. Research on Tax Collection and Administration Application and Legal Issues Based on Big Data Analysis. Hindawi Journal of Sensors, https://doi.org/10.1155/2022/6578964

Mapesa, H. J., Nyalle, M. A., & Masunga, F. J. (2020). Influence of E-tax System on Tax Revenue Collection in Tanzania Large Taxpayers: A Prior and Posterior Analysis. Journal of Accounting Finance and Auditing Studies (JAFAS), 6(4), 44–63. https://doi.org/10.32602/jafas.2020.027

Mawarni, A., Jamali, H., Mispa, S. 2024. The Effect of E-Filling Application and Knowledge of Taxation on Taxpayer Compliance at The Tax Service Office. The Accounting Journal of BINANIAGA Vol. 07, No. 02, December, p-ISSN: 2527-4309, e-ISSN: 2580-1481

Mayori, E., Halimah, S. N. (2022). Analysis of Application of E-Tax Invoice in Improving Compliance with PP23Tax Report 2018 (Case Study at CV Pusaka Utama). Budapest International Research and Critics Institute-Journal (BIRCI-Journal), Volume 5, No 2, May 2022, Page: 10343-10350, e-ISSN: 2615-3076(Online), p-ISSN: 2615-1715(Print)

Nadiminti, V., Bapa, B., Othlyn, T. (2023). The Negative Effects of Application of Increase Taxes: a Case of Residents of Papua New Guinea University of Technology. International Journal of Reseacrh in Business and Social Science, Vol. 12 No. 5, ISSN: 2147-4478, hal. 437-445

Nanda, A. W. M., Nafi'ah, B. A. 2025. Analysis of The Application Withholding, Depositing, and Reporting Income Tax Article 23 on Services in Private Companies. Amnesty: Jurnal Riset Perpajakan, vol. 8 No. 1, Mei

Nuryani, Mutiara, A. B., Wiryana I M. Purnamasari, D., Nurafrianto, S. 2024. Artificial Intelligence Model for Detecting Tax Evasion Involving Complex Network Schemes. APTISI Transactions on Technopreneurship (ATT), Vol. 6 No. 3, November, hal. 339 – 356

Ochengo, J., Deng, H. 2023. Evaluating Literature Review Methods in Plantae Studies: A Critical Analysis for Improved Research Synthesis. International Journal of Engineering, Economic, Social Politic and Government, Vol. 1, No. 3, ISSN 2503-3212

Purnamasari, D., Aji, A. B., Putri, D. W. A., Reza, F. A., Oktiana, M. S., Yanda, N., Hidayati, U. 2023. Pengantar Metode Analisis Sentimen. ISBN: 978-602-0764-57-3, Gunadarma Penerbit

Purnamasari, D., Siregar, M. M., Reza, F. A., Hidayati, U., Oktiana, M. S. 2024. Aplikasi Cluster Wajib Pajak dengan Algoritma K-Means. Manual Book, HAKI no. EC002024229191, November

Putri, S. K., Rukmana, E. N., Saepudin, E. 2022. Narrative Literature Review Penelitian Perpustakaan Digital sebagai Sumber Pembelajaran Saat Covid-19 dalam Database Google Scholar. Jurnal Perpustakaan Universitas Airlangga: Media Informasi dan Komunikasi Kepustakawanan, vol. 12 no. 2, ISSN 2356-1408

Sucahyo, F. R., Santi, I. H., Rahmat, M. F., Fahrizal, D. 2025. Comparing k-Means and K-medoids Algorithms for Clustering Hamlet Region by Tax Liabilities in Tax Determination Documents. International Journal of Science and Technology Research Archive, vol 8 no. 1, ISSN 0799-6632

Syahril, M. A. F., Hasan, H. (2024). A Comparative Research on The Effectiveness of Progressive versus Proportional Tax Systems in Enhancing Social Justice. Adminstrative and Environmental Law Review, P-ISSN 2723-2484, Vo. 5 Issue 2, July

Tarmidi, D., Fitria, G. N., Purwaningsih, S. 2017. Tax Compliance: Impact of Implementation Online Tax Application (Empirical Study Tax Payer in KPP Madya Jakarta Timur). International Journal of Business, Economics and Law, Vol. 14, Issue 1 (December), ISSN 2289-1552

Villanueva, O. I., Ponce, V. G., Linan, F. S., Clemente, S. B., Carbonell, M. C., Sentiment Analysis of Tweets using Unsupervised Learning Techniques and the K-Means Algorithm, International Journal of Advanced Computer Science and Applications, Vol. 13, No. 6, 2022