Scientax
Jurnal Kajian Ilmiah Perpajakan Indonesia

djp

VOLUME 6 NO. 2 | APRIL 2025 Journal Page: ejurnal.pajak.go.id ISSN 2686-5718

Dealing with Last-Mile Analytics: Evidence from Indonesian Tax Administration through Practice Research

Agung Darono
Center for Tax Education and Training, Jakarta, Indonesia. Email: agungdarono@kemenkeu.go.id

* Corresponding author: agungdarono@kemenkeu.go.id

ABSTRACT

This study employs a "practice-as-research" approach to investigate the role of analytics as a pivotal component of Indonesia's tax administration system, specifically focusing on addressing tax gaps. Tax analytics systems and applications operate centrally to generate data that support several tax administration functions, including taxpayer registration, compliance monitoring, dispute resolution, and law enforcement. However, the tax officers—who serve as end-users of this data—frequently encounter the "last-mile problem", where the data provided is not immediately actionable. Consequently, such tax officers are often required to develop their own data pipelines to further process and analyze the data before it can be effectively used for decision-making. This study identifies two categories of last-mile issues: those that can be eliminated and those that can only be mitigated to a limited extent. Two key recommendations are proposed to address these challenges. First, existing analytics applications should enhance taxpayer profile data by integrating the most comprehensive analytics outcomes, including compliance risk profiling. This can be achieved by implementing a "reverse-ETL" approach to improve existing analytics applications, facilitating the seamless flow of processed data back into operational system data. Second, the study advocates for more flexible self-service analytics platform for scenarios where last-mile challenges are unavoidable. This could be an analytics sandbox or a data-as-a-product approach that leverages containerization to enable tax officers to process and analyze data independently. These recommendations aim to improve the efficiency and effectiveness of Indonesia's tax administration system by addressing the critical last-mile challenges faced by tax officers, thereby enhancing the overall utility of analytics in supporting tax-related decision-making processes.

Keywords: analytics, data pipeline, compliance risk, function, tax gaps

1. INTRODUCTION

Analytics has become an essential issue in the field of taxation, not only from the perspective of authorities and taxpayers but also from various think tank institutions, academics, consulting firms, and information technology vendors (IOTA, 2016; OECD, 2017; EY, 2017; Microsoft and PwC, 2017; CIAT, 2020; ADB, 2022; Deloitte, 2020; Pijnenburg, 2020; Haenen & Emens, 2021; KPMG, 2020; Pianko, 2018). For tax authorities, one of the main objectives of implementing analytics is to reduce the gap between the amount of tax that has been paid and the amount of tax that the taxpayer should pay. From the taxpayer's perspective, analytics are helpful for better tax management and planning. (Cleary, 2011; Martikainen, 2012; Wua et al., 2012; Wessels, 2014; Gregg, 2015; Stankeviciusa and Leonas, 2015; Pijnenburg et al., 2017; Pijnenburg, 2020).

The Directorate General of Taxes (DGT), the tax authority in Indonesia, has also determined that data analytics is one of the strategic initiatives

DOI: 10.52869/st.v6i2.592
Received: August 14, 2022; Revised: February 24, 2025; Accepted: March 13, 2025; Published: April 30, 2025

How to Cite:
Darono, A. (2025). Dealing with last-mile analytics: Evidence from Indonesian tax administration through practice research. *Scientax: Jurnal Kajian Ilmiah Perpajakan Indonesia*, *6*(2), 107–123. https://doi.org/10.52869/st.v6i2.592

to achieve its primary goal: sustainable tax compliance. In various situations, data analytics have been used to support multiple operational activities or to prepare and evaluate tax policies. (DJP, 2015; 2022). Referring to Sakti (2021) and Djuniardi (2016), end-users can use the data analysis results presented by the DGT's Directorate of Tax Data and Information (DIP) as the core function of data management in DGT using two endpoints. The endpoint is an application system called Approweb, which is accompanied by an information retrieval function. Data governance manages the mechanisms and data access rights inherent in each endpoint.

However, previous studies highlight that tax officers often require additional tools, such as data analysis applications, to fulfill their duties (Iqbalsah, 2023; Winata & Hadi, 2023; Rosid et al., 2022; Rosid, 2023; Prastuti & Lasmin, 2021; DAA, 2019). The available endpoints frequently do not meet their data needs, especially when tax supervision or administration tasks require data matching between internal application systems and external data sources. For instance, an account representative may need to compare tax returns with comparative data from an e-marketplace when supervising the taxpayer operating an online sales channel (DGT 2019). Similarly, a tax auditor must test accounting data in the form of a general ledger obtained from the taxpayer to validate the tax return data (DAA, 2019). The consequence of this situation is that the officer must develop his/her own techniques, procedures, and data analysis pipelines that may not be in line with the organization's data governance policy.

As a result, tax officers need to develop their own techniques, procedures, and data analysis pipelines, which may conflict with the organization's data governance policies (Darono & Pratama, 2022; Ilhamsyah, 2020; Pratama & Darono, 2022; TETC, 2024). This situation is often called the "last mile problem" in analytics, where there is a gap between the information presented at the end-point of an analytics platform and the decisions or actions taken by the users of that information (Brahm & Sherer, 2017; Brownlow, 2022). The consequence of this last mile gap is that the end user must take several additional steps to utilize the information so that they can make decisions or actions according to the tasks they are doing (Dykes, 2019; Earley, 2015; Verhulst et al., 2024). There are several efforts to overcome this last mile, including self-service analytics (or self-service business intelligence) (Darono, 2023; Davidson, 2023; Nentwich, 2022; WeAreNetflix, 2018), enriching data in the core system with data analytics results through the Reverse-ETL mechanism (Dash & Swayamsiddha, 2022; George, 2023; Manohar & Kline, 2024), or providing data to end users using the data-as-a-product approach (Blohm et al., 2024; Dehghani, 2021; Hasan & Legner, 2023).

The challenges associated with the last mile in tax administration motivate the author to explore ways to improve its management, ensuring the optimal implementation of tax analytics. This study uses the practice as a research approach (Candy, 2006; Gherardi, 2019; Bispo, 2015) to provide an in-depth understanding (Costantino, 2008; Bhattacharya, 2008; Fox, 2008; Nickerson, 2024) related to how tax officers deal with last-mile analytics situations that arise in data processing and analysis activities that must be carried out as part of their daily work. This study aims to contribute by (1) providing additional knowledge to understand last-mile problems that arise in analytical practices from a practice-led perspective; (2) providing recommendations from a practice-based perspective on techniques or tools to improve tax officers' effectiveness in addressing last-mile analytics issues.

The systematics of this paper are as follows. In the first section, author establishes the research background and articulate the significant contributions this study intends to make. The second section thoroughly review the existing literature and prior studies, creating an analytical framework to enhance understanding of the findings. Next, the third section delves into practice research, outlining it as an effective research methodology employed in this study. The fourth section presents and discusses the research findings, highlighting their implications. Finally, the fifth section synthesizes the insights gained, providing robust conclusions and actionable recommendations for future research.

2. THEORETICAL FRAMEWORK

Analytics is a framework of knowledge and practice related to organizational actions. It uses the data it manages to produce various insights to support decision-making and the actions of organizational actors (Davenport & Harris, 2007; Power et al., 2018; White & Imhoff, 2010). The term analytics has various variants because of differences in the emphasis on the purpose or process. However, the substance of the meaning remains the same, namely, gaining insights from the analyzed data. These terms include OLAP, business intelligence, data mining, and data science. OLAP (Online Analytical Processing) is a term that differentiates it from OLTP (Online Transaction Processing), where the relationship between the two is that data obtained from transaction processing should be further analyzed to gain insights. Along with OLAP, the term Business Intelligence (BI) also emerged, emphasizing business more than technology. In general, data analysis techniques are carried out by aggregation categorization without any models or forecasts (Codd et al., 1993; Chaudhuri & Dayal, 1997 ; Delen & Ram, 2018)

However, there is also data mining. This term expands the data analysis approach with associations, clustering, classification, estimation, and forecasting to predict what might happen and even provide prescriptions (recommendations) for actions should be taken based on the earlier prediction results. This approach is also known as machine learning because with the data that has been collected, the computer, as a machine, is taught to recognize patterns in the data that has been collected to be then used to perform associations and clustering, which is called unsupervised learning or classification, estimation, forecasting, which is called supervised learning. The approach that emphasizes the existence of predictive and prescriptive models is then known as data science (Gartner, 2014; Jin, 2017; Larose & Larose, 2015 ; Atwal, 2020; Donoho, 2017).

The term "data analytics" then emerged as an attempt to summarize all these understandings into a body of knowledge about data analysis. In understanding data analytics, the descriptive and diagnostic approaches are almost the same as the scope of OLAP and BI. In contrast, advanced analytics, including predictive and prescriptive analytics, has almost the same scope as data science (Aasheim et al., 2015; Tesch, 2020). However, for the scope of the study, Larose & Larose (2015) used the term "data mining." Figure 1 shows the relationships among the various terms described.From another perspective, some studies have developed a taxonomy of terms related to analytics to sharpen or clarify the position of analytics as a body of knowledge. Analytics can also be understood by dividing the data sources, the business functions it supports, and the actors. Based on the area or business function it supports,

Figure 1

Position of "analytics" term compared to its synonym or alternative terms.

Data analytics, data mining, and data science (Davenport, 2013; Larose & Larose, 2014; Gartner, 2014; Aasheim et al., 2015; Richardson et al., 2019)	Business intelligence (or OLAP)	Descriptive	aggregation, categorization, recapitulation, cross-tabulation
		Diagnostics	
	Data science (or machine learning)	Predictive	Unsupervised learning: clustering, correlation, association Supervised learning: classification, estimation, forecasting
		Prescriptive	

Note. Processed by Authors

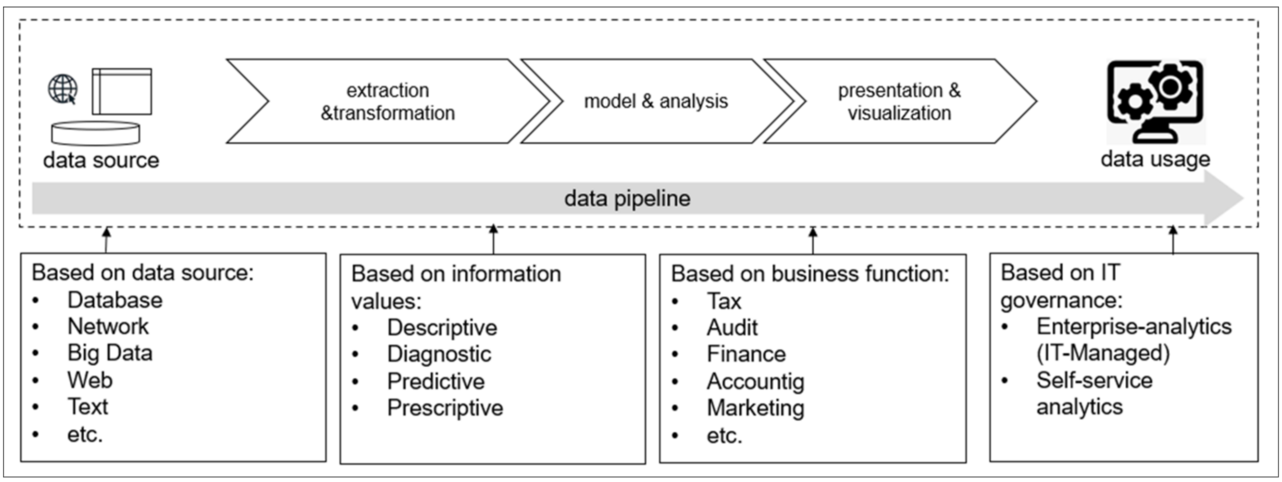
it is known by several terms, such as accounting analytics, audit analytics, marketing analytics, and human resource analytics. From the actor perspective, it is known by several terms: enterprise-analytics (also called corporate-analytics or IT-managed analytics) and self-service analytics or end-user analytics as an analytical mechanism developed independently by the business line/end-user (Chen et al., 2012; Gartner, 2014; Richardson et al., 2019; Duan & Xiong, 2015; Schroeck et al., 2015; SAS, 2016; Delen and Ram, 2018).

Other considerations for understanding analytics include the stages of work, data flows, and devices or technology stacks required. Hotz (2022; 2023) stated that de facto CRISP-DM is the most widely used framework for developing and implementing analytics in organizations. The development and implementation framework also causes terms such as data pipelines and data stacks. These two aspects are part of the development and implementation of analytics, explaining how data flow from source for user benefits and what technological support is needed to enable data flow (Raj et al., 2020). The primary purpose of each stage in the data pipeline is to form a virtual pipe in which data flows, originating from the data source, through various extract-transform-load stages until it reaches a point where users will use the data as insights to support decision-making or action (Microsoft, 2010; Hevo, 2022; Kitching et al., 2021). All stages of data flow require data technology stack support in the form of a combination of devices suitable for each data processing and analysis stage. In this context, ideal means suitability for a given price. Figure 2 shows the relationship between various analytics, BI, and data mining terms.

Globally, tax authorities stated that they must embrace analytics as a function of overall tax administration. (IOTA, 2016; OECD, 2017; CIAT, 2020). The forum held by the OECD (2016, 2017) forum directly mentioned the function of advanced analytics (or data science in Tesch (2020) terms) as part of tax administration. Advanced analytics demonstrate how tax authorities will embrace such analytics. It is not just aggregation, it will lead to classification, estimation, or forecasting. The OECD (2016) proposed the term "advanced analytics" to represent predictive and prescriptive analytics as part of data processing and analysis for the benefit of tax administration.

The report emphasizes that advanced analytics have been introduced previously. This term is more directed at reducing human judgment by relying more on the results of data analysis to determine various decisions in tax administration (ADB, 2022)). In the context of tax

Figure 2
Flow of work and types of analytics.



Note. Processed by Authors

administration, tax data analytics is a series of techniques for obtaining information to support decision-making or reviewing tax policies in various forms, such as the selection of audited taxpayers, management of tax receivables, improving service quality or evaluating the impact of different tax policies that have been set (OECD 2016). Deloitte (Deloitte, 2020) emphasized the use of the term "tax analytics." It defines a combination of technical knowledge of taxation and advanced information technology to identify patterns and anomalies to support organizations in managing strategies related to the impact of taxes on overall organizational performance. Data analysis techniques include pattern recognition, outlier detection, cluster analysis, experimental design, network analysis, and text mining. From a different perspective, ANAO (2008) stated that data analytics in tax administration includes all computer-based methodologies that enable tax authorities to handle large amounts of data from various data sources, make comparisons, present relationship patterns, or construct functional or hypothetical relationships between data.

Furthermore, data analytics is divided into two parts: (1) basic analytics, which includes data exploration and aggregation and statistical profiling and analysis; and (2) advanced analytics, which includes the use of data mining technology to find and develop a model and decision making. In practice, all organizations implementing analytics, including tax authorities, must determine the technology stack and implementation methodology (Altexsoft, 2023; Brunthaler, 2022; Vago, 2022). The technology stack and methodology choice will instantiate a data pipeline that will flow data from its source to be processed, analyzed, and then used by its users (Alley, 2019; DiCostanzo et al., 2023; Raj et al., 2020). The available data pipeline should meet user requirements as much as possible. However, it is not that easy because user needs are dynamic. Some companies, referred to as "data-driven nirvana" (WeAreNetflix, 2018; Thusoo & Sarma, 2017), provide sufficient space for users to perform self-service analytics.

This situation triggers the emergence of "last mile analytics problem". There is a "distance" between the results of the data analysis process produced at the enterprise level, which at the time of use wants to obtain the insights such user needs immediately; it turns out that it still requires further processing and analysis steps according to the needs of the user. This last-mile gap is almost impossible to eliminate. It will be tough to meet all users' very dynamic data needs. What must be done is to identify the gaps that occur and then the data analytics methodology solutions needed to overcome these gaps (Earley, 2015; Bridgewater, 2018; Dykes, 2019; Nentwich, 2022; Brownlow, 2022).

Such a gap is also found in the implementation of analytics at the DGT. This can be identified from several documents and study results. The DGT (2019) paper presents a last-mile problem related to e-commerce. A similar situation is also found in the document presented by DAA (DAA, 2019) that tax auditors must compare taxpayer report data submitted via the DGT online application with data obtained by tax auditors when conducting fieldwork plus various data obtained by third-party DGT, including AEoI data with tax authorities in other countries. The DAC (DAC, 2015) report highlights numerous instances where users must build their own data pipelines and data stacks, as the standard pipelines available fail to comprehensively meet their data needs.

3. METHODOLOGY

The research design is a series of research strategies and methods explicitly applied to one study (Creswell, 2013). The design set for this study was a practice-as-research. This choice was made by considering the purpose of the study, which is to understand in depth (verstehen) the internal perspective of the actor organization in the research locus. The new knowledge expected from practice-as-research is not through surveys or interviews but rather through hands-on experience with everyday organizational artefacts (Candy,

Table 1

Community of practice as practice-as-research loci

Training programs as a basis for establishing a community of practice (CoP)	The last mile refers to the data analysis process of completing a project or task.	Code for CoP
Training of tax data analysts and technical assistance toward tax compliance supervision	how officers in the Data Quality Assurance Section require data pipelines to provide the data required by all sections at the tax office (KPP)	CoP-1
Training for digital economy tax compliance supervision	The supervision carried out by account representatives requires data matching and analysis tools obtained from internal or external sources, including e-marketplaces.	CoP-2
Training for e-Audit (basic, intermediate, advanced)	Every tax auditor needs a data pipeline to compare tax return data with audited taxpayer financial and other related data to produce a tax audit report.	CoP-3
Training for digital economy tax audits	Every tax auditor needs a data pipeline to compare tax return data with e-commerce merchants as audited taxpayer financial and related data to produce a tax audit report.	CoP-4
Training of tax digital forensics (basic, intermediate)	How tax officers, as digital forensics specialists, develop data pipelines to analyze data obtained from fieldwork.	CoP-5

Note. Processed by Authors

2006; Schrag, 2019; Gherardi, 2019; Douglas et al., 2000; Bulley & Şahin, 2021). Candy (2006) identified two aspects of practice research: practice-based and practice-led research. The research is considered practice-based when a creative artifact forms a contribution to knowledge. It is deemed practice-led if it primarily generates new insights about practice. Hoveid and Hoveid (2007) stated that practice is one method of gaining understanding so that it ultimately produces (a framework) of knowledge related to the field being practiced.

The practice areas are identified during detailed discussions of training materials that require data analysis practices to resolve last-mile problems in tax analytics. The goal is to develop a deep understanding and present it as a framework of knowledge expected to serve as an additional reference in discussions and practices related to tax data analytics. The research focuses on the community of practice (CoP) formed through training activities and technical guidance at the Indonesian Tax Education and Training Center, where the author participated as a training program officer (Error! Reference source not found.. Data collected using documentation studies, interviews, and case study analyses presented during various training sessions (Bohnsack, 2014; Coffey, 2014; Bowen, 2009; Erich et al., 2017; Erickson et al., 2016; Darono, 2016).

4. RESULTS AND DISCUSSIONS

The results and discussion section will be structured using the following flow: First is the case context related to how the last mile in tax data analytics occurs. This section outlines the stages of the work, data flow, and technology support. Based on this review, the last mile related to a particular business function is identified. Based on the identified last mile, this study compiles a practice containing steps for developing a data pipeline. The purpose of this development is to (1) gain an in-depth understanding of the research problem and (2) find alternative references that are expected to be used to overcome the last-mile problem that occurs up to a permanent solution in the form of a data flow that can be used directly without users having to develop their data pipeline.

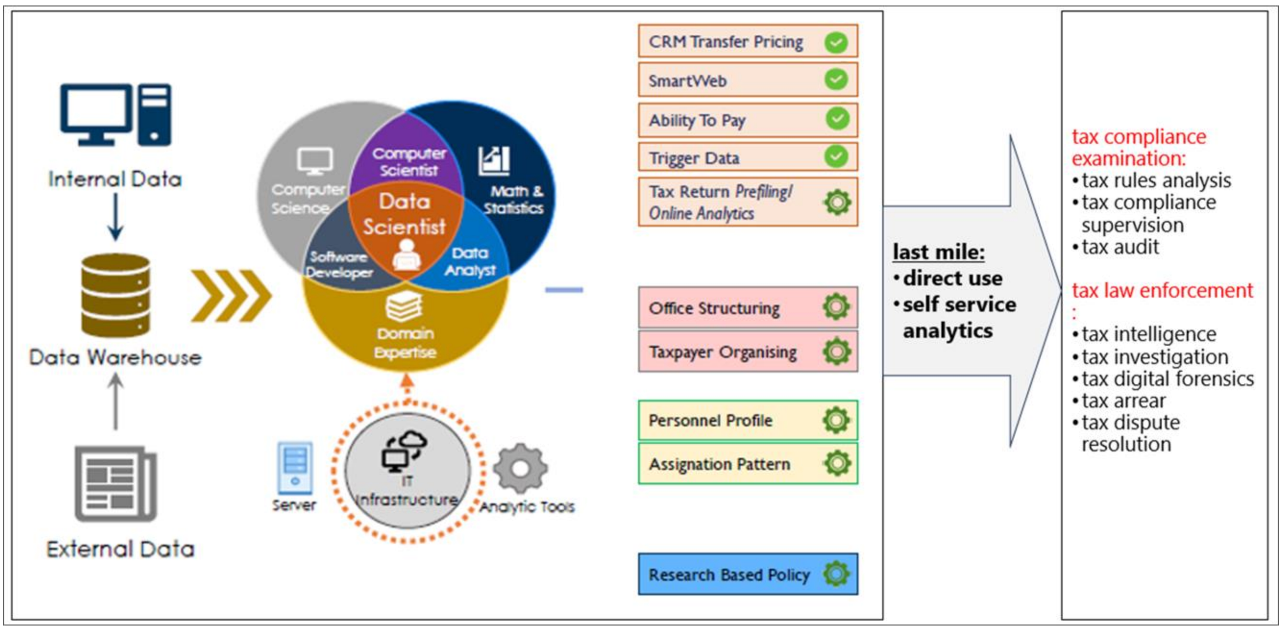
4.1 Context of the case: How does the last mile exist in the Indonesian tax administration data exist?

The tax collection system in Indonesia, as explained in Article 12 of the Tax General Provision and Procedure Law, uses self-assessment. Taxpayers are required to report the implementation of rights and obligations through a Tax Return. It can be re-calculated if the tax authority has evidence that the tax return (SPT) needs to be corrected. The re-calculation can be through an appeal to improve the tax return (SPT) or law enforcement actions broadly, starting from research, examination, and collection to investigation of tax crimes (Darono & Nuruliman, 2017; Deloitte, 2023)). Analytics used at all stages using CRM where taxpayers are based on their risk profile for non-compliance (DGT, 2022; Ilhamsyah, 2020; Sakti, 2021). In this case, DGT (2022) preferred to use the term Business Intelligence (BI) in implementing analytics in the CRM. Thus, the use of BI and analytics in the context of this CRM replaced each other.

Actions that can be taken against taxpayers as referred to in the Regulation of the Minister of Finance Number 132 of 2023 concerning Implementation Guidelines and Technical Guidelines for Functional Positions in the State Finance Sector in conjunction with Minister of Finance's Regulation Number 131/PMK.03/2022 concerning Implementation Guidelines for Tax Auditor Functional Positions related to the core functions of tax administration, namely compliance supervision and law enforcement. At this point, the last mile appears because there are several additional steps that users of their data information must take to use the data to complete their tasks. Note that Figure 3 describes the tax data analytics workflow of the tax administration function for compliance supervision and tax law enforcement.

In this last-mile stage, there are two situations. First, the available data can be utilized directly. Second, the user must take several steps to perform self-service analytics (SSA) or self-service BI (SSBI). In this case, the author can identify at least two types of SSA: simple and complex.

Figure 3
Tax data analytics workflow (adapted from Sakti 2021; DJP 2022)



Note. Processed by Authors

Simple SSA can occur when the analytics include only a few diagnostic steps to fill in missing columns or match data. Complex SSA must form a relatively long data pipeline with several stages or even all stages of the data analytics flow from the data source, extract, transform, load, and data analysis to finally be used using data stacks that cover many types of devices.

Furthermore, this analysis will take a "practice in research" approach (Hoveid & Hoveid, 2007), where the author, based on data from the research locus (see Table 1), performs the development of a data pipeline according to the identified data analysis use case. Users can also implement this data pipeline in their daily work by adjusting several settings, such as data sources, transformation steps, and data analysis techniques. Based on the author's observations from the existing research locus, the practices that will be carried out include the following:

(1) Use of "e-Faktur" data for tax compliance;
(2) Matching third-party data (for example, e-marketplace) with master file data, VAT, and income tax return;
(3) Developing a data pipeline for e-Audit and digital tax forensics.

The selection of the third-use case was based on the availability of dummy data that researchers could access and the magnitude of benefits that could be obtained if the practice could be implemented in a real organizational situation, especially in the context of this study, namely how supporting tax administration can reduce tax gaps. The following description describes our attempt to form a knowledge framework through practices related to overcoming the last-mile problem by conducting self-service analytics (self-service BI). For this purpose, the author only uses dummy data for research and not actual data.

4.2 Practice 1: Tax compliance supervision strategy - use case of "e-Faktur" data

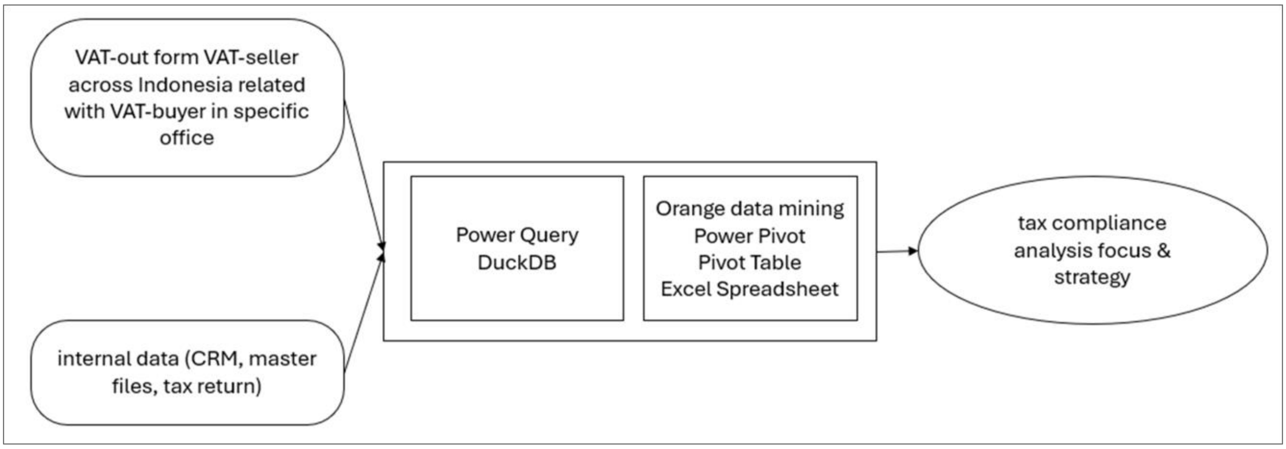
In the current Indonesian tax self-assessment system, two main functions are to determine whether taxpayers are compliant: compliance testing and law enforcement. One step in compliance testing is supervision. Related to this supervision, there are already instructions for its implementation through the Circular of the Director General of Taxes Number SE-05/PJ/2022 concerning Supervision of Taxpayer Compliance (hereinafter referred to as SE-05). Referring to the provisions in this SE-05, it can be concluded that supervisory actions are highly dependent on tax data analysis activities, namely, analysis activities to identify modes of non-compliance that arise and estimates of potential tax obligations that have not been met and then determine recommendations for follow-up actions to support the implementation of supervision. One step in the analysis is to determine the focus on the sector that will be the target of the analysis implementation: manufacturing, trade, services, natural resources, government spending, the digital economy, or other sectors. This means that the analysis will use sectoral data, which will later lead to a more detailed analysis of taxpayer compliance in the sector. Of course, this focus will differ from one tax office to another.

One of the internal data that is quite complete in determining the focus of analysis is the "e-Faktur" data generated by the DGT's tax invoice application system. The tax regime for goods and services in Indonesia is the value-added tax, and the tax credit system is based on tax invoices. This system requires VAT taxpayers to issue invoices when making taxable deliveries. For sellers, it is VAT-out, and buyers are VAT-in. Every transaction must be made using an electronic tax invoice. In addition to VAT compliance in determining how much VAT is underpaid or overpaid due to the VAT-in vs. VAT-out mechanism, these e-invoice data can be used to determine the focus of tax supervision. The "e-Faktur" data itself contains quite complete data, including the identity of the seller and buyer (Taxpayer identification number/TIN, name, address), amount of VAT paid, and description of goods or services as VAT objects. Based on the data structure, the author in Practice-1, as in CoP-1, carried out the following practices:

(1) diagnostic analytics in the form of a recapitulation of the amount and number of

Figure 4

Tax compliance supervision strategy: use case of "e-Faktur" data



Note. Processed by Authors

taxable goods or taxable services per sector and address to be compared with VAT-return data and income tax returns;

(2) diagnostic analytics in the form of a recapitulation of the amount and number of taxable goods or taxable services per taxpayer reported by VAT-sellers from all over Indonesia to VAT-buyers to be compared with VAT-out reported by VAT-buyers or sales subject to income tax by taxpayers registered at the tax office where the VAT-out taxpayer is registered;

(3) diagnostic analytics in the form of a recapitulation of the amount and number of taxable goods or taxable services per taxpayer reported by VAT-sellers from all over Indonesia to VAT-buyers to be compared with VAT-out (in VAT-return) reported by VAT-buyers registered at a particular tax office;

(4) Predictive analytics with time series analysis to forecast taxable deliveries per sector and region;

(5) Supply chain analysis determines the distribution layer of goods and benchmarks the price of goods and services per sector per region.

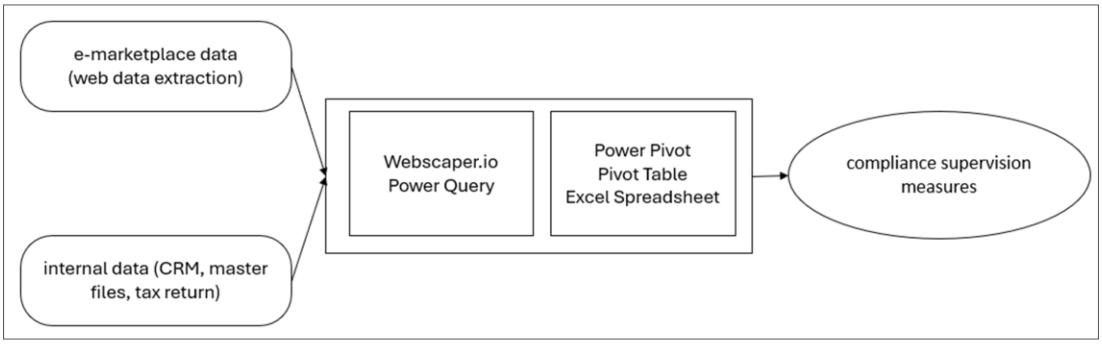
To perform Practice-1, the developed data pipeline and stacks are illustrated in Figure 4.

4.3 Practice 2: Data matching in tax compliance supervision

Practice 2, as previously described, emphasizes the "micro-analytics" aspect as a follow-up of data found at the macro level directed at individual taxpayer compliance. The use case in this practice is to search for taxpayer data from various sources. However, it is not equipped with a unique identity such as TIN or "NIK" (identity number in Indonesian citizen administration). This data matching technique is an old technique that still has to be carried out by many tax authorities (ATO, 2023; HMRC, 2023; IRS, 2020; OECD, 2018) because the comparative data obtained do not necessarily have unique identity attributes that make it easier to match them with tax reporting data held by the tax authority.

The use case in Practice 2 comes from discussions in CoP-1 and CoP-2, which raise the problem of how to match data obtained from several sources between e-marketplaces or ownership of movable assets such as motor vehicles obtained by tax authorities without unique identity attributes that can be directly matched with reporting data that taxpayers have submitted. The feature used in this data matching action is a fuzzy match that allows matching between non-unique attributes such as name and address, with a

Figure 5
Data pipeline and stack for data matching with e-marketplace data for the use case.



Note. Processed by Authors

particular match threshold. Thus, data that previously did not have unique attributes can be determined by TIN or NIK to test its tax compliance. The pipeline and stack data developed for Practice 2 are shown in Figure 5.

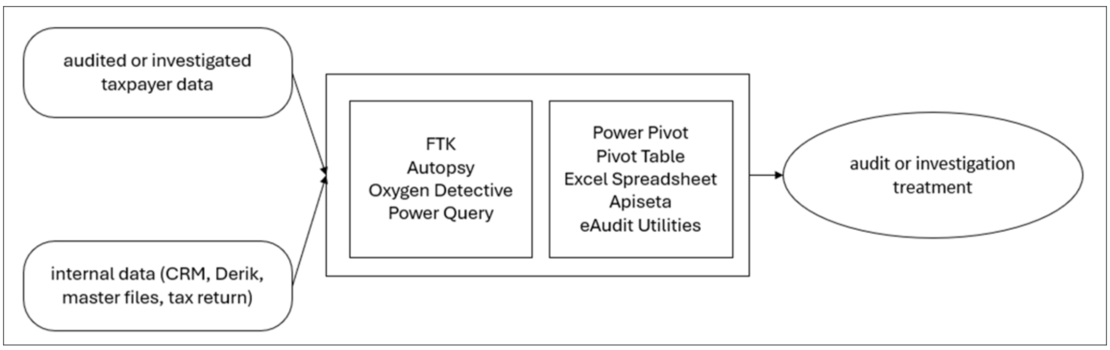
4.4 Practice 3: Data extraction and analysis in e-Audit and tax digital forensics

Procedurally, compliance testing actions in the form of supervision can be upgraded to a tax audit. Even if there are indications of tax fraud, the case can be upgraded to an investigation into alleged tax crimes. The DGT has issued several provisions related to the procedures for handling evidence in the form of electronic data in both tax audits and tax criminal investigations, namely:

(1) SE-36/PJ/2017 on Digital Forensic Guidelines for Taxation Interests;
(2) SE-25/PJ/2013 concerning the e-Audit Guidelines;
(3) SE-10/PJ/2017 concerning Technical Instructions for Field Examination in the context of Examination to Test Compliance with Tax Obligations;
(4) SE-1/PJ/2024 concerning Technical Instructions for Examination of Initial Evidence of Tax Crimes.

As discussed in CoP-3, CoP-4, and CoP-5, the last mile in tax audits or investigations can be considered "inherent" or “native”—an unavoidable aspect of the process. This is because auditors or investigators inevitably need to develop their own data pipelines to manage ongoing cases effectively. The data pipeline is created to conduct audit tests so that conclusions related to compliance testing activities or law enforcement can be drawn. Regarding the data stack, e-Audit practitioners have developed their in-house devices, the Apiseta application and e-Audit Utilities. To run Practice 3, the developed data pipeline is represented in Figure 6.

Figure 6
e-Audit and tax fraud forensics data pipeline and stack



Note. Processed by Authors

4.5 Reflection on the practice

DGT's analytics operate centrally from its head office by the Directorate of Information and Communication Technology and the Directorate of Tax Data and Information to produce data ("insight") that can be used for the benefit of implementing every tax function, from registration, compliance monitoring, dispute handling to law enforcement. This means that outside the use cases outlined above, many other analytics use cases to support related functions. Practices 1-3 were chosen because they were found by the author in the CoP that was formed (see Table 1 again).

The development of an analytical platform seeks to ensure that at the end (last mile) of the data analysis results, there are no gaps that cause users to require additional steps in utilizing information to support decision-making or action. In this regard, several techniques and procedures have been proposed, including the following:

(1) reversed-ETL (Manohar & Kline, 2024): completing taxpayer profile data with risk analysis results (CRM) that are already attached to each taxpayer profile record so that the need for additional analysis that requires simple SSA/SSBI will be reduced;
(2) analytics sandbox or analytics-as-a-service (Russom et al., 2010): the DGT server provides a container through specific governance procedures as a robust computing environment to perform data analysis with large amounts of data accompanied by complex computing;
(3) data-as-a-product (Blohm et al., 2024); (Hasan & Legner, 2023): creating a "data-product" that is stored as a container equipped with a log feature for data security, where all data related to the audit or law enforcement assignment is contained, which can easily be downloaded for tax inspectors so that they no longer need to scrape data using applications Apiseta or e-Audit Utilities.

The last mile should be addressed using standard menus and reverse-ETL processes, eliminating the need for users to develop their own data pipelines. The "inherent" or "native" nature of the last mile is ad hoc, as it involves specific types of analyses tailored to unique needs, business units, or functions. However, any approach to resolving last-mile issues must adhere to data governance principles outlined in KEP-215/PJ/2021 on Data Governance at DGT.

5. CONCLUSION

The DGT has determined that data analytics is a strategic step toward sustainable tax compliance. Data analytics have been used to support various operational activities or to prepare and evaluate tax policies to minimize tax gaps. The purpose of implementing analytics is to provide insights that can be used to support decisions and actions in tax administration.

A key challenge in tax data analytics lies in the last-mile stage, where analysis results cannot be directly utilized. It forces end users to build their own data pipelines and stacks. This last-mile problem can be solved by improving the analysis process using "simple SSA" and "complex SSA" approaches.

Simple SSA refers to a situation where tax officers develop their analytic pipeline as little as possible. Tax officers directly obtain data according to their work by their function as end users without taking additional steps or their data pipeline. This can be achieved through a reverse ELT approach, where data analysis output (for example, CRM results) are attached into each taxpayer profile record.

On the other hand, complex SSA arises from the inherent nature of business processes and data flows, which often require tax officers to build and manage their own data pipelines as the "last mile". This enables them to process data from various sources to generate insights relevant to their tasks. Organizations can adopt a data-as-a-product or analytics-as-a-service approach to address the challenges posed by this complexity. In addition to simplifying the data pipeline, these two approaches allocate computing resources efficiently required for analytics. The analytics process is performed in an environment with more powerful computing power, and the implementation of optimal data governance is guaranteed.

This study is limited using dummy data as input data for practice-as-research. While this limitation is unavoidable due to data governance constraints, it represents the best possible approach under the circumstances. The implication is that the results of the practices must first be adjusted to the actual situation in the work unit that conducts data analysis. The expected contribution of the results of this research on a broader scale is the availability of a conceptual framework that can be used to understand how the last mile can occur in the implementation of data analytics within public administration and government organizations, along with ways to overcome these problems.

REFERENCES

Aasheim, C. L., Williams, S., Rutner, P., & Gardiner, A. (2015). Data analytics vs. data science: A study of similarities and differences in undergraduate programs based on course descriptions. *Journal of Information Systems Education*, *26*, 103–115.

Asian Development Bank. (2022). *Launching a digital tax administration transformation—What you need to know*. https://www.adb.org/sites/default/files/publication/792586/digital-tax-administration-transformation.pdf

Alley, G. (2019). Data integration vs. data pipeline: What's the difference? https://dzone.com/articles/data-integration-vs-data-pipeline-whats-the-differ

Altexsoft. (2023). *The modern data stack: What it is, how it works, use cases, and ways to implement*. https://www.altexsoft.com/blog/modern-data-stack/

Australian Taxation Office. (2023). *How we use data matching*. https://www.ato.gov.au/About-ATO/Commitments-and-reporting/In-detail/Privacy-and-information-gathering/How-we-use-data-matching/

Atwal, H. (2020). *Practical DataOps—Delivering agile data science at scale*. Springer. https://doi.org/10.1007/978-1-4842-5104-1

Bhattacharya, H. (2008). Interpretive research. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 467–474). SAGE Publications.

Bispo, M. (2015). Methodological reflections on practice-based research in organization studies. *BAR - Brazilian Administration Review*, *12*(3), 309–323. https://doi.org/10.1590/1807-7692bar2015150026

Blohm, I., Wortmann, F., Legner, C., & Köbler, F. (2024). Data products, data mesh, and data fabric. *Business & Information Systems Engineering*. https://doi.org/10.1007/s12599-024-00876-5

Bohnsack, R. (2014). Documentary method. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 217–232). SAGE Publications.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27–40.

Brahm, C., & Sherer, L. (2017). *Closing the results gap in advanced analytics: Lessons from the front lines*. Bain & Company, Inc. https://www.bain.com/insights/closing-the-results-gap-in-advanced-analytics-lessons-from-the-front-lines/

Bridgwater, A. (2018, August 17). Why business fails to travel the "last mile" of analytics. *Forbes*. https://www.forbes.com/sites/adrianbridgwater/2018/08/17/why-business-fails-to-travel-the-last-mile-of-analytics/

Brownlow, T. (2022). The who, what & how of the last mile of analytics—Re-framing the problem you're probably already trying to solve. *Medium*. https://medium.com/geekculture/the-who-what-how-of-the-last-mile-of-analytics-693c860e4ac8

Brunthaler, M. (2022). What is a modern data stack? *Adverity*. https://www.adverity.com/blog/what-is-a-modern-data-stack

Bulley, J., & Şahin, Ö. (2021). *Practice research—Report 1: What is practice research? And Report 2: How can practice research be shared?* UKRI (Research England).

Candy, L. (2006). *Practice based research: A guide* (Creativity and Cognition Studios Report No. 1). Creativity and Cognition Studios.

Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM SIGMOD Record*, *26*(1), 65–74.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165–1188.

CIAT. (2020). *ICT as a strategic tool to leapfrog the efficiency of tax administrations*. Inter-American Center of Tax Administrations (CIAT).

Cleary, D. (2011). Predictive analytics in the public sector: Using data mining to assist better target selection for audit. *Electronic Journal of E-Government*, *9*(2), 132–140.

Codd, E. F., Codd, S. B., & Salley, C. T. (1993). *Providing OLAP to user-analysts: An IT mandate*. E.F. Codd Associates.

Coffey, A. (2014). Analysing documents. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 367–376). SAGE Publications.

Costantino, T. E. (2008). Constructivism. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 116–119). SAGE Publications.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications.

Directorate of Audit and Arrear (DAA) – DGT. (2019). *Pedoman umum e-audit Direktorat Jenderal Pajak (DGT e-audit's general guidelines)*.

Directorate of Audit and Arrear (DAC) – DGT. (2015). *E-audit utilities manual guide*.

Darono, A. (2016). Pembelajaran business analytics dan big data dalam pendidikan ekonomi dan bisnis. In *National Seminar on Accounting and Finance 2016* (pp. 8–16).

Darono, A. (2023). DataOps dalam analitika data keuangan negara: Studi eksploratif. *Indonesian Treasury Review: Jurnal Perbendaharaan, Keuangan Negara dan Kebijakan Publik*, *8*(2), 125–136.

Darono, A., & Nuruliman, A. (2017). Tax compliance monitoring as a discursive practice: Evidence from Indonesia. In *Tax Research Network 26th Annual Conference 2017*, Bournemouth.

Darono, A., & Pratama, A. (2022). Tax audit in the era of big data: The case of Indonesia. *Journal of Tax Administration*, *7*(2), 27–50.

Dash, B., & Swayamsiddha, S. (2022, December). Reverse ETL for improved scalability, observability, and performance of modern operational analytics: A comparative review. *2022 International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (OCIT)*. https://doi.org/10.1109/OCIT56763.2022.00097

Davenport, T. H. (2013). Introduction: The new world of enterprise analytics. In T. H. Davenport (Ed.), *Enterprise analytics: Optimize performance, process, and decisions through big data*. Pearson Education.

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Review Press.

Davidson, S. (2023). A brief history of business intelligence, modern day data transformation and building trust along the way. *Hashboard*. https://hashboard.com/blog/a-brief-history-of-business-intelligence-modern-day-data-transformation-and-building-trust-along-the-way

Dehghani, Z. (2021). *Data mesh*. O'Reilly Media.

Delen, D., & Ram, S. (2018). Research challenges and opportunities in business analytics. *Journal of Business Analytics*, *1*(1), 2–12. https://doi.org/10.1080/2573234X.2018.1507324

Deloitte. (2020). *My data is a mess. Is analytics out of reach?* Deloitte Tax LLP.

Deloitte. (2023). *Indonesia tax guide 2023–2024*. Deloitte Touche Solutions. https://www2.deloitte.com/content/dam/Deloitte/id/Documents/tax/id-tax-guide-2023-2024.pdf

Directorate General of Taxes (DGT). (2019). *Economy digital audit guidelines (Modul penanganan dan pemeriksaan transaksi ekonomi digital)*.

Directorate General of Taxes (DGT). (2022). *CRMBI: Langkah awal menuju data driven organization*. Direktorat Jenderal Pajak (DJP).

DiCostanzo, D. J., Ayan, A. S., Jhawar, S. R., Allen, T. T., & Patterson, E. S. (2023). Machine learning data pipeline for the democratization of AI. *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care, 12*(1), 120–124. https://doi.org/10.1177/2327857923121029

Direktorat Jenderal Pajak (DJP). (2015). *Cetak biru teknologi informasi dan komunikasi Direktorat Jenderal Pajak Kementerian Keuangan Republik Indonesia tahun 2015–2019*.

Direktorat Jenderal Pajak (DJP). (2022). *CRMBI: Langkah awal menuju data driven organization*.

Djuniardi, I. (2016). Next generation data analysis: The implementation of big data in Directorate General of Taxes Republic of Indonesia. *13th ATAIC Annual Technical Conference*, Melaka, Malaysia. http://ataic13.hasil.gov.my/SLOT3/IndonesiaDGTBigData.pdf

Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745–766. https://doi.org/10.1080/10618600.2017.1384734

Douglas, A., Scopa, K., & Gray, C. (2000). Research through practice: Positioning the practitioner as researcher. *Working Papers in Art & Design*.

Dykes, B. (2019, December 18). Two keys to conquering the last mile in analytics. *Forbes*. https://www.forbes.com/sites/brentdykes/2019/12/18/two-keys-to-conquering-the-last-mile-in-analytics/?sh=18bf744094fb

Earley, S. (2015). Visualisation: The last mile of analytics. *Applied Marketing Analytics*, *1*(4), 296–298.

Erich, F. M. A., Amrit, C., & Daneva, M. (2017). A qualitative study of DevOps usage in practice. *Journal of Software: Evolution and Process*, *29*(6), Article e1885. https://doi.org/10.1002/smr.1885

Erickson, A., Noonan, P., Brussow, J., & Carter, K. (2016). Measuring the quality of professional development training. *Professional Development in Education*, *43*(1), 1–4. https://doi.org/10.1080/19415257.2016.1179665

Ernst & Young (EY). (2017). *The tax authority of the future: How tax authorities are using analytics to deliver new levels of value*.

Fox, N. J. (2008). Postpositivism. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 659–664). SAGE Publications.

Gartner. (2014, October 21). Gartner says advanced analytics is a top business priority. *Gartner Inc.* https://www.gartner.com/en/newsroom/press-releases/2014-10-21-gartner-says-advanced-analytics-is-a-top-business-priority

Gelvanovska-Garcia, N., Mačiulė, V., & Rossotto, C. M. (2024). *Advancing cloud and data infrastructure markets: Strategic directions for low- and middle-income countries*. World Bank.

George, J. (2023). Harnessing the power of real-time analytics and reverse ETL: Strategies for unlocking data-driven insights and enhancing decision-making. *World Journal of Advanced Engineering Technology and Sciences*, *10*(1), 190–201. https://doi.org/10.30574/wjaets.2023.10.1.0247

Gherardi, S. (2019). Practice as a collective and knowledgeable doing. In *Medien der Kooperation*. https://doi.org/10.25969/mediarep/12641

Gregg, R. (2015). Passion for change and good data: The right combination for the DATA Act. *Journal of Government Financial Management*, *64*(2), 48–52.

Haenen, R., & Emens, R. (2021). Squaring away the total cost of tax compliance using SAP S/4 HANA. *Compact – KPMG Netherlands*. https://www.compact.nl/articles/squaring-away-the-total-cost-of-tax-compliance-using-sap-s-4-hana/

Hasan, M. R., & Legner, C. (2023). *Understanding data products: Motivations, definition, and categories*.

Hevo. (2022). *Pipelines*. Hevo Data Inc. https://docs.hevodata.com/pipelines/

HM Revenue & Customs (HMRC). (2023). *The tax administration framework review – information and data*. https://www.gov.uk/government/consultations/the-tax-administration-framework-review-information-and-data/the-tax-administration-framework-review-information-and-data#third-party-information-and-data-gathering-powers-and-taxpayer-safeguards

Hotz, N. (2022). What is a data science life cycle? *Data Science Process Alliance*. https://www.datascience-pm.com/data-science-life-cycle/

Hotz, N. (2023). *What is CRISP DM?* Data Science Process Alliance. https://www.datascience-pm.com/crisp-dm-2

Hoveid, M., & Hoveid, H. (2007). Research in, on or with practice? In *Education and the spirit of time* (pp. 43–56). https://doi.org/10.1163/9789087903190_005

Ilhamsyah. (2020). Mekanisme pengujian pengawasan kepatuhan wajib pajak (Studi kasus Kantor Pelayanan Pajak Perusahaan Masuk Bursa). *Simposium Nasional Keuangan Negara*, *2*(1). https://jurnal.bppk.kemenkeu.go.id/snkn/article/view/592

Intra-European Organisation of Tax Administrations (IOTA). (2016). *Applying data and analytics in tax administration (IOTA good practice guide)*. https://www.iota-tax.org/good-practice-guide-applying-data-and-analytics-tax-administrations

Iqbalsah, R. (2023). Machine learning: Classifying taxpayer's supervising zone based on the street address using Natural Language Processing algorithm. *Scientax*, *4*, 233–242. https://doi.org/10.52869/st.v4i2.486

Internal Revenue Service (IRS). (2020). *Improve the filing process*. https://www.taxpayeradvocate.irs.gov/wp-content/uploads/2020/07/ARC18_PurpleBook_02_ImproveFiling.pdf

Jin, J. (2017). OLAP and machine learning. *Encyclopedia with Semantic Computing and Robotic Intelligence*, *1*, 1630019. https://doi.org/10.1142/S2425038416300196

Kitching, K. A., Coble, E., & Phillips, A. (2021). DATA Act dashboard: An instructional case using data visualization. *Journal of Governmental & Nonprofit Accounting*, *10*(1), 157–174. https://doi.org/10.2308/JOGNA-2021-004

KPMG. (2020). *Data and indirect tax*. KPMG International Cooperative. https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2020/02/Indirect-tax-webcast-web.pdf

Larose, D., & Larose, C. (2015). *Data mining and predictive analytics* (2nd ed.). Wiley.

Larose, D. T., & Larose, C. D. (2014). Discovering knowledge in data: An introduction to data mining. In *Discovering knowledge in data* (pp. 1–15). https://doi.org/10.1002/9781118874059.ch1

Manohar, T., & Kline, L. (2024). *What is reverse ETL? The definitive guide*. Hightouch. https://hightouch.com/blog/reverse-etl

Martikainen, J. (2012). *Data mining in tax administration—Using analytics to enhance tax compliance* [Master's thesis, Aalto University School of Business].

Microsoft. (2010). *Microsoft EDW architecture, guidance and deployment best practices*. Microsoft Corp. https://docs.microsoft.com/en-us/previous-versions/sql/sql-server-2008-r2/hh146875(v=msdn.10)

Microsoft, & PwC. (2017). *Digital transformation of tax administration*. Microsoft and PricewaterhouseCoopers Belastingadviseurs N.V.

Nentwich, C. (2022). Now is the time to conquer the last mile to full data automation. *Journal of Securities Operations & Custody, 14*. https://doi.org/10.69554/DKHB7905

Nickerson, C. (2024). *Verstehen in sociology: Empathetic understanding (Sociology)*. Simply Psychology. https://www.simplypsychology.org/verstehen.html

Organisation for Economic Cooperation and Development (OECD). (2016). *Advanced analytics for better tax administration: Putting data to work*. https://read.oecd-ilibrary.org/taxation/advanced-analytics-for-better-tax-administration_9789264256453-en#page1

Organisation for Economic Cooperation and Development (OECD). (2017). "Advanced analytics." In *Tax administration 2017: Comparative information on OECD and other advanced and emerging economies*.

Organisation for Economic Cooperation and Development (OECD). (2018). *Implications of the digital transformation for the business sector—Summary*. Conference on Implications of the Digital Transformation for the Business Sector, London. http://www.oecd.org/sti/ind/digital-transformation-business-sector-summary.pdf

Pianko, D. (2018). *Analytics to fight tax fraud*. SAS Institute Inc.

Pijnenburg, M. G. F. (2020). *Data science for tax administration* [Doctoral dissertation, Leiden University].

Pijnenburg, M., Kowalczyk, W., & van der Lissette, H. D. (2017). A roadmap for analytics in taxpayer supervision. *The Electronic Journal of E-Government*, *15*(1), 19–32.

Power, D. J., Heavin, C., McDermott, J., & Daly, M. (2018). Defining business analytics: An empirical approach. *Journal of Business Analytics*, *1*(1), 40–53. https://doi.org/10.1080/2573234X.2018.1507605

Prastuti, G., & Lasmin. (2021). Assessing analytics maturity level in the Indonesian tax administration: The case of compliance risk management. *Jurnal Scientax*, *2*(2), 199–217. https://doi.org/10.52869/st.v2i2.157

Pratama, A., & Darono, A. (2022). Exploring Indonesian customs and tax data using text analytics approach. In *Proceedings of the International Conference on Customs and Tax Cooperation*. Jakarta.

Raj, A., Bosch, J., Olsson, H. H., & Wang, T. J. (2020). Modelling data pipelines. In *2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)* (pp. 13–20). https://doi.org/10.1109/SEAA51224.2020.00014

Richardson, V. J., Teeter, R., & Terrell, K. L. (2019). *Data analytics for accounting*. McGraw-Hill Education.

Rosid, A. (2023). Artificial neural networks for predicting taxpaying behaviour of Indonesian firms. *Scientax*, *4*(2), 174–204. https://doi.org/10.52869/st.v4i2.526

Rosid, A., Ardin, G., & Sanjaya, T. B. (2022). Prediksi keikutsertaan pelaku usaha dalam pemanfaatan insentif pajak dengan artificial neural network. *Jurnal Ekonomi Indonesia*, *11*(2), 109–142. https://doi.org/10.52813/jei.v11i2.178

Russom, P., Halper, F., Stodder, D., & Berberich, M. (2010). *Three tiers of analytic sandboxes: New techniques to empower business analysts*. TDWI Blog. https://tdwi.org/blogs/tdwi-blog/2010/03/analytic-sandboxes.aspx

Sakti, N. W. (2021). Implementation of data analytics in Directorate General of Taxes. Paper presented at *The 2nd Conference of Belt and Road Initiative Tax Administration Cooperation Forum*. https://www.britacom.org/news/2ndOnlineMeeting/files/22Slides%20of%20Indonesia.pdf

SAS. (2016). *What is analytics?* SAS. http://www.sas.com/en_us/insights/analytics/what-is-analytics.html

Schrag, A. (2019). What is practice research and why is it important to anyone other than artists? *Queen Margaret University*. https://www.qmu.ac.uk/campus-life/blogs/staff-dr-anthony-schrag/what-is-practice-research-and-why-is-it-important-to-anyone-other-than-artists/

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2015). *Analytics: The real-world use of big data—How innovative enterprises extract value from uncertain data (IBM Global Business Services Business Analytics and Optimization Executive Report)*. IBM Institute for Business Value and Saïd Business School at the University of Oxford.

Stankeviciusa, E., & Leonas, L. (2015). Hybrid approach model for prevention of tax evasion and fraud. In *Proceedings of the 20th International Scientific Conference Economics and Management – 2015 (ICEM-2015)*, Kaunas, Lithuania.

Tableau. (n.d.). *Business intelligence: A complete overview*. Tableau Software LLC. https://www.tableau.com/learn/articles/business-intelligence

Tesch, M. (2020). Modelling of causes and effects. *LeanBI Blog*. https://leanbi.ch/en/blog/modelling-of-causes-and-effects/

Tax Education & Training Center. (2024). *Web data analytics (Training materials – Digital Economy Compliance Supervision)*.

Thusoo, A., & Sarma, J. S. (2017). *Creating a data-driven enterprise with DataOps*. O'Reilly Media, Inc.

Vago, I. (2022). Legacy data stacks vs modern data stacks: Understanding the modern data stack. *MuttData Blog*. https://muttdata.ai/blog/2022/12/07/faq-01.html

Verhulst, S., Cattuto, C., Paolotti, D., & Vespignani, A. (2024, February). Outpacing pandemics: Solving the first and last mile challenges of data-driven policy making. *Zenodo*. https://doi.org/10.5281/zenodo.10610765

WeAreNetflix. (2018, March 14). *Netflix research: Analytics* [Video]. YouTube. https://www.youtube.com/watch?v=sAQmj8a_al8

Wessels, E. (2014). *Towards a conceptual decision support systems framework aimed at narrowing the tax gap in South Africa: A narrative case study* [Doctoral dissertation, University of South Africa].

White, C., & Imhoff, C. (2010). Advanced analytics and business intelligence: Term abuse? *BeyeNETWORK*. http://www.b-eye-network.com/view/13797

Winata, Y. G., & Hadi, M. (2023). Application of data mining techniques for VAT-registered business compliance. *Scientax*, *4*(2), 243–260. https://doi.org/10.52869/st.v4i2.317

Wu, R.-S., Ou, C. S., Lin, H., Chang, S.-I., & Yen, D. C. (2012). Using data mining technique to enhance tax evasion detection performance. *Expert Systems with Applications*, *39*(10), 8769–8777.