

Sains Data Perpajakan di Indonesia: Implementasi Praktis, Pembelajaran, dan Agenda Kajiannya

Agung Darono
Pusdiklat Pajak

ABSTRACT

As a relatively recently developed discipline, Tax data science requires a study that focuses on important themes related to how the discipline is applied by both taxpayers and tax authorities (central or regional). This research using an action research strategy is an attempt to present a map of the practical implementation of data science and then relate it to the learning needs for the development of human resource capacity of related tax practitioners as well as the necessary research agenda so that this practice works in line with the needs of the organization. This study, from the aspect of tax research methodology, contributes by providing a result that can be an additional reference regarding how action research strategies can produce practical knowledge of taxation data science from both the perspective of taxpayers and tax authorities.

ABSTRAK

Sebagai sebuah disiplin yang relatif baru saja berkembang, sains data perpajakan memerlukan kajian yang dapat memetakan berbagai hal yang berkaitan dengan bagaimana disiplin ini diimplementasikan baik oleh wajib pajak maupun otoritas pajak (pusat ataupun daerah). Penelitian ini dengan menggunakan strategi penelitian tindakan sebuah merupakan upaya untuk menyajikan peta dari implementasi praktis sains data kemudian mengaitkannya dengan kebutuhan pembelajaran untuk pengembangan kapasitas sumber daya manusia praktisi perpajakan yang terkait sebagaimana juga agenda penelitian yang diperlukan sehingga praktik ini bekerja sejalan dengan kebutuhan organisasi. Kajian ini, dari aspek metodologi penelitian perpajakan, berkontribusi dengan menyediakan sebuah hasil yang dapat menjadi tambahan rujukan berkaitan dengan bagaimana strategi penelitian tindakan dapat menghasilkan pengetahuan-praktis sains data perpajakan baik dari perspektif wajib pajak maupun otoritas pajak.

1. PENDAHULUAN

Data merupakan komponen utama sebuah sistem pemungutan pajak. Wajib pajak melaksanakan pemenuhan hak dan kewajiban perpajakan dengan menggunakan data yang disampaikannya melalui surat pemberitahuan. Pada sisi lain, otoritas pajak akan menggunakan data yang berasal dari wajib pajak tersebut dan kemudian juga berdasarkan data yang dimilikinya akan menetapkan apakah memang pemenuhan hak dan kewajiban perpajakan yang dilaksanakan oleh wajib pajak telah sesuai dengan ketentuan yang berlaku.

Perkembangan teknologi informasi dan komunikasi membawa dampak besar terhadap manajemen data perpajakan baik sisi wajib pajak ataupun otoritas pajak. Berbagai laporan terkait hak dan kewajiban perpajakan beserta data pendukung lainnya untuk kepentingan administrasi pajak tidak ada saja berubah menjadi data elektronik, namun lebih dari itu, ternyata pertambahan datanya sangat pesat dan format datanya sangat beragam. Hal inilah yang juga sering disebut sebagai *big data* perpajakan (Houser dan Sanders 2017; Carina dan Thompson 2019). Keberadaan *big data* perpajakan ini kemudian mendorong kepada adanya perkembangan sebuah area kajian dan praktik baru yang dikenal dengan sains data perpajakan (*tax data science*). Atau pun, beberapa praktisi/akademisi menyebutkan dengan *tax data analytics*. Kedua istilah ini pada tataran praktik sering dipertukarkan penggunaannya karena memang pada prinsipnya mempunyai tujuan yang sama yaitu bagaimana menggunakan data untuk mendapatkan berbagai wawasan (*insight*) baru untuk mendukung pengambilan keputusan (Aasheim et al. 2015; IOTA 2016; Kurtz 2018; Shan 2019; Microsoft dan PwC 2018; Pijnenburg 2020).

Situasi di Indonesia menunjukkan bahwa keberadaan sains data perpajakan dari sisi otoritas pajak telah mulai diimplementasikan bahkan berdasarkan hasil kajian yang dilakukan oleh Prastuti dan Lasmin (2021) tingkat *maturity* telah mencapai level yang cukup tinggi. Sementara itu dari sisi wajib pajak, Hardyana (2020) menguraikan bahwa fungsi perpajakan wajib pajak dapat memanfaatkan sains data untuk manajemen dan perencanaan pajak sehingga pelaksanaan hak dan kewajiban pajaknya tidak akan bertentangan dengan ketentuan perpajakan yang berlaku.

Rumusan masalah penelitian ini adalah bagaimana sains data perpajakan telah diimplementasikan

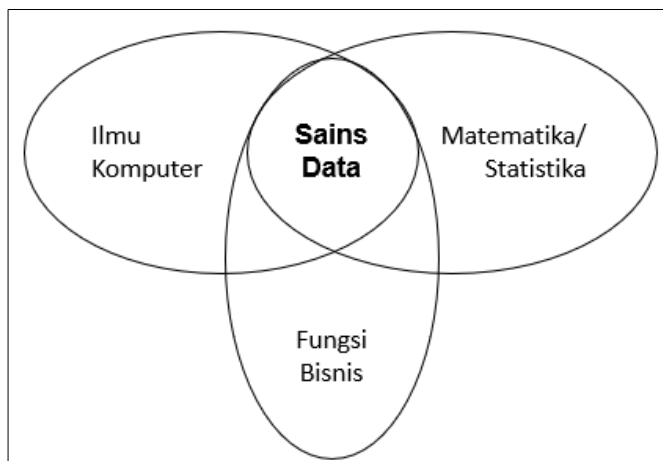
dalam praktik perpajakan baik di sisi otoritas pajak ataupun wajib pajak, dan bagaimana kemudian praktik tersebut berkaitan dengan proses pembelajaran maupun agenda kajian/penelitian yang berhubungan dengan praktik dan juga proses pembelajaran tersebut. Penelitian ini, dengan menggunakan strategi penelitian tindakan (*action research*), bertujuan untuk memetakan implementasi praktis berbagai prinsip dan metodologi sains data di bidang perpajakan. Hasil pemetaan tersebut kemudian dicoba untuk dikaitkan dengan kebutuhan pembelajaran untuk pengembangan kapasitas sumber daya manusia praktisi perpajakan yang terkait serta agenda kajian yang diperlukan sehingga praktik sains data perpajakan ini dapat berkembang dengan lebih baik.

Sistematika penyampaian gagasan dalam tulisan ini adalah sebagai berikut. Bagian ini adalah latar belakang masalah yang mendorong perlunya kajian ini dilakukan. Bagian kedua akan berisi tinjauan literatur tentang disiplin sains data perpajakan terutama yang terkait dengan ruang lingkup dan metodologi pengembangan dan implementasinya. Bagian ketiga akan memaparkan rancangan penelitian. Bagian keempat akan menyajikan analisis data dan pembahasan. Bagian terakhir adalah penutup dengan menyajikan kesimpulan dan saran.

2. SAINS DATA PERPAJAKAN: BATASAN DAN RUANG LINGKUPNYA

Secara agak sedikit anekdotal, Donoho (2017) menyebutkan bahwa disiplin sains data sudah ada kurang lebih 50 tahun yang lalu. Pada sisi lain, artikel dengan judul "*Data Scientist: The Sexiest Job of the 21st Century*" (Davenport dan Patil 2012) juga sempat menimbulkan euforia kecil yang mendorong orang untuk tahu lebih jauh tentang sains data ini. Disiplin ini mendapatkan momentum pengembangan dan penggunaannya kembali terutama karena perkembangan teknologi pengolahan data yang semakin maju namun dengan biaya yang lebih terjangkau sehingga memunculkan istilah *big data*. Keberadaan *big data* pada gilirannya membutuhkan kerangka pengetahuan yang secara khusus dapat menganalisisnya. Dalam hal inilah, disiplin sains data kemudian berperan besar (lihat misalnya Press 2013). Mengacu artikel Press (2013), istilah *data science* pertama kali disebutkan oleh Peter Naur dan didefinisikan sebagai "*the science of dealing with data, once they have been established, while the relation of the data to what they represent is delegated to other fields and sciences.*" Dari definisi ini dapat dilihat bahwa memang dari sejak awal sains data merupakan kajian multidisiplin yang akan melibatkan disiplin pengolahan dan analisis data satu sisi dengan disiplin lain yang akan memanfaatkan hasil analisis tersebut.

Pada perkembangannya terdapat cukup banyak pendapat yang mencoba memberikan definisi tentang sains data ini. Das (2016) menyatakan bahwa sains data merupakan kajian yang melibatkan beragam teknologi, proses, dan metodologi untuk mendapatkan wawasan dan nilai dari data dalam berbagai bentuk, struktur, dan ukuran. Sedangkan Provost and Fawcett (2013) menyatakan bahwa sains data adalah bidang kajian yang mencakup prinsip, proses, dan teknik untuk memahami fenomena melalui analisis data secara otomatis. Sementara itu Pierson (2015) mengajukan definisi bahwa sains data adalah praktik menggunakan metode komputasi untuk memperoleh wawasan berharga dari kumpulan data mentah untuk kemudian dapat ditindaklanjuti. Definisi yang lain adalah yang dikemukakan oleh Zhu and Xiong (2015), sains data adalah teori, metode, dan teknologi mempelajari dan mengeksplorasi sifat data (*datanature*) sehingga dari hal tersebut data dapat digunakan untuk memahami gejala alam ataupun manusia. David Smith, sebagaimana dikutip oleh Press (2013), mengemukakan definisi ringkas yang cenderung mudah diingat, sains data adalah "*combination of computer hacking, data analysis, and problem solving*". Mirip dengan definisi ini, Worsley (2017) memberikan pengertian sains data sebagai kombinasi dari ilmu komputer, matematika/statistika untuk kemudian dipadukan kerangka pengetahuan bisnis/organisasi. Definisi ini dapat diperjelas pengertiannya dengan melihat Gambar 1.



Gambar 2.1 Sains Data sebagai kajian multidisiplin
sumber: diadaptasi dari Worsley (2017)

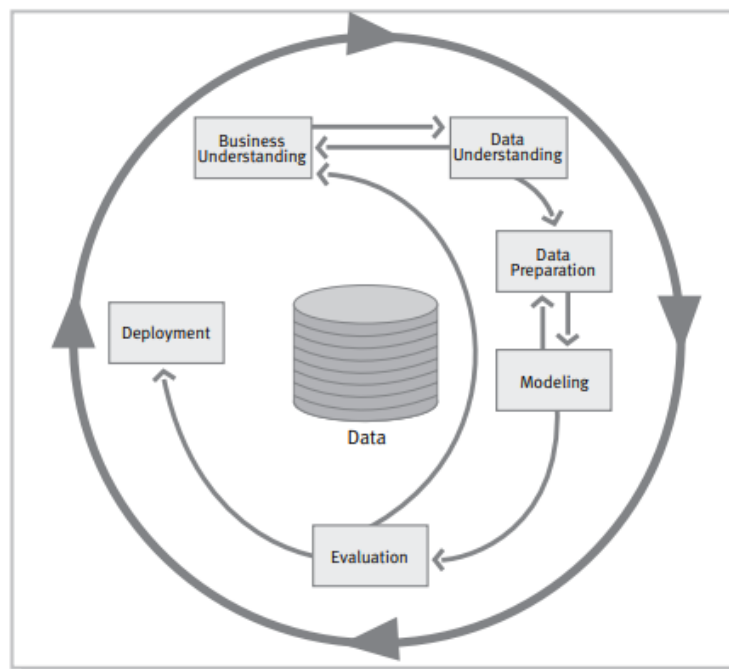
Dalam beberapa situasi istilah sains data ini terkadang dipertukarkan dengan analitika data (*data analytics*)¹. Cuadrado-Gallego dan Demchenko (2020:2) menyatakan bahwa analitika data berbeda dengan sains data karena analitika data adalah proses inspeksi, transformasi, dan pemodelan data dengan tujuan untuk menemukan tren, pola, atau hubungan yang menggambarkan fenomena kehidupan nyata yang dapat diamati dan dapat digunakan untuk pengambilan keputusan yang tepat. Sementara itu sains data adalah bidang interdisipliner yang menggunakan metode, proses, algoritma, dan sistem ilmiah untuk mengekstraksi pengetahuan dan wawasan dari data terstruktur dan tidak terstruktur.

Sementara itu Aasheim *et al.* (2015) mencoba membedakan pengertian dan praktik sains data dengan analitika data dengan menggunakan perbandingan kurikulum mata kuliah yang diajarkan pada program studi pada beberapa perguruan tinggi. Kesimpulannya adalah sains data cenderung menekankan pada pengetahuan pengembangan model matematika dan statistika untuk pengambilan keputusan sedangkan analitika data lebih menekankan sisi aplikasi untuk kepentingan bisnis. Dalam pandangan penulis, kedua istilah ini dalam tataran konseptual bisa saja dibedakan namun dalam praktik dan aplikasinya terutama untuk pengolahan data bisnis/organisasi, kedua istilah tersebut dapat dipertukarkan penggunaannya

¹ istilah “analitika” diserap dari “analytics”, jika mengikuti kaidah *mathematics, economics, statistics* diserap ke dalam bahasa Indonesia menjadi matematika, ekonomika, statistika.

bahkan secara arbitrer, bergantung pada konteks situasi yang melingkupi penggunaan istilah tersebut.

Selain definisi dan ruang lingkupnya, hal penting yang harus diungkapkan terkait dengan sains data adalah metodologi atau kerangka kerja pengembangan dan implementasi. Metodologi implementasi sains data adalah strategi umum yang memandu proses dan aktivitas untuk mengimplementasi prinsip, metode dan pendekatan teknis sehingga sains data dapat diterapkan sesuai dengan fungsi dan tujuannya. Salah satu metodologi yang paling awal dibentuk dan banyak diadopsi adalah metodologi yang dikenal dengan CRISP-DM (Rollins 2015; Martinez *et al.* 2021). Metodologi ini akan memandu organisasi yang akan mengimplementasikan sains data untuk mengikuti urutan langkah tertentu sehingga sains data dapat diterapkan sesuai dengan fungsi yang diinginkan. Perhatikan Gambar 2 yang menjelaskan tahapan (siklus hidup/*life cycle*) yang harus dilalui jika menggunakan CRISP-DM sebagai metodologi/kerangka kerja implementasi sains data.



Gambar 2.2 Metodologi impelentasi sains data dengan CRISP-DM
sumber: Chapman *et al.* (2000)

Berdasarkan uraian di atas dapat disimpulkan bahwa sains data perpajakan merupakan penerapan konsep, prinsip, dan metodologi sains data untuk kepentingan perpajakan baik di sisi wajib pajak ataupun otoritas pajak. Sains data perpajakan merupakan bidang kajian dan praktik organisasional yang bersifat multidisiplin dengan melibatkan bidang statistika dan matematika, ilmu komputer, serta perpajakan. Bidang perpajakan itu sendiri sebenarnya sudah berupa kajian multidisiplin juga yang setidaknya melibatkan bidang kajian ekonomi, akuntansi, dan hukum. Suatu organisasi yang ingin menerapkan sains data perpajakan dapat mengacu kepada kerangka kerja ini sebagai rujukan awal walaupun pada pelaksanaannya bisa saja melakukan beberapa modifikasi.

3. RANCANGAN PENELITIAN

Creswell (2013:22) mendefinisikan pendekatan penelitian (*research approach*) sebagai kerangka kerja yang tersusun dari cara pandang peneliti terhadap topik penelitian (*paradigm*) yang kemudian digunakan untuk menetapkan rancangan penelitian (*research design* atau disebut juga strategi penelitian) beserta metode penelitiannya (*research method*). Penelitian ini menggunakan rancangan penelitian tindakan (*action research*). Rancangan ini memungkinkan sebuah penelitian dilakukan dengan menggunakan dua aktivitas yang saling berkaitan yaitu meneliti (sebagai cara untuk memperluas cakupan pengetahuan tentang sebuah tema tertentu) dan pada saat sama juga melakukan tindakan tertentu yang didorong oleh proses penelitian yang sedang dilakukan (Heller 2004; Altrichter *et al.* 2002; Somekh 2006; French 2009). Merujuk Heller (2004) serta Baskerville dan Wood-Harper (1996), prinsip penelitian tindakan yang diadopsi oleh kajian ini adalah:

- (1) rangkaian prosedur penelitian yang dilakukan terdiri dari mendiagnosis masalah, merencanakan tindakan, melakukan tindakan, mengevaluasi hasil dan menspesifikasikan hasil penelitian berdasarkan tindakan yang telah diberikan;
- (2) hubungan yang erat antara akuisisi pengetahuan dan tindakan sebagai bagian integral dari metodologi lapangan;
- (3) pengetahuan dan tindakan berasal dari interaksi hasil diagnosis dan dapat berlaku sebaliknya penelitian.

4. ANALISIS DATA DAN PEMBAHASAN

Bagian ini akan terdiri dua subbagian yaitu analisis data dan pembahasan. Subbagian analisis data akan memaparkan proses pengumpulan, pengolahan dan analisis data. Selanjutnya, berdasarkan analisis data yang sudah dilakukan tersebut akan dilakukan pembahasan untuk mendapatkan jawaban atas pertanyaan penelitian, dan berupaya untuk mengajukan beberapa sumbangan pemikiran yang berkaitan dengan dengan pengembangan sains data perpajakan di Indonesia.

Pengumpulan dan Analisis Data

Berdasarkan data yang dikumpulkan dari *action research* yang dilakukan kemudian penulis (sebagai peneliti) menyusun jawaban atas pertanyaan penelitian yang berkaitan dengan urgensi sains data perpajakan dan kemudian bagaimana hal tersebut ditindaklanjuti dari sisi kajian, pembelajaran dan implementasi praktis. Partisipan² dari *action research* yang dilaksanakan berasal dari komunitas praktisi (*community of practice*) yang terbentuk dari interaksi penulis dalam beberapa rekan praktisi dalam berbagai kegiatan antara lain pembelajaran klasikal, fasilitator pelatihan *online* ataupun narasumber seminar/webinar. Penanganan terhadap masalah pengolahan dan analisis data inilah yang penulis ajukan sebagai *action research* dalam kajian ini. Dari serangkaian tindakan (*action*) yang penulis sampaikan kepada para partisipan, dengan seizin mereka sebagai pemilik data³ dan aktor-organisasi terlibat untuk melakukan tindakan pengolahan pada data yang sesungguhnya, penulis kemudian menyajikan proses dan hasil tindakan tersebut menjadi data penelitian ini. Teknik analisis data yang digunakan untuk kepentingan ini adalah *thematics analysis* (lihat misalnya Ayres 2008; Lapadat 2010)

Berdasarkan berbagai interaksi awal dari berbagai forum tersebut, peneliti dalam beberapa kesempatan mendapatkan interaksi lanjutan dengan para peserta kegiatan, dengan berbagai saluran

² istilah yang digunakan adalah partisipan (lihat misalnya Heller 2004) bukan responden, informan atau narasumber sebagaimana beberapa jenis strategi penelitian lain seperti *grounded theory*, etnografi atau studi kasus.

³ dengan tetap menjalankan protokol untuk menjaga kerahasiaan dan keamanan data, penulis hanya mendapatkan *dummy data* (bukan *real data*) dalam memberikan tindakan

komunikasi seperti *instant messaging*, email ataupun telepon sebagai diskusi lanjutan dari berbagai topik yang disajikan dalam kegiatan tersebut. Berangkat dari beberapa komunikasi ini selanjutnya menjadi ajang tanya-jawab dan diskusi untuk menyelesaikan berbagai kasus pengolahan dan analisis data yang mereka hadapi untuk dimintakan solusi pengolahan datanya kepada penulis.

PEMBAHASAN

Penerapan metodologi CRISP-DM: tindakan dalam penelitian

Interaksi awal antara penulis dengan partisipan merupakan saat untuk memahami kebutuhan analisis data yang diperlukan. Berangkat dari situasi ini, penulis memutuskan untuk mulai menerapkan metodologi CRISP-DM sebagai bagian dari tindakan yang rencananya akan penulis berikan sebagai jawaban atas permasalahan analisis data yang dihadapi oleh partisipan. Artinya, atas semua permasalahan penelitian (*use case*⁴) sebagaimana terpampang dalam Tabel 1, penulis mengikuti mencoba untuk melaksanakan tahapan dalam kerangka CRISP-DM.

Tabel 4.1 Use case yang digunakan sebagai action research dalam penelitian ini

Kode	Deskripsi singkat	Tindakan yang disarankan
UC-1	merelasikan data PDRB yang tersedia dalam situs BPS kabupaten/kota untuk mengetahui korelasi dan analisis jalur antara PDRB dengan penerimaan setiap jenis pajak daerah	<i>Web data extraction</i> dengan menggunakan perangkat aplikasi Excel 365 - Power Query dan IBM SPSS Statistics
UC-2	mengolah dan analisis data <i>general ledger</i> dalam bentuk <i>text report file</i> dengan ukuran file + 2GB	<i>Data extraction</i> dengan menggunakan perangkat aplikasi Power BI Desktop
UC-3	merelasikan data penjualan dengan faktur PPN di mana pada setiap data tidak terdapat <i>unique relation key</i> yang dapat digunakan untuk menghubungkan data tersebut	menerapkan fitur <i>fuzzy match</i> dalam perangkat aplikasi Excel 365 - Power Query
UC-4	mengolah dan analisis data pembayaran gaji tenaga medis dari sebuah rumah sakit dalam format <i>multiple XLSX report file</i> yang dibuat oleh sebuah aplikasi akuntansi keuangan	menerapkan fitur <i>transform & combine</i> dalam perangkat aplikasi Excel 365 - Power Query
UC-5	mengolah dan analisis data untuk melakukan analisis risiko wajib pajak daerah	menggunakan teknik <i>decision tree</i> dan <i>risk quadrant</i> untuk memvisualisasikan hasil

⁴ *use case* adalah yaitu serangkaian aktivitas yang dilaksanakan oleh aktor organisasi untuk mengeksekusi proses bisnis, prosedur, serta sistem aplikasi dalam satu atau lebih organisasi sehingga dapat menghasilkan keluaran tertentu (lihat misalnya Booch *et al.* 1998:185)

Kode	Deskripsi singkat	Tindakan yang disarankan
		analisis risiko dan disajikan dalam <i>executive dashboard</i> dengan menggunakan perangkat aplikasi IBM SPSS Statistics dan Excel 365 - Power Query

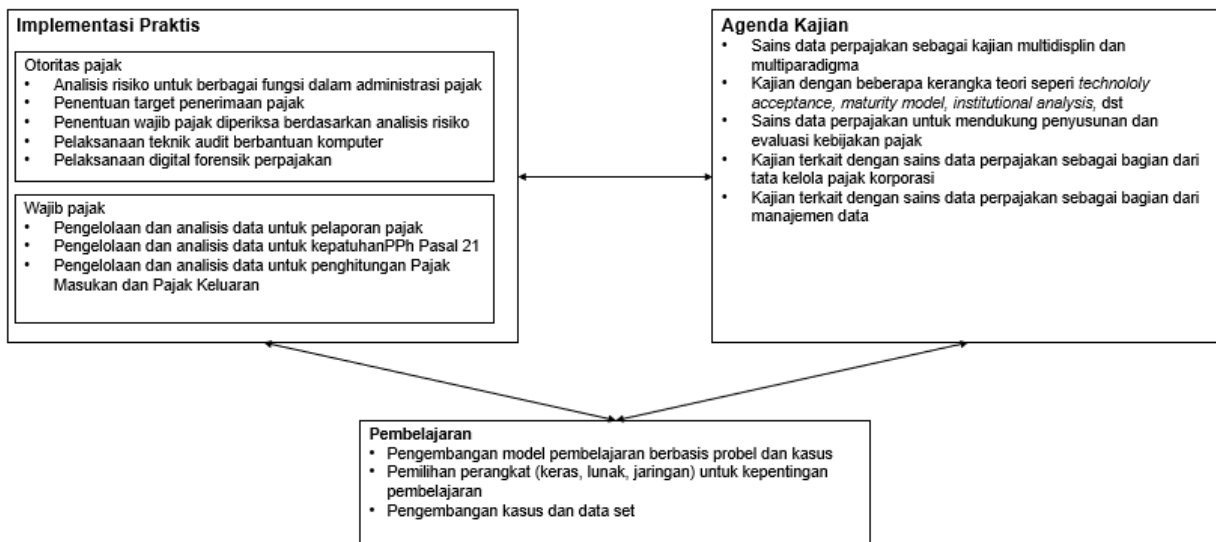
sumber: hasil olahan penulis

Berdasarkan tahapan dalam CRISM-DM (perhatikan kembali Gambar 2), penulis mengajak partisipan untuk terlibat dalam setiap tahap tersebut. Penulis menekankan kepada partisipan bahwa tindakan ini merupakan simulasi dari apa yang seharusnya akan dilakukan oleh partisipan terhadap data yang sesungguhnya. Beberapa hal yang peneliti temukan dalam pemberian tindakan ini adalah sebagai berikut:

- (1) menyadari perlunya pemahaman metodologi penerapan sains data terutama untuk analisis data yang berada pada level *enterprise*;
- (2) kemampuan untuk menentukan perangkat pengolahan data yang sesuai dengan berbagai format, sumber, dan ukuran data yang dihadapi;
- (3) kemampuan untuk mengotomasikan pekerjaan pengolahan data yang berulang;
- (4) kemampuan untuk melakukan analisis data secara deskriptif dan diagnostik;
- (5) kemampuan untuk melakukan visualisasi data dengan teknik informasi *dashboard*;
- (6) kemampuan untuk merelasikan data dengan menggunakan teknik *fuzzy match*;
- (7) tahapan deployment sains data perpajakan baru pada individual partisipan karena keterbatasan wewenang partisipan.

Keterkaitan antara implementasi praktis dengan pembelajaran dan agenda kajian sains data perpajakan

Berdasarkan hasil analisis terhadap tindakan berdasarkan *use case* yang diajukan oleh partisipan, penulis kemudian menelaah lebih lanjut untuk mengaitkannya dengan bagaimana proses dan model sains data perpajakan ini dapat dilaksanakan serta agenda kajian/penelitian apa yang diperlukan sehingga nantinya dapat digunakan untuk memperbaiki implementasi praktis itu sendiri. Ketiga hal tersebut (implementasi praktis, pembelajaran dan agenda kajian harus dilihat sebagai hal yang saling memengaruhi. Untuk itu, tulisan ini mengajuka kerangka pemikiran yang digunakan untuk membahas dan menjelaskan keterkaitan antara penelitian, pembelajaran serta implementasi praktis sains data perpajakan sebagaimana terpampang dalam Gambar 3.



Gambar 4.1 Keterkaitan antara implementasi praktis dengan pembelajaran dan agenda kajian sains data perpajakan

sumber: analisis penulis

5. PENUTUP

Sains data perpajakan merupakan disiplin dan praktik organisasional yang melibatkan bidang statistika dan matematika, ilmu komputer, dan perpajakan dengan tujuan untuk mendapatkan berbagai wawasan (*insight*) baru untuk mendukung pengambilan keputusan untuk kepentingan perpajakan baik di sisi wajib pajak ataupun otoritas pajak. Penelitian ini, berdasarkan tindakan yang diberikan kepada partisipan, menemukan bahwa (1) sains data perpajakan merupakan kebutuhan nyata baik untuk kepentingan wajib pajak maupun otoritas pajak; (2) perlunya mengaitkan implementasi praktis sains data perpajakan dengan pembelajaran dan agenda kajiannya.

Kajian ini merekomendasikan perlunya para praktisi perpajakan baik di sisi wajib pajak ataupun otoritas pajak (pusat maupun daerah) untuk mengeksplorasi berbagai kemungkinan mengimplementasikan sains data perpajakan menjadi bagian dari proses manajerial-organisasional sehingga keputusan yang berkaitan dengan tujuan perpajakan diperoleh melalui proses sistematis dan yang didukung dengan data dan argumentasi yang logis serta sesuai dengan aturan hukum yang berlaku. Dari sisi metodologi penelitian di bidang perpajakan, kajian ini berkontribusi dengan menyediakan sebuah hasil yang dapat menjadi tambahan rujukan berkaitan dengan bagaimana strategi penelitian action research dipilih sehingga dapat menghasilkan pengetahuan-praktis sains data perpajakan baik dari perspektif otoritas maupun wajib pajak. Kajian ini mempunyai keterbatasan dari variasi tindakan yang dapat dilakukan karena peneliti tidak secara proaktif menemukan situasi/organisasi yang kemungkinan memerlukan tindakan berupa implementasi sains data perpajakan. Merujuk pada kerangka CRISP-DM, tindakan yang dilakukan masih parsial dan belum mencakup tahap evaluasi dan deployment dengan tingkat skalabilitas yang cukup besar.

DAFTAR PUSTAKA

- Aasheim, C.L., S. Williams, Paige Rutner, and A. Gardiner. 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 (January): 103–15.
- Altrichter, Herbert, Stephen Kemmis, Robin Mctaggart, and Ortrun Zuber-Skerritt. 2002. "The Concept of Action Research." *Learning Organization, The* 9 (August): 125–31. <https://doi.org/10.1108/09696470210428840>.
- Ayres, Lioness. 2008. "Thematic Coding and Analysis." In *Encyclopedia of Qualitative Research Methods*, 867–68.
- Baskerville, Richard, and Trevor Wood-Harper. 1996. "A Critical Perspective on Action Research as a Method for Information Systems Research." *Journal of Information Technology* 11 (September): 235–46. <https://doi.org/10.1080/026839696345289>.
- Booch, Grady, James Rumbaugh, and Ivar Jacobson. 1998. *Unified Modeling Language User Guide*. Reading, Massachusetts: Addison Wesley.
- Carina, Federico, and Travis Thompson. 2019. "Do IRS Computers Dream About Tax Cheats? Artificial Intelligence and Big Data in Tax Enforcement and Compliance." *Journal of Tax Practice & Procedure*, no. Feb-Mar: 43–47. <https://www.crowell.com/files/2019-Feb-March-Do-IRS-Computers-Dream-About-Tax-Cheats-Federico.pdf>.
- Chapman, Peter, Janet Clinton, Randy Kerber, Tom Khabaza, Thomas P. Reinartz, Colin Shearer, and Richard Wirth. 2000. "CRISP-DM 1.0: Step-by-Step Data Mining Guide." In .
- Creswell, John W. 2013. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, California: SAGE Publications.
- Cuadrado-Gallego, Juan J., and Yuri Demchenko. 2020. "Introduction to the Data Science Framework." In , 1–7. Springer.
- Das, Sibbanjan. 2016. *Data Science Using Oracle Data Miner and Oracle R Enterprise*. New York: Apress Media.
- Davenport, Thomas H., and D.J. Patil. 2012. "Data Scientist: The Sexiest Job of the 21st Century" 2012 (October).
- Donoho, David. 2017. "50 Years of Data Science." *Journal of Computational and Graphical Statistics* 26: 745–66. <https://doi.org/10.1080/10618600.2017.1384734>.
- French, Steven. 2009. "Action Research for Practising Managers." *Journal of Management Development* 28 (March): 187–204. <https://doi.org/10.1108/02621710910939596>.
- Hardyana, Yan. 2020. "Best Practice for A Data Analytics Tax Strategy." Presented at the Seminar Nasional

- Perpajakan, FIA UB.
- Heller, Frank. 2004. "Action Research and Research Action: A Family of Methods." In *Essential Guide to Qualitative Methods in Organizational Research*, edited by Catherine Cassell and Gillian Symon, 349–60. London: SAGE Publications Ltd.
- Houser, Kimberly A., and Debra Sanders. 2017. "The Use of Big Data Analytics by the IRS: Efficient Solutions or the End of Privacy as We Know It?" *Vanderbilt Journal of Entertainment & Technology Law* 19 (4): 871–72.
- IOTA. 2016. "Applying Data and Analytics in Tax Administration." IOTA Good Practice Guide. Intra-European Organisation of Tax Administrations (IOTA). <https://www.iota-tax.org/good-practice-guide-applying-data-and-analytics-tax-administrations>.
- Kurtz, Jennifer. 2018. "Understanding Data's Impact on the Future of Tax Functions." Vertex. <https://www.europeanceo.com/finance/understanding-datas-impact-on-the-future-of-tax-functions/>.
- Lapadat, Judith C. 2010. "Thematic Analysis." In *Encyclopedia of Case Study Research*, edited by Albert J. Mills, Gabrielle Durepos, and Elden Wiebe, 925–27. Thousand Oaks, California: SAGE Publications, Inc.
- Martinez, Iñigo, Elisabeth Viles, and Igor Olaizola. 2021. "Data Science Methodologies: Current Challenges and Future Approaches." *Big Data Research* 24 (January): 100183. <https://doi.org/10.1016/j.bdr.2020.100183>.
- Microsoft, and PwC. 2018. "The Data Intelligent Tax Administration - Meeting the Challenges of Big Tax Data and Analytics." Netherland: Microsoft and PricewaterhouseCoopers Belastingadviseurs N.V. <https://www.pwc.nl/nl/assets/documents/the-data-intelligent-tax-administration-whitepaper.pdf>.
- Pierson, Lillian. 2015. *Data Science For Dummies*. John Wiley & Sons, Inc.
- Pijnenburg, Mark Gerardus Franciscus. 2020. "Data Science for Tax Administration." Ph.D. thesis, Netherland: Leiden University.
- Prastuti, Gitarani, and Lasmin. 2021. "Assessing Analytics Maturity Level in The Indonesian Tax Administration: The Case of Compliance Risk Management" 2 (2): 199–217. <https://doi.org/10.52869/st.v2i2.157>.
- Press, Gil. 2013. "A Very Short History Of Big Data," 2013. <https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#2135d2fe65a1>.
- Provost, Foster, and Tom Fawcett. 2013. *Data Science for Business*. Sebastopol, CA: O'Reilly Media, Inc.
- Rollins, John B. 2015. "Foundational Methodology for Data Science." Route 100 Somers, NY 10589: IBM Analytics.
- Shan, JingJing. 2019. "Optimization Strategy of Tax Planning System in the Context of Artificial Intelligence and Big Data." *Journal of Physics: Conference Series* 1345 (November): 052006. <https://doi.org/10.1088/1742-6596/1345/5/052006>.
- Somekh, Bridget. 2006. *Action Research: A Methodology for Change and Development*. Open University Press McGraw-Hill Education.
- Worsley, Rob. 2017. "Data Science for Business." IBM Corporation. <https://www.ibm.com/downloads/cas/YLORK8WL>.
- Zhu, Yangyong, and Yun Xiong. 2015. "Defining Data Science." CoRR abs/1501.05039. <http://arxiv.org/abs/1501.05039>.