

ANALISIS POTENSI E-COMMERCE MELALUI IMPLEMENTASI DATA MINING DALAM PERPAJAKAN: SEBUAH STUDI KOMPARASI

Oleh

Dewi Noor Fatikhah Rokhimakhumullah^{1*}, Devi Nur Cahaya Ningsih², Aji Widya Firmansyah³, Ayun Hanafiyah⁴

^{1,2,3,4}Business Administration, Universitas Brawijaya

Email: ¹dewi.noor@ub.ac.id, ²devinurcahaya@ub.ac.id, ³ajiwidya3@student.ub.ac.id,
⁴ayunhanafiyah@student.ub.ac.id

Abstract

The digital economy is experiencing quite rapid development, not to mention the e-commerce sector which is showing growth. From 2025 to 2030 it is projected that there will be a three-fold increase, mainly from the e-commerce sector. A number of transactions and data are abundant, causing DGT to have difficulty processing data. In addition, determining the involvement of e-commerce actors who are not easily detected is a challenge in itself. Based on these problems, this study raises the application of data mining in taxation to analyze the potential of e-commerce through comparative studies. This study uses a qualitative descriptive approach with literature study and interview methods through the acquisition of primary data from informants and secondary data through relevant literature sources. This study uses content analysis techniques. The results of the study show that Indonesia needs to adopt a data mining system implemented by several countries based on benchmark analysis. Comparison and effectiveness of data mining, especially in preventing, detecting, and overcoming tax evasion behavior by taxpayers. In addition, reviewing the potential to be implemented in Indonesia is very large. Application of data in large sets will be processed using machine learning, random forest, artificial intelligence, and data mining. Based on analysis with other countries, this method minimizes the number of problems in computerized systems.

Keywords: E-Commerce, Data Mining, Tax, Digital Economy

PENDAHULUAN

Era digitalisasi akibat pandemi telah mengubah lanskap ekonomi Indonesia secara masif. Mayoritas bisnis telah bertransformasi dari pola konvensional menjadi digital. Hal itu membuat ekonomi digital Indonesia tumbuh pesat. Laporan e-economy SEA 2022 menunjukkan bahwa dengan pertumbuhan *Gross Merchandise Value* (GMV) sebesar 22% *year on year* (YoY), ekonomi digital Indonesia bernilai \$77 miliar pada 2022. Lalu, pada 2025 diproyeksikan dapat tumbuh hingga \$130 miliar dan dapat meningkat tiga kali lipat sekitar \$220-\$360 miliar pada 2030, dengan e-commerce sebagai promotor utama (Google, Temasek, & Bain Company, 2022).

E-commerce telah berkembang secara luas di Indonesia, seperti tokopedia, shopee, dan marketplace lainnya. Indonesia memiliki probabilitas dan potensi yang baik dalam menghadapi *free-market* dan ekonomi digital. Pesatnya ekonomi digital dapat memberikan keuntungan yang besar bagi perekonomian, mulai dari transaksi hingga investasi industri digital di Indonesia. Data menunjukkan bahwa 8 juta masyarakat Indonesia berbelanja secara *online* dan diprediksi akan terus meningkat (Prastyaningtyas, 2019). Hal itu menunjukkan bahwa terdapat potensi pajak yang besar dari pesatnya perkembangan ekonomi digital dan e-commerce di Indonesia.

Berbeda dengan perdagangan konvensional, transaksi *e-commerce* memiliki karakteristik tersendiri sehingga berimplikasi pada pengenaan pajak yang cukup kompleks. Perkembangan *e-commerce* juga tidak terjangkau oleh *tax regulation*, sekalipun yang terbaru (S & Y, 2022). Kompleksitas itulah yang membuat kepatuhan pajak pelaku usaha digital rendah. Kompleksitas perpajakan seperti rumitnya administrasi dan peraturan perpajakan yang sering berubah-ubah, dapat berpengaruh terhadap kepatuhan wajib pajak (Dharmawan & Adi, 2021). Hal itu mengharuskan Direktorat Jenderal Pajak (DJP) selaku otoritas pajak untuk adaptif dan melakukan penyesuaian.

Sebenarnya DJP telah melakukan berbagai langkah untuk menangkap potensi pajak agar tidak hilang. Hal itu dilakukan dengan cara menjadikan *marketplace* sebagai mitra dalam memungut pajak. Selain itu DJP dapat melakukan pengawasan dan ekualisasi pajak melalui *account representative* di masing-masing unit vertikal. Selanjutnya, DJP telah berupaya untuk mengoptimalkan administrasi pajak melalui aplikasi dan website perpajakan, seperti misalnya *e-filing*. Baru-baru ini, DJP juga telah melakukan terobosan baru melalui simplifikasi administrasi dengan menjadikan NIK sebagai NPWP (S & Y, 2022). Namun, lanskap ekonomi digital lebih luas daripada itu, DJP masih perlu langkah yang lebih efektif melalui pemanfaatan dan penggalan data pelaku usaha digital.

DJP mempunyai rencana untuk melakukan kajian lebih lanjut dalam sektor *e-commerce* dengan melibatkan beberapa pihak yang terkait. Tentunya tidak mudah untuk diterapkan karena tantangan yang harus dihadapi berupa resistensi dari *merchant* pada *marketplace*. Hal ini berimplikasi pada penyesuaian regulasi yang dibuat oleh DJP, misalnya melalui Peraturan Menteri Keuangan (PMK) untuk menangkap potensi dari *e-commerce* (S & Y, 2022). Kebijakan penetapan peraturan perpajakan mempunyai korelasi yang

positif dengan tingkat kepatuhan dari wajib pajak dalam ekonomi digital (Hasanah, 2016).

Sektor *e-commerce* dapat meningkatkan penerimaan negara dalam penyeteroran pajak penghasilan. Akan tetapi, data-data keberadaan dan sumber dana untuk bisnis dalam *e-commerce* masih belum diketahui secara pasti dari mana asalnya (BPPK, 2019). DJP mengalami kesulitan untuk penggalan potensi untuk pemajakan dari kondisi tersebut. Sifat *e-commerce* yang tidak terlihat oleh mata, hanya mengandalkan bukti transaksi dalam prosesnya. Misalnya, pelaku usaha di dalam *platform* yang tidak diketahui secara jelas jumlah kategori PKP dan Non-PKP. Selain itu, tantangan DJP dalam menentukan subjek pajak di media sosial karena tidak terdapat nama pemilik akun yang jelas (S & Y, 2022).

Pengelolaan sejumlah data dan transaksi yang membutuhkan sistem secara terintegrasi untuk memudahkan DJP dalam penggalan potensi pajak atas *e-commerce*, salah satunya dengan menggunakan *data mining*. Menurut Davies (dalam Aminuddin, 2015) *data mining* sebagai pencarian suatu pola tertentu yang berasal dari data yang sangat kompleks dan berjumlah besar dengan tujuan memperoleh suatu informasi penting. *Data mining* membutuhkan ilmu statistik sekaligus komputasi tingkat tinggi yang bertujuan memperoleh informasi dalam suatu data yang nantinya akan diletakkan dalam *database* ataupun jenis *storage* lainnya yang hasilnya akan digunakan untuk mengambil keputusan atau kebijakan. Penggunaan *Data mining* mempermudah membaca data-data yang terdapat di dalam *database* sehingga menstimulasi pekerjaan DJP dalam memindai keberadaan data-data *e-commerce*.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengetahui potensi pajak atas *e-commerce* melalui penerapan *data mining* dalam perpajakan. Selain itu, penelitian ini menggunakan analisis *benchmark* untuk membandingkan penggunaan *data mining* dalam perpajakan yang telah diimplementasikan oleh negara lain yang nantinya dapat menjadi solusi dalam menghadapi tantangan kompleksitas *e-commerce*. Penelitian ini dapat menjadi acuan atas kebijakan pajak *e-commerce* di Indonesia yang dikombinasikan dengan implementasi *data mining* sebagai rekomendasi kebijakan pemerintah kedepannya.

LANDASAN TEORI

Ekonomi Digital

Ekonomi digital melingkupi tingkah laku manusia ketika dihadapkan pada pilihan untuk pemenuhan kebutuhan hidup yang tiada batasnya. Hal tersebut didukung oleh adanya teknologi. Implementasi kegiatan tersebut dengan adanya sentuhan jari tangan pada *gadget* atau ekonomi digital ketika melakukan perbelanjaan maupun hal lainnya. Aktivitas ini memberikan keuntungan bagi manusia untuk mendapatkan barang atau jasa dalam waktu singkat. Pemenuhan keinginan manusia ini tidak melibatkan pada kondisi dimana harus ke pasar ketika memenuhi keinginannya (Permana dan Puspitaningsih, 2019). Misalnya kawasan Asia Tenggara, ekonomi digital sedang berkembang secara aktif dengan besarnya peluang pasar. Teknologi yang menyulut pertumbuhan ekonomi digital tersebut diakibatkan oleh *mobile internet*, *data mining*, *automation of knowledge*, *internet of things*, dan *cloud technology*.

E-commerce

E-commerce atau perdagangan elektronik sebagai wadah transaksi jual beli barang dan jasa melalui internet (Khan, 2016). *E-commerce* kegiatan bisnis yang menggunakan teknologi informasi untuk pengoptimalan penjualan, efisiensi bisnis dan

mengenalkan produk serta layanan yang baru (Isoraité dan Neringa, 2018). *E-commerce* berkorelasi dengan aktivitas perdagangan barang dan jasa yang pada bidang elektronik, transaksi keuangan dalam perbankan, sewa guna usaha, asuransi dan jasa keuangan lainnya, investasi, transaksi mata uang dan sekuritas; hotel, pariwisata, pendidikan, konsultasi, pembayaran utilitas, iklan, dan lainnya. *E-commerce* jika dibandingkan dengan toko *offline*, situs web dalam *e-commerce* lebih dekat dengan permintaan pasar dan lebih mudah diterapkan untuk harga diferensial untuk pembeli. Misalnya, harga tinggi dibebankan kepada konsumen di daerah berpenghasilan tinggi, dan harga rendah diberikan kepada konsumen yang sering menggunakan situs pembandingan harga. Tak luput harga yang ditawarkan dapat dibeli oleh masyarakat dengan penghasilan yang berbeda.

Data mining

Data mining digunakan untuk menangkap penemuan mengenai pengetahuan dalam *database* (Turban *et al*, 2005). Data mining sering digunakan untuk mengekstraksi pola yang sebelumnya tidak diketahui dan berpotensi tersedia dari sejumlah besar data. Hal tersebut dapat mengkombinasikan dari teknologi kecerdasan buatan (AI) dan teknologi basis data. Pada implementasinya juga melibatkan pembelajaran mesin, pengenalan pola, statistik, basis data cerdas, akuisisi pengetahuan, visualisasi data, sistem pakar, dll. Sisi lainnya penggunaan data mining melalui web dapat memperoleh informasi berharga seperti perilaku dan preferensi pengguna yang kemudian dianalisis dengan lebih efektif. Selanjutnya digabungkan dengan sarana bisnis lain, sehingga berdampak pada manfaat bisnis bagi perusahaan. Meluasnya penggunaan perangkat pintar, sejumlah besar informasi mengalir ke Internet setiap hari.

Penelitian Terdahulu

Penelitian yang dilakukan oleh Hong Pan & Hanxun Zhou (2020) dengan judul “Study on convolutional neural network and its

application in data mining and sales forecasting for E-commerce” menjelaskan jika *convolutional neural network* digunakan untuk data mining dari *e-commerce* guna mencapai prediksi penjualan atas barang. Penggunaan konsep tersebut secara otomatis mengelola fitur efektif dari data pengaturan waktu terstruktur, yang dapat menghindari rekayasa fitur artifisial yang menghabiskan waktu, tenaga, dan membutuhkan pengetahuan lapangan dari personel yang menjalankan fitur tersebut rekayasa. Metode tersebut akan memudahkan para pengguna data mining dalam mencapai perkiraan penjualan yang efektif.

Penelitian lain yang dilakukan oleh Lei He & Yixuan He (2021) dengan judul “Cross-border E-commerce logistics development model based on data mining and QoS” menjelaskan bahwa metode peramalan deret waktu atau *forecasting method* yang biasa digunakan untuk memproyeksikan QoS layanan *e-commerce*. Bermula dari adanya pelanggaran QoS layanan *e-commerce* setiap waktu sehingga berdampak pada pengguna dan penyedia layanan *e-commerce*. Solusi untuk mengurangi kelemahan dari metode peramalan yang ada, seperti tingkat akurasi yang rendah, pengelolaan jumlah informasi yang besar dan biaya prediksi yang tinggi, penggunaan metode peramalan QoS layanan *e-commerce* berdasarkan data mining menunjukkan implikasi yang positif.

Penelitian yang dilakukan oleh Lixia Zhao & Wei Jin (2022) dengan judul “Application of Web Data Mining and Information Combination Technology in E-commerce” menunjukkan bahwa algoritme rekomendasi dari gabungan teknologi pemfilteran kolaboratif berbasis pengguna dan algoritme pemfilteran kolaboratif berbasis item untuk digunakan dalam sistem *e-commerce* dapat mempertajam kekuatan dan menghindari kelemahan daya saing inti situs web. Penelitian dan pengembangan data mining saat ini menunjukkan bahwa data mining perlu mencakup berbagai tugas aplikasi yang

berbeda, mulai dari preprocessing data hingga penemuan aturan asosiasi, analisis klaster, klasifikasi data, pengecekan deviasi, analisis pola urutan dan tugas spesifik lainnya.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif deskriptif dengan metode studi kepustakaan dan wawancara. Penelitian kualitatif bertujuan untuk menganalisis gejala, fenomena, maupun realitas secara mendalam melalui serangkaian proses (Semaiwan, 2010). Penggunaan metode studi kepustakaan dilakukan melalui pengumpulan bahan literatur lalu dilanjutkan dengan literasi dan mencatat informasi penting yang mana pada akhirnya diolah menjadi bahan penelitian (Zed, 2004). Selanjutnya, penelitian ini juga melakukan wawancara dengan informan dalam rangka memperbanyak data informasi sebagai bahan literatur tambahan penelitian.

Penelitian ini menggunakan data primer yang berasal dari informan wawancara untuk mendapatkan informasi yang diperlukan oleh peneliti. Informan wawancara berasal dari DJP sub Direktorat Potensi, Kepatuhan, dan Penerimaan Perpajakan dengan inisial S & Y. Sebagai mendukung analisis data, peneliti membutuhkan sumber data sekunder yang berasal dari berbagai pustaka, seperti jurnal, artikel, prosiding seminar maupun laporan yang berkaitan dengan topik penelitian, seperti *data mining* dan *e-commerce*.

Penelitian ini menggunakan teknik analisis isi (*content analysis*). Analisis isi merupakan sebuah teknik untuk memperoleh kesimpulan dengan mengidentifikasi berbagai partikularitas suatu informasi secara objektif, sistematis, dan generalis (Asfar & Taufan, 2019). Teknik ini digunakan untuk menganalisis secara mendalam dari informasi tertulis yang dimiliki. Hal itu dilakukan agar data dapat tersusun secara sistematis dan mendapatkan hasil analisis data kualitatif yang valid, kredibel, serta berkualitas.

Tahapan penelitian ini diawali dengan melakukan identifikasi permasalahan terlebih dahulu terkait ekonomi digital dan *e-commerce* serta masalah penggalian data perpajakan untuk menggali potensi penggunaan *data mining* di Indonesia. Selanjutnya, melakukan studi pustaka dari berbagai literatur yang relevan, absah, dan kredibel terkait permasalahan yang diteliti, khususnya *data mining*. Lalu, penelitian ini dilanjut dengan proses pengumpulan data melalui sumber data primer dengan wawancara kepada informan, yaitu S & Y yang berasal dari sub Direktorat PKP dan sumber data sekunder dengan menggunakan jurnal, artikel, laporan, maupun prosiding yang diperoleh dari google scholar, sciencedirect, dan sumber kredibel lainnya. Nantinya, data yang telah dikumpulkan akan dianalisis secara mendalam dan dilakukan interpretasi data agar dapat memilah informasi mana yang dapat disajikan secara sistematis dan inklusif serta dapat tercipta kesimpulan penelitian. Pada akhirnya, hasil penelitian akan disajikan naratif dalam bentuk jurnal dan akan dilakukan publikasi karya ilmiah.

HASIL DAN PEMBAHASAN

Implementasi Data Mining dalam Perpajakan di Berbagai Negara

1. Korea Selatan

Penggunaan data mining pada negara Korea Selatan berporos pada peningkatan efisiensi produksi terutama sektor perpajakan. Sektor ini membutuhkan kemampuan yang profesional, berangkat dari komputerisasi saat proses pajak dilakukan. Selanjutnya sejumlah data dalam kategori besar masuk dan bertemu pada titik yang sama sehingga tahapan data mining dapat dilakukan. Data mining membantu negara Korea Selatan ketika para akuntan pajak yang berkecimpung di dalam profesi yang relevan menerapkan data mining ketika mengakumulasi data pajak tersebut. Teknik data mining diimplementasikan

dalam bentuk random forest dan dikembangkan menggunakan f1-score. Random Forest atau hutan acak digunakan berdasarkan data yang diberikan oleh perusahaan terkait pajak dalam negeri. Metode ini berkaitan dengan teknik ansambel yang digunakan untuk analisis regresi klasifikasi. Algoritma random forest yang secara acak dapat dikelompokkan sesuai dengan kategori data.

Istilah skor-f1 atau f1-score diterapkan untuk menunjang proses data mining, rata-rata data yang harmonik dengan presisi dan perolehan kembali membuat data secara acak. Hal ini dapat terselesaikan dengan penggunaan f1-score untuk memberikan ukuran data yang akurat ketika data tidak seimbang. Tipe data yang masuk berupa faktur pajak elektronik yang beragam jumlahnya sehingga dilakukan analisis silang untuk mereduksi data menjadi informasi yang mudah terbaca. Klasifikasi dilakukan terlebih dahulu menurut jenis transaksinya, kemudian diproses terlebih dahulu. Data yang diproses sebelumnya diubah menjadi bentuk yang sesuai untuk dimasukkan ke dalam model hutan acak, dan mengeksekusi prediksi. Hasil yang didapatkan ketika menerapkan sistem tersebut selama 2 tahun menunjukkan akurasi yang tinggi sesuai dengan prediksi (Choi et al, 2019).

Penelitian yang dilakukan oleh Francesco et al (2017) yang menjelaskan masalah dalam sistem pendukung keputusan kelompok melalui algoritma k-means. Dikombinasikan dengan proses kerja penelitian hirarkis menghasilkan keputusan yang yang berimplikasi secara positif dari tugas yang diberikan. Keputusan tersebut menjadi alternatif

penyaringan dan diurutkan berdasarkan tingkat kepentingannya. *Bezier curve fitting* digunakan untuk mendapatkan fungsi preferensi pembuat keputusan. Model k-means yang digunakan dalam penelitian ini memilih alternatif berdasarkan kelas dan kendala anggaran yang menghasilkan kinerja yang lebih baik.

2. Chile

Penggunaan *data mining* dalam mendeteksi adanya penggunaan faktur pajak yang palsu pada sistem pembayaran wajib pajak. Algoritma *data mining* yang digunakan dari pengelompokan SOM dan gas saraf untuk mengidentifikasi kelompok wajib pajak dengan perilaku yang sama ketika membayar pajak. Selanjutnya *data mining* berupa *decision tree*, *neural network*, dan *bayesian network* digunakan untuk mengidentifikasi variabel-variabel yang terkait dengan perilaku penipuan atau tidak ada penipuan. Mendeteksi pola perilaku yang relevan dan menentukan proses kasus penipuan atau tidak ada penipuan dapat dideteksi (Gonzalez et al, 2013).

Informasi yang diperoleh akan membantu mengidentifikasi pola penipuan dan menghasilkan pengetahuan yang dapat digunakan dalam pekerjaan audit dalam administrasi pajak Negara Chile. Pendeteksian ini juga menggunakan metode pohon keputusan yang diterapkan pada kasus-kasus di mana hasil kecurangan dan tidak adanya kecurangan diketahui. Hal ini dikarenakan ketika menganalisis distribusi variabel pada masing-masing kelompok, diketahui bahwa kasus yang ada cenderung mengambil nilai variabel yang strategis. Dampaknya dapat dibedakan rentang antara peluang terjadinya dan peluang tidak terjadinya

perilaku penipuan. Menurut beberapa ahli hasilnya konsisten sesuai dengan yang terjadi.

Penelitian yang dilakukan Hwang et al (2018) mengulas mengenai sistem pendukung keputusan berbasis data berkonsentrasi pada proses pengumpulan data. Pengumpulan informasi yang penting dapat membantu untuk mendapatkan keputusan yang lebih baik dengan durasi yang singkat. Berawal dari pencarian halaman web dan mengidentifikasi dokumen yang relevan untuk kumpulan tertentu dalam model berbasis dokumen. Pengambilan dan analisis dokumen bergantung pada penyimpanan dan adaptasi teknologi dalam model ini. Selanjutnya data dikumpulkan, data dimanipulasi berdasarkan kebutuhan pengguna, karena sifat data yang terstruktur atau tidak terstruktur akibat sumber internal dan eksternal yang berbeda. Sistem ini dapat membantu kinerja *data mining* dalam sistem pendukung keputusan berbasis data, komunikasi, dokumen, model, dan pengetahuan.

Penggunaan *data mining* dilakukan melalui pohon keputusan dan jaringan syaraf tiruan. Respon yang diberikan dianalisis menggunakan matriks pohon keputusan dan dilatih menggunakan jaringan syaraf tiruan untuk mendapatkan strategi klasifikasi yang lebih baik. Hasil eksperimen kemudian diamati dan dibandingkan dengan Hidden Markov Model konvensional serta sistem pendukung keputusan berbasis vektor pendukung. Model yang diusulkan mencapai hasil klasifikasi yang lebih baik dibandingkan dengan model lain yang dapat diterapkan untuk menyelidiki berbagai domain. Hasil penelitian menunjukkan peningkatan dengan

memasukkan model optimasi untuk optimalisasi kinerja penilaian (Kumar, 2020).

3. Argentina

Di Argentina, penggunaan data dalam mekanisme untuk menyilangkan informasi melalui matriks risiko telah dievaluasi sejak 2010. Model pemrosesan data yang diusulkan untuk menentukan dan menilai kelompok pembayar pajak tertentu, menurut analisis risiko wajib pajak. Metode data mining ini melibatkan keseluruhan cakupan tipe data dari perusahaan dengan sumber data yang berbeda, kemudian perancangan Proses Standar Lintas Industri untuk Data Mining (CRISP-DM). Data mining data memberikan perlindungan hak-hak pembayar pajak, akan tetapi regulasi perpajakan negara Argentina tidak menyebutkan mengenai hak pembayar pajak. Ketika transparansi mengenai file administrasi yang digunakan untuk mendukung model pemrosesan data. Pemerintah Argentina juga menyatakan hak untuk mengetahui alasan suatu pemeriksaan sehingga untuk mempertimbangkan pengetahuan tentang operasi logis atau aritmatika yang mendukung pemeriksaan tersebut (Ugalde et al, 2020).

Penelitian yang dilakukan oleh Pazo & Hallo (2019) mengulik mengenai peran *data mining* dalam administrasi perpajakan yang menunjukkan telah menangkap dan mendaftarkan data wajib pajak karena kepatuhan terhadap kewajiban wajib pajak. Informasi tersebut dikumpulkan agar mampu memberikan informasi mengenai pola perilaku Wajib Pajak ketika berinteraksi dengan administrasi perpajakan. Konsekuensinya, data mining telah membantu mengeksploitasi informasi dan diadopsi

oleh administrasi perpajakan untuk menyelidiki, mengendalikan, mengelola risiko, dan membuat keputusan dalam tahap pengendaliannya. Namun, teknik analitik informasi menghadapi tantangan untuk mengeksploitasi data dengan cara yang efisien sudah tepat karena jumlah transaksi dan jumlah variabel yang digabungkan.

Data mining dalam hal ini menjadi solusi dan tantangan apabila pengelolaan dan pengelompokan data sulit dijelaskan dan divisualisasikan. Biasanya hal tersebut dipengaruhi oleh pengumpulan informasi mengenai transaksi wajib pajak, tetapi database tidak dirancang untuk proses analitik. Misalnya, otoritas pajak harus melalui proses persiapan informasi seperti melabeli data wajib pajak. Selain itu, data yang beragam menyebabkan ledakan kombinatorial ketika teknik *data mining* menemukan pola dalam data. Pendeteksian penggelapan dan pola dalam *database* administrasi pajak menjadi nilai yang berharga secara komputasi, sehingga beberapa proses pengolahan data agar tidak menyalahi aturan perlu adanya dasar hukum untuk dipertimbangkan dalam analisis informasi (Wang et al, 2016).

Efektivitas Data Mining dalam Mendeteksi Tax Evasion di Berbagai Negara

1. Taiwan

Para pelaku bisnis di Taiwan seringkali melakukan tindakan *tax evasion* yang berdampak krusial bagi penerimaan negara. Tindakan *tax evasion* membuat pemerintah harus menanggung biaya pendeteksian dan pencegahan aktivitas penghindaran pajak ilegal. Akibatnya, cara-cara efektif untuk menanggulangi *tax evasion* selalu menjadi isu kritis dan menantang bagi otoritas pajak.

Umumnya, masalah *tax evasion* di taiwan lebih cenderung kearah Pajak Pertambahan Nilai (PPN). Tahun 1999-2001, lebih dari 35% kasus di Taiwan melibatkan penggelapan PPN (Huang & Lin, 2009). Oleh karena itu, penggelapan PPN merupakan salah satu isu penting bagi otoritas pajak di Taiwan.

Penelitian yang dilakukan oleh Wu et al (2012) mengulas khusus bagaimana *data mining* dapat secara efektif mengatasi masalah penggelapan PPN. Penelitian tersebut menggunakan aturan asosiasi teknik *data mining* ke database PPN untuk mengungkap pola dan hubungan antar atribut yang berguna untuk mengidentifikasi laporan *tax evasion*. Model penyaringan ini digunakan untuk menyeleksi kasus-kasus yang diduga sebagai ketidaksesuaian laporan PPN untuk pemeriksaan audit lebih lanjut. Dengan kata lain, tujuan penggunaan *data mining* sebagai teknik untuk mendeteksi penggelapan PPN meningkatkan produktivitas pemeriksa pajak dalam memulihkan kerugian penerimaan pajak (Wu et al, 2012).

Hasil penelitian Wu et al (2012) menunjukkan bahwa penggunaan *data mining* metode *association rule* lebih efektif dan hemat sumber daya dibandingkan dengan penggalan data secara manual. Teknik *data mining* bisa mendapatkan pola *tax evasion* dari melimpahnya data dengan tingkat akurasi yang kuat. Hal itu efektif untuk mencegah adanya potensi *tax evasion*. Dengan demikian, teknik *data mining* dapat digunakan untuk menyaring semua laporan pajak dan mengurangi pemborosan sumber daya staf audit yang tidak perlu. Selain itu, dengan menginformasikan wajib pajak bahwa laporan pajak mereka dapat dianalisis

secara cepat dan ilmiah, tingkat kepatuhan sukarela dapat meningkat dimana akan dapat berimplikasi pada peningkatan pemulihan kerugian pajak.

Data mining tidak hanya digunakan untuk mencegah *tax evasion* di Taiwan, namun bisa juga dimanfaatkan untuk mendeteksi *fraud* laporan keuangan. Penelitian yang dilakukan oleh Lin et al (2015) menggunakan *data mining* dengan metode *logistic regression*, *artificial neural network*, dan *decision tree*. Hasil menunjukkan bahwa ketiga metode tersebut dapat efektif mendeteksi *fraud* dalam laporan keuangan yang dilakukan oleh perusahaan. *Data mining* berkinerja cukup baik dan mungkin menjadi alat pendukung bagi para praktisi. Ini juga menyiratkan bahwa inovasi berkelanjutan dalam *artificial intelligence* diperlukan dan dapat diterapkan untuk memfasilitasi evaluasi bukti audit (Lin et al, 2015).

2. Chile

Urgensi permasalahan tindakan *tax evasion* terkait PPN juga dialami oleh negara Amerika Latin, khususnya Chile. PPN mewakili 45% sebesar USD \$18,7 miliar dolar dan menghasilkan lebih dari 400 juta faktur per tahun, dimana 56% diterbitkan dalam format kertas dan 44% dalam format elektronik (Bergman, 2010). Kasus utamanya adalah pemalsuan faktur PPN. Kasus tersebut sangat kompleks dan sulit dideteksi karena ada keterlibatan transaksi fiktif didalamnya. Akibatnya, otoritas pajak Chile harus menanggung beban yang cukup besar dan memakan waktu yang lama (Gonzalez & Velazquez, 2013).

Penelitian yang dilakukan oleh Gonzales & Velazquez (2013) menjelaskan bagaimana *data mining* dapat mengkarakterisasi dan

mendeteksi perilaku *tax evasion* wajib pajak yang menggunakan faktur palsu. *Data mining* dipercaya memiliki potensi yang besar untuk mengkarakterisasi dan mendeteksi perilaku *tax evasion* dari sumber data yang luas dan melimpah. Chile dapat belajar dari Peru yang merupakan negara Amerika Latin pertama dalam menerapkan *data mining* untuk mengatasi *tax evasion*. (García & Valderrama, 2007). Otoritas pajak disana menggunakan *artificial intelligence* berbasis *data mining* dengan teknik *neural network*, *fuzzy*, dan *association rule* serta *classification tree*. Hal itu ditujukan untuk memilih dan memilah variabel yang relevan dengan permasalahan *tax evasion*.

Hasil penelitian Gonzales & Velazquez (2013) menunjukkan bahwa Chile menerapkan *data mining* pertama kali pada tahun 2007 dengan teknik SOM dan K-means. Pada tahun 2009, dilakukan pengembangan sesuai tren internasional berupa penambahan teknik *neural network*, *decision tree*, dan *logistic regression*. Metode klasifikasi yang digunakan untuk mencirikan wajib pajak yang memiliki perilaku fiskal baik atau buruk terkait dengan penggunaan faktur palsu menunjukkan bahwa beberapa karakteristik pembeda antara satu kelompok dengan kelompok lainnya dapat diidentifikasi, sesuai dengan apa yang terjadi di lapangan. Metode *decision tree* yang diterapkan merupakan teknik yang baik untuk mendeteksi variabel yang dapat membedakan antara kecurangan dan tidak ada kecurangan. Sehubungan dengan model deteksi, yang berkinerja lebih baik adalah model *neural network perceptron multilayer*, dimana dapat mendeteksi kasus penipuan faktur palsu

oleh usaha kecil dan menengah dengan benar adalah 92%, sedangkan dalam kasus usaha menengah dan besar, persentase ini adalah 89%.

Sejalan dengan penelitian Faundez-Ugalde et al (2020) yang menjelaskan bahwa teknik karakterisasi *data mining* penting bagi Chile dan negara Amerika Latin lainnya. Hal itu dikarenakan untuk mendeteksi sumber *tax evasion* dan meningkatkan kepatuhan sukarela wajib pajak. Teknik yang paling banyak digunakan untuk memvalidasi model perilaku kontributor adalah algoritma *Self Organizing Feature Map* (SOFM), K-means, *neural network*, dan *bayesian network*. Data yang menjadi dasar teknik tersebut antara lain berasal dari Surat Pemberitahuan (SPT) yang terkait dengan siklus hidup wajib pajak, rasio pendapatan dan profitabilitas perusahaan dengan likuiditasnya dan indikator yang terkait dengan perilaku historis, perilaku terkait dan karakteristik khususnya.

3. Iran

Perilaku *tax evasion* di Iran membuat potensi penerimaan pajak hilang. Padahal, pajak merupakan sumber utama pendapatan Iran. Penelitian yang dilakukan oleh Rad & Shahbahrami (2015) mengulas tentang bagaimana implementasi kinerja *data mining* dengan menggunakan teknik algoritma khusus dapat mengatasi dan mendeksi *tax fraud* di Iran. Peneliti menggunakan teknik *bayesian* dalam mendeteksi *fraud* melalui prosedur paralel.

Hasil penelitian Rad & Shahbahrami (2015) menunjukkan bahwa *smart system* yang disebut sistem deteksi *tax fraud* dengan menggunakan teknik dari *data mining* dapat mengidentifikasi penipuan di antara

data pajak yang tersedia dengan koefisien kesalahan yang lebih kecil. Jumlah atau volume data pajak yang luas, besar dan melimpah, menjalankan program dalam proses manual memakan waktu yang lama. Penerapan teknologi paralelisme baru pada bagian-bagian prosedur pajak efektif untuk mengurangi waktu dalam menjalankan program. Sistem ini diterapkan pada struktur sistem perpajakan di Iran untuk mengatasi *tax fraud*.

Sejalan dengan penelitian Mojahedi et al (2022) yang memaparkan bahwa *tax evasion* merupakan masalah utama dari sistem perpajakan di sebagian besar negara yang harus diatasi dengan menggunakan metode yang dapat mengidentifikasi kasus *tax evasion* dalam administrasi perpajakan. Sejak mesin algoritma pembelajaran memiliki prediktif dan klasifikasi fitur, proses pengambilan keputusan dalam masalah keuangan dapat difasilitasi. Selain itu, *neural network* menyediakan biaya rendah solusi algoritmik dan memfasilitasi analisis karena mereka tidak memerlukan asumsi statistik yang berbeda. Sebelumnya, sistem ini diterapkan oleh administrasi pajak dataset menggunakan metode validasi silang 10 kali lipat dan metode pelatihan berulang, pengujian, dan validasi.

Hasil paper pada 1500 sampel pajak di Iran menunjukkan bahwa pajak penghindaran dapat dideteksi menggunakan metode pembelajaran mesin. Keakuratan model IPSO-MLP lebih dari 93%. Sangat kecil kemungkinan model tersebut mengalami kesalahan. Evaluasi dari *hidden layer active neuron* dan pelatihan model *artificial neural network* menunjukkan bahwa 30 iterasi siklus dengan sepuluh neuron lapisan

tersembunyi sebagai optimalisasi *artificial neural network* cocok untuk mendeteksi perilaku *tax evasion*. Selanjutnya model IPSO-SVM, SVM, dan MLP menunjukkan hasil yang baik juga dibandingkan penelitian sebelumnya (Mojahedi et al, 2022).

Potensi dan Adopsi Data Mining bagi Perpajakan di Indonesia

Beragam *benchmarking* telah dilakukan, meninjau potensi untuk diterapkan di Indonesia sangat besar. Pertama, masuknya data dalam kumpulan yang besar akan diproses menggunakan *machine learning*, *random forest*, *artificial intelligence*, dan *data mining*. Berdasarkan analisis dengan negara lain, cara tersebut meminimalisir sejumlah masalah dalam sistem komputerisasi. Penerapan teknologi dengan sistem komputerisasi tersebut, akan berpeluang dalam meminimalisir beban kerja yang ditanggung petugas pajak. Kedua, sistem administrasi perpajakan menjadi lebih efektif dan efisien. Pengumpulan data yang dikelola secara cepat dalam wujud *data mining* yang bersifat kesederhanaan dan memberikan kemudahan.

Penggunaan utama *data mining* dalam administrasi perpajakan adalah untuk mengkarakterisasi, mengidentifikasi kelompok kepentingan untuk mengendalikan penggelapan, membuat model, dan merancang metode untuk mengendalikan pembayar pajak (Pazo et al, 2019). Ketiga, penerapan data mining dapat mendukung peningkatan target penerimaan pajak. Misalnya, dengan menggunakan algoritma apriori dalam analisis permasalahan pencapaian target dengan meninjau *rule interesting* atau aturan menarik (Eka et al, 2018). Setiap daerah di Indonesia apabila menggunakan sistem *data mining* dalam pencapaian target penerimaan pajak atas tiap daerahnya mempunyai potensi yang tinggi. Skema kebijakan dari pemerintah berawal dari daerah di masing-masing provinsi kemudian dapat ditransfer ke pemerintah pusat. Keempat, potensi pengendalian dalam audit sehingga

mengurangi adanya kasus penggelapan pajak maupun penghindaran pajak dalam implementasi *data mining*. Kelima, peluang untuk menjadi sarana yang mendukung penyediaan layanan informasi yang sesuai dengan kebutuhan wajib pajak. *Data mining* sebagai alat layanan menguntungkan kepatuhan sukarela wajib pajak.

Melihat komparasi dan efektivitas *data mining* dalam mencegah, mendeteksi, dan menanggulangi perilaku *tax evasion* oleh wajib pajak, Indonesia dapat melakukan *benchmarking* melalui kebijakan beberapa negara untuk diadopsi. Tujuannya adalah sebagai kombinasi untuk mengoptimalkan besarnya potensi *e-commerce* di Indonesia. Terlebih lagi, ketidakpatuhan, penghindaran, dan penggelapan pajak yang marak dilakukan oleh pelaku usaha digital seringkali terjadi. Masalah tersebut dapat teratasi dengan mengadopsi *data mining* beserta tekniknya, seperti *classification tree*, *neural network*, *bayesian network*, dan *association rule*.

Sejauh ini Indonesia telah menggunakan teknologi *data mining* dalam beberapa bidang saja dan tidak terlalu kompleks dalam aspek perpajakan. Indonesia telah mengadopsi untuk gempa bumi terhadap potensi tsunami, *data mining* untuk pengelompokan wisata asing metode K-means, menentukan status provinsi dengan metode *clustering*, dan data transaksi penjualan dengan metode *frequent pattern growth* (Utomo & Purba, 2019; Ardiansyah et al, 2020; Simanjuntak & Windarto, 2020; Sari & Hartama, 2018). Dalam aspek perpajakan, baru sebagian kecil saja penerapan yang dilakukan oleh otoritas pajak, seperti penggunaan *data mining* metode *clustering* untuk mendeteksi kepatuhan wajib pajak bumi dan bangunan dan *data mining* untuk memprediksi besarnya pajak kendaraan (Putri et al, 2022; Suhada et al, 2021). Maka dari itu diperlukan adopsi kebijakan yang lebih luas lagi dengan melihat Chile, Korea, Taiwan maupun Iran.

PENUTUP

Kesimpulan

Pengelolaan sejumlah data dan transaksi yang membutuhkan sistem secara terintegrasi untuk memudahkan DJP dalam penggalian potensi pajak atas *e-commerce*, salah satunya dengan menggunakan *data mining*. DJP selaku otoritas pajak dapat melakukan *benchmark* terhadap beberapa negara yang telah maksimal menerapkan *data mining*. Penggunaan utama *data mining* dalam administrasi perpajakan adalah untuk mengkarakterisasi, mengidentifikasi *tax evasion*, membuat model, dan merancang metode untuk mengendalikan pembayar pajak. Hal itu dikarenakan penggunaan *data mining* melalui beberapa teknik seperti *random forest*, *neural network*, *decision tree*, *bayesian network*, dan *association rule* dapat secara efektif mencegah, mendeteksi, dan menanggulangi perilaku *tax evasion* oleh wajib pajak. Terlebih lagi, ketidakpatuhan, penghindaran, dan penggelapan pajak yang marak dilakukan oleh pelaku usaha digital seringkali terjadi.

Terdapat beberapa rekomendasi untuk pemerintah dan penelitian selanjutnya:

1. Pemerintah perlu melakukan kajian ulang untuk proses pengelolaan data-data dari *e-commerce* yang tidak terdeteksi keberadaannya sehingga dapat menangkap potensi pihak-pihak yang harus dipajaki. Berkaca lanskap ekonomi digital yang luas dengan melimpahnya data dari berbagai sumber, menjadikan *data mining* sebagai alternatif terbaik untuk pengelolaannya.
2. Pemerintah perlu merumuskan regulasi yang sesuai dengan penerapan *data mining* untuk mengantisipasi pengaksesan algoritma dalam *data mining* yang diakses wajib pajak diluar kendali. Pengoperasian harus di bawah mekanisme pengamanan yang terjaga

dengan penerapan yang efektif dan efisien.

3. Penelitian selanjutnya dapat melakukan analisis *data mining* yang lebih komprehensif untuk penentuan sampel dan pengelolaan data secara sistematis sehingga dapat mengurangi adanya ketidaksesuaian dengan hasil.
4. Penelitian selanjutnya dapat melakukan penilaian performa kinerja yang rendah untuk mengklasifikasikan data yang masuk dengan data yang ada dengan memperbaiki model atau bentuk prediksi.

DAFTAR PUSTAKA

- [1] Aminudin, M. (2015). Data Mining Aplikasi Analisis Asosiasi Data Pasien Menggunakan Algoritma Apriori Di Puskesmas Kabupaten 'X' (Doctoral dissertation, Universitas Muhammadiyah Gresik).
- [2] Aprilia, N. D. (2021). PERKEMBANGAN EKONOMI DIGITAL INDONESIA. *Ekonomi Pertahanan*, 7(2), 245-259.
- [3] Ardiansyah, A. H., Nugroho, W., Alfiah, N. H., Handoko, R. A., & Bakhtiar, M. A. (2020, August). Penerapan Data Mining Menggunakan Metode Clustering untuk Menentukan Status Provinsi di Indonesia 2020. In *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)* (Vol. 4, No. 3, pp. 329-333).
- [4] Asfar, I. T., & Taufan, I. (2019). Analisis Naratif, Analisis Konten, Dan Analisis Semiotik (Penelitian Kualitatif). *no. January*, 1-13.
- [5] Bergman, M. (2010). *Tax evasion and the rule of law in Latin America: the political culture of cheating and compliance in Argentina and Chile*. Penn State Press.
- [6] Choi, D. Bin., Jo, In. S., & Park, Y. B. (2019). Taxation Analysis Using Machine Learning. *Journal of Semiconductor Display Technology*, 18(2), 73-77.
- [7] Dharmawan, Y. E., & Adi, P. H. (2021). Kompleksitas Pajak, Moral Wajib Pajak dan Norma Subyektif Terhadap Kepatuhan Perpajakan. *Jurnal Penelitian dan Pengembangan Sains dan Humaniora*, 5(2), 212-219.
- [8] Djufri, M (2020). Penerapan Teknik Web Scraping Untuk Penggalan Potensi Pajak (Studi Kasus Pada Online Market Place Tokopedia, Shopee Dan Bukalapak). *Jurnal BPPK: Badan Pendidikan Dan Pelatihan Keuangan*, 13(2), 65-75.
- [9] Faúndez-Ugalde, A., Mellado-Silva, R., & Aldunate-Lizana, E. (2020). Use of artificial intelligence by tax administrations: An analysis regarding taxpayers' rights in Latin American countries. *Computer Law & Security Review*, 38, 105441.
- [10] Francesco, L., Alessio I., Rita, G., Bianca, R.,, Laura, P. (2017). Requalifying public buildings and utilities using a group decision support system. *Journal of Cleaner Production*, 164, 1081-1092.
- [11] García, V., & Valderrama, J. (2007). Toward a more efficient tax policy. *An Opportunity for a Different*, 103.
- [12] Gonzales, P. C., & Velasquez, Juan. D. (2013). Expert Systems with Applications. *Elsevier*, 40, 1427-1436.
- [13] Google, Temasek, dan Bain & Company. (2022). *e-Conomy SEA 2022 Report*.
- [14] Hasanah, R. A. (2016). *Pengaruh pemahaman peraturan pajak, tarif pajak, lingkungan, dan kesadaran wajib pajak terhadap kepatuhan wajib pajak pengguna e-commerce: Studi Kasus Pada Pengusaha Online Shopping* (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).
- [15] He, L., & He, Y. (2021). Cross-border E-commerce logistics development model based on data mining and QoS.

-
- Proceedings of the 6th International Conference on Communication and Electronics Systems*, 1376-1379.
- [16] Huang, M. J., & Lin, S. M. (2009). Tax evasion and audit selection. *The International Journal of Accounting Studies*, 48, 35-66.
- [17] Hwang, B. G., Shan, M., & Looi, K. Y. (2018). Knowledge-based decision support system for prefabricated prefinished volumetric construction. *Automation in Construction*, 94, 168-178.
- [18] Kumar, T. S. (2020). Data mining based marketing decision support system using hybrid machine learning algorithm. *Journal of Artificial Intelligence*, 2(03), 185-193.
- [19] Lin, C. C., Chiu, A. A., Huang, S. Y., & Yen, D. C. (2015). Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments. *Knowledge-Based Systems*, 89, 459-470.
- [20] Mojahedi, H., Babazadeh Sangar, A., & Masdari, M. (2022). Towards Tax Evasion Detection Using Improved Particle Swarm Optimization Algorithm. *Mathematical Problems in Engineering*, 2022.
- [21] Nurdin, N., & Astika, D. (2019). Penerapan Data Mining Untuk Menganalisis Penjualan Barang Dengan Menggunakan Metode Apriori Pada Supermarket Sejahtera Lhokseumawe. *TECHSI-Jurnal Teknik Informatika*, 7(1), 132-155.
- [22] Pan, H., & Zhou, H. (2020). Study on convolutional neural network and its application in data mining and sales forecasting for E-commerce. *Electronic Commerce Research*, 20(2), 297-320.
- [23] Pazo, J. O., & Hallo, Maria. (2019). Data Mining Techniques Applied in Tax Administrations: A Literature Review. *IEEE*, 224-229.
- [24] Prastyaningtyas, E. W. (2019). Dampak Ekonomi Digital bagi Perekonomian Indonesia. In *seminar nasional manajemen, Ekonomi, Akuntansi* (Vol. 1, No. 1, pp. 103-108).
- [25] Pusdiklat Pajak. (2019). E-Commerce Dalam Pajak Penghasilan. Badan Pendidikan dan Pelatihan Keuangan: Jakarta Selatan.
- [26] Putri, M. A., Rahaningsih, N., Basysyar, F. M., & Nurdiaawan, O. (2022). Penerapan Data Mining Menggunakan Metode Clustering Untuk Mengetahui Kelompok Kepatuhan Wajib Pajak Bumi dan Bangunan. *Jurnal Accounting Information System (AIMS)*, 5(2), 145-156.
- [27] Rad, M. S., & Shahbahrami, A. (2015, December). High performance implementation of tax fraud detection algorithm. In *2015 Signal Processing and Intelligent Systems Conference (SPIS)* (pp. 6-9). IEEE.
- [28] S & Y. (2022). Personal communication, August 10, 2022.
- [29] Sari, R. W., & Hartama, D. (2018, July). Data Mining: Algoritma K-Means Pada Pengelompokan Wisata Asing ke Indonesia Menurut Provinsi. In *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)*, 1(1).
- [30] Semiawan, C. R. (2010). *Metode penelitian kualitatif*. Grasindo.
- [31] Simanjuntak, H. E., & Windarto, W. (2020). Analisa Data Mining Menggunakan Frequent Pattern Growth pada Data Transaksi Penjualan PT Mora Telematika Indonesia untuk Rekomendasi Strategi Pemasaran Produk Internet. *Jurnal Media Informatika Budidarma*, 4(4), 914-923.
- [32] Suandy, E. (2008). *Perencanaan Pajak (ed. 4) HVS*. Penerbit Salemba.
- [33] Suhada, S., Ginting, G. L., & Hondro, R. K. (2021). Penerapan Data Mining Untuk Memprediksi Besarnya Pembayaran
-

- Pajak Kendaraan Pada:(Badan Pengelolaan Pajak Dan Retribusi Daerah Upt Samsat Medan Selatan) Menggunakan Algoritma Expectation Maximization. *Bulletin of Information Technology (BIT)*, 2(2), 69-75.
- [34] Tang, Y., Li, Y., Qu, M., & Yang, Q. (2021, December). Research on Risk Early Warning Methods of Enterprise Finance and Taxation Management Based on Big Data Mining. In *2021 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS)* (pp. 157-161). IEEE.
- [35] Utomo, D. P., & Purba, B. (2019, September). Penerapan data mining pada data gempa bumi terhadap potensi tsunami di Indonesia. In *Prosiding Seminar Nasional Riset Information Science (SENARIS)* (Vol. 1, pp. 846-853).
- [36] Wang, J., Yu, X., Li, P., & Xia, J. (2016). The research and application of tax risk identification based on pso-bp algorithm. *International Conference on Audio, Language and Image Processing (ICALIP)*, 120–125.
- [37] Wicaksono, B. (2018). Meningkatkan potensi pajak UMKM online melalui data e-commerce. *Symposium Nasional Keuangan Negara*, 141-161.
- [38] Wu, R. S., Ou, C. S., Lin, H. Y., 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.
- [39] Zed, M. (2004). *Metode penelitian kepustakaan*. Yayasan Obor Indonesia.
- Zhao, L., & Jin, W. (2022). Application of Web Data Mining and Information Combination Technology in E-commerce. *Proceedings of the International Conference on Electronics and Renewable Systems*, 1636-163