

Prediksi Keikutsertaan Pelaku Usaha dalam Pemanfaatan Insentif Pajak dengan *Artificial Neural Network*

Arifin Rosid^{a,b,*}, Galih Ardin^a, & Tri Bayu Sanjaya^a

^aDirektorat Jenderal Pajak, Kementerian Keuangan Republik Indonesia

^bFakultas Ekonomi dan Bisnis, Universitas Indonesia

Abstract

The provision of tax incentives is one of critical fiscal policies amidst the Covid-19 pandemic. The characteristics of businesses related to the use of incentives are essential factors in policy formulation. This study offers an Artificial Neural Network (ANN) approach to predict the participation of businesses in tax incentives based on their characteristics. The ANN model in this study uses survey data on the number of workers, primary market share, annual turnover, the nature of main business activities, and main source of supply from 12,361 businesses. The ANN approach in this study predicts with an accuracy rate of 70%. The results demonstrate that the number of employees, the annual turnover, and the primary market share are the three most important variables that determine the participation of Indonesian businesses in utilising tax incentives.

Keywords: *Indonesia; Covid-19; tax incentive; Artificial Neural Network*

Abstrak

Pemberian insentif pajak adalah salah satu kebijakan fiskal yang penting selama masa pandemi Covid-19. Karakteristik dari pelaku usaha yang terkait dengan pemanfaatan insentif adalah informasi penting di dalam perumusan kebijakan. Studi ini menawarkan pendekatan *Artificial Neural Network* (ANN) untuk memprediksi keikutsertaan pelaku usaha dalam insentif pajak berdasarkan karakteristik yang dimiliki. Model ANN dalam studi ini menggunakan data empiris jumlah pekerja, pangsa pasar utama, besaran omzet tahunan, sifat usaha utama, dan sumber utama pasokan dari 12.361 pelaku usaha hasil survei. Pendekatan ANN dalam studi ini memprediksi dengan tingkat akurasi sekitar 70%. Hasil studi ini menunjukkan bahwa jumlah pekerja, omzet tahunan, dan pangsa pasar utama adalah tiga variabel terpenting yang menentukan keikutsertaan pelaku usaha dalam pemanfaatan insentif pajak.

Kata Kunci: *Indonesia; Covid-19; insentif pajak; Artificial Neural Network*

Kode Klasifikasi JEL: D04; H20; H25; H30

*Alamat Korespondensi Penulis: Kantor Pusat Direktorat Jenderal Pajak, Jalan Gatot Subroto Kav. 40-42 Jakarta 12190 Indonesia. *Email:* arifin.rosid@pajak.go.id; arifin.rosid@ui.ac.id.

1. Pendahuluan

1.1. Latar Belakang

Badan Pusat Statistik (2020a)–yang selanjutnya disingkat BPS–mencatat bahwa selama tahun 2020 terjadi penurunan aktivitas ekonomi yang disebabkan oleh pandemi Covid-19. Penurunan aktivitas ini tercermin dari anjloknya pertumbuhan ekonomi pada triwulan II sampai dengan triwulan IV tahun 2020 yang terjadi pada hampir semua sektor ekonomi. Lebih lanjut, berdasarkan survei yang dilakukan oleh BPS (2020b) juga diketahui bahwa pandemi Covid-19 telah menyebabkan penurunan omzet, kenaikan biaya, dan peningkatan angka pemutusan hubungan kerja pada tahun 2020.

Banyak negara telah mengambil kebijakan yang tegas untuk menahan dan mengurangi dampak ekonomi dari pandemi Covid-19 (OECD, 2020a). Guna mengurangi dampak negatif yang diakibatkan oleh pandemi Covid-19 khususnya pada sektor ekonomi, pemerintah Indonesia juga telah menerbitkan berbagai kebijakan ekonomi salah satunya pemberian insentif pajak. Insentif pajak diperlukan sebagai peredam awal untuk guncangan ekonomi yang disebabkan oleh anjloknya permintaan dan penawaran agregat.¹ Selain itu, insentif pajak juga diperlukan untuk membantu Wajib Pajak terdampak pandemi Covid-19 dan membantu pemulihan ekonomi nasional.

Setidaknya, ada lima jenis insentif pajak yang diberikan pemerintah bagi Wajib Pajak terdampak pandemi Covid-19 sebagaimana diatur dalam Peraturan Menteri Keuangan Nomor PMK-23/PMK.03/2020 sebagaimana telah diubah terakhir dengan PMK-82/PMK.03/2021. Jenis insentif pajak dimaksud adalah (i) insentif PPh Pasal 21 ditanggung Pemerintah (DTP), (ii) insentif pembebasan PPh Pasal 22 impor, (iii) insentif PPh Final peredaran bruto tertentu DTP, (iv) insentif pengurangan angsuran PPh Pasal 25, dan (v) insentif pengembalian pendahuluan PPN dipercepat.

Berdasarkan *Laporan Keuangan Pemerintah Pusat (LKPP) Tahun 2020 (audited)* (Kementerian Keuangan RI, 2021) diketahui bahwa sampai dengan akhir Desember 2020, dari keseluruhan Wajib Pajak badan dan orang pribadi yang memenuhi kriteria, jumlah Wajib Pajak yang mengajukan permohonan atas 5 jenis insentif tersebut adalah sejumlah 495.817 Wajib Pajak, dengan rincian: 146.068 Wajib Pajak mengajukan permohonan insentif PPh Pasal 21 DTP, 18.753 Wajib Pajak mengajukan permohonan insentif pembebasan PPh Pasal 22 impor, 79.796 Wajib Pajak mengajukan permohonan insentif pengurangan angsuran PPh Pasal 25, 2.593 Wajib Pajak mengajukan permohonan insentif pengembalian pendahu-

¹Konsepsi skema kebijakan selama dan pasca pandemi menurut OECD (2020a) dibagi dalam tiga fase: (i) respons yang segera; (ii) meredam dampak ekonomi dan mempertahankan kapasitas usaha; dan (iii) pemulihan. Salah satu fokus kebijakan pajak dalam fase pertama dan kedua adalah sisi likuiditas (OECD, 2020a). Secara umum, insentif pajak yang diberikan oleh pemerintah melalui Peraturan Menteri Keuangan Nomor PMK-23/PMK.03/2020 sebagaimana telah diubah terakhir dengan PMK-82/PMK.03/2021 berkaitan erat dengan upaya membantu sisi likuiditas pelaku usaha.

luan PPN dipercepat, dan 248.607 Wajib Pajak mengajukan permohonan insentif PPh Final peredaran bruto tertentu DTP. Total nilai permohonan insentif pajak selama tahun 2020 adalah sebesar Rp46,11 triliun (Kementerian Keuangan RI, 2021). Membandingkan jumlah Wajib Pajak yang memanfaatkan insentif dengan jumlah Wajib Pajak yang ada, besar kemungkinan tingkat keikutsertaan para Wajib Pajak atau para pelaku usaha belum maksimal.²

Terkait hal ini, OECD (2020a) merekomendasikan agar insentif pajak ditujukan kepada pelaku usaha yang paling membutuhkan dukungan. Meskipun secara administratif tidak mudah, pemberian stimulus ke target yang tepat akan membantu proses pemulihan bagi pelaku usaha yang paling membutuhkan dengan lebih cepat. Dengan demikian, kemampuan pemerintah untuk mengidentifikasi dan mendorong pelaku usaha agar mendapatkan insentif yang paling diperlukan menjadi prasyarat penting.

Sayangnya, mengidentifikasi faktor-faktor yang relevan dengan keikutsertaan pelaku usaha dalam pemanfaatan insentif pajak bukan pekerjaan yang mudah. Selain membutuhkan ketersediaan data empiris yang memadai diperlukan juga metode analisis data yang tepat. Selain itu, faktor-faktor yang berkaitan dengan keikutsertaan pelaku usaha dalam memanfaatkan insentif pajak banyak berkaitan dengan faktor-faktor yang bersifat non-linier, misalnya pangsa pasar, sumber utama pasokan, sifat usaha, atau lokasi usaha.

Di sisi lain, masih terdapat keterbatasan literatur terkait studi teoritis maupun empiris yang secara spesifik membahas mengenai faktor-faktor yang memengaruhi keikutsertaan pelaku usaha dalam mengikuti insentif pajak di masa pandemi Covid-19. Beranjak dari sini, penulis mencoba menggunakan teknis analisis yang relevan untuk memprediksi dan mengidentifikasi karakteristik atau faktor-faktor yang menentukan tingkat keikutsertaan pelaku usaha dalam pemanfaatan insentif pajak. Salah satu metode analisis data yang populer digunakan dalam area kecerdasan buatan adalah *Artificial Neural Network* (ANN). ANN adalah alat pemodelan non-parametrik yang dapat melakukan pemetaan fungsi-fungsi yang kompleks dengan tingkat akurasi yang memadai (Zhang *et al.*, 1999). Selain itu, di antara kelebihan dari neural network, selain bersifat adaptif terhadap data, adalah kapabilitasnya dalam non-linieritas (Haykin, 1999). Dalam area yang bersifat non-linier, tingkat akurasi dari metode yang ada dalam kecerdasan buatan (*artificial intelligence*) cenderung lebih unggul dari pada metode statistik tradisional (Bahrammirzaee, 2010).

Tulisan ini berkontribusi di dua tataran. Pertama, pada tataran literatur perpajakan, tulisan ini berkontribusi dalam memberikan pengetahuan tambahan mengenai bagaimana pendekatan *artificial intelligence* dapat diterapkan dalam area studi perpajakan. Implikasi empiris dari studi ini dapat dibilang cukup signifikan karena merupakan studi yang pertama kali mencoba memprediksi

²Sebagai gambaran, pada tahun 2019 terdapat 3.323.971 pelaku usaha berbentuk Wajib Pajak badan (Direktorat Jenderal Pajak [DJP], 2021). Dalam tulisan ini penggunaan istilah pelaku usaha lebih sering digunakan untuk menggantikan istilah Wajib Pajak.

keikutsertaan pelaku usaha dalam memanfaatkan insentif pajak selama pandemi menggunakan data observasi level nasional dengan jumlah cukup besar. Kedua, dalam tataran praktis, hasil analisis ini menghasilkan informasi yang relevan bagi pengambil kebijakan, baik dalam fase desain maupun implementasi kebijakan, khususnya terkait dengan pemanfaatan insentif pajak.

Tulisan ini terdiri dari lima bagian utama. Bagian pertama menjelaskan latar belakang dan tujuan studi. Tinjauan literatur dijabarkan pada bagian kedua. Bagian ketiga membahas mengenai metode penelitian, meliputi data yang digunakan dalam studi dan pendekatan yang digunakan untuk menganalisis data. Selanjutnya, bagian keempat menguraikan hasil analisis dan pembahasan dari hasil analisis. Simpulan menjadi bagian penutup dari tulisan ini.

1.2. Tujuan Studi

Studi empiris ini bertujuan untuk menjawab tiga pertanyaan riset. Pertama, seberapa akurat pendekatan ANN dapat memprediksi status keikutsertaan pelaku usaha dalam pemanfaatan insentif pajak? Kedua, tiga faktor apa yang paling relevan dalam memprediksi keikutsertaan pelaku usaha dalam memanfaatkan insentif pajak?³ Ketiga, seberapa kokoh (*robust*) hasil prediksi ANN jika sampel data yang digunakan dilakukan *sampling variability* berbasis lokasi pelaku usaha? Pertanyaan ketiga ini sekaligus bertujuan untuk mengetahui apakah karakteristik utama dari pelaku usaha yang teridentifikasi memiliki hubungan kuat dengan keikutsertaan dalam insentif pajak bersifat homogen di enam kelompok pulau (Jawa, Sumatra, Kalimantan, Sulawesi, Bali – Nusa Tenggara, dan Papua – Maluku).

1.3. Tinjauan Literatur

Peran kebijakan pajak untuk mengurangi dampak pandemi cenderung berubah dari satu fase ke fase berikutnya dalam kurun waktu berlangsungnya pandemi dimaksud. Pada fase pertama pandemi Covid-19, yang ditandai dengan berlangsungnya pembatasan aktivitas masyarakat (*lockdown*) dan disrupsi yang akut baik pada aspek ekonomi maupun kesehatan, insentif pajak diberikan untuk menyediakan tambahan arus kas (*improving cash flow*) bagi dunia usaha maupun rumah tangga untuk mencegah kolapsnya ekonomi secara keseluruhan (Collier *et al.*, 2020; Devereux *et al.*, 2020; OECD, 2020a). Hal itu karena kebijakan fiskal yang diambil di masa-masa awal krisis ditujukan untuk menyerap dampak negatif jangka pendek dari krisis tersebut (Alberola *et al.*, 2021).

³Sebagai sebuah *predictive model*, ANN dapat dikombinasikan dengan model *Decision Tree* untuk analisis lebih lengkap (*International Business Machines* [IBM], 2019). Salah satu pendekatan yang populer dalam model *Decision Tree* adalah *Chi-Squared Automatic Interaction Detection* (CHAID). Dalam pendekatan CHAID, hanya ada tiga tingkatan *tree*—yang dalam hal ini merujuk pada variabel independen—yang dihasilkan (IBM, 2017). Atas dasar pertimbangan ini, penulis hanya fokus pada tiga variabel terpenting.

Pada fase kedua yang ditandai dengan diangkatnya kebijakan *lockdown* dan mulai terlihat tanda-tanda pemulihan, insentif pajak serupa masih perlu diberikan yang perlu dilengkapi dengan insentif pajak untuk menstimulasi permintaan maupun penawaran (Collier *et al.*, 2020; Devereux *et al.*, 2020; OECD, 2020a). Pada fase ketiga yaitu fase jangka panjang, kebijakan pajak perlu didominasi kembali oleh fungsi tradisionalnya yaitu mengumpulkan penerimaan negara. Bahkan, kebijakan pajak pada fase setelah pandemi berakhir dapat ditujukan untuk mengumpulkan penerimaan negara dalam rangka mengganti belanja negara yang telah keluar untuk penanggulangan dampak pandemi (Collier *et al.*, 2020; OECD, 2020a).

Terkait peran insentif pajak yang esensial di masa krisis, sudah banyak literatur yang membahas jenis-jenis insentif pajak yang jamak diaplikasikan untuk menanggulangi dampak pandemi. Secara garis besar insentif pajak dimaksud dapat dibagi menjadi tiga kategori. Kategori pertama adalah insentif untuk dukungan dunia usaha, meliputi di antaranya adalah penyusutan aset dipercepat, kompensasi yang lebih lama atas rugi selama krisis, biaya-biaya perusahaan untuk mengatasi pandemi dapat dibiayai secara fiskal, penurunan tarif pajak, penundaan batas waktu pelaporan Surat Pemberitahuan (SPT), penundaan pembayaran pajak dan sanksi perpajakan, serta restitusi dipercepat. Kategori kedua adalah insentif untuk dukungan rumah tangga yang memiliki jenis insentif yang pada umumnya sama dengan jenis insentif pada kategori dukungan dunia usaha. Kategori ketiga adalah insentif untuk dukungan sektor kesehatan, yang mana jenis insentif yang paling sering digunakan pada kategori ini adalah penurunan atau penghapusan sementara tarif pajak seperti Bea Masuk dan PPN impor atas alat-alat kesehatan dan insentif pajak bagi tenaga kesehatan (OECD, 2020b; Sen, 2021; World Bank, 2021). Semakin inovatif suatu negara dalam mengembangkan kebijakan perpajakan, termasuk insentif pajak, dalam merespons pandemi Covid-19, maka negara tersebut akan cenderung lebih cepat bangkit dari resesi ekonomi (Asian Development Bank [ADB], 2020).

Sementara itu, terkait faktor-faktor yang mempengaruhi keikutsertaan pelaku usaha dalam program insentif pajak, literatur yang ada masih sangat terbatas. Salah satu studi di area ini adalah yang dilakukan oleh Busom *et al.* (2017) ketika menguji faktor-faktor penentu terjadinya partisipasi kembali atau partisipasi berulang (*persistent participation*) dari sebuah perusahaan dalam program insentif pajak untuk *research and development* (R&D), yang dapat dibagi menjadi dua kelompok faktor, yaitu faktor *heterogeneity* atau karakteristik perusahaan dan faktor *true state dependence* yaitu pengalaman pernah berpartisipasi dalam program serupa sebelumnya. Dengan menggunakan data longitudinal level perusahaan dari perusahaan-perusahaan manufaktur di Spanyol untuk tahun 2001–2008 dan metode *random effects dynamic bivariate probit model*, studi tersebut menemukan bahwa faktor *true state dependence* memiliki pengaruh positif dan signifikan terhadap keikutsertaan perusahaan dalam program insentif pajak. Sementara itu, terkait faktor *heterogeneity*, studi Busom *et al.* (2017) menemukan bahwa sementara variabel tingkat produktivitas, umur usaha, dan kendala pembiayaan

tidak mempengaruhi keikutsertaan, variabel jumlah pekerja dan pangsa pasar perusahaan mempengaruhi keikutsertaan perusahaan dalam program insentif pajak untuk R&D.

Berdasarkan hal tersebut, penulis ingin berkontribusi terhadap keterbatasan literatur dengan berfokus pada prediksi faktor-faktor yang mempengaruhi keikutsertaan pelaku usaha di Indonesia dalam insentif pajak penanggulangan dampak pandemi Covid-19. Metode yang dipilih dalam studi ini adalah ANN. ANN cukup banyak digunakan dalam penelitian lintas disiplin ilmu. Misalnya, di bidang keuangan, Sánchez-Serrano *et al.* (2020) menggunakan pendekatan ANN untuk membuat model prediksi dari opini audit khusus untuk laporan keuangan terkonsolidasi. Studi ini dapat memprediksi opini audit dengan tingkat akurasi 83%. Penggunaan pendekatan *artificial intelligence* di bidang keuangan mulai banyak digunakan mengingat sering kali perilaku bersifat non-linier dan penuh ketidakpastian (Bahrammirzaee, 2010). Di bidang pendidikan misalnya, Aryadoust & Baghaei (2016) menguji hubungan antara kemampuan membaca, pengetahuan leksikal, dan tata bahasa dari kelompok siswa yang menggunakan bahasa Inggris sebagai bahasa asing. Dalam studi ini, ANN dapat secara akurat mengklasifikasikan secara akurat sekitar 78% siswa.

Pendekatan ANN dapat dikatakan mengungguli model linier dan non-linier lainnya dalam hal akurasi dan kemampuan prediksi (Murorunkwere *et al.*, 2022). Dari sudut pandang kuantitatif, *neural network* sering terdiri dari kombinasi optimal yang memungkinkan prediksi dan perkiraan yang lebih akurat daripada model lain (Pérez López *et al.*, 2019). Selain itu, pendekatan ANN dapat meniru berbagai model statistik tanpa mengharuskan peneliti untuk membuat asumsi sebelumnya tentang hubungan antara variabel dependen dan independen. Ini karena proses pembelajaran dalam ANN yang akan menentukan bagaimana koneksi akan terbentuk (IBM, 2021).

Pendekatan ANN juga digunakan dalam studi perpajakan (lihat misalnya, Denton *et al.*, 1995; Chen *et al.*, 2011; Lin *et al.*, 2012; Jupri & Sarno 2018; Jang, 2019; Pérez López *et al.*, 2019). Misalnya, pendekatan ANN diaplikasikan untuk melakukan proyeksi besarnya penerimaan pajak di Korea Selatan (Jang, 2019). Lin *et al.* (2012) menggunakan pendekatan *artificial intelligence* untuk mendeteksi penyelundupan pajak (*tax evasion*) di Taiwan. Meski demikian, sepanjang pengetahuan penulis, belum ada studi empiris menggunakan pendekatan ANN untuk memprediksi keikutsertaan pelaku usaha dalam insentif pajak, khususnya yang diberikan selama pandemi Covid-19. Merujuk pada hasil penelitian terdahulu (lihat misalnya, Aryadoust & Baghaei, 2016; Pérez López *et al.*, 2019; Sánchez-Serrano *et al.*, 2020; Bekesiene *et al.*, 2021), penulis menduga pendekatan ANN dapat digunakan untuk memprediksi keikutsertaan pelaku usaha dalam memanfaatkan insentif pajak yang diberikan pemerintah bagi Wajib Pajak terdampak pandemi Covid-19.

2. Metodologi

2.1. Data dan Strategi Empiris

Studi ini menguji data dari 12.361 pelaku usaha yang dikumpulkan melalui survei.⁴ Survei dilakukan secara daring dengan cara mengirimkan surat elektronik (surel) atau *e-mail* ke alamat surel Wajib Pajak strategis, dengan mencantumkan tautan survei. Survei dilakukan dengan pendekatan *stratified random sampling* dan bersifat anonim; responden tidak diberikan pertanyaan apa pun yang berkaitan dengan identitas individu atau entitas. Elemen populasi dari survei ini adalah pelaku usaha dengan kategori Wajib Pajak strategis yang tersebar di 352 Kantor Pelayanan Pajak (KPP) yang tersebar di 34 provinsi di seluruh Indonesia.⁵ *Response rate* dari survei ini adalah 8,79%. Hasil uji keterwakilan (*representativeness tests*) responden survei menunjukkan bahwa responden survei adalah representasi yang sangat baik dari target populasi baik dari segi sebaran lokasi usaha, sebaran sektor usaha, dan sebaran omzet tahunan.

Studi ini menggunakan data respons dari pelaku usaha atas pertanyaan-pertanyaan yang diajukan dalam survei. Penting untuk dicatat bahwa data survei tidak menyediakan informasi mengenai *unobserved heterogeneity* terkait dengan faktor-faktor lain yang mungkin mempengaruhi keikutsertaan pelaku usaha dalam insentif pajak—misalnya, perbedaan tingkat sosialisasi yang dilakukan oleh kantor pajak di mana para pelaku usaha terdaftar, perbedaan tingkat literasi dari pelaku usaha, perbedaan status penggunaan *intermediary* atau konsultan pajak oleh pelaku usaha. Berdasarkan ketersediaan data hasil survei, dan untuk menguji temuan di dalam penelitian Busom *et al.* (2017) yang menyimpulkan bahwa variabel jumlah pekerja dan variabel pangsa pasar utama berpengaruh signifikan terhadap keikutsertaan pelaku usaha dalam program insentif pajak, terdapat lima karakteristik usaha yang secara khusus dianalisis dalam penelitian ini: (i) jumlah pekerja; (ii) pangsa pasar utama; (iii) omzet setahun; (iv) sifat usaha utama; dan (v) sumber utama pasokan. Variabel jumlah pekerja diukur dengan skala ordinal dan memiliki tujuh pilihan: (i) satu orang/pemilik, (ii) 2–10 orang, (iii) 11–50 orang; (iv) 51–100 orang, (v) 101–250 orang, (vi) 251–500 orang, dan (vii) di atas 500 orang. Variabel *pangsa pasar utama* merupakan variabel bersifat kategori dan memiliki tiga pilihan: (i) lokal, (ii) ekspor, dan (iii) campuran lokal

⁴Survei dilakukan selama tiga minggu (21 Juli sampai dengan 7 Agustus 2020). Survei ini merupakan bagian kegiatan dari Kelompok Kerja Penerimaan Program Pemulihan Ekonomi Nasional (Pokja Penerimaan PEN) Kementerian Keuangan Republik Indonesia yang diketuai oleh Staf Ahli Menteri Keuangan Bidang Kepatuhan Pajak.

⁵Wajib Pajak strategis adalah Wajib Pajak orang pribadi dan Wajib Pajak badan berstatus pusat yang terdaftar di Kantor Pelayanan Pajak (KPP) di lingkungan Kantor Wilayah (Kanwil) DJP Wajib Pajak Besar, KPP di lingkungan Kanwil DJP Jakarta Khusus, seluruh KPP Madya, dan Wajib Pajak dengan kriteria tertentu—umumnya terkait dengan besarnya nilai pembayaran pajak—yang terdaftar di tiap-tiap KPP Pratama. Secara umum terdapat sekitar 300 s.d. 500 Wajib Pajak strategis dalam setiap KPP. Dengan demikian, pengertian pelaku usaha dalam penelitian ini hanya merujuk pelaku usaha yang teradministrasi dalam sistem perpajakan DJP.

dan ekspor.

Selanjutnya, variabel *omzet tahunan* diukur dengan skala ordinal dalam enam pilihan: (i) kurang dari Rp5 miliar, (ii) antara Rp5 miliar–Rp10 miliar, (iii) antara Rp10 miliar–Rp25 miliar; (iv) antara Rp25 miliar–Rp50 miliar, (v) antara Rp50 miliar–Rp100 miliar, dan (vi) di atas Rp100 miliar.⁶ Variabel *sifat usaha utama* merupakan variabel bersifat kategori yang memiliki dua pilihan: (i) produsen dan (ii) non-produsen. Terakhir, variabel *sumber utama pasokan* merupakan variabel bersifat kategori dengan empat pilihan skala nominal: (i) dalam negeri, (ii) luar negeri (impor), (iii) campuran dalam dan luar negeri, dan (iv) jasa murni (tidak membutuhkan bahan baku/barang dagangan).

Strategi empiris jamak diartikan sebagai keseluruhan proses yang dilakukan oleh peneliti dalam membangun pertanyaan penelitian dan mendapatkan jawaban. Strategi empiris dalam studi ini mencakup sembilan tahapan. Seperti terlihat dalam Gambar 1, studi ini dimulai dengan menetapkan tujuan studi dan merumuskan pertanyaan penelitian. Studi ini secara spesifik menggunakan pendekatan ANN sebagai alat analisis utama. Hasil analisis akan diikuti dengan uji *robustness* untuk melihat tingkat konsistensi dari temuan, sebelum dilakukan interpretasi dan pembahasan hasil analisis.

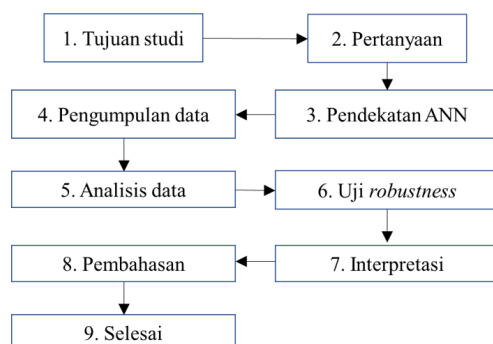
2.2. Metode Analisis

Merujuk pada dua tujuan utama penelitian menurut Babbie (2010), fokus tulisan ini adalah memberikan gambaran (*description*), bukan penjelasan (*explanation*). Untuk mencapai tujuan ini, studi ini mengadopsi pendekatan *applied research*. *Applied research* didesain untuk menjawab permasalahan tertentu atau menawarkan solusi dari permasalahan praktis yang terjadi (Neuman, 2014).⁷ Seperti yang telah dijelaskan sebelumnya, keikutsertaan pelaku usaha dalam insentif pajak, secara intuisi, banyak berkaitan dengan faktor-faktor yang bersifat non-linier. Dalam area yang bersifat non-linier, tingkat akurasi dari metode yang ada dalam *artificial intelligence* cenderung lebih unggul dari pada metode statistik tradisional (Bahrammirzaee, 2010).⁸ Berdasarkan pertimbangan ini—dan merujuk pada

⁶Penting dicatat, studi ini tidak mengklasifikasikan secara spesifik mana kelompok pelaku usaha yang termasuk Usaha Mikro, Kecil, dan Menengah (UMKM) dan mana yang tidak termasuk UMKM karena bukan merupakan tujuan analisis dari studi ini. Sebagai catatan, berdasarkan Undang-Undang Nomor 20 Tahun 2008 tentang UMKM, batasan (*threshold*) UMKM dibentuk oleh dua alternatif pengukuran, yaitu besaran omzet (mikro s.d. Rp300 juta; kecil s.d. Rp2,5 miliar; menengah s.d. Rp50 miliar) atau nilai kekayaan bersih (mikro s.d. Rp50 juta; kecil s.d. Rp500 juta; menengah s.d. Rp10 miliar). Sementara itu, dalam regulasi perpajakan Indonesia, besaran omzet setahun tidak melebihi *threshold* Rp4,8 miliar jamak dikaitkan dengan Wajib Pajak dengan peredaran bruto tertentu yang berhak atas tarif PPh Final 0,5% sesuai Peraturan Pemerintah Nomor 23 Tahun 2018, dan Pengusaha Kecil yang belum wajib mengukuhkan usahanya sebagai Pengusaha Kena Pajak yang wajib memungut PPN sesuai Peraturan Menteri Keuangan Nomor 197/PMK.03/2013.

⁷Implikasinya, *applied social research* jarang memiliki keterkaitan kuat dengan aktivitas membangun, menguji, atau menghubungkan teori secara mendalam (Neuman, 2014).

⁸Model regresi linier memiliki struktur model yang kaku dan serangkaian asumsi yang diterapkan sebelum belajar dari data. Sebaliknya, *neural network* dapat memperkirakan berbagai model statistik



Gambar 1: Strategi Empiris

Keterangan: Gambar ini menunjukkan keseluruhan proses empiris yang dilakukan dalam studi ini. Dimulai dari tujuan studi dan identifikasi pertanyaan riset, studi ini menggunakan pendekatan ANN sebagai alat analisis utama. Data empiris yang dianalisis berasal dari data survei. Sebelum dilakukan interpretasi hasil dan pembahasan, studi ini melakukan uji *robustness* dengan melakukan *sampling variability* berdasarkan lokasi geografis.

tujuan studi, penulis menggunakan pendekatan *Artificial Neural Network* (ANN).⁹

ANN tersusun dari metode matematika yang banyak digunakan dalam penelitian prediksi dan klasifikasi (Aryadoust & Baghaei, 2016). ANN adalah pendekatan yang banyak digunakan untuk analisis *predictive data mining* karena memiliki tingkat akurasi, fleksibilitas, dan kemudahan penggunaan—khususnya dalam situasi di mana proses yang mendasarinya rumit (IBM, 2019). Keunggulan utama dari ANN adalah pengenalan dan sekaligus klasifikasi pola (*pattern*) karena sifat pembelajaran adaptif non-parametrik non-linier yang dimiliki (Zhang *et al.*, 1999). ANN dibangun pada premis bahwa hubungan yang tepat antara variabel independen dan variabel dependen dapat diestimasi menggunakan fungsi matematika non-linier (Aryadoust & Baghaei, 2016).¹⁰

Lebih spesifik, dalam studi ini penulis menggunakan modul *Multilayer Per-*

tanpa mengharuskan kita berhipotesis terlebih dahulu mengenai hubungan tertentu antara variabel dependen dan independen—karena bentuk hubungan ditentukan selama model melakukan proses *learning*. *Neural network* akan secara fleksibel memilih apakah model linier atau non-linier dalam proses ini. Konsekuensinya, sebagai akibat fleksibilitas ini, *synaptic weight* dari *neural network* tidak mudah diinterpretasikan. Jika penafsiran dari hubungan linier antara variabel dependen dan independen merupakan tujuan utama, maka pendekatan model statistik yang tradisional lebih disarankan (IBM, 2019).

⁹ANN adalah salah satu metode *artificial intelligence* yang paling populer digunakan di bidang keuangan (Bahrammirzaee, 2010).

¹⁰Pembahasan mengenai ANN melibatkan model matematis yang cukup kompleks. Pembaca dapat merujuk pada Ripley (1996) dan Haykin (1999) untuk pembahasan model teoritis dan matematis yang komprehensif. Pembaca yang tertarik aspek aplikatif dari ANN ini dapat mempelajari lebih lanjut dalam *'IBM SPSS Neural Networks 26'* (IBM, 2019).

ceptron (MLP) dari IBM SPSS.¹¹ Konsepsi MLP dalam ANN beranjak dari *backpropagation learning error*—sebuah algoritma yang paling sering digunakan dalam ANN (Pérez López *et al.*, 2019). Pendekatan MLP lebih populer digunakan dalam ANN dibanding pendekatan lain (Zhang *et al.*, 1999). Studi yang dilakukan oleh Sánchez-Serrano *et al.* (2020) juga menemukan bahwa metode MLP menghasilkan tingkat akurasi yang lebih tinggi daripada pendekatan *Radial Basis Function* (RBF).¹²

Merujuk pada formasi partisi yang dilakukan oleh Bekesiene *et al.* (2021), data yang dianalisis diklasifikasikan secara acak ke dalam tiga kelompok: (i) data *training* sebanyak 60%, (ii) data *testing* sebanyak 20%, dan (iii) data *holdout* sebanyak 20%—atau jamak disebut formasi 60%-20%-20%.¹³ Data *training* digunakan untuk mencari bobot dan membangun model. Data *testing* digunakan untuk menemukan kesalahan dan mencegah *overtraining* selama mode pelatihan. Sedangkan data *holdout* digunakan untuk memvalidasi model (IBM, 2019).¹⁴

3. Hasil dan Pembahasan

3.1. Hasil Analisis

3.1.1. Statistik Deskriptif

Terdapat 12.361 data pelaku usaha dari seluruh Indonesia yang dianalisis dalam studi ini. Komposisi jumlah pelaku usaha berdasarkan lokasi di enam pulau utama adalah sebagai berikut: Jawa (70%, $n=8.681$), Sumatra (13%; $n=1.633$), Kalimantan (6%; $n=706$), Sulawesi (5%; $n=612$), Bali dan Nusa Tenggara (5%; $n=573$), Papua dan Maluku (1%; $n=156$). Statistik deskriptif mengenai variabel yang diuji dapat dilihat di panel A Tabel 1.

Panel B Tabel 1 menyajikan koefisien korelasi (r) dari variabel yang diuji. Nilai koefisien korelasi paling besar dihasilkan oleh hubungan antara variabel *jumlah pekerja* dan *omzet setahun* ($r = 0,644$). Meski demikian, kekuatan hubungan antara dua variabel ini bersifat sedang (*moderate*) karena memiliki nilai $r < 0,69$

¹¹Terdapat dua modul aplikasi prediksi berbasis ANN dalam IBM SPSS: MLP dan *Radial Basis Function* (RBF) (IBM, 2019). MLP dan RBF adalah dua arsitektur *neural network* yang paling banyak digunakan (Ripley, 1996).

¹²Perlu dicatat, metode MLP tidak selalu lebih akurat. Misalnya, Jupri & Sarno (2018) melakukan komparasi empat algoritma klasifikasi C4.5, *Support Vector Machine* (SVM), *K-Nearest Neighbour* (KNN), dan MLP untuk mengklasifikasi tingkat kepatuhan dari Wajib Pajak dan menyimpulkan bahwa algoritma C4.5 merupakan algoritma klasifikasi yang paling akurat.

¹³Studi yang dilakukan oleh Bekesiene *et al.* (2021) menunjukkan bahwa format partisi 60%-20%-20% lebih optimal dibandingkan dengan format 50%-30%-20% maupun 70%-20%-10%.

¹⁴Formasi partisi yang cukup jamak dilakukan adalah 70% untuk data *training* dan 30% untuk data *testing* (70%-30%). Formasi ini biasanya digunakan untuk observasi yang relatif tidak banyak (misalnya kurang dari 1.000) dan oleh karenanya tidak memiliki alokasi data yang digunakan untuk memvalidasi model, yaitu data *holdout*.

Tabel 1: Descriptive Statistics dan Hubungan Antarvariabel yang Diuji

Panel A. Descriptive Statistics						
	N	Min.	Max.	Mean	SD	Variance
Status pemanfaatan insentif	12.361	0	1	0,49	0,500	0,250
Jumlah pekerja	12.361	1	7	3,11	1,475	2,176
Sifat usaha utama	12.361	1	2	1,26	0,439	0,193
Sumber utama pasokan	12.361	1	4	2,04	1,212	1,468
Pangsa pasar utama	12.361	1	3	1,25	0,634	0,402
Omzet setahun	12.361	1	6	2,59	1,744	3,042
Valid N (<i>listwise</i>)	12.361					

Panel B. Hubungan Antarvariabel yang Diuji						
	Status pemanfaatan insentif	Jumlah pekerja	Sifat usaha utama	Sumber utama pasokan	Pangsa pasar utama	Omzet setahun
N	12.361	12.361	12.361	12.361	12.361	12.361
Status pemanf. insentif	1	.356**	.240**	.019*	.179**	.345**
Jumlah pekerja	.356**	1	.382**	.072**	.288**	.644**
Sifat usaha utama	.240**	.382**	1	-.088**	.233**	.266**
Sumber utama pasokan	.019*	.072**	-.088**	1	.143**	.024**
Pangsa pasar utama	.179**	.288**	.233**	.143**	1	.219**
Omzet setahun	.345**	.644**	.266**	.024**	.219**	1

Keterangan: **. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

(Schober *et al.*, 2018).¹⁵ Hubungan korelasi status pemanfaatan insentif dengan variabel *jumlah pekerja*, *sifat usaha utama*, *sumber pasokan utama*, dan *pangsa pasar utama* memiliki nilai koefisien yang bervariasi dengan kecenderungan tingkat hubungan yang lemah karena memiliki nilai < 0,40.

3.1.2. Artificial Neural Network (ANN)

Tujuan pertama dari studi ini adalah untuk menguji seberapa akurat pendekatan *multilayer perceptron* (MLP) *Neural Network* dapat memprediksi keikutsertaan pelaku usaha dalam pemanfaatan insentif pajak (memanfaatkan atau tidak memanfaatkan) dengan menganalisis data lima karakteristik utama dari pelaku usaha dan mengetahui urutan tingkat kepentingan dari variabel bebasnya.

Tabel 2 menunjukkan bahwa sebanyak 12.361 observasi digunakan untuk membangun model ANN dalam studi ini. Seperti yang telah dijelaskan sebelumnya, mengikuti hasil studi yang dilakukan oleh Bekesiene *et al.* (2021), data pelaku usaha hasil survei dalam studi ini dibagi dalam tiga kelompok dengan proporsi: (i) data *training* sebanyak 7.415 observasi (60%); (ii) data *testing* sebanyak 2.460 observasi (19,9%); dan (iii) data *holdout* sebanyak 2.486 observasi

¹⁵Menurut Schober *et al.* (2018), koefisien korelasi 0,40–0,69 menunjukkan korelasi sedang, 0,70–0,89 menunjukkan korelasi kuat, dan 0,90–1,00 menunjukkan korelasi yang sangat kuat.

(20,1%). Tidak ada observasi yang dikeluarkan (*excluded*) oleh aplikasi dalam analisis ini.

Tabel 2: Case Processing Summary

		N	Percent
Sample	Training	7415	60.0%
	Testing	2460	19.9%
	Holdout	2486	20.1%
Valid		12361	100.0%
Excluded		0	
Total		12361	

Tabel 3 menunjukkan jumlah *neuron* dalam tiap *layer* dan lima variabel bebas yang digunakan dalam analisis (*input layer*): (i) *sifat usaha utama*; (ii) *pangsa pasar utama*; (iii) *sumber utama pasokan*; (iv) *jumlah pekerja*; dan (v) *omzet setahun*. Dalam analisis ini, 3 variabel kategorikal masuk dalam kelompok *factors* dan 2 variabel ordinal masuk dalam kategori *covariates*, dengan jumlah total 11 unit.

Tabel 3: Network Information

Input Layer	Factors	1	Sifat usaha utama
		2	Pangsa pasar utama
		3	Sumber utama pasokan
	Covariates	1	Jumlah pekerja
		2	Omzet setahun
	Number of Units ^a		11
Rescaling Method for Covariates		Standardized	
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1 ^a		8
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Apakah sudah/sedang memanfaatkan insentif?
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy

a. Excluding the bias unit

Fitur *automatic architecture* dari aplikasi menunjukkan ada 8 unit di *hidden layers* yang terbentuk, sementara untuk *output layer* terdapat 2 unit yang merepresentasikan status pemanfaatan insentif pajak.¹⁶ Aktivasi fungsi *hidden layer*

¹⁶*Hidden layers* memungkinkan ANN untuk melakukan emulasi terhadap pola non-linier dalam data secara lebih akurat. Tanpa *hidden layer*, ANN akan berperilaku seperti model linier biasa yang tidak memiliki kemampuan untuk mendeteksi pola non-linier (Aryadoust & Baghaei, 2016).

dalam analisis ini menggunakan *hyperbolic tangent*, sementara untuk *output layer* menggunakan *Softmax*. *Cross-entropy* digunakan sebagai *error function* karena penggunaan metode *Softmax* sebagai fungsi aktivasi.

Grafik 1 menunjukkan diagram jaringan yang digunakan SPSS untuk memprediksi keikutsertaan pelaku usaha (*tidak ikut insentif=0, ikut insentif=1*) berdasarkan lima karakteristik utama dari pelaku usaha: (i) jumlah pekerja; (ii) pangsa pasar utama; (iii) omzet setahun; (iv) sifat utama usaha; dan (v) sumber utama pasokan.

Diagram jaringan di Grafik 1 menunjukkan terdapat 11 *input nodes*, 8 *hidden nodes*, dan 2 *output nodes* yang menunjukkan status keikutsertaan pelaku usaha dalam insentif pajak.¹⁷

Selanjutnya, Tabel 4 menyajikan ringkasan informasi yang berkaitan dengan hasil training (dan testing) dan hasil uji terhadap *holdout sample*. Nilai dari *cross-entropy error* juga disajikan baik untuk *training sample* maupun *testing sample* karena ini menunjukkan nilai *error function* yang diminimalkan oleh model ANN selama fase *training*. Nilai *cross-entropy error* yang lebih kecil (1.449) untuk *testing sample* dibandingkan *training sample* (4.472) mengindikasikan bahwa tidak terjadi *overfitted* terhadap data *training* dalam model jaringan.

Tabel 4: Model Summary

Training	Cross Entropy Error	4472.528
	Percent Incorrect Predictions	32.6%
	Stopping Rule Used	10 consecutive step(s) with no decrease in error ^a
	Training Time	0:00:00,35
Testing	Cross Entropy Error	1449.531
	Percent Incorrect Predictions	30.3%
Holdout	Percent Incorrect Predictions	31.1%

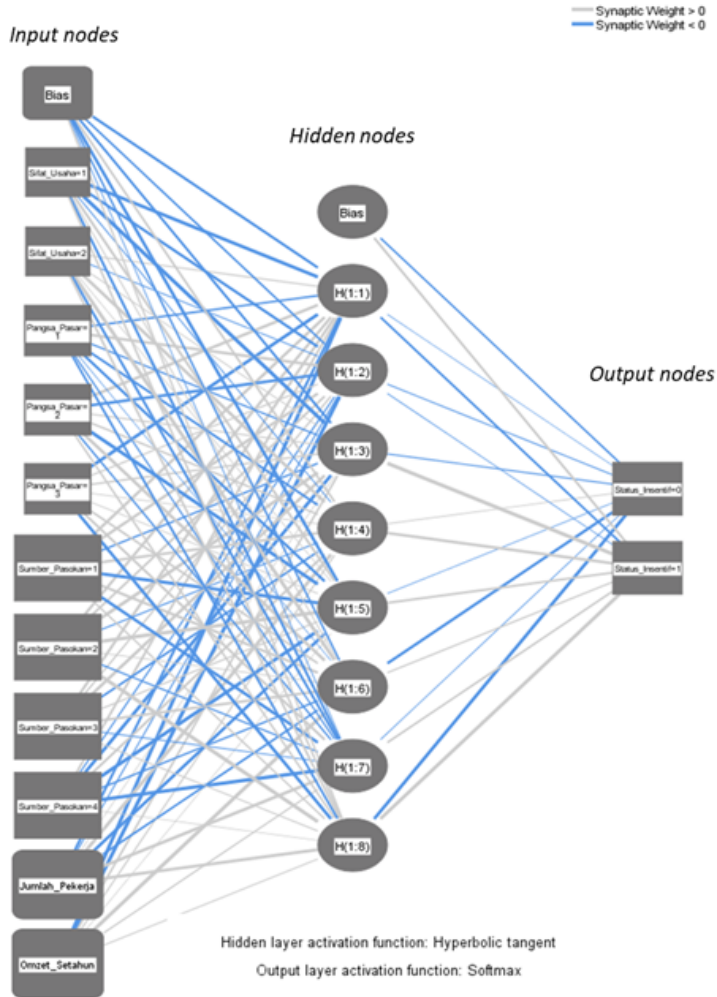
Dependent Variable: Apakah sudah/sedang memanfaatkan insentif?

a. Error computations are based on the testing sample.

Hasil ini memberikan justifikasi mengenai peran dari *testing sample*, yaitu mencegah terjadinya *overtraining*. Berdasarkan informasi dalam Tabel 4, proporsi dari prediksi yang tidak tepat dari *training sample* dan *testing sample* masing-masing adalah 32,6% dan 30,3%. Untuk *holdout sample*, proporsi prediksi yang tidak tepat adalah 31,1%.

Tabel 5 menyajikan besaran dari *synaptic weights* antara tiga *layer* seperti terlihat dalam Grafik 1. *Synaptic weights* adalah estimasi koefisien yang menunjukkan

¹⁷Struktur ini dikenal juga sebagai '*feedforward architecture*' karena hubungan dalam jaringan bergerak maju dari lapisan input menuju lapisan *output* tanpa *feedback loop*. Lapisan input terdiri dari prediktor, lapisan *hidden* terdiri dari unit yang tidak dapat terobservasi (*unobservable*), dan lapisan *output* berisi respons. Unit dalam *output* merupakan kumpulan dari beberapa fungsi dari unit di lapisan *hidden* (IBM, 2019).



Grafik 1: Network Diagram

hubungan antara unit pada suatu layer dengan unit pada layer berikutnya. Estimasi dari *synaptic weights* ini hanya didasarkan pada data *training*,¹⁸ dan oleh karenanya umumnya tidak digunakan untuk menginterpretasikan hasil uji ANN (IBM, 2019).

Tabel 6 menyajikan informasi mengenai tingkat akurasi dari model ANN

¹⁸Meskipun data dipartisi dalam tiga kategori: *training*, *testing*, dan *holdout*, estimasi *synaptic weights* hanya didasarkan pada data *training* (IBM, 2019).

Tabel 5: Parameter Estimates

Predictor	Predicted								Output Layer		
	Hidden Layer 1								[Status_Insentif =0]	[Status_Insentif =1]	
	H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)			
Input Layer	(Bias)	-.300	-.211	-.266	.240	-.363	-.207	-.039	.168		
	[Sifat_Usaha=1]	-.427	-.316	-.535	-.054	.270	.057	-.284	.212		
	[Sifat_Usaha=2]	.116	-.052	.047	.399	.023	.492	-.275	.246		
	[Pangsa_Pasar=1]	-.212	.342	-.161	-.257	-.305	-.012	-.144	-.250		
	[Pangsa_Pasar=2]	.353	-.350	.118	.204	-.442	.398	-.170	.200		
	[Pangsa_Pasar=3]	-.405	.139	.189	.036	.042	.230	.268	-.372		
	[Sumber_Pasokan=1]	.352	.346	-.179	.308	-.413	.315	-.360	.167		
	[Sumber_Pasokan=2]	.317	.255	.418	-.105	.532	.040	-.152	.767		
	[Sumber_Pasokan=3]	.324	.166	.374	-.235	.143	.300	-.055	.088		
	[Sumber_Pasokan=4]	.255	-.131	-.410	.266	-.479	-.191	-.529	.019		
	Jumlah_Pekerja	-.296	-.079	.012	.449	-.366	-.231	.402	.394		
	Omzet_Setahun	-.477	-.545	.225	.119	.008	.565	.151	.071		
	Hidden Layer 1	(Bias)								-.195	.265
H(1:1)									-.005	-.247	
H(1:2)									-.063	-.029	
H(1:3)									-.093	.846	
H(1:4)									.042	.347	
H(1:5)									-.022	.249	
H(1:6)									-.342	.158	
H(1:7)									-.030	.219	
H(1:8)								-.407	.530		

untuk sampel data *training*, *testing*, dan *holdout*. Masing-masing kelompok data memiliki tingkat akurasi secara berurutan 67,4%, 69,7%, dan 68,9%. Secara agregat, pelaku usaha yang diprediksi tidak ikut insentif adalah sebesar 56,3% untuk data *training*, 58,8% untuk data *testing*, dan 57,2% untuk data *holdout*. Sedangkan pelaku usaha yang ikut insentif adalah 43,7% untuk data *training*, 41,2% untuk data *testing*, dan 42,8% untuk data *holdout*. Tingkat akurasi klasifikasi dari masing-masing kelompok data mendekati 70%: data *training* (67,4%), data *testing* (69,7%), dan data *holdout* (68,9%).

Lebih spesifik, dalam data *training* 2.758 pelaku usaha yang tidak ikut insentif terklasifikasi secara akurat (*true negative*) dan terdapat 999 pelaku usaha yang tidak ikut insentif, tetapi diprediksi sebagai ikut insentif (*false positive*), menghasilkan tingkat *sensitivity* 73,4%. Dalam kelompok data ini terdapat 2.238 pelaku usaha yang terklasifikasi secara akurat ikut insentif (*true positive*), tetapi terdapat 1.420 pelaku usaha ikut insentif, tetapi terklasifikasi sebagai tidak ikut insentif (*false negative*). Hal ini menghasilkan tingkat *specificity* 61,2%.

Selanjutnya, dalam data *testing* 967 pelaku usaha yang tidak ikut insentif terklasifikasi secara akurat (*true negative*) dan terdapat 226 pelaku usaha yang tidak ikut insentif, tetapi diprediksi sebagai ikut insentif (*false positive*), menghasilkan tingkat *sensitivity* 69,7%. Dalam kelompok data ini terdapat 747 pelaku usaha yang terklasifikasi secara akurat ikut insentif (*true positive*), tetapi terdapat 480 pelaku usaha yang ikut insentif, tetapi terklasifikasi sebagai tidak ikut insentif (*false negative*). Komposisi ini menghasilkan tingkat *specificity* 60,9%.

Tabel 6: Accuracy of Classification

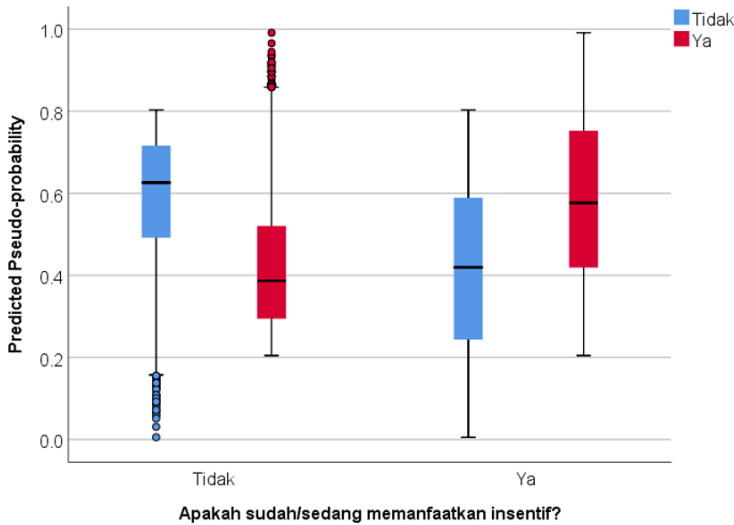
Sample	Observed	Predicted		Percent Correct
		Tidak	Ya	
Training	Tidak	2758	999	73.4%
	Ya	1420	2238	61.2%
	Overall Percent	56.3%	43.7%	67.4%
Testing	Tidak	967	266	78.4%
	Ya	480	747	60.9%
	Overall Percent	58.8%	41.2%	69.7%
Holdout	Tidak	957	307	75.7%
	Ya	465	757	61.9%
	Overall Percent	57.2%	42.8%	68.9%

Dependent Variable: Apakah sudah/sedang memanfaatkan insentif?

Terakhir, dalam data *holdout* 957 pelaku usaha yang tidak ikut insentif terklasifikasi secara akurat (*true negative*) dan terdapat 266 pelaku usaha yang tidak ikut insentif, tetapi diprediksi sebagai ikut insentif (*false positive*), menghasilkan tingkat *sensitivity* 75,7%. Dalam kelompok data ini terdapat 757 pelaku usaha yang terklasifikasi secara akurat ikut insentif (*true positive*), tetapi terdapat 465 pelaku usaha yang ikut insentif, tetapi terklasifikasi sebagai tidak ikut insentif (*false negative*). Hal ini menghasilkan tingkat *specificity* 61,9%.

Aplikasi IBM SPSS juga menampilkan grafik *predicted pseudo-probability* untuk dua kelompok pelaku usaha dari status pemanfaatan insentif dalam sebuah diagram *box-plot* (lihat Grafik 2). Grafik ini secara spesifik menggambarkan prediksi dari dua kategori dari variabel tergantung status pemanfaatan insentif. Perlu dicatat, grafik ini menunjukkan dua *box-plot* yang mewakili *predicted pseudo-probability* berdasarkan hasil analisis dari seluruh data ($n = 12.361$). Untuk setiap *box-plot* di masing-masing kategori, nilai di atas 0,5 mengindikasikan prediksi yang akurat. Sebaliknya, nilai di bawah 0,5 menunjukkan prediksi yang tidak akurat.

Penafsiran yang lebih detail dari Grafik 2 dapat dimulai dengan melihat bagian kiri dari grafik. Sisi kiri dari sumbu-x menunjukkan kategori pelaku usaha yang tidak memanfaatkan insentif. Terdapat dua *box-plot* di kategori ini: pemanfaat insentif dan non-pemanfaat. Selanjutnya, sumbu-y menunjukkan nilai *predicted pseudo-probability*. Grafik 2 menggambarkan prediksi yang cukup akurat untuk pelaku usaha yang tidak ikut insentif karena mayoritas area *box-plot* dari kelompok ini berada di atas angka 0,5 untuk kategori non-pemanfaat dan berada di bawah angka 0,5 untuk kategori pemanfaat insentif. Meski demikian, seperti terlihat di sisi kanan dari grafik, tingkat akurasi prediksi untuk pelaku usaha pemanfaat insentif lebih rendah dibandingkan dengan tingkat akurasi prediksi dari pelaku usaha non-pemanfaat. Hal ini dapat terlihat dari luas area yang berada di atas nilai 0,5 untuk kategori pelaku usaha non-pemanfaat insentif yang menunjukkan keberadaan *false positive*. Di sisi lain, luas area yang berada di bawah nilai 0,5 untuk kategori pelaku usaha pemanfaat insentif juga cukup jelas



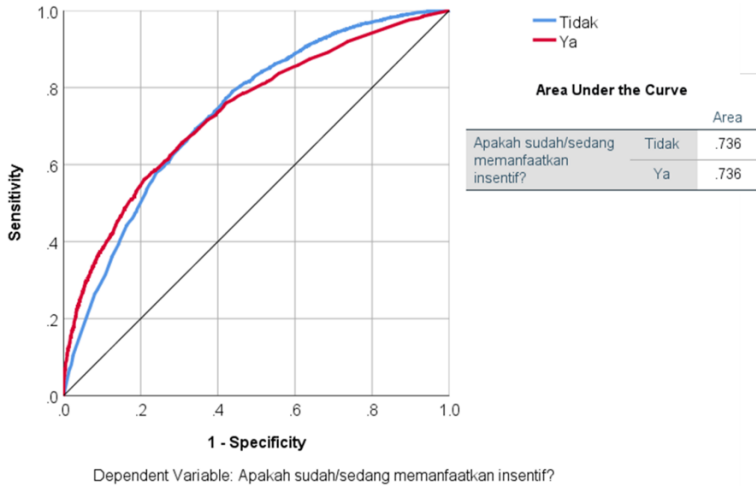
Grafik 2: *Predicted-by-observed Chart*

terlihat. Hal ini menunjukkan keberadaan *false negative*.

Grafik 3 menunjukkan kurva *Receiver Operating Characteristic* (ROC). Kurva ROC adalah penggambaran kinerja klasifikasi dua dimensi (Fawcett, 2006). Grafik ini memberikan gambaran tingkat *sensitivity* dan *specificity* berdasarkan gabungan antara data *training* dan data *sampling*. Garis diagonal 45-derajat dari bagian kiri bawah ke bagian atas kanan menunjukkan garis *no-discrimination*. Titik yang berada di bawah di bawah garis *no-discrimination* mengindikasikan klasifikasi yang tidak akurat, dan titik di atas garis *no-discrimination* menunjukkan hasil klasifikasi yang efektif (Fawcett, 2006).

Dalam Grafik 3 juga disajikan nilai *area under the curve* (AUC) sebesar 0,74. Nilai AUC sebesar 0,74 mengindikasikan bahwa jika satu pelaku usaha dari kelompok pemanfaat insentif dan satu pelaku usaha dari kelompok non-pemanfaat insentif terpilih secara acak, maka terdapat 74% kemungkinan bahwa model *predicted pseudo-probability* untuk pelaku usaha yang pertama masuk dalam kategori pemanfaat insentif (status insentif = 1). Secara umum, nilai AUC tergolong dalam lima kategori: 0,50–0,60 (gagal); 0,60–0,70 (kurang akurat); 0,70–0,80 (cukup akurat); 0,80–0,90 (akurat); dan 0,90–1,00 (sangat akurat). Dengan demikian, nilai AUC dari studi menunjukkan bahwa hasil prediksi ini cukup akurat.

Terkait dengan tujuan kedua dari studi ini, Grafik 4 menunjukkan seberapa besar pengaruh variabel independen dalam model ANN, yang diukur berdasarkan tingkat kepentingan relatif dan tingkat kepentingan yang dinormalisasi (*normalised importance*). Seperti terlihat dalam grafik, variabel *jumlah pekerja* memiliki skor tertinggi (0,399; *normalised importance* = 100%), disusul oleh *omzet setahun*



Grafik 3: ROC Curve dan Area Under the Curve

(0,273; *normalised importance* = 68,3%) dan *pangsa pasar* (0,147; *normalised importance* = 36,7%). Dua variabel dengan pengaruh terendah adalah *sifat usaha* (0,131; *normalised importance* = 32,8%) dan *sumber pasokan* (0,050; *normalised importance* = 12,5%).

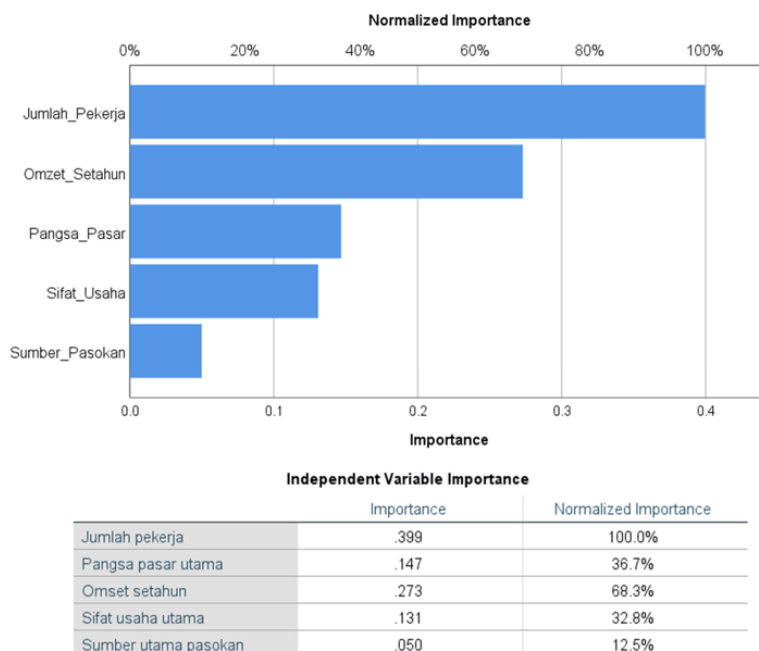
Dari Grafik 4 dapat disimpulkan bahwa jumlah pekerja yang dimiliki pelaku usaha, besaran omzet tahunan, dan pangsa pasar utama dari pelaku usaha memiliki efek yang paling besar dalam bagaimana ANN mengklasifikasi dan memprediksi keikutsertaan pelaku usaha dalam pemanfaatan insentif pajak.

3.1.3. Robustness Tests

Untuk menguji seberapa robust temuan sebelumnya, penulis melakukan *sampling variability* dan mereplika pendekatan ANN dengan menguji data berdasarkan lokasi pelaku usaha di enam kelompok pulau. Uji *robustness* ini juga berkaitan dengan tujuan ketiga dari studi ini yaitu untuk mengetahui seberapa kuat (*robust*) hasil prediksi ANN jika sampel data yang digunakan berbasis lokasi pelaku usaha. Ringkasan hasil analisis akhir berupa nilai *variable independent importance* dari masing-masing kelompok pulau kemudian dibandingkan dan disajikan di Tabel 7.¹⁹

Tabel 7 menunjukkan bahwa variabel *jumlah pekerja* konsisten memiliki skor *Independent Variable Importance* (IVI) yang paling besar di seluruh kelompok

¹⁹Dengan tujuan menjaga keringkasn tulisan, penjelasan teknis dan detail terkait hasil analisis seperti yang dijelaskan sebelumnya tidak dibahas dalam bagian *robustness tests* ini.



Grafik 4: Independent Variable Importance

Tabel 7: Komparasi Tiga Terbesar Nilai Independent Variable Importance Berdasarkan Lokasi Pelaku Usaha

	Jawa (n=8.681)	Sumatra (n=1.633)	Kalimantan (n=706)	Sulawesi (n=612)	Bali – NT (n=573)	Papua– Maluku (n=156)
Jumlah pekerja	0,383 (100%)	0,402 (100%)	0,465 (100%)	0,275 (100%)	0,493 (100%)	0,410 (100%)
Omzet setahun	0,333 (86,90%)	0,268 (66,50%)	0,119 (25,70%)	0,255 (92,80%)	0,149 (30,20%)	0,291 (71%)
Pangsa pasar utama	0,090 (23,60%)	0,124 (30,70%)	0,240 (51,60%)	0,228 (83,10%)	0,225 (45,60%)	0,089 (21,80%)
Sifat usaha utama	0,170 (44,30%)	0,117 (29%)	0,104 (22,40%)	0,169 (61,40%)	0,052 (10,50%)	0,122 (29,70%)
Sumber utama pasokan	0,024 (6,20%)	0,090 (22,30%)	0,072 (15,40%)	0,073 (26,60%)	0,082 (16,60%)	0,088 (21,50%)

Keterangan: Persentase dalam kurung menunjukkan nilai yang dinormalisasi (*normalised importance*).

Angka dalam kotak menunjukkan deviasi posisi menurut hasil prediksi ANN menggunakan sampel data nasional (n=12.361).

Variabel disusun secara berurutan berdasarkan urutan di Grafik 4.

pulau: Jawa (0,383; *normalised importance* = 100%), Sumatra (0,402; *normalised importance* = 100%), Kalimantan (0,465; *normalised importance* = 100%), Sulawesi (0,275; *normalised importance* = 100%), Bali – Nusa Tenggara (0,493; *normalised*

importance = 100%), dan Papua – Maluku (0,410; *normalised importance* = 100%).²⁰

Variabel *omzet setahun* menduduki posisi kedua dalam skor IVI di empat kelompok pulau, kecuali pulau Kalimantan, dan Bali – Nusa Tenggara. Nilai IVI dari variabel *omzet setahun* untuk kelompok pulau Jawa, Sumatra, Sulawesi, dan Papua–Maluku masing-masing secara berurutan sebesar 0,333 (*normalised importance* = 86,9%), 0,268 (*normalised importance* = 66,5%), 0,255 (*normalised importance* = 92,8%), dan 0,291 (*normalised importance* = 71%). Sedangkan untuk kelompok pulau Kalimantan dan Bali – Nusa Tenggara, masing-masing nilai IVI dari variabel *omzet setahun* adalah 0,119 (*normalised importance* = 25,7%) dan 0,149 (*normalised importance* = 30,2%).

Terakhir, variabel *pangsa pasar utama* menduduki posisi ketiga dalam skor IVI di dua kelompok pulau—Sumatra (0,124; *normalised importance* = 30,7%) dan Sulawesi (0,228; *normalised importance* = 83,1%) dan posisi kedua di Kalimantan (0,240; *normalised importance* = 51,6%) dan Bali–Nusa Tenggara (0,225; *normalised importance* = 45,6%). Yang menarik, untuk kelompok pulau Jawa dan Papua – Maluku, variabel *pangsa pasar utama* menduduki posisi keempat dengan besaran IVI secara berurutan 0,090 (*normalised importance* = 23,6%) dan 0,089 (*normalised importance* = 21,8%). Di dua kelompok pulau ini variabel *sifat usaha utama* menduduki posisi ketiga dengan besaran IVI: pulau Jawa (0,170; *normalised importance* = 44,3%) dan pulau Bali – Nusa Tenggara (0,122; *normalised importance* = 29,7%). Hasil analisis *robustness test* menunjukkan bahwa faktor penentu keikutsertaan pemanfaatan insentif pajak oleh pelaku usaha homogen di seluruh kelompok pulau, khususnya untuk variabel jumlah pekerja.²¹

3.2. Pembahasan

Menguatkan studi dari Aryadoust & Baghaei (2016), Pérez López *et al.* (2019), Sánchez-Serrano *et al.* (2020), dan Bekesiene *et al.* (2021), hasil analisis menunjukkan bahwa tingkat akurasi dari pendekatan ANN dalam studi ini cukup tinggi yaitu sekitar 70%. Hal ini sesuai dengan kesimpulan Bahrammirzaee (2010) yang menyatakan bahwa studi empiris komparatif menunjukkan bahwa keberhasilan penggunaan ANN di bidang keuangan sangat menjanjikan, meskipun tidak bisa dibilang sangat akurat. Misalnya, hasil uji empiris yang dilakukan Lin *et al.* (2012) menunjukkan bahwa pendekatan *decision tree* memiliki akurasi yang lebih tinggi untuk mendeteksi perilaku ketidakpatuhan pajak dari Wajib Pajak orang pribadi, sementara ANN mendeteksi secara lebih akurat untuk Wajib Pajak badan.

²⁰Nilai *normalized* IVI dalam bentuk persentase disajikan untuk menunjukkan peringkat dari variabel ini secara relatif terhadap variabel lain yang dikomparasi. Nilai 100% menunjukkan bahwa nilai IVI dari variabel ini adalah paling besar dibandingkan variabel yang lain.

²¹Terkait dengan determinan keikutsertaan pelaku usaha, perlu dicatat bahwa selain berdasarkan hasil uji *robustness*, penulis juga melakukan pendekatan *unsupervised machine learning* berupa *cluster analysis*. Hasilnya juga menunjukkan bahwa variabel *jumlah pekerja* adalah faktor terpenting penentu *cluster*, disusul variabel *omzet setahun*, dan *pangsa pasar utama*. Detail dari hasil *cluster analysis* dapat dilihat dalam Lampiran 2.

Hasil studi ini juga menunjukkan adanya kecenderungan bahwa tingkat akurasi prediksi untuk pelaku usaha yang tidak ikut insentif (*true negative*) lebih tinggi dibandingkan dengan tingkat akurasi untuk pelaku usaha yang ikut insentif (*true positive*). Secara berurutan, tingkat akurasi dari prediksi *true negative* untuk data *training*, *testing*, dan *holdout* adalah 72,4%, 72,7%, dan 73,5%. Sedangkan untuk prediksi *true positive* untuk data *training*, *testing*, dan *holdout* adalah 61,5%, 64,6%, dan 60,7%.

Selain tingkat prediksi, hasil analisis menunjukkan bahwa *jumlah pekerja* adalah karakteristik utama dari pelaku usaha yang memiliki bobot paling tinggi dalam memprediksi status keikutsertaan pemanfaatan insentif pajak, diikuti oleh besaran omzet tahunan dan pangsa pasar utama dari pelaku usaha. Untuk membahas lebih detail bagaimana ketiga variabel ini berhubungan dengan status pemanfaatan insentif pajak sehingga secara intuitif lebih mudah dipahami, penulis akan membahas secara deskriptif.²²

Data terkait *jumlah pekerja*, *omzet tahunan*, dan *pangsa pasar utama* pelaku usaha dapat dikelompokkan menjadi dua klaster berdasarkan status pemanfaatan insentif.²³ Untuk memudahkan interpretasi, komparasi visual dari dua klaster status pemanfaatan insentif pajak berdasarkan jumlah pekerja, omzet tahunan, dan pangsa pasar utama dari pelaku berdasarkan status pemanfaatan insentif disajikan secara visual sebagaimana terlihat dalam Grafik 5 sampai dengan Grafik 7.

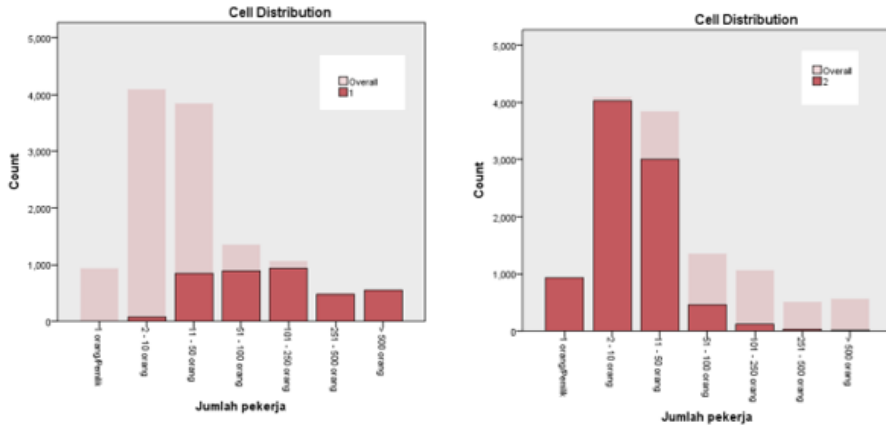
Grafik 5 menunjukkan secara visual proporsi dua klaster proporsi pelaku usaha yang ikut dan tidak ikut insentif (sumbu y) berdasarkan kelompok jumlah pekerja dari pelaku usaha (sumbu x). Menampilkan distribusi absolut (Panel A) dan distribusi relatif (Panel B), grafik ini menunjukkan bahwa semakin besar jumlah pekerja, kecenderungan pelaku usaha tersebut ikut insentif akan semakin tinggi. Misalnya, di bagian kiri dari Panel B (*cluster 1*) terlihat bahwa untuk pelaku usaha dengan jumlah pekerja 51 orang ke atas, maka proporsi pelaku usaha yang memanfaatkan insentif melebihi 60% dan jumlah ini semakin progresif seiring dengan semakin besarnya jumlah pekerja yang dimiliki.

Sebaliknya, proporsi jumlah pelaku usaha yang tidak ikut insentif ini semakin menurun (*cluster 2*), seiring dengan besarnya jumlah pekerja yang dimiliki pelaku usaha. Misalnya, terlihat bahwa untuk pelaku usaha dengan jumlah pekerja antara kurang dari 51 orang, maka proporsi pelaku usaha yang belum/tidak

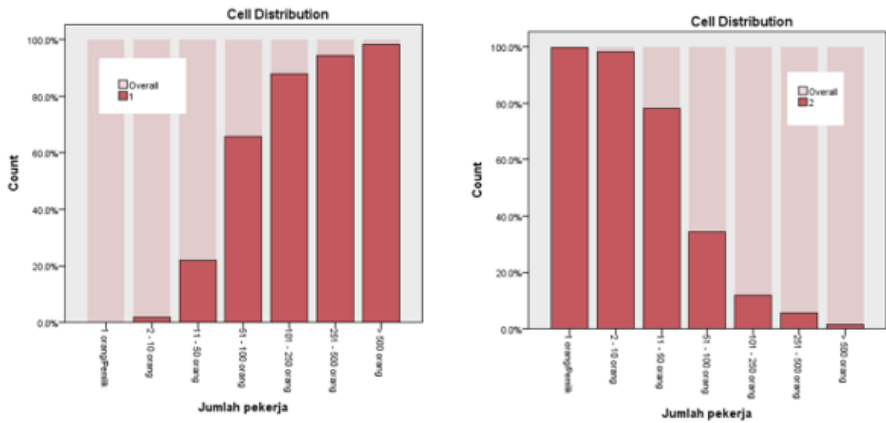
²²Meski menyediakan informasi mengenai urutan variabel yang paling penting, hasil analisis ANN belum menyediakan informasi mengenai beberapa hal penting, misalnya: (i) bagaimana hubungan antara hubungan antara jumlah pekerja pelaku usaha dengan tingkat partisipasi pelaku usaha dalam insentif pajak? Apakah semakin besar jumlah pekerja maka kecenderungan pelaku insentif juga akan semakin tinggi?, (ii) bagaimana hubungan antara tingkat omzet pelaku usaha dengan tingkat partisipasi pelaku usaha dalam insentif pajak? Apakah sifat hubungan juga positif?, dan (iii) bagaimana hubungan antara pangsa pasar utama status pelaku usaha dengan tingkat partisipasi pelaku usaha dalam insentif pajak?

²³Dua klaster merupakan hasil dari *cluster analysis*. Lihat Lampiran 2 untuk penjelasan yang lebih detail.

Panel A: Distribusi absolut kluster pemanfaat dan non-pemanfaat insentif



Panel B: Distribusi relatif kluster pemanfaat dan non-pemanfaat insentif



Grafik 5: Proporsi Keikutsertaan Pelaku Usaha Berdasarkan Jumlah Pekerja (n=12.361)

Keterangan: Grafik ini menunjukkan proporsi jumlah pelaku usaha pemanfaat insentif pajak dan non-pemanfaat insentif pajak berdasarkan jumlah pekerja. Panel A menyajikan distribusi absolut dari kluster pelaku usaha mayoritas pemanfaat insentif (*cluster 1*) dan distribusi absolut dari dari kluster pelaku usaha mayoritas non-pemanfaat insentif (*cluster 2*). Dari Panel A dapat kita lihat bahwa pelaku usaha dengan jumlah pekerja di atas 500 orang mayoritas merupakan pemanfaat insentif (ditunjukkan dengan area berwarna coklat tua)—*cluster 1*, sebaliknya pelaku usaha dengan jumlah pekerja kurang dari 11 orang mayoritas menyatakan status non-pemanfaat insentif. Untuk memudahkan interpretasi, distribusi absolut ditampilkan secara visual dalam bentuk distribusi relatif seperti disajikan dalam Panel B.

memanfaatkan insentif melebihi 60% dan jumlah ini semakin progresif seiring dengan semakin kecilnya jumlah pekerja yang dimiliki.

Selanjutnya, Grafik 6 menunjukkan secara visual proporsi dua kluster pro-

porsi pelaku usaha yang ikut dan tidak ikut insentif berdasarkan kelompok besaran omzet dari pelaku usaha. Menampilkan distribusi absolut (Panel A) dan distribusi relatif (Panel B), grafik ini menunjukkan bahwa semakin besar nilai omzet, kecenderungan pelaku usaha tersebut ikut insentif akan semakin tinggi. Misalnya, di bagian kiri dari Panel B (*cluster 1*) terlihat bahwa untuk pelaku usaha dengan nilai omzet tahunan di atas Rp25 miliar, maka proporsi pelaku usaha yang memanfaatkan insentif melebihi 50% dan jumlah ini semakin progresif seiring dengan semakin besarnya nilai omzet tahunan dari pelaku usaha.

Sebaliknya, proporsi jumlah pelaku usaha yang tidak ikut insentif ini semakin menurun (*cluster 2*), seiring dengan semakin kecilnya omzet tahunan dari pelaku usaha. Misalnya, terlihat bahwa untuk pelaku usaha dengan nilai omzet kurang dari Rp25 miliar, maka proporsi pelaku usaha yang tidak/belum memanfaatkan insentif melebihi 60% dan jumlah ini semakin progresif seiring dengan semakin kecilnya omzet tahunan yang dimiliki.

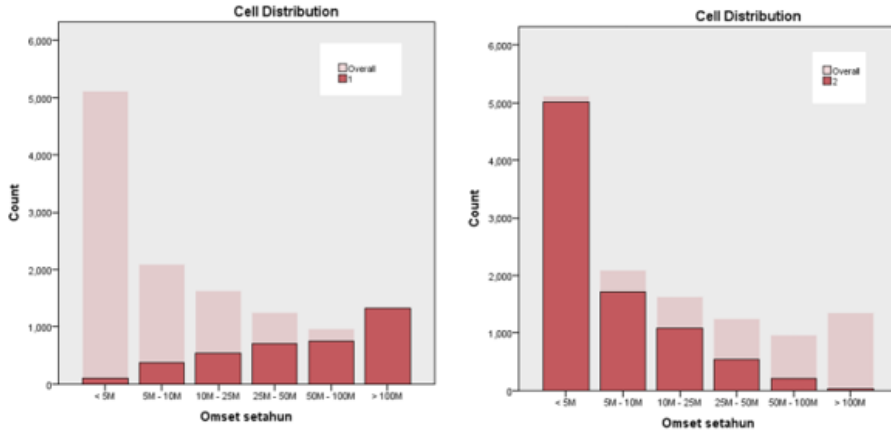
Terakhir, Grafik 7 menunjukkan secara visual proporsi dua klaster proporsi pelaku usaha yang ikut dan tidak ikut insentif berdasarkan pangsa pasar utama dari pelaku usaha.

Menampilkan distribusi absolut (Panel A) dan distribusi relatif (Panel B), grafik ini menunjukkan bahwa kecenderungan pelaku usaha tersebut yang ikut insentif lebih tinggi saat pangsa pasar utama dari pelaku usaha adalah campuran atau murni ekspor. Misalnya, di bagian kiri dari Panel B (*cluster 1*) terlihat bahwa 60% pelaku usaha dengan pangsa pasar campuran mengikuti insentif dan sekitar 83% pelaku usaha ekspor murni melaporkan status sedang/telah memanfaatkan insentif. Sebaliknya, proporsi jumlah pelaku usaha dengan pangsa pasar utama lokal (*cluster 2*) melaporkan proporsi yang lebih rendah dibandingkan pelaku usaha dengan pangsa pasar campuran maupun ekspor murni. Misalnya, terlihat bahwa sekitar 77% pelaku usaha dengan pangsa pasar lokal murni melaporkan status tidak memanfaatkan insentif pajak.

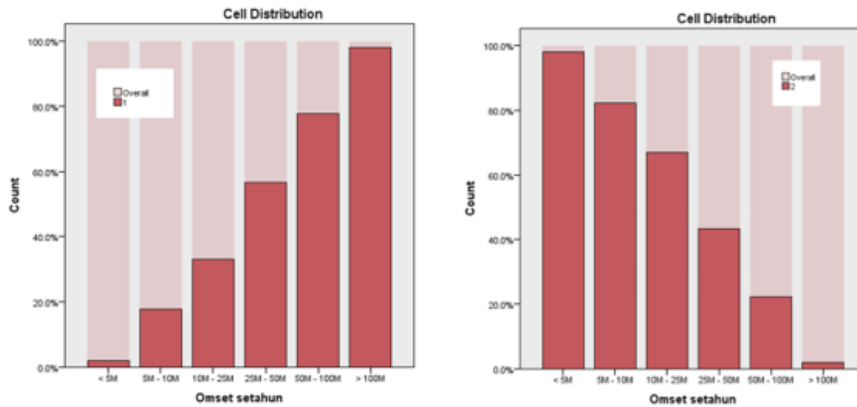
Hal lain yang menarik dari hasil studi ini adalah konsistensi variabel *jumlah pekerja* dalam *robustness tests* berdasarkan lokasi pelaku usaha. Hasil uji menunjukkan bahwa variabel ini memiliki nilai IVI yang paling besar di seluruh kelompok pulau (lihat kembali Tabel 7). Artinya, prediktor paling kuat keikutsertaan pelaku usaha di Pulau Jawa, Sumatra, Kalimantan, Sulawesi, Bali – Nusa Tenggara, dan Papua – Maluku adalah jumlah pekerja yang dimiliki pelaku usaha. Sementara itu, besaran omzet tahunan sebagai variabel terpenting kedua berdasarkan hasil *robustness tests* menunjukkan hasil yang sedikit berbeda. Variabel ini merupakan variabel terpenting kedua di kelompok Pulau Jawa, Sumatra, Sulawesi, dan Papua – Maluku. Untuk kelompok Pulau Kalimantan dan Bali – Nusa Tenggara, variabel terpenting kedua adalah pangsa pasar utama dari pelaku usaha. Di dua kelompok pulau ini, omzet tahunan pelaku usaha merupakan variabel terpenting ketiga.

Hasil *robustness tests* juga menunjukkan hasil yang sedikit berbeda untuk variabel pangsa pasar utama. Variabel ini menduduki posisi ketiga di kelompok

Panel A: Distribusi absolut kluster pemanfaat dan non-pemanfaat insentif



Panel B: Distribusi relatif kluster pemanfaat dan non-pemanfaat insentif



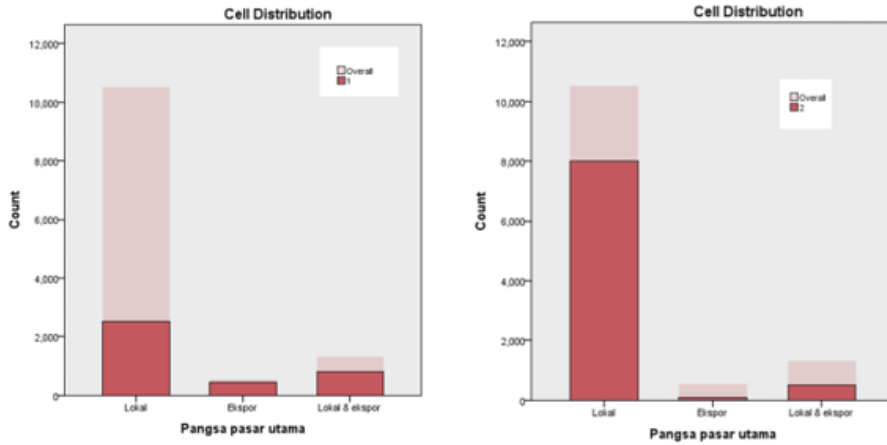
Grafik 6: Proporsi Keikutsertaan Pelaku Usaha Berdasarkan Omzet Tahunan (n=12.361)

Keterangan: Grafik ini menunjukkan proporsi jumlah pelaku usaha pemanfaat insentif pajak dan non-pemanfaat insentif pajak berdasarkan besaran omzet tahunan. Panel A menyajikan distribusi absolut dari kluster pelaku usaha mayoritas pemanfaat insentif (*cluster 1*) dan distribusi absolut dari kluster pelaku usaha mayoritas non-pemanfaat insentif (*cluster 2*). Dari Panel A dapat kita lihat bahwa pelaku usaha dengan nilai omzet di atas Rp25 miliar mayoritas merupakan pemanfaat insentif (ditunjukkan dengan area berwarna coklat tua)—*cluster 1*, sebaliknya pelaku usaha dengan nilai omzet kurang dari Rp25 miliar mayoritas menyatakan status non-pemanfaat insentif. Untuk memudahkan interpretasi, distribusi absolut ditampilkan secara visual dalam bentuk distribusi relatif seperti disajikan dalam Panel B.

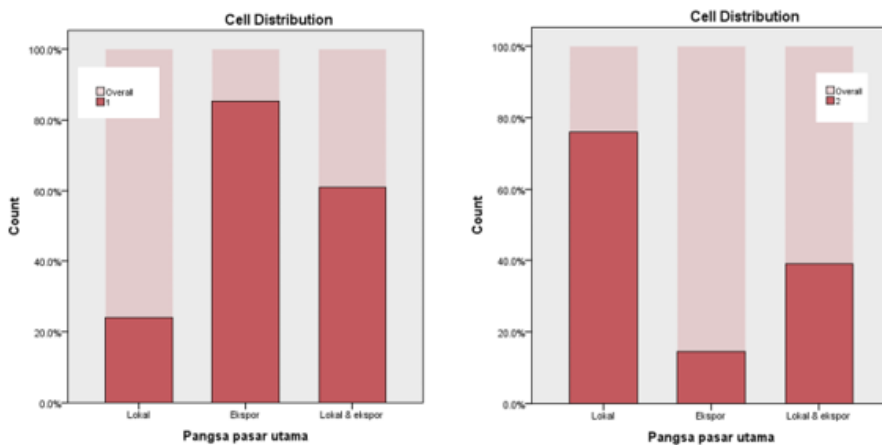
Pulau Sumatra dan Sulawesi.²⁴ Yang menarik, variabel terpenting ketiga di

²⁴Perlu dicatat, di kelompok Pulau Kalimantan dan Bali – Nusa Tenggara, variabel ini merupakan

Panel A: Distribusi absolut kluster pemanfaat dan non-pemanfaat insentif



Panel B: Distribusi relatif kluster pemanfaat dan non-pemanfaat insentif



Grafik 7: Proporsi Keikutsertaan Pelaku Usaha Berdasarkan Pangsa Pasar Utama (n=12.361)

Keterangan: Grafik ini menunjukkan proporsi jumlah pelaku usaha pemanfaat insentif pajak dan non-pemanfaat insentif pajak berdasarkan jenis pangsa pasar utama. Panel A menyajikan distribusi absolut dari kluster pelaku usaha mayoritas pemanfaat insentif (*cluster 1*) dan distribusi absolut dari kluster pelaku usaha mayoritas non-pemanfaat insentif (*cluster 2*). Dari Panel A dapat kita lihat bahwa pelaku usaha dengan pangsa pasar ekspor mayoritas (80%) merupakan pemanfaat insentif (ditunjukkan dengan area berwarna coklat tua)—*cluster 1*, sebaliknya pelaku usaha dengan pangsa pasar utama lokal mayoritas menyatakan status non-pemanfaat insentif. Untuk memudahkan interpretasi, distribusi absolut ditampilkan secara visual dalam bentuk distribusi relatif seperti disajikan dalam Panel B.

terpenting kedua, bertukar posisi dengan variabel *omzet setahun* yang berada di posisi ketiga.

kelompok Pulau Jawa dan Papua – Maluku bukan pangsa pasar utama dari pelaku usaha, melainkan sifat usaha utama dari pelaku usaha (yaitu produsen atau non-produsen).

Informasi yang dihasilkan dari hasil analisis ini relevan dalam pengambilan keputusan yang bersifat strategis—misalnya hasil analisis ini dapat digunakan untuk membuat segmentasi kelompok pelaku usaha yang dianggap perlu untuk dinaikkan tingkat keikutsertaannya. Misalnya, otoritas pajak atau yang terkait dapat menggunakan pendekatan-pendekatan ekonomi perilaku (*behavioural economics*) untuk secara tersegmentasi melakukan intervensi kebijakan.²⁵ Meski demikian, hasil analisis ini tidak dapat menjelaskan lebih spesifik mengenai penyebab beberapa variasi dari hasil temuan ini. Selain karena di luar cakupan dari tujuan studi, hal ini juga disebabkan karena, misalnya, status ‘pemanfaatan insentif pajak’ dalam studi ini bersifat umum; dalam arti tidak dibedakan secara spesifik jenis insentif pajak apa yang dimanfaatkan atau berapa jumlah total insentif yang dimanfaatkan.

4. Kesimpulan

Studi ini memiliki tiga tujuan. Tujuan pertama adalah mengetahui seberapa akurat pendekatan ANN dapat memprediksi keikutsertaan pelaku usaha dalam pemanfaatan insentif pajak. Untuk mendapatkan jawaban, studi ini menggunakan modul MLP dari ANN dan membangun model dengan formasi 60%-20%-20%. Hasil prediksi dengan modul MLP memiliki tingkat akurasi hampir 70% untuk data *testing* maupun *holdout*. Mendukung beberapa studi empiris yang menguji kemampuan prediktif ANN (lihat misalnya, Aryadoust & Baghaei, 2016; Pérez López *et al.*, 2019; Sánchez-Serrano *et al.*, 2020; Bekesiene *et al.*, 2021), hasil studi ini menunjukkan bahwa ANN dapat digunakan untuk memprediksi apakah pelaku usaha akan memanfaatkan insentif pajak atau tidak berdasarkan karakteristik tertentu yang dimilikinya dengan tingkat akurasi yang cukup bagus. Meski demikian, perlu dicatat bahwa prediksi yang dihasilkan lebih akurat untuk kelompok pelaku usaha yang tidak ikut memanfaatkan (*true negative*)—yang berada di kisaran 73% sampai dengan 74%. Prediksi untuk kelompok pelaku usaha yang ikut insentif (*true positive*) berada di 61% sampai dengan 65%.

Tujuan kedua dari studi ini adalah menentukan tiga faktor yang paling berpengaruh terhadap tingkat keikutsertaan pelaku usaha dalam pemanfaatan insentif pajak. Hasil studi ini menunjukkan bahwa faktor pertama yang paling menen-

²⁵Sebagai ilustrasi, untuk pelaku usaha yang belum memanfaatkan insentif dapat disampaikan pesan yang *customised*. Misalnya, untuk pelaku usaha dengan jumlah pekerja yang banyak atau memiliki omzet tahunan yang besar, namun belum memanfaatkan insentif pajak, maka pesan yang disampaikan dapat menggunakan pendekatan *peer-pressure* dengan merujuk pada hasil analisis ini. Sebagai contoh, untuk pelaku usaha dengan jumlah pekerja antara 251 s.d. 500 orang yang belum ikut insentif pajak, narasi pesan yang dapat disampaikan: “8 dari 10 pelaku usaha yang memiliki jumlah pekerja seperti perusahaan Anda sudah memanfaatkan insentif fiskal. . . .” Diskusi lebih lanjut mengenai penggunaan ekonomi perilaku di area perpajakan, misalnya, dapat dilihat dalam Weber *et al.* (2014).

tukan keikutsertaan pelaku usaha adalah jumlah pekerja yang dimiliki. Variabel *jumlah pekerja* memiliki nilai *normalised IVI* tertinggi. Hasil analisis menunjukkan bahwa semakin banyak jumlah pekerja yang dimiliki pelaku usaha, semakin besar kemungkinan pelaku usaha tersebut memanfaatkan insentif pajak. Faktor terpenting kedua yang menentukan keikutsertaan pelaku usaha adalah nilai omzet tahunan. Dalam hal ini, semakin tinggi omzet pelaku usaha, semakin besar kemungkinan pelaku usaha tersebut untuk memanfaatkan insentif pajak. Penting untuk dicatat bahwa terdapat hubungan korelasional positif antara jumlah pekerja dengan omzet, meskipun hubungan ini tidak kuat (*moderate*). Selanjutnya, pangsa pasar utama dari pelaku usaha adalah faktor terpenting ketiga yang menentukan tingkat pemanfaatan insentif pajak. Pelaku usaha dengan pangsa pasar utama ekspor memiliki kecenderungan yang paling tinggi dalam pemanfaatan insentif pajak, disusul oleh pelaku usaha dengan pangsa pasar utama campuran antara ekspor dan lokal. Hasil studi ini yang menggunakan pendekatan ANN juga mengkonfirmasi temuan di dalam penelitian Busom *et al.* (2017) yang menggunakan pendekatan ekonometri bahwa variabel jumlah pekerja dan variabel pangsa pasar utama berpengaruh signifikan terhadap keikutsertaan pelaku usaha dalam program insentif pajak.

Tujuan ketiga studi ini adalah untuk mengetahui apakah faktor-faktor penentu yang teridentifikasi dalam model ANN bersifat *robust* ketika dilakukan *sampling variability*. *Sampling variability* dilakukan dengan mereplika model ANN dengan data dari enam kelompok pulau. Hasil analisis menunjukkan bahwa jumlah pekerja yang dimiliki pelaku usaha konsisten memiliki nilai *normalised IVI* tertinggi di seluruh kelompok pulau (Jawa, Sumatra, Kalimantan, Sulawesi, Bali – Nusa Tenggara, dan Papua – Maluku). Hasil analisis juga menunjukkan bahwa besarnya omzet tahunan pelaku usaha juga cukup konsisten sebagai faktor terpenting kedua. Nilai *normalised IVI* dari variabel *omzet setahun* konsisten berada di urutan kedua di empat kelompok pulau (Jawa, Sumatra, Sulawesi, dan Papua – Maluku). Hal ini menunjukkan bahwa karakteristik utama dari pelaku usaha yang teridentifikasi memiliki hubungan kuat dengan keikutsertaan pelaku usaha dalam insentif pajak di enam kelompok pulau (Jawa, Sumatra, Kalimantan, Sulawesi, Bali – Nusa Tenggara, dan Papua – Maluku) bersifat homogen.

Dapat disimpulkan bahwa, dengan pendekatan ANN, keikutsertaan pelaku usaha dalam pemanfaatan insentif pajak selama masa pandemi Covid-19 dapat diprediksi dengan tingkat akurasi yang cukup memadai. Informasi yang dihasilkan juga relevan dalam pengambilan keputusan yang bersifat strategis—misalnya hasil analisis ini dapat digunakan untuk membuat segmentasi kelompok pelaku usaha yang dianggap perlu untuk dinaikkan tingkat keikutsertaannya. Meski demikian, studi ini tidak dapat memberikan informasi lebih jauh mengenai, misalnya, mengapa—atau bagaimana—faktor seperti jumlah pekerja, nilai omzet, atau pangsa pasar utama merupakan faktor terpenting dalam menentukan keikutsertaan pelaku usaha dalam pemanfaatan insentif pajak. Untuk mengetahui hal tersebut diperlukan data dan strategi empiris yang berbeda dengan studi ini.

Daftar Pustaka

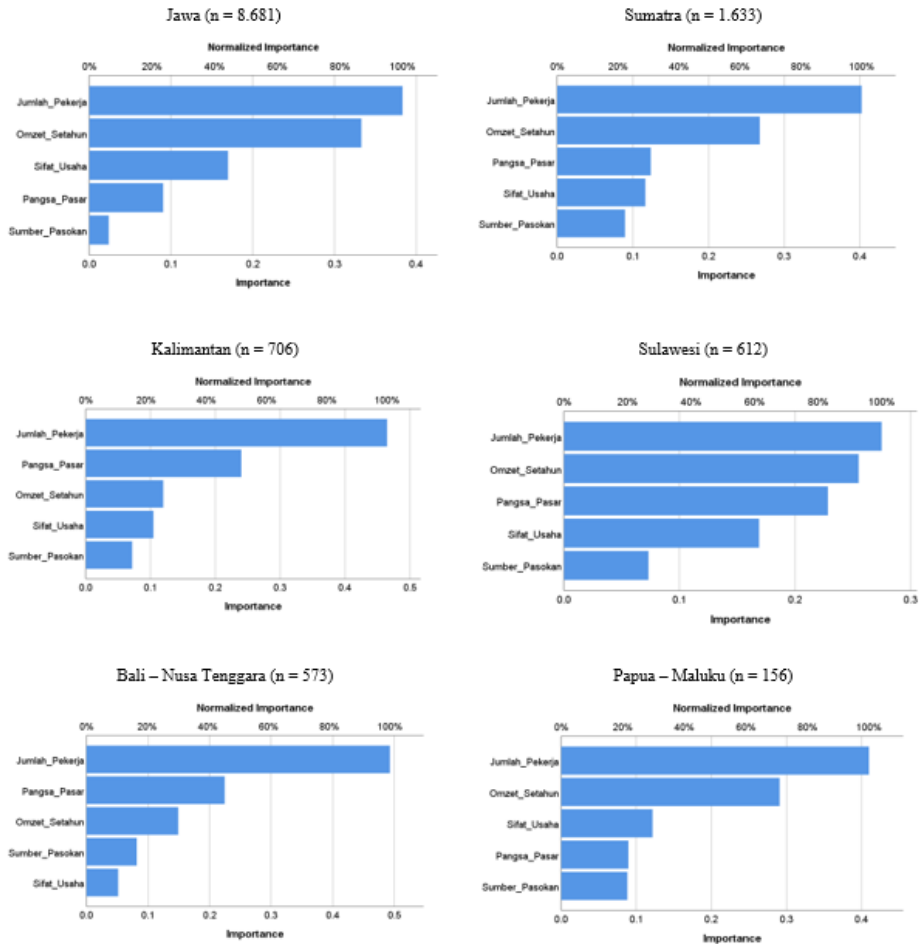
- [1] ADB (2020). *Asian Development Outlook (ADO) 2020: What Drives innovation in Asia?*. Asian Development Bank. doi: <http://dx.doi.org/10.22617/FLS200119-3>.
- [2] Alberola, E., Arslan, Y., Cheng, G., & Moessner, R. (2021). Fiscal response to the COVID-19 crisis in advanced and emerging market economies. *Pacific Economic Review*, 26(4), 459-468. doi: <https://doi.org/10.1111/1468-0106.12370>.
- [3] Aryadoust, V., & Baghaei, P. (2016). Does EFL readers' lexical and grammatical knowledge predict their reading ability? Insights from a perceptron artificial neural network study. *Educational Assessment*, 21(2), 135-156. doi: <https://doi.org/10.1080/10627197.2016.1166343>.
- [4] Babbie, E. R. (2010). *The practice of social research* (12th edition). Wadsworth Cengage, Belmont, CA.
- [5] BPS. (2020a). *Laju pertumbuhan PDB menurut pengeluaran (persen) 2020*. Badan Pusat Statistik. Diakses 22 Juli 2021 dari <https://www.bps.go.id/indicator/169/108/2/-seri-2010-laju-pertumbuhan-pdb-menurut-pengeluaran.html>.
- [6] BPS. (2020b). Analisis hasil survei dampak Covid-19 terhadap pelaku usaha. Badan Pusat Statistik. Diakses 21 Juli 2021 dari <http://data.bimakota.go.id/sites/default/files/publikasi/Analisis%20Hasil%20Survei%20Dampak%20Covid-19%20Terhadap%20Pelaku%20Usaha-min.pdf>.
- [7] Bahrammirzaee, A. (2010). A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing and Applications*, 19(8), 1165-1195. doi: <https://doi.org/10.1007/s00521-010-0362-z>.
- [8] Bekesiene, S., Smaliukiene, R., & Vaicaitiene, R. (2021). Using artificial neural networks in predicting the level of stress among military conscripts. *Mathematics*, 9(6), 626. doi: <https://doi.org/10.3390/math9060626>.
- [9] Busom, I., Corchuelo, B., & Martínez-Ros, E. (2017). Participation inertia in R&D tax incentive and subsidy programs. *Small Business Economics*, 48(1), 153-177. doi: <https://doi.org/10.1007/s11187-016-9770-5>.
- [10] Chen, J. H., Su, M. C., Chen, C. Y., Hsu, F. H., & Wu, C. C. (2011). Application of neural networks for detecting erroneous tax reports from construction companies. *Automation in Construction*, 20(7), 935-939. doi: <https://doi.org/10.1016/j.autcon.2011.03.011>.
- [11] Collier, R., Pirlot, A., & Vella, J. (2020). COVID-19 and fiscal policies: Tax policy and the COVID-19 crisis. *Intertax*, 48(8/9), 794-804. doi: <https://doi.org/10.54648/taxi2020078>.
- [12] Denton, J. W., Sayeed, L., Perkins, N. D., & Moorman, A. H. (1995). Neural networks to classify employees for tax purposes. *Accounting, Management and Information Technologies*, 5(2), 123-138. doi: [https://doi.org/10.1016/0959-8022\(95\)00008-W](https://doi.org/10.1016/0959-8022(95)00008-W).
- [13] Devereux, M. P., Güçeri, İ., Simmler, M., & Tam, E. H. (2020). Discretionary fiscal responses to the COVID-19 pandemic. *Oxford Review of Economic Policy*, 36(Supplement 1), S225-S241. doi: <https://doi.org/10.1093/oxrep/graa019>.
- [14] DJP. (2021). *Buku statistik perpajakan 2020*. Direktorat Jenderal Pajak, Jakarta.
- [15] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874. doi: <https://doi.org/10.1016/j.patrec.2005.10.010>.
- [16] Haykin, S. (1999). *Neural networks: A comprehensive foundation* (2nd edition). Macmillan College Publishing, New York.
- [17] [IBM] International Business Machines. (2017). *IBM SPSS Decision Trees* 25. IBM

- Corporation. Diakses 15 Juni 2021 dari https://www.ibm.com/docs/en/SSLVMB_25.0.0/pdf/en/IBM.SPSS.Decision.Trees.pdf.
- [18] [IBM] International Business Machines. (2019). *IBM SPSS Neural Networks 26*. IBM Corporation. Diakses 21 Juli 2021 dari https://www.ibm.com/docs/SSLVMB_26.0.0/pdf/en/IBM.SPSS.Neural.Network.pdf.
- [19] [IBM] International Business Machines. (2021). *IBM SPSS Neural Networks 28*. IBM Corporation. Diakses 14 Juli 2022 dari https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/IBM.SPSS.Neural.Network.pdf.
- [20] Jang, S. B. (2019). A design of a tax prediction system based on artificial neural network. In *2019 International Conference on Platform Technology and Service (PlatCon)* (pp. 1-4). IEEE. doi: 10.1109/PlatCon.2019.8669416.
- [21] Jupri, M., & Sarno, R. (2018). Taxpayer compliance classification using C4. 5, SVM, KNN, Naive Bayes and MLP. In *2018 International Conference on Information and Communications Technology (ICOIACT)* (pp. 297-303). IEEE. doi: 10.1109/ICOIAC-CT.2018.8350710.
- [22] Kementerian Keuangan RI. (2021). *Laporan Keuangan Pemerintah Pusat Tahun 2020 (Audited)*. Diakses 21 Juli 2021 dari <https://klc2.kemenkeu.go.id/kms/knowledge/laporan-keuangan-pemerintah-pusat-tahun-2020-audited-443ee9d2/detail/>.
- [23] Lin, C. H., Lin, I. C., Wu, C. H., Yang, Y. C., & Roan, J. (2012). The application of decision tree and artificial neural network to income tax audit: the examples of profit-seeking enterprise income tax and individual income tax in Taiwan. *Journal of the Chinese Institute of Engineers*, 35(4), 401-411. doi: <https://doi.org/10.1080/02533839.2012.655901>.
- [24] Murorunkwere, B. F., Tuyishimire, O., Haughton, D., & Nzabanita, J. (2022). Fraud detection using neural networks: a case study of income tax. *Future Internet*, 14(6), 168. doi: <https://doi.org/10.3390/fi14060168>.
- [25] Neuman, W. L. (2014). *Social research methods: qualitative and quantitative approaches* (7th edition), Pearson new international edition. Pearson Education Limited.
- [26] OECD. (2020a). *Tax and fiscal policy in response to the Coronavirus crisis: strengthening confidence and resilience*. Organisation for Economic Co-operation and Development. Diakses 21 Juli 2021 dari <https://www.oecd.org/tax/tax-policy/tax-and-fiscal-policy-in-response-to-the-coronavirus-crisis-strengthening-confidence-and-resilience.htm>.
- [27] OECD. (2020b). *Revenue statistics in Asian and Pacific economies 2020*. Organisation for Economic Co-operation and Development. doi: <https://doi.org/10.1787/d47d0ae3-en>.
- [28] Pérez López, C., Delgado Rodríguez, M. J., & de Lucas Santos, S. (2019). Tax fraud detection through neural networks: An application using a sample of personal income taxpayers. *Future Internet*, 11(4), 86. doi: <https://doi.org/10.3390/fi11040086>.
- [29] Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge University Press.
- [30] Sánchez-Serrano, J. R., Alaminos, D., García-Lagos, F., & Callejón-Gil, A. M. (2020). Predicting audit opinion in consolidated financial statements with artificial neural networks. *Mathematics*, 8(8), 1288. doi: <https://doi.org/10.3390/math8081288>.
- [31] Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5), 1763-1768. doi: <https://doi.org/10.1213/ANE.0000000000002864>.
- [32] Sen, T. K. (2021). Exit strategy to ease or eliminate tax responses to the

- COVID-19 pandemic. *The Governance Brief*, 43. Asian Development Bank. doi: <http://dx.doi.org/10.22617/BRF210142>.
- [33] Weber, T. O., Fooker, J., & Herrmann, B. (2014). Behavioural economics and taxation. *Taxation Papers*, 41. European Commission. Diakses 21 Juli 2021 dari https://taxation-customs.ec.europa.eu/system/files/2016-09/taxation_paper_41.pdf.
- [34] World Bank. (2021). *A review of fiscal policy responses to COVID-19*. Macroeconomics, trade and investment: equitable growth, finance & institutions insight. Diakses 19 Agustus 2022 dari <https://openknowledge.worldbank.org/handle/10986/35904>.
- [35] Zhang, G., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, 116(1), 16-32. doi: [https://doi.org/10.1016/S0377-2217\(98\)00051-4](https://doi.org/10.1016/S0377-2217(98)00051-4).

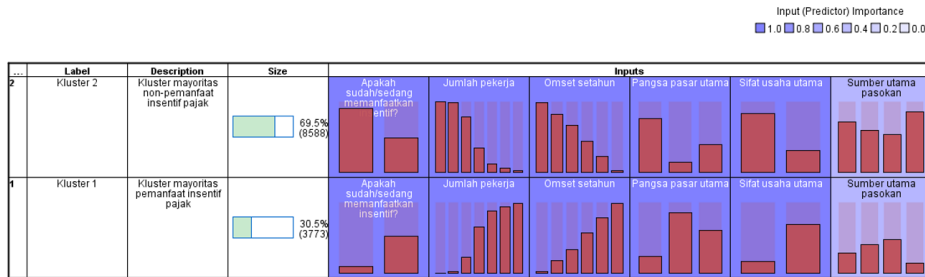
Lampiran

Lampiran 1. Hasil *Robustness Test* di Enam Kelompok Pulau: *Independent Variable Importance*

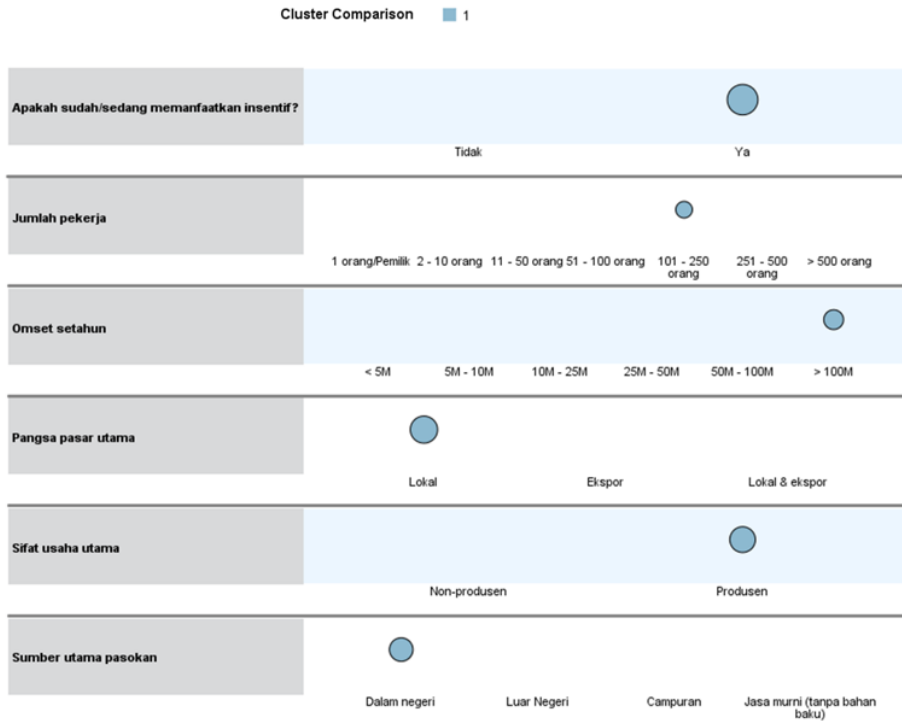


Keterangan: Grafik di atas merupakan hasil analisis ANN secara terpisah menggunakan data pelaku usaha dari enam pulau utama di Indonesia: Jawa, Sumatra, Kalimantan, Sulawesi, Bali - Nusa Tenggara, dan Papua - Maluku. Nilai IVI menunjukkan bahwa variabel *jumlah pekerja* konsisten memiliki urutan tertinggi sebagai variabel independen yang memiliki tingkat kepentingan paling besar dalam prediksi ANN di enam kelompok pulau. Sedangkan *omzet setahun* merupakan variabel independen terpenting kedua untuk kelompok pulau Jawa, Sumatra, Sulawesi, dan Papua - Maluku.

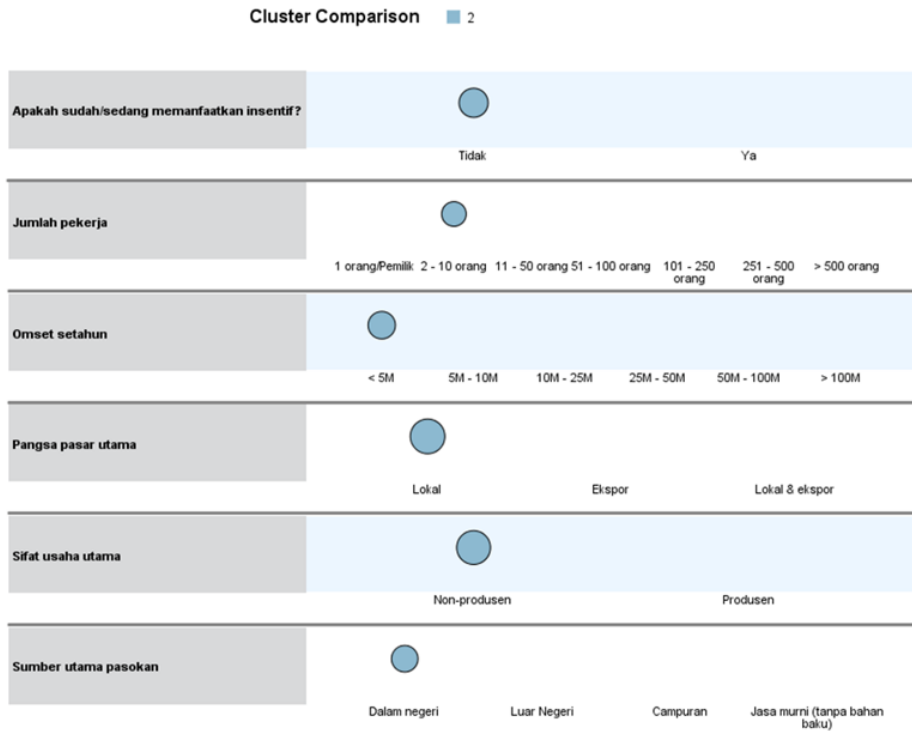
Lampiran 2. Hasil *Cluster Analysis*



Keterangan: Diagram ini merupakan *output* dari *cluster analysis*. *Cluster analysis* merupakan pendekatan *unsupervised machine learning*. Diagram ini menunjukkan bahwa terdapat dua kluster yang terbentuk. Kluster pertama adalah kelompok pelaku usaha yang mayoritas merupakan pemanfaat insentif dan kluster kedua adalah kelompok pelaku usaha yang mayoritas non-pemanfaat insentif. Dari grafik ini terlihat bahwa karakteristik pelaku usaha berupa *jumlah pekerja*, *omzet tahunan*, dan *pangsa pasar utama* adalah tiga input atau prediktor yang terpenting dalam proses klusterisasi.



Keterangan: Diagram ini merupakan *output* dari *cluster analysis*. Diagram ini secara visual menunjukkan bahwa *cluster 1* adalah kelompok dengan pelaku usaha yang dominan memiliki status ‘sudah/sedang memanfaatkan’ insentif pajak. Secara visual—ditunjukkan dengan posisi dari lingkaran—terlihat bahwa jumlah pekerja adalah karakteristik pelaku usaha yang paling menentukan dalam proses klusterisasi, disusul kemudian variabel omzet tahunan dari pelaku usaha. Secara absolut, pelaku usaha dengan jumlah pekerja antara 101 s.d. 250 orang merupakan jumlah terbanyak dalam kluster ini. Untuk omzet, pelaku usaha dengan omzet di atas Rp100 miliar merupakan jumlah terbanyak dalam kluster ini.



Keterangan: Diagram ini merupakan *output* dari *cluster analysis*. Diagram ini secara visual menunjukkan bahwa *cluster* 2 adalah kelompok dengan pelaku usaha yang dominan memiliki status ‘belum memanfaatkan’ insentif pajak. Secara visual—ditunjukkan dengan posisi dari lingkaran—terlihat bahwa jumlah pekerja adalah karakteristik pelaku usaha yang paling menentukan dalam proses klusterisasi, disusul kemudian variabel omzet tahunan dari pelaku usaha. Secara absolut, pelaku usaha dengan jumlah pekerja antara 2 s.d. 10 orang memiliki jumlah terbanyak dalam klaster ini. Untuk omzet, pelaku usaha dengan omzet tidak lebih dari Rp5 miliar merupakan jumlah terbanyak dalam klaster ini.