



PENERAPAN TEKNIK WEB SCRAPING UNTUK PENGALIAN POTENSI PAJAK (Studi Kasus pada Online Market Place Tokopedia, Shopee dan Bukalapak)

Mohammad Djufri

Pusdiklat Pajak, BPPK, Kementerian Keuangan, Email : mhd.djufri@gmail.com

INFO ARTIKEL

SEJARAH ARTIKEL

Diterima Pertama
10 November 2020

Dinyatakan Dapat Dimuat
11 Desember 2020

KATA KUNCI:

Web scraping,
E-commerce,
Informetric,
Business Intelligence.

ABSTRAK

Saat ini, jutaan data transaksi tersaji di internet, dapat ditarik dan dijadikan bahan analisis untuk penggalan potensi pajak. Tulisan ini bertujuan untuk mengkaji apakah pencarian data melalui teknik *web scraping* dapat diterapkan dalam upaya penggalan potensi pajak oleh para Account Representative. Tulisan ini menggunakan pendekatan studi informetrika, dimana akan dikaji informasi kuantitatif berupa data transaksi para penjual yang terekam pada tiga situs online marketplace (OMP) yaitu Tokopedia, Shopee dan Bukalapak. Hasil penelitian menunjukkan bahwa teknik web scraping dapat digunakan untuk penggalan potensi pajak, dan teknik web scraping terbaik yang dapat dilakukan oleh Direktorat Jenderal Pajak (DJP) adalah mengembangkan aplikasi web scraping sendiri yang terintegrasi sebagai suatu sistem Business Intelligence. Hasil Penelitian ini diharapkan dapat memberi kontribusi secara akademis berupa cara penggunaan *web scraping* dalam ekstraksi data untuk penggalan potensi pajak dan implikasi kebijakan dalam hal pencarian data melalui internet oleh DJP.

Currently, millions of transaction data are available on the internet, which can be retrieved and analyzed for excavating potential taxes. This article aims to examine whether the search data through web scraping techniques can be applied in an attempt to excavate the potential tax by the Account Representative. This paper uses an informetric approach, which will be examined quantitative information in the form of transaction data of sellers recorded on the three online marketplace (OMP) namely Tokopedia, Shopee and Bukalapak. The results show that web scraping techniques can be used for extracting potential taxes, and the best web scraping technique that can be done by the Directorate General of Taxation (DJP) is to develop its own integrated web scraping application as a Business Intelligence system. The results of this research are expected to contribute academically in the form of the use of web scraping in data extraction for the excavation of potential taxes and policy implications in terms of data search through the internet by the Directorate General of Taxation.

1. PENDAHULUAN

1.1 Latar Belakang

Pesatnya perkembangan teknologi informasi dan komunikasi mendorong pertumbuhan di sektor *e-commerce* di Indonesia. Laporan *GlobalWebIndex* dalam Datareportal (2020) menyatakan bahwa Indonesia memiliki tingkat penggunaan *e-commerce* tertinggi di antara negara-negara di dunia, dengan 90% pengguna internet yang sudah melakukan pembelian produk dan layanan online.

Badan Pusat Statistik (2019) mencatat, selama tahun 2018 terdapat 24,8 juta transaksi *online* dengan nilai transaksi mencapai 17,21 triliun rupiah. Statista menyebutkan jumlah pengguna *e-commerce* di Indonesia akan mengalami pertumbuhan yang meningkat dari 139 juta dolar pada 2017 menjadi lebih dari 212 juta dolar pada 2023. Begitu pun pendapatan ritel *e-commerce* Indonesia diperkirakan

akan meningkat lebih dari enam kali lipat dari hanya 7,3 miliar dolar pada 2017 menjadi 47,1 miliar dolar pada 2023. Data tersebut menunjukkan bahwa potensi transaksi *e-commerce* semakin berkembang, dan hal ini menjadi peluang bagi peningkatan penerimaan pajak di masa-masa mendatang. Berdasarkan pesatnya perkembangan transaksi online tersebut, menjadi menarik untuk membahas bagaimana menghimpun data transaksi *e-commerce* khususnya untuk tujuan perpajakan.

Data terkait transaksi *e-commerce* telah banyak tersedia di sistem informasi yang dimiliki DJP, seperti data faktur pajak, data bukti potong dan data alat keterangan. Namun kenyataannya, para *Account Representative* (AR) di lapangan, masih memerlukan data lain selain data yang telah tersedia di DJP, sebagai data pembanding dan pelengkap dalam melakukan penggalan potensi pajak.

Untuk memperkuat pendapat ini, dilakukan pengambilan data pendahuluan terhadap seluruh peserta Pelatihan Teknis *Account Representative* (AR) Angkatan IV Tahun 2020 yang berlangsung dari tanggal 24 Februari hingga 28 Februari 2020 di Pusdiklat Pajak. Para peserta pelatihan adalah para AR yang bertugas di beberapa Kantor Pelayanan Pajak yang tersebar di beberapa wilayah Indonesia. Pengambilan data berupa respons terhadap kuesioner yang dilakukan pada tanggal 28 Februari 2020. Kuesioner yang berisi beberapa pertanyaan tertutup, diberikan kepada seluruh peserta pelatihan yang berjumlah 30 orang dan keseluruhan peserta memberikan responsnya. Respons peserta terhadap pertanyaan terkait kualitas/kuantitas data yang saat ini dimiliki DJP, menghasilkan bahwa sebesar 60% peserta menyatakan bahwa data cukup lengkap dan valid, dan 40% menyatakan kurang lengkap dan valid. Terkait kebutuhan data lain selain yang telah dimiliki DJP, yaitu selain data faktur pajak, bukti potong ataupun alat keterangan, 77% peserta menyatakan sangat membutuhkan data lain, 23% membutuhkan data lain dan tidak ada satu peserta pun yang menyatakan tidak membutuhkan data lain. Cara mencari data tambahan yang dilakukan para peserta, sebesar 47% menyatakan mencarinya lewat internet, 33% melalui pengamatan di lapangan dan 20% melalui kerja sama dengan pihak ketiga. Dalam hal kemudahan pencarian data ini, peserta menyatakan bahwa yang paling mudah adalah melalui teknik *web scraping* sebesar 40%, melalui pengamatan sebesar 37% dan melalui kerja sama dengan Pihak Ketiga sebesar 23%.

Hingga saat ini telah terdapat beberapa penelitian yang berkaitan dengan pemanfaatan data di internet menggunakan *web scraping*. Dari penelitian yang telah dilakukan tersebut, *web scraping* telah diterapkan dalam berbagai bidang sebagai teknik pengambilan datanya seperti pencarian artikel ilmiah (Josi dkk, 2014), pengambilan data untuk produksi data *e-commerce* (Suwiyanto, 2018), teknologi *web crawling* untuk melakukan ekstraksi dan analisis produk di *marketplace* (Fernando, 2019), analisis untuk data bencana alam (Sonya dan Prihandoko, 2016). Wicaksono (2018) melakukan penelitian dengan teknik *web crawling* pada situs *olx.com* untuk mengestimasi tingkat potensi penerimaan pajak sektor Usaha Mikro Kecil dan Menengah (UMKM) melalui *Big Data. The Revenue Departement* (RD) Thailand (Bangkokpost, 2019), otoritas pajak Austria, Belgia, Inggris dan Finlandia menggunakan *web scraping* dan sejenisnya untuk mencari data di internet (OECD, 2017,2018).

Berbeda dengan penelitian sebelumnya, dalam penelitian ini akan dilakukan teknik *web scraping* untuk menghimpun data tiga besar *online marketplace* (OMP) di Indonesia yaitu Tokopedia, Shopee dan Bukalapak, dan apakah hasilnya dapat digunakan untuk menggali potensi perpajakan?

Sejalan dengan perkembangan teknologi, di mana transaksi online semakin berkembang, dan jutaan data transaksi ataupun petunjuk yang mengarah kepada potensi perpajakan tersaji di internet, maka data tambahan yang dibutuhkan AR dalam penggalian potensi pajak pun dapat dicari melalui internet. Mengingat sangat banyaknya jumlah data yang tersaji di internet, khususnya di tiga *online marketplace* (OMP) besar di Indonesia, yaitu Tokopedia, Shopee dan Bukalapak, tidak mungkin dicari secara manual satu persatu data transaksi online dari penjual di OMP-OMP tersebut, namun membutuhkan suatu teknik yang disebut *web scraping* untuk mempermudah ekstraksi/penarikan data tersebut. Untuk itu, permasalahan yang akan dibahas dalam penelitian ini adalah bagaimana teknik *web scraping* ini digunakan dalam ekstraksi data untuk tujuan penggalian potensi pajak.

Tujuan penelitian ini adalah untuk menjawab pertanyaan penelitian (*question research*) apakah hasil ekstraksi data menggunakan teknik *web scraping* khususnya pada ketiga OMP di atas berguna untuk penggalian potensi pajak? Juga teknik *web scraping* terbaik seperti apa yang dapat dilakukan DJP dalam rangka penggalian potensi pajak.

Hasil penelitian ini, diharapkan dapat memberi kontribusi secara akademis berupa cara penggunaan *web scraping* dalam ekstraksi data khususnya pada OMP Tokopedia, Shopee dan Bukalapak dan memberi kontribusi berupa kebijakan dalam hal pencarian data melalui internet oleh DJP.

2. KERANGKA TEORITIS

2.1 Pendekatan Konseptual *Web Scraping*

Web scraping adalah proses pengambilan sebuah dokumen semi-terstruktur dari halaman web, untuk diambil data tertentu saja dari halaman tersebut (Turland, 2010). *Web scraping* adalah teknik untuk mengekstraksi data dari internet dan disimpan ke *file* atau *database* untuk kebutuhan analisis data (Zhao, 2017). *Web scraping* sebagai teknik yang efisien dan kuat untuk mengumpulkan big data (Mooney et al. 2015; Bar-Ilan 2001 dalam Zhao, 2017). *Web scraping* dapat melakukan pencarian apa-apa yang tidak dapat dilakukan oleh mesin pencari tradisional seperti *Google Search* (Mitchell, 2015).

Beberapa pengertian di atas menjelaskan bahwa *web scraping* adalah metode untuk mengekstraksi informasi dari situs web, sehingga menjadi data yang dapat dianalisis dan dimanfaatkan untuk berbagai tujuan.

Proses *web scraping* dari internet pada dasarnya terbagi menjadi dua langkah yang berurutan, yaitu menemukan *web* yang akan diekstrak datanya dan kemudian mengekstraksi data/informasi yang dibutuhkan dari *web* tersebut (Zhao, 2017).

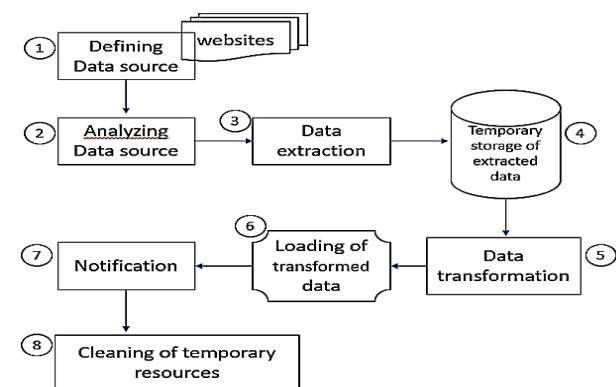
Beberapa *tools/software* yang dapat digunakan untuk melakukan *web scraping* adalah pemrograman

python (Mitchell, 2015), PHP (Thurland, 2010), Bahasa R (Munzert, et.al, 2015), Microsoft Excel (Phillips, 2016), Botsol (Botsol.com, 2020), Outwit Hub, Spinn3r, Fminer, ParseHub, OctoParse, Import.io, VisualScaper, Dexi.io, 80legs dan Web Scraper (Milev, 2017). Dalam penelitian ini digunakan *Web Scraper* dan *Instant Data Scraper* sebagai ekstensi Google Chrome. Walaupun menurut Milev (2017) *Web Scraper* dan *Instant Data Scraper* ini termasuk aplikasi *web scraping* yang tradisional, namun aplikasi ini menawarkan kemudahan bagi penggunaanya, dan tanpa harus memahami bahasa pemrograman (*coding*).

Berbeda dengan sumber-sumber di atas, Milev (2017) menyebutkan *web scraping* hanyalah langkah awal dalam *data mining*, sementara *data mining* adalah bagian dari *business intelligence* (BI).

Menurut beberapa sumber (dalam Milev, 2017) *business intelligence* adalah seperangkat konsep, metode, proses, alat, dan teknologi yang digunakan untuk mengubah informasi menjadi data dan knowledge yang dibutuhkan untuk mendukung keputusan bisnis. Stefanova (2008 dalam Milev, 2017), menyatakan bahwa ada dua elemen yang membedakan sistem BI dari sistem lain, yaitu: 1) Adanya integrasi data, yang berarti menggabungkan data dari berbagai sumber data yang berbeda dan format data yang berbeda, dan menyediakan akses data yang koheren; dan 2) Tersedianya teknik analisis data dan visualisasi data yang dapat mudah dipahami oleh pengguna.

Terdapat beberapa pendekatan berbeda dalam desain dan pengembangan aplikasi BI ini. Sistem BI adalah suatu sistem dinamis yang terintegrasi yang terus menerus berkembang di mana data dan fungsionalitas dikembangkan dalam beberapa iterasi berdasarkan umpan balik yang diperoleh dari penggunaanya, sementara pendekatan tradisional termasuk *web scraper* tidak sedinamis BI. Oleh karenanya Milev (2017) mengusulkan pendekatan konseptual untuk pengembangan aplikasi web



scraping dengan delapan langkah sebagai berikut:

Gambar 1. Pendekatan Konseptual Pengembangan *Web Scraping*

Sumber (Milev, 2017)

- 1) Melakukan pendefinisian situs web atau menentukan *url* dari *website* yang akan diekstrak datanya;
- 2) Menganalisis sumber data yang didefinisikan untuk memastikan proses *web scraping* itu dapat dilakukan pada web tersebut dengan menyesuaikan bahasa yang digunakan dalam web tersebut atau menambahkan perintah tertentu dalam *script web scraping* untuk proses ekstraksinya.
- 3) Memulai proses ekstraksi data secara otomatis atau terjadwal;
- 4) Menyimpan data yang diekstraksi menjadi basis data temporary selama proses *web scraping*;
- 5) Transformasi hasil ekstraksi data ke dalam suatu bentuk yang bisa di-load ke dalam struktur data warehouse;
- 6) Load data yang telah ditransformasi ke dalam data warehouse;
- 7) Setelah data ada dalam *data warehouse*, dimungkinkan pencarian data berdasarkan daftar kata kunci yang diinginkan dan sistem akan memberikan notifikasi; dan
- 8) Pembersihan *temporary data*, sehingga hanya data yang telah ditransformasi yang di-load ke dalam *data warehouse* akan digunakan.

2.2 Aspek Legalitas *Web scraping*

Menurut beberapa pakar melakukan *web scraping* terhadap suatu web masih kontroversial. Menurut Fisher et al. (2010), Ketentuan Layanan (*term of service-ToS*) menciptakan hambatan spesifik untuk mengumpulkan informasi sebagai contoh ToS Facebook yang menyebutkan "*You will not collect users' information, or otherwise access Facebook, using automated means (such as harvesting bots, robots, spiders, or scrapers) without our permission.*"

Hirschey (2014) mengatakan *scraping* yang dapat merusak web dapat diadukan sebagai pelanggaran hak akses. O'Reilly (2006) menyebutkan *scraping* terkait dengan hak cipta (*intellectual property*). Krotov dan Silva (2018) menyebutkan bahwa aspek legalitas dan etika melakukan *web scraping* ini masih *grey area*.

Pada dasarnya teknik *web scraping* adalah proses *copy-paste* dari suatu web namun dilakukan secara otomatis dan terorganisir (Saurkar, et.all, 2018). Teknik *web scraper* mengambil data yang tersaji (*publish*) dalam suatu web, bukan merupakan teknik *hacking* yang mengambil data langsung dari server/database pemilik web (*backward database*). Dengan demikian, baik pemilik web maupun yang melakukan *scraping*, kedua belah pihak sebenarnya mendapat keuntungan, pihak pemilik web mendapat *traffics* yang dapat meningkatkan valuasinya sementara pihak yang melakukan *scraping* mendapat data yang dibutuhkan. Walaupun di sisi lain, penggunaan *web scraping* dapat mempengaruhi performa akses bagi pengguna lainnya.

2.3 Gambaran Online Marketplace

E-commerce atau perdagangan elektronik adalah jual beli barang dan jasa melalui internet (Khan, 2016). *E-commerce* aktivitas bisnis yang menggunakan teknologi informasi untuk meningkatkan penjualan, efisiensi bisnis dan memperkenalkan produk dan layanan baru (Išoraitė dan Neringa, 2018).

E-commerce melibatkan banyak kegiatan perdagangan barang dan jasa yang dapat diterapkan secara luas di bidang elektronik; transaksi keuangan dalam penyediaan perbankan, leasing, asuransi dan jasa keuangan lainnya, investasi, transaksi mata uang dan sekuritas; hotel, pariwisata, pendidikan, konsultasi, pembayaran utilitas, iklan, dan lainnya. (Misevičiūtė, 2001 dalam Išoraitė dan Neringa, 2018).

Online marketplace ataupun toko online merupakan bagian kecil dari *e-commerce* ini, keduanya adalah tempat terjadinya transaksi jual beli antara penjual dan pembeli secara *online*. Jika dianalogikan, marketplace adalah ibarat mall atau pasar yang menyediakan sejumlah fasilitas berupa lapak-lapak yang bisa digunakan penjual untuk berjualan. Dalam hal ini marketplace berperan sebagai pihak ketiga yang menyediakan tempat berjualan dan fasilitas pembayarannya. Toko *online* hampir sama dengan marketplace, namun satu hal yang mendasar yang membedakannya adalah lapak yang digunakan untuk berjualan. Jika berjualan di *marketplace* diibaratkan dengan menyewa lapak, maka berjualan di toko *online* lapaknya milik sendiri. Namun untuk memperluas pemasarannya, saat ini tidak jarang ditemui pemilik toko online yang membuka lapak di *marketplace* juga.

Perilaku konsumen dalam berbelanja telah berubah. Konsumen cenderung menyukai belanja online karena cenderung lebih praktis, aman, dapat dilakukan setiap waktu tanpa harus keluar rumah serta dapat membandingkan harga dari toko yang satu dengan toko lainnya. Berubahnya perilaku berbelanja konsumen yang bergeser ke pasar online tersebut membuat semakin pesatnya pertumbuhan *marketplace* di Indonesia (Lestari, 2015).

Direktur Pemberdayaan Informatika, Direktorat Jenderal Aplikasi Informatika Kementerian Kominfo, Septiana Tangkary menyatakan pertumbuhan nilai perdagangan elektronik di Indonesia mencapai 78 persen, tertinggi di dunia (Kemkominfo, 2019). Data analisis Ernst & Young dalam Kemkominfo, (2015), menunjukkan bahwa pertumbuhan nilai bisnis online di Indonesia berkembang sebesar 40% setiap tahunnya. Dari data tersebut tampak bahwa perkembangan bisnis online termasuk transaksi melalui *marketplace* cukup tinggi.

IPrice (2020) mengurutkan pemain besar marketplace di Indonesia berdasarkan rata-rata pengunjung website di setiap kuartal. Berdasarkan Peta *E-commerce* Indonesia yang dirilis iPrice

tersebut, sepanjang tahun 2019 Kuartal I-IV, yang datanya terakhir diperbaharui pada tanggal 3 Maret 2020, lima pemain besar adalah Shopee, Tokopedia, Bukalapak, Lazada dan Blibli.

2.4 Asas Pemungutan Pajak

Penghasilan adalah setiap tambahan kemampuan ekonomis yang diterima atau diperoleh Wajib Pajak, baik yang berasal dari Indonesia maupun dari luar Indonesia, yang dapat dipakai untuk konsumsi atau untuk menambah kekayaan Wajib Pajak yang bersangkutan, dengan nama dan dalam bentuk apa pun (Undang-Undang Nomor 36 Tahun 2008). Penghasilan ini lah yang menjadi objek Pajak Penghasilan (PPh).

PPh adalah jenis pajak subjektif sehingga penentuan subjek pajak atau pihak yang menjadi sasaran pengenaan pajak, menjadi hal yang sangat penting dalam penerapan sistem PPh, apakah suatu pihak memiliki kewajiban atau tidak dalam pemenuhan kewajiban perpajakan atas penghasilan yang diterima atau diperolehnya. Subjek pajak yang dikenai kewajiban untuk memenuhi kewajiban perpajakan inilah yang disebut Wajib Pajak (WP).

Subjek pajak dibedakan menjadi subjek pajak dalam negeri dan subjek pajak luar negeri. Subjek pajak orang pribadi dalam negeri akan menjadi WP apabila telah menerima atau memperoleh penghasilan yang besarnya melebihi Penghasilan Tidak Kena Pajak (PTKP). Subjek pajak badan dalam negeri menjadi WP sejak saat didirikan, atau bertempat kedudukan di Indonesia dan sekaligus menjadi WP karena menerima dan/atau memperoleh penghasilan yang bersumber dari Indonesia atau melalui bentuk usaha tetap di Indonesia. Dengan perkataan lain, Wajib Pajak adalah orang pribadi atau badan yang telah memenuhi kewajiban subjektif dan objektif, sehingga kepadanya akan dilakukan pemungutan pajak.

Terkait pemungutan pajak ini, Adam Smith (dalam Guzy, 2017), yang dikenal luas sebagai pencetus prinsip perpajakan menyebutkan ada empat prinsip yang harus dipegang teguh dalam melakukan pemungutan pajak, yaitu 1) *Equality*, atau asas keadilan yaitu pemungutan pajak yang dilakukan oleh negara harus sesuai dengan kemampuan dan penghasilan wajib pajak. Negara tidak boleh bertindak membedakan terhadap wajib pajak; 2) *Certainty*, atau asas kepastian hukum yaitu semua pungutan pajak harus berdasarkan Undang-Undang sehingga ada kepastian hukum, yaitu kepastian tentang subyek pajak, obyek pajak, tarif pajak yang berlaku dan kepastian tentang prosedur pajak; 3) *Convenience of Payment*, atau asas kenyamanan yaitu pajak harus dipungut pada saat yang tepat bagi wajib pajak; dan 4) *Efficiency*, asas efisien atau asas ekonomis yaitu biaya pemungutan pajak diusahakan seekonomis mungkin, tidak lebih besar dari hasil pemungutan pajak itu sendiri.

Keadilan pajak atau asas *equality* menurut Adam Smith tersebut adalah adanya kesebandingan beban pajak yang dikenakan berdasarkan setiap kemampuan individu dalam membayar pajak (*ability to pay*) dan sesuai dengan manfaat yang diterima. Terdapat dua macam kriteria yaitu kemampuan membayar dan manfaat yang diterima. Kedua kriteria tersebut menurut Musgrave dan Musgrave (1989) terbagi menjadi dua pendekatan untuk mencapai asas keadilan, yaitu Pendekatan Manfaat (*Benefit Approach*) dan Pendekatan Kemampuan Membayar (*Ability-to-Pay Approach*). Pada *Benefit Approach*, Wajib Pajak dibebankan sesuai dengan manfaat yang diterima dari layanan yang diberikan pemerintah, sehingga setiap Wajib Pajak membayar berbeda sesuai dengan manfaat yang diterima dari pengeluaran pemerintah untuk melakukan layanannya. Sedangkan pada *Ability-to-Pay Approach*, Wajib Pajak dibebankan sesuai dengan kemampuan membayar masing-masing.

Berdasarkan pendekatan kemampuan membayar ini, dikenal istilah keadilan horizontal dan keadilan vertikal. Keadilan horizontal menurut Devas (1988) adalah beban pajak haruslah sama benar antara berbagai kelompok yang berbeda tetapi dengan kedudukan ekonomi yang sama. Sedangkan keadilan vertikal adalah kelompok yang memiliki sumber daya yang besar membayar lebih banyak daripada yang memiliki sumber daya kecil.

Asas *certainty*, ditunjukkan dengan jelas dalam Undang-undang Ketentuan Umum Perpajakan yang mengatur hukum formal, di mana di dalamnya diatur subjek, objek dan prosedur perpajakan. Undang-undang Pajak Penghasilan (PPh) dan Undang-undang Pajak Pertambahan Nilai dan Pajak Penjualan atas Barang Mewah (PPN dan PPnBM) yang mengatur hukum material, di dalamnya diatur tarif pajak. Tarif pajak sebagaimana disebutkan dalam undang-undang tersebut terdiri dari : 1) tarif pajak progresif merupakan tarif pungutan pajak yang mana persentase akan naik sebanding dengan dasar pengenaan pajaknya. Tarif pajak progresif ini diterapkan untuk pajak penghasilan (PPh) wajib pajak orang pribadi; 2) Tarif proporsional, yang diterapkan di PPh Wajib Pajak Badan, Pajak Pertambahan Nilai (PPN) dan Wajib Pajak yang memiliki peredaran bruto tertentu, di mana besarnya persentase tarif tetap meski terjadi perubahan terhadap dasar pengenaan pajak; 3) Tarif flat/terap yang diartikan sebagai tarif pajak yang akan selalu tetap sesuai dengan peraturan yang telah diberlakukan, seperti Bea Meterai dengan nominal tarif sebesar Rp3.000 dan Rp6.000.

Begitupun asas Convenience of Payment, dalam Undang-undang PPh maupun PPN dan PPnBM, diatur saat terutangnya pajak dan saat pembayaran pajak, yang memberi kejelasan kepada Wajib pajak kapan saatnya melakukan pemotongan/pemungutan dan penyetoran/pembayarannya.

Terkait asas *Efficiency*, dalam ketentuan perundang-undangan perpajakan menyebutkan bahwa sistem pemungutan adalah *self assessment*, di mana wajib pajak sendiri yang akan melakukan penghitungan, pemotongan/pemungutan, penyetoran/pembayaran dan pelaporan pajaknya, sehingga tentunya akan memperkecil biaya pemungutan karena wajib pajak sendiri yang aktif dalam melakukan kewajiban perpajakannya.

2.5 Penelitian Terdahulu

Josi, dkk (2014) melakukan penelitian implementasi web scraping pada sejumlah mesin pencari yang gratis seperti portal Garuda, Indonesian Scientific Journal Database (ISJD), dan Google Scholar, dengan hasil yang menyatakan bahwa dengan menerapkan teknik web scraping mampu mengekstrak informasi mengenai artikel jurnal ilmiah dari portal-portal tersebut sekaligus mampu menyimpan data hasil scraping pada database secara otomatis.

Sonya dan Prihandoko, (2016) melakukan penelitian pada tiga media online yaitu, Detikcom, Liputan6 dan VivaNews untuk mencari dan menganalisis informasi terkait bencana alam, dan menemukan bahwa Detikcom lebih banyak menghasilkan data yang relevan dibandingkan dengan Liputan6 dan VivaNews dan metode teknik Breadth-First Search sangat membantu dalam hal pencarian record dengan menelusuri URL utama hingga ke bagian terdalam link.

Suwiyanto, (2018) melakukan penelitian pada beberapa marketplace menggunakan aplikasi Kofax® Kapow® sebuah *tools* yang melakukan *scraping* secara proses robotik (Robotics Process Automation-RPA) atau menghasilkan informasi yang real-time. Hasil penelitiannya menghasilkan bahwa metode *scraping* dari snapshot marketplace menggunakan aplikasi tersebut mampu untuk menghasilkan data *e-commerce*. Ada 9 variabel yang berhasil dikumpulkan yaitu: 1) lokasi penjual (provinsi dan kota / kota); 2) nama produk; 3) kategori produk; 4) stok produk; 5) harga produk; 6) jumlah barang yang dijual; 7) Nama Penjual; 8) periode waktu (tahun dan bulan); dan 9) URL produk. Informasi yang dihasilkan membantu Badan Pusat Statistik untuk mendapatkan informasi tentang aktual dan fluktuasi harga dan transaksi produk secara bulanan, triwulanan, dan tahunan.

Fernando, et.al (2019) melakukan ekstraksi data produk di *marketplace* menggunakan web crawling dan *web scraping* dan memvisualisasikan hasilnya dengan aplikasi Power BI. Penelitian Fernando menunjukkan bahwa metode *web crawling* dan *web scraping* ini berguna untuk melakukan analisis data.

Ayani, et.al., (2019) melakukan pengujian untuk mengekstraksi data di tiga situs web *marketpalce*, yaitu Bukalapak, Elevenia, dan JD.id. Hasil pengujian dengan menggunakan *white box testing* dan *black box testing*, sistem telah mampu memberikan hasil terbaik

produk dari gabungan hasil pencarian di tiga situs *web marketplace* tersebut.

Wicaksono (2018) melakukan penelitian dengan teknik *web crawling* pada situs *olx.com* untuk mengestimasi tingkat potensi penerimaan pajak sektor Usaha Mikro Kecil dan Menengah (UMKM) melalui Big Data. Hasil penelitian Wicaksono tersebut menunjukkan adanya estimasi peningkatan penerimaan pajak UMKM bersumber data *e-commerce* untuk wilayah Jawa adalah sebesar 23.43% dari Pajak Penghasilan Orang Pribadi dan Badan Tahun 2016.

Direktur Jenderal *The Revenue Departement* (RD) Thailand, Ekniti Nitithanprapas, dalam Bangkokpost (2019) mengatakan bahwa RD berencana untuk meluncurkan uji *web scraping*, dengan tujuan untuk mencegah penghindaran pajak oleh pelaku bisnis online murni dan sebagai informasi awal penjual online seperti lokasi dan jumlah pengunjung di suatu situs web.

OECD (2017) menyebutkan Austria melakukan pemantauan melalui internet menggunakan *tools scraping*, baik menggunakan *tools* yang *open source* maupun *tools* yang dibuat khusus, menghasilkan data sebagai *feeding* dalam hal pemantauan kepatuhan pajak dan edukasi perpajakan. Upaya kepatuhan ini menghasilkan penerimaan PPN sebesar EUR 10 juta, dan pemasukan SPT secara sukarela sejumlah 44 SPT dengan nilai PPN sebesar EUR 5,5 juta. Belgia menggunakan internet *scraping* bersamaan dengan alat analisis lain seperti toolkit forensik untuk mengumpulkan data, memilah data, menganalisis data dan melakukan beberapa pemeriksaan standar. Inggris menggunakan *tools* yang disebut COSAIN yang mengotomatisasikan pengumpulan data bersumber dari media sosial dan situs *web* lainnya. *Tools* yang termasuk *web scraping* ini dapat menyusun profil wajib pajak, dapat digunakan untuk memantau tren transaksi online dalam area geografis/wilayah tertentu atau sektor bisnis tertentu.

OECD, 2018 menyebutkan beberapa otoritas pajak telah berupaya meningkatkan jumlah informasi melalui teknik *web scraping*, walaupun efektivitasnya bergantung pada sistem di platformnya masing-masing web. Sebagai contoh, *The Finnish Tax Administration* (FTA), otoritas pajak Finlandia menggunakan teknik *web scraping* untuk memperoleh informasi pajak langsung dari platform, namun, data yang diperoleh dengan cara ini sering tidak lengkap dan menghadapi hambatan administratif.

3. METODOLOGI PENELITIAN

Metodologi dari penelitian ini dibagi dalam desain penelitian, metode pengumpulan data dan metode analisis data.

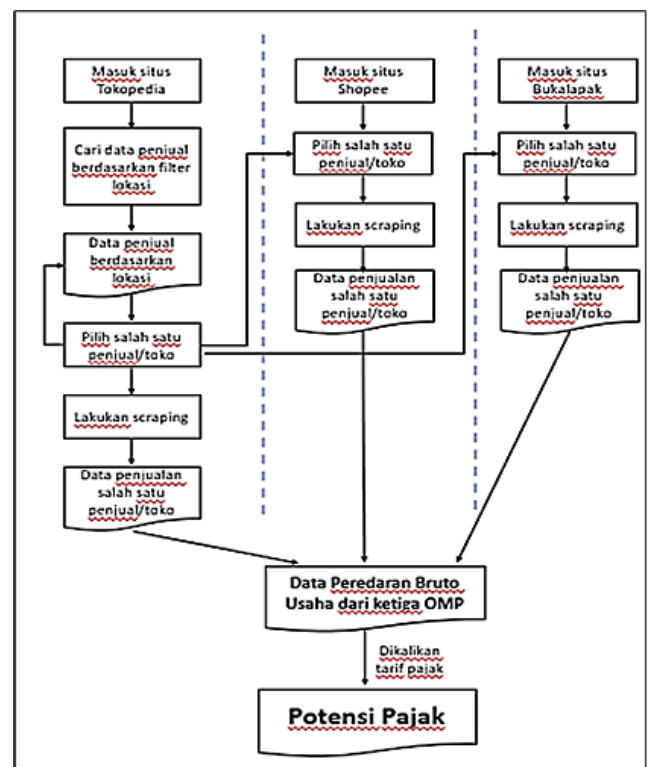
3.1 Desain Penelitian

Tulisan ini menggunakan studi informetrika, yaitu studi yang mencakup studi empiris literatur dan dokumen (Hood dan Wilson, 2001). Dalam artikel

Hood dan Wilson tersebut juga disebutkan beberapa pengertian informetrika ini, di antaranya (1) Tague-Sutcliffe yang menyatakan informetrika adalah studi tentang aspek kuantitatif informasi dalam bentuk apa pun, bukan hanya catatan atau bibliografi, dan dalam kelompok sosial apa pun, bukan hanya ilmuwan; (2) Ingwersen & Christensen yang menyatakan informetrika adalah ekstensi terbaru dari analisis bibliometrik tradisional juga meliputi komunitas non-ilmiah di mana informasi yang dihasilkan, dikomunikasikan dan digunakan; dan (3) Wilson yang menyatakan informetrik adalah studi kuantitatif dari satuan ukuran berpotensi informatif, diarahkan ke pemahaman ilmiah menginformasikan proses di tingkat sosial.

Studi informetrika mencakup kajian terhadap aspek kuantitatif informasi dalam bentuk apapun termasuk catatan, bibliografi, dari kelompok manapun, dalam hal ini termasuk informasi yang tersaji di internet. Dalam penelitian ini akan dikumpulkan informasi kuantitatif berupa data transaksi para penjual yang terekam pada sejumlah situs online marketplace sebagai berikut: 1) Tokopedia; 2) Shopee; dan 3) Bukalapak. Pemilihan ketiga marketplace tersebut karena ketiganya menduduki tiga besar marketplace di Indonesia sepanjang tahun 2019 versi iPrice (2020).

Desain penelitian mengikuti alur sebagaimana Gambar 2. Situs Tokopedia akan menjadi rujukan awal dalam pencarian data, karena di situs ini fasilitas filteringnya lebih lengkap dibanding Shopee dan Bukalapak dan dalam situs Tokopedia ini terdapat informasi nama deskripsi toko dan alamat toko.



Gambar 2 Desain Penelitian

3.2 Metode Pengumpulan Data dan Instrumen Penelitian

Penelitian ini menggunakan data yang tersaji pada ketiga marketplace menggunakan tools/software Web Scraper dari webscraper.io dan Instant Data Scraper dari webrobots.io, keduanya merupakan extension dari Chrome. Data yang dikumpulkan berupa daftar penjual yang berlokasi di wilayah Kota Cirebon pada ketiga marketplace tersebut, info toko yang berisi deskripsi toko, alamat dan tanggal buka toko, serta info produk yang meliputi nama produk, jumlah terjual dan harga masing-masing produk.

Instrumen Penelitian meliputi:

- a. Kuesioner, untuk justifikasi dan memperkuat masalah penelitian, dilakukan pengambilan data pendahuluan terhadap seluruh peserta Pelatihan Teknis *Account Representative* (AR) Angkatan IV Tahun 2020 yang berlangsung dari tanggal 24 Februari hingga 28 Februari 2020 di Pusdiklat Pajak. Para peserta pelatihan adalah para AR yang bertugas di beberapa Kantor Pelayanan Pajak yang tersebar di beberapa wilayah Indonesia. Pengambilan data berupa respon terhadap kuissoner yang dilakukan pada tanggal 28 Februari 2020.

Pertanyaan yang diajukan dalam kuesioner tersebut adalah (1) terkait kuantitas/kualitas data yang diturunkan oleh Kantor Pusat DJP ataupun Kanwil masing-masing, (2) Ada atau tidaknya kebutuhan tambahan data selain data yang telah diturunkan tersebut? (3) upaya yang telah dilakukan dalam mencari tambahan data, dan (4) metode apa yang digunakan dalam pencarian data.

- b. Aplikasi Web Scraper dan Instant Data Scraper yang di install sebagai *plugin* Chrome, untuk pengambilan data nama toko, alamat toko, nama barang, jumlah terjual dan harga barang. Dengan menggunakan teknik *web scraping*, data tersebut yang akan ditarik sebagai data yang akan digunakan dalam penggalan potensi pajak.

Untuk mengambil data di Tokopedia dan Bukalapak, digunakan aplikasi *Web Scraper*, dengan hasil berupa *file Comma Separated Values* (CSV) atau file teks yang berisi daftar data yang masing-masing baris (*row*) datanya dipisahkan oleh koma, sehingga perlu di *convert* ke dalam file berekstensi xls atauxlsx menggunakan Microsoft Excel. Sementara, untuk mengambil data di Shopee, digunakan aplikasi *Instant Data Scraper*, dengan hasil berupa data dalam bentuk file berekstensi xlsx.

Kedua aplikasi yang digunakan dalam penelitian ini, masing-masing memiliki keunggulan dan kelemahan. Aplikasi Web Scraper memiliki keunggulan dalam mengambil data hingga semua level dalam suatu website, namun aplikasi ini memerlukan waktu yang lama dalam proses *scraping*nya. Dalam Tokopedia ataupun Bukalapak, data yang akan

diambil terdapat dalam level berikutnya setelah halaman muka.

Aplikasi Instant Data Scraper memiliki keunggulan dalam hal penggunaan yang sederhana (*user friendly*) dan kecepatan proses *scraping*, namun kelemahannya, data yang dapat diambil hanyalah data yang tersaji di halaman muka, dan tidak bisa menarik data untuk level berikutnya. Di Shopee, data yang akan diambil, seluruhnya tersaji dalam halaman muka.

Pengambilan data dilakukan pada OMP Tokopedia dan Bukalapak dilakukan tanggal 4 hingga 6 Juli 2020, sementara pada OMP Shopee dilakukan tanggal 6 Juli 2020.

3.3 Analisis Data

Data yang diperoleh ditampilkan dalam bentuk tabel dan di urutkan berdasarkan alamat, sehingga dapat diketahui data toko yang alamatnya lengkap atau tidak. Data yang lengkap akan mempermudah untuk penggalan potensi pajak.

Selanjutnya terhadap data yang lengkap, dilakukan *scraping* untuk mendapatkan data penjualan bruto pada masing-masing OMP. Dari data penjualan bruto inilah akan dihitung berapa potensi pajak toko-toko tersebut.

Atas potensi pajak tersebut selanjutnya akan dibahas terkait dengan asas pemungutan pajak sebagaimana Adam Smith. Apakah memang seharusnya para penjual di ketiga OMP tersebut terdaftar dalam administrasi perpajakan Indonesia dan membayar pajaknya atau tidak. Terkait teknis ekstraksi data dari ketiga OMP tersebut juga akan dibahas apakah ada kendala yang ditemui dan apakah sudah sesuai dengan pendekatan konseptual untuk pengembangan aplikasi *web scraping* yang disampaikan Milev. Dengan demikian dapat diketahui di bagian mana yang perlu dilakukan penyesuaian jika diperlukan.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Hasil ekstraksi data di *website* Tokopedia dengan tools Web Scraper yang dilakukan pada tanggal 4-6 Juli 2020, menunjukkan bahwa jumlah penjual di Tokopedia dengan dengan filter lokasi di wilayah Kota Cirebon berjumlah 4.913 toko.

Pada bagian alamat, seluruh penjual tidak menyebutkan alamat yang jelas, sebagian besar penjual menuliskan alamat hanya sebatas kecamatan dan kode pos. Sebaliknya beberapa penjual menuliskan alamat jelas di bagian deskripsi, bahkan beberapa diantaranya menuliskan nomor kontak yang bisa dihubungi. Walaupun proses *scraping* menggunakan filter alamat Kota Cirebon, namun dari hasil *scraping* ternyata tidak semua penjual beralamat di Kota Cirebon, ada juga yang beralamat di Kabupaten Cirebon, Jakarta, Bekasi dan Bandung. Hasil ini bukan merupakan kesalahan proses *scraping*,

namun data yang tersaji dalam OMP tersebut memang demikian adanya.

Selanjutnya, dari data penjual di OMP Tokopedia di atas, diambil tiga contoh penjual untuk diambil data detail penjualannya, yang meliputi nama barang, jumlah terjual dan masing-masing harganya. Ketiga contoh yang diambil datanya tersebut tidak mengikuti kaidah pengambilan *sample*, namun hanyalah contoh yang dipilih berdasarkan *professional judgement* peneliti. Pemilihan contoh ini, semata-mata karena ketiga contoh tersebut memiliki data alamat yang dapat ditelusuri keberadaannya, untuk mempermudah tindak lanjut data yang diperoleh tersebut. Selanjutnya, dari ketiga contoh penjual tersebut, diekstrak data penjualan dari OMP Shopee dan Bukalapak. Hasil dari ekstraksi data penjualan pada tiga situs OMP tersebut adalah sebagai berikut:

Tabel 1. Penjualan Bruto di Beberapa Online Marketplace

Profil Penjual	Online marketplace	Penjualan Bruto
Nama Toko: Jer***ID Deskripsi: Menjual batik tulis, mesin kasir dan mixer heavy duty, dll Alamat: Pekalipan, Kota Cirebon, 45116 Buka Sejak: Oktober 2014	Tokopedia	1.428.691.000
	Bukalapak	0
	Shopee	0
	Jumlah	1.428.691.000
Nama Toko: Ab***ni Deskripsi: Menyediakan Produk-produk Ikan asin khas Cirebon dan Indramayu. (WA 0838-2393-7829) Alamat: Cirebon Buka Sejak: Agustus 2014	Tokopedia	225.126.000
	Bukalapak	44.719.000
	Shopee	45.883.500
	Jumlah	315.728.500
Nama Toko: Mel***on Deskripsi: Menjual Pakaian Wanita, Pria, Muslim, Anak, Pesta, dan Pengantin Alamat: Jl Gunung Pangrango I No 17 RT 03/01 Larangan Harjamukti Cirebon Buka Sejak: Januari 2015	Tokopedia	855.000
	Bukalapak	0
	Shopee	7.252.906.222
	Jumlah	7.253.761.222

Sumber : Hasil olahan Penulis

Ketentuan perpajakan menyatakan bahwa atas penghasilan dari usaha yang diperoleh Wajib Pajak (WP) dengan jumlah peredaran bruto tertentu, dikenai Pajak Penghasilan (PPh) yang bersifat final dengan tarif proporsional sebesar 0,5% (nol koma lima persen). Jumlah peredaran bruto dimaksud adalah jumlah yang tidak melebihi Rp 4.800.000.000,00 (empat miliar delapan ratus juta rupiah) dalam satu tahun pajak (Peraturan Pemerintah (PP) Nomor 23 Tahun 2018). Namun apabila melebihi Rp 4.800.000.000,00, maka terhadap WP tersebut dikenai PPh dengan tarif progresif berdasarkan Pasal 17 Undang-Undang PPh dengan tarif bervariasi minimal 5% hingga 30% berdasarkan Penghasilan Kena Pajak yang diperoleh penjual tersebut (Undang-Undang Nomor 36 Tahun 2008).

Sebagai contoh dalam tulisan ini, digunakan perhitungan menurut skema PP 23 tahun 2018, dengan tarif progresif yang bersifat final sebesar 0,5%. Sehingga terdapat potensi PPh pada Toko Jer***ID sebesar $Rp1.428.691.000 \times 0,5\% = Rp. 7.143.455$; Potensi pajak Toko Ab***ni sebesar $Rp315.728.500 \times 0,5\% = Rp. 1.578.643$; Begitupun untuk Toko Mel***on terdapat potensi pajak dari penjualan broto sebesar Rp. 7.253.761.222.

Seharusnya tidak dikenai tarif sesuai PP 23, namun sebagai gambaran saja jika dikenai tarif sebesar 0,5% terdapat potensi pajak sebesar Rp. 36.268.806.

Potensi Pajak tersebut selanjutnya oleh AR akan ditelusuri pada Surat Pemberitahuan (SPT) PPh yang telah disampaikan penjual tersebut ke kantor Pelayanan Pajak (KPP). Ada tiga kemungkinan terjadi :

- 1) Jumlah peredaran bruto yang dilaporkan di SPT lebih besar potensi yang ada, maka hal ini bukan prioritas untuk ditindaklanjuti.
- 2) Jumlah peredaran bruto yang dilaporkan di SPT tidak lebih besar dari potensi yang ada, maka hal ini menjadi prioritas untuk ditindaklanjuti.
- 3) Penjual tidak terdaftar dalam administrasi Direktorat Jenderal Pajak atau tidak memiliki NPWP, sehingga tidak ada data SPTnya, maka hal ini menjadi prioritas untuk ditindaklanjuti.

Walaupun teknik *web scraping* ini dapat membantu dalam ekstraksi data dari situs OMP ini, namun terdapat juga beberapa kendala dalam implementasinya, sehingga tidak seluruh data yang diharapkan dapat terekstraksi. Beberapa kendala tersebut di antaranya:

- 1) Sangat membutuhkan akses internet yang stabil, sehingga ketika tiba-tiba akses terputus, proses ekstraksi data pun terhenti.
- 2) Data yang dihasilkan tidak terlalu rapi, menyisakan beberapa kolom yang berisi text yang tidak diinginkan seperti tag HTML
- 3) Beberapa data yang diharapkan diproteksi oleh *web developer* situs yang bersangkutan, sehingga tidak dapat diekstraksi.
- 4) Data yang tersaji tidak lengkap, misalnya ada beberapa penjual yang info tokonya lengkap berisi deskripsi nama pemilik toko, deskripsi toko dan alamat tokonya, dan ada beberapa info toko yang tidak lengkap datanya.
- 5) Data yang tersaji adalah data saat proses scraping dilakukan, sehingga tidak dapat menelusuri perubahan data dari sebelumnya. Misalnya data harga produk adalah harga saat proses scraping dilakukan, dapat saja harga tersebut telah berubah dari harga ketika produk tersebut sebelumnya telah terjual.

4.2 Pembahasan

Sejatinya pernah terbit ketentuan berupa Peraturan Menteri Keuangan Nomor 210/PMK.010/2018 (PMK 210). Dalam ketentuan tersebut dinyatakan bahwa penjual di OMP melaksanakan kewajiban Pajak Penghasilan sesuai dengan ketentuan perundang-undangan di bidang Pajak Penghasilan. Selain itu, bagi penjual yang berstatus sebagai Pengusaha Kena Pajak (PKP) wajib memungut, menyetor, dan melaporkan Pajak Pertambahan Nilai (PPN) atau Pajak Pertambahan Nilai dan Pajak Penjualan atas Barang Mewah (PPnBM) yang terutang. Penyedia Platform Marketplace (OMP) juga wajib melaporkan rekapitulasi transaksi perdagangan yang dilakukan oleh penjual ke DJP. Namun sebelum dilaksanakan,

ketentuan PMK 210 tersebut dicabut dan dinyatakan tidak berlaku oleh PMK nomor 31/PMK.010/2019, sehingga pelaksanaan kewajiban perpajakan atas penjual di OMP menjadi tertunda. Begitupun dengan data transaksi di OMP urung didapatkan DJP.

Para penjual di OMP tentunya mendapatkan penghasilan dari transaksi yang dilakukannya di OMP tersebut. Sesuai dengan asas equality Adam Smith, maka para penjual tersebut harus membayar pajak atas penghasilan yang diperolehnya, sebagaimana halnya para penjual yang bertransaksi secara offline melalui toko fisik. Akan sangat tidak adil jika penjual yang bertransaksi melalui toko fisik dikenai pajak, namun penjual yang bertransaksi melalui OMP tidak dikenai pajak.

Sesuai asas certainty dan efficiency, sebagaimana Adam Smith (dalam Guzy, 2017), dimana menurut ketentuan perpajakan sudah diatur tatacara pemenuhan kewajiban perpajakannya secara self assesment, para penjual wajib mendaftarkan diri pada administrasi DJP, menghitung, memperhitungkan, menyeter dan melaporkan pajaknya secara mandiri. Pada kenyataannya, para penjual di OMP tersebut masih belum melaksanakan kewajiban pajaknya secara mandiri atau belum sepenuhnya melaksanakan kewajiban pajak dengan benar sesuai ketentuan perundang-undangan perpajakan, maka para AR harus bertindak sesuai kewenangan yang dimilikinya.

Dan sesuai azas convenience of payment Adam Smith, dimana para wajib pajak akan lebih nyaman membayar pajak ketika mereka masih memiliki uang, maka para AR harus segera menagihnya ketika toko masih aktif di OMP.

DJP secara institusi masih hanya menunggu data yang diberikan oleh penyedia platform OMP, belum secara aktif mencari dan mengekstraksi data yang tersaji di internet. Oleh karenanya para AR harus segera dan aktif mencari data secara mandiri, tidak hanya menunggu feeding dari DJP. Salah satu cara yang dapat dilakukan AR adalah dengan melakukan teknik web scraping pada OMP. Hasil kegiatan web scraping pada ketiga OMP ini menunjukkan bahwa data hasil ekstraksi tersebut dapat dimanfaatkan untuk tujuan penggalan potensi pajak.

Walaupun begitu, karena implementasi teknik *web scraping* ini masih menemui banyak kendala, dan untuk dapat menghasilkan data yang baik, dapat diaplikasikan delapan tahapan pendekatan konseptual untuk pengembangan aplikasi web scraping yang ditawarkan Milev (2017) yang sudah diintegrasikan dengan Business Intelligent (BI).

Teknik *web scraping* tradisional, seperti Web Scraper ataupun Instant Data Scraper, ternyata 1) baru memenuhi empat tahapan pertama sebagaimana pendekatan konseptual pengembangan *web scraping* Milev, atau baru sampai tahap mengunduh data hasil ekstraksi yang tersimpan di *temporary* aplikasi. Untuk dapat digunakan untuk penggalan potensi pajak, data

tersebut masih perlu divalidasi dengan data yang dimiliki DJP seperti Master File Wajib Pajak. 2) Proses scraping menggunakan teknik web scraping ini dieksekusi secara manual oleh AR, tidak secara robotik/otomatis, sehingga ketika AR membutuhkan data dari suatu penjual di OMP, maka AR akan melakukan kembali proses *scraping*.

Pada konsep Milev (2017), data hasil ekstraksi akan ditransformasi sesuai struktur *database* yang dimiliki DJP, dibersihkan dan di-load ke dalam suatu data warehouse secara robotik dan kontinyu, sehingga data tersebut dapat secara langsung terintegrasi dengan data internal yang dimiliki DJP. Data yang sudah terintegrasi ini akan mempermudah AR untuk mengeksekusi data potensi tersebut menjadi penerimaan pajak. Konsep yang ditawarkan Milev ini adalah bagian dari suatu sistem *business intelligence*, dimana sistem ini sangat dinamis dan terus menerus berkembang serta terintegrasi dengan berbagai sumber data dan format data yang berbeda.

5. KESIMPULAN

Data yang tersaji di internet begitu banyaknya, tidak mungkin dilakukan secara manual diekstraksi satu persatu. Sementara para AR tidak bisa menunggu data dari feeding yang diberikan DJP. Teknik web scraping dapat membantu para AR untuk mengekstraksi data dari OMP secara otomatis untuk digunakan dalam kegiatan penggalan potensi pajak. Dari contoh ekstraksi data dari OMP Tokopedia, Shopee dan Bukalapak, tampak bahwa data tersebut mengandung potensi pajak yang dapat dieksekusi oleh para AR untuk menambah jumlah Wajib Pajak terdaftar dan menambah penerimaan pajak.

Walaupun begitu, teknik *web scraping* ini masih termasuk teknik tradisional dalam ekstraksi data, memerlukan *effort* yang cukup besar disamping banyak kendala dalam teknik *web scraping* sendiri. Untuk melakukan teknik web scraping yang terbaik, yang dapat dilakukan oleh Direktorat Jenderal Pajak (DJP) adalah mengembangkan aplikasi web scraping sendiri yang terintegrasi sebagai suatu sistem Business Intelligence, sehingga memungkinkan data hasil ekstraksi dari ketiga OMP tersebut terintegrasi dengan database yang dimiliki DJP.

6. IMPLIKASI DAN KETERBATASAN

Banyaknya data dapat dimanfaatkan oleh DJP untuk penggalan potensi pajak, untuk memanfaatkan data yang melimpah di internet tersebut, DJP dapat mengeluarkan kebijakan seperti penerbitan Surat Edaran Direktur Jenderal Pajak (SE DJP) yang menginstruksikan para AR secara aktif dan mandiri melakukan pencarian dan ekstraksi data dari internet. DJP pun secara institusi, sudah saatnya secara aktif mencari dan mengekstraksi data yang tersaji di internet, bukan hanya menunggu data yang diberikan penyedia platform OMP. DJP dapat mengembangkan teknik ekstraksi data dari internet dengan mengaplikasikan sistem Business Intelligence sesuai

konsep Milev. Dalam hal ini DJP juga bisa belajar dari institusi pajak di negara lain atau belajar ke BPS yang telah menggunakan teknik scraping menggunakan aplikasi Kofax® Kapow® yang dapat mengekstraksi data secara proses robotik dan menghasilkan informasi yang *real-time*.

Penelitian ini dilakukan pada penjual di wilayah Kota Cirebon, semata-mata karena data yang ditarik cukup sebagai data penelitian dan tidak terlalu besar. Tiga penjual yang datanya diekstraksi hanyalah sebagai contoh, peneliti berikutnya dapat menambah contoh-contoh penjual lain baik di ketiga OMP tersebut di atas atau OMP-OMP lainnya.

Penelitian menggunakan aplikasi yang sederhana, tidak berbayar dan tidak memerlukan bahasa pemrograman (*coding*) yaitu Aplikasi Web Scraper dan Instant Data Scraper sebagai ekstensi dari Chrome, semata-mata karena unsur kemudahan, peneliti berikutnya dapat menggunakan tools atau aplikasi lain untuk melakukan *web scraping* ini.

DAFTAR PUSTAKA

Sumber Buku

- Ali, Muhammad. 1982, Penelitian Pendidikan Prosedur dan Strategi, Bandung, Angkasa.
- Mitchell, Ryan, 2015, Web scraping with Python Collecting Data from the Modern Web, O'Reilly, USA.
- Munzert, Simon, Christian Rubba, Peter Meißner, Dominic Nyhuis, 2015, Automated Data Collection with R: A Practical Guide to Web scraping and Text Mining, Wiley.
- Turland, M., 2010, php| architect's Guide to Web scraping, Marco Tabini & Associates, Inc, USA.

Jurnal

- Ayani, Dhita Deviacita, et.al, 2019, "Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace", Jurnal Sistem dan Teknologi Informasi (Justin), Vol. 7, No. 4, Oktober 2019.
- Fisher, D., McDonald, D. W., Brooks, A. L., & Churchill, E. F., 2010, "Terms of service, Ethics, and Bias: Tapping the social web for CSCW research", CSCW 2010, February 6–10, 2010, Savannah, Georgia, USA.
- Fernando, Edward Hanafi, et.al, 2019, "Ekstraksi dan Analisis Produk di Marketplace Secara Otomatis dengan Memanfaatkan Teknologi Web Crawling", Jurnal Teknik Informatika dan Sistem Informasi, Volume 5 Nomor 3 Desember 2019.
- Guzy, Magdalena Jarczok, 2017, The principles of tax law equality in the context of direct taxation, Journal of Economics and Management, Vol. 30 (4), University of Economics in Katowice, Poland.
- Hirschey, J. K., 2014. "Symbiotic Relationships: Pragmatic Acceptance of Data Scraping", Berkeley Technology Law Journal, Vol. 29, 2014.

- Hood, William W and Wilson, Concepción S, 2001, "The Literature Of Bibliometrics, Scientometrics, And Informetrics", Scientometrics, Vol. 52, No. 2 (2001) 291–314, Akadémiai Kiadó, Budapest.
- Josi, A., Abdillah L.A., Suryayusra, 2014, Penerapan Teknik Web scraping pada Mesin Pencari Artikel Ilmiah, dalam laman https://www.researchgate.net/publication/267214300_Penerapan_teknik_web_scraping_pada_mesin_pencari_artikel_ilmiah diakses tanggal 29 Januari 2020
- Khan, Abdul Ghafar, 2016, "Electronic Commerce : A Study on Benefits and Challenges in an Emerging Economy", Global Journal of Management and Business Research: B Economics and Commerce Volume 16 Issue 1 Version 1.0 Year 2016.
- Krotov, Vlad dan Leiser Silva, 2018, "Legality And Ethics Of Web Scraping", Emergent Research Forum (ERF), Twenty-fourth Americas Conference on Information Systems, New Orleans,
- Lestari, Sri Budi, 2015, Shopping Online sebagai Gaya Hidup, Jurnal Ilmu Sosial, Vol.14 No. 2, November 2015.
- Milev, Plamen, 2017, "Conceptual Approach for Development of Web scraping Application for Tracking Information", Economic Alternatives, 2017, Issue 3, pp. 475-485.
- O'Reilly, S., (2006, "Nominative Fair Use and Internet Aggregators: Copyright and Trademark Challenges Posed by Bots, Web Crawlers and Screen-Scraping Technologies", Loyola Consumer Law Review Volume 19 Issue 3.
- Phillips, David M. W, 2016, Web scraping with Excel : How to Use VBA to Write Simple and Effective Web Scrapes, CreateSpace Independent Publishing Platform
- Prowebscraper.com, 2020, A Must-Have Guide to Web scraping Using Chrome dalam <https://prowebscraper.com/blog/a-must-have-guide-to-web-scraping-using-chrome/> diakses tanggal 21 Maret 2020.
- Saurkar, Anand V., Kedar G. Pathare, Shweta A. Gode, 2018, "An Overview On Web scraping Techniques And Tools", International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE), Volume: 4 Issue: 4, April 2018.
- Sonya, Izatul Putri dan Prihandoko, 2016, "Analisis Web scraping Untuk Data Bencana Alam Dengan Menggunakan Teknik Breadth-First Search Terhadap 3 Media Online", Jurnal Informatika dan Komputer Volume 21 No. 3, Desember 2016.
- Wicaksono, Banu, 2018, Meningkatkan Potensi Pajak UMKM Online Melalui Data E-commerce, Simposium Nasional Keuangan Negara 2018.
- Zhao, Bo, 2017, Web scraping, Springer International Publishing AG (outside the USA) 2017.

Naskah dari Internet

- Arni, Ulti Desi, 2018, Bukalapak vs Tokopedia vs Shopee, dalam laman <https://garudacyber.co.id/artikel/1246-bukalapak-vs-tokopedia-vs-shopee> diakses tanggal 7 April 2020.
- Badan Pusat Statistik, 2019, Statistik E-commerce 2019 diakses tanggal 7 April 2020.
- Bangkokpost, 2019, Web scraping a new tool in revenue hunt <https://www.bangkokpost.com/business/1604558/web-scraping-a-new-tool-in-revenue-hunt> diakses tanggal 8 April 2020.
- Botsol.com, 2020, The Ultimate Web scraping Service dalam botsol.com diakses tanggal 12 Maret 2020.
- Datareportal, 2020, Digital 2019 Spotlight: Ecommerce In Indonesia, dalam laman <https://datareportal.com/reports/digital-2019-ecommerce-in-indonesia> diakses tanggal 30 Januari 2020.
- Hiremath, Omkar S, 2020, A Beginner's Guide to learn web scraping with python dalam <https://www.edureka.co/blog/web-scraping-with-python/#whyweb> diakses tanggal 13 Maret 2020.
- Kemkominfo, 2015, Indonesia Akan Jadi Pemain Ekonomi Digital Terbesar di Asia Tenggara dalam https://www.kominfo.go.id/content/detail/6441/indonesia-akan-jadi-pemain-ekonomi-digital-terbesar-di-asia-tenggara/0/berita_satker diakses tanggal 13 Maret 2020
- Kemkominfo, 2019, Pertumbuhan e-commerce Indonesia Capai 78 Persen dalam

https://kominfo.go.id/content/detail/16770/kemkominfo-pertumbuhan-e-commerce-indonesia-capai-78-persen/0/sorotan_media diakses tanggal 13 Maret 2020.

- IPrice, 2020, Peta E-Commerce Indonesia, dalam laman <https://iPrice.co.id/insights/mapofecommerce/> diakses tanggal 7 April 2020.
- OECD, 2017, Technology Tools to Tackle Tax Evasion and Tax Fraud, OECD.
- , 2018, Tax Challenges Arising from Digitalisation – Interim Report 2018: Inclusive Framework on BEPS, OECD/G20 Base Erosion and Profit Shifting Project, OECD Publishing, Paris.
- Statista, Retail E-commerce Revenue Forecast from 2017 to 2024 (in billion U.S. dollars). In Statista – The Statistics Portal. Dari laman <https://www.statista.com/statistics/280925/e-commerce-revenue-forecast-in-indonesia/> diakses tanggal 30 Januari 2020
- Suwiyanto, Viktor dan Gusnisa Siswayu, 2018, "Snapshot Crawling to Produce and Disseminate Time Series Data of E-commerce: A Case Study in Market Place Site in Indonesia", Asia-Pacific Economic Statistics Week 2018.
- Webscraper.io, 2020, Web Scraper Making web data extraction easy and accessible for everyone dalam <https://webscraper.io/> diakses tanggal 13 Maret 2020.

Naskah Produk Kebijakan

- Peraturan Menteri Keuangan Republik Indonesia Nomor 210/PMK.010/2018 tentang Perlakuan Perpajakan atas Transaksi Perdagangan Melalui Sistem Elektronik (E-Commerce).