

Pengklasifikasian Surat Pemberitahuan Pajak Daerah Menggunakan Metode Regresi Logistik Biner Untuk Mengetahui Patuh Dan Tidak Patuh Dalam Pembayaran Pajak Daerah

Mohamad Ripai¹, Umi Hayati², Wita Widyawati³, Heliyanti Susana⁴, Fathurrohma⁵.

Program Studi Teknik Informatika, STMIK IKMI Cirebon, Kota Cirebon Indonesia^{1,2,4}
Program Studi Komputerisasi Akuntansi, STMIK IKMI Cirebon, Kota Cirebon Indonesia³
Program Studi Rekayasa Perangkat Lunak, STMIK IKMI Cirebon, Kota Cirebon Indonesia⁵

Email : ripaimohamad98@gmail.com¹, umihyati11@gmail.com², wita.widya202@gmail.com³, helyaynti12@yahoo.co.id⁴, fathru12@gmail.com⁵.

Email Penulis Korespondensi : ripaimohamad98@gmail.com

Submitted : 09-06-2022; Accepted 25-06-2022; Published 25-06-2022

Abstract— Taxes are an obligation of citizens in the form of service and active participation of citizens for national and regional development based on statutory regulations. However, the level of taxpayer compliance is still low because the target of local tax revenue has not been achieved, therefore researchers are interested in examining how the level of taxpayer compliance in the Regional Revenue and Finance Management Agency (BPKPD) Cirebon City. The purpose of this study is to classify which types and taxpayers are still in arrears on taxpayer compliance. The method used in this study is to use the Logistics Regression method, to find the relationship between the response variable (y) which is binary and the predictor variable (x) from 2 categories, namely on time and not which is denoted by $y=1$ (compliance) and $y=0$ (did not comply). The expected result is to know the level of taxpayer compliance based on the type of tax.

Keyword: Machine Learning, Local Taxes, Logistics Regression Biner, Data Mining, Classification.

Abstrak— Pajak adalah sebagai suatu kewajiban warga negara berupa pengabdian serta peran aktif warga negara untuk pembangunan nasional maupun daerah berdasarkan Peraturan Undang-Undang. Tetapi Tingkat kepatuhan wajib pajak yang masih rendah karena belum tercapainya target penerimaan pajak daerah, oleh karena itu peneliti tertarik untuk meneliti bagaimana tingkat kepatuhan wajib pajak di Badan Pengelola Keuangan dan Pendapatan Daerah (BPKPD) Kota Cirebon. Tujuan dari penelitian ini adalah untuk mengklasifikasikan jenis dan wajib pajak mana saja yang masih banyak menunggak terhadap Kepatuhan Wajib Pajak. Metode yang digunakan dalam penelitian ini yaitu menggunakan metode Regresi Logistik, untuk mencari hubungan antara variabel respon (y) yang bersifat biner dengan variabel prediktor (x) dari 2 kategori yaitu tepat waktu dan tidak yang dinotasikan dengan $y=1$ (patuh) dan $y=0$ (tidak patuh). Hasil yang diharapkan yaitu diketahuinya tingkat kepatuhan wajib pajak berdasarkan jenis pajaknya.

Kata Kunci: Machine Learning, Pajak Daerah, Regresi Logistik, Data Mining, Klasifikasi

I. PENDAHULUAN

Pajak daerah merupakan sumber pendapatan daerah yang penting guna membiayai pelaksanaan pemerintah

daerah. Undang-undang Nomor 28 tahun 2009 tentang Pajak Daerah dan Retribusi Daerah (UU PDRD) yang berlaku mulai tanggal 1 Januari 2010 memberikan perluasan objek pajak dan pemberian diskresi dalam penetapan tarif agar kemandirian daerah meningkat. Kemandirian daerah dicerminkan oleh besarnya kontribusi Pendapatan Asli Daerah (PAD) terhadap Anggaran Pendapatan dan Belanja Daerah (APBD). Masyarakat negara bagian dan local (Pemerintah Daerah) harus memiliki kebijakan pajak aktif untuk berkontribusi pada pencapaian tujuan ekonomi dan social mereka, dimana mereka harus menemukan ukuran yang memadai agar tidak membahayakan prinsip ekonomi dan pasar dan sebuah lingkungan ekonomi yang menguntungkan. UU PDRD mengatur bahwa jenis pajak yang dapat dipungut oleh pemerintah daerah Kabupaten/Kota meliputi 11 jenis pajak yaitu: Pajak Hotel, Pajak Restoran, Pajak Hiburan, Pajak Reklame, Pajak Parkir, Pajak Air Bawah Tanah, Pajak Penerangan Jalan Umum, Pajak Sarang Burung Walet, Pajak Pajak Mineral Bukan Logam dan Batuan, Pajak Bumi dan Bangunan Perdesaan dan Perkotaan (PBB P2), dan Bea Perolehan Hak atas Tanah dan Bangunan (BPHTB).

Beberapa kewenangan diberikan kepada pemerintah daerah agar pemungutan pajak daerah dapat berjalan optimal, antara lain penetapan tarif pajak, pemberian pengurangan pajak, pemberian insentif, dan penentuan nilai jual objek pajak. Dari sebelas jenis pajak kabupaten/kota tersebut, terdapat tiga jenis pajak yang dipungut secara official assessment, yaitu besarnya pajak yang terutang (pajak yang harus dibayar) ditentukan oleh pejabat pajak, meliputi Pajak Reklame, PBB P2, Pajak Air Bawah Tanah. Delapan jenis pajak lainnya dipungut secara self assessment, artinya Wajib Pajak sendiri yang menghitung berapa besarnya pajak yang harus dibayar. Sistem ini berlaku untuk Pajak Hotel, Pajak, Restoran, Pajak Hiburan, Pajak Parkir, Pajak Penerangan Jalan Umum, Pajak, Sarang Burung Walet, Pajak Mineral Bukan Logam, dan Batuan serta BPHTB. Dalam pemungutan pajak dengan system self assessment, peran kesadaran dan kejujuran Wajib Pajak sangat besar [1].

Menurut Peraturan Daerah Kota Cirebon Nomor 3 Tahun 2012 Surat Pemberitahuan Pajak Daerah (SPTPD) merupakan salah satu jenis surat yang ada dalam pajak daerah.

Surat ini digunakan oleh wajib pajak (WP) untuk melakukan pelaporan, perhitungan, dan pembayaran terhadap pajak, objek pajak maupun bukan objek pajak, harta, serta kewajiban lainnya sesuai dengan Undang-Undang pajak daerah.

Pada tahun 2020 di Kota Cirebon terdapat 1.383 wajib pajak yang terbagi menjadi beberapa jenis pajak, yaitu Pajak Restoran, Pajak Hotel, Pajak Hiburan, dan Pajak Parkir di Luar Badan Jalan. Jumlah tersebut cukup besar untuk peningkatan Pendapatan Asli Daerah Kota Cirebon. Namun masih banyak wajib pajak yang belum sadar akan kewajibannya dalam membayar pajak. Oleh karena itu Badan Pengelola Keuangan dan Pendapatan Daerah Kota Cirebon memerlukan tindakan yang tepat untuk mengetahui faktor-faktor yang mempengaruhi ketidak patuhan wajib pajak dalam membayarkan pajaknya melalui pengelompokan surat pemberitahuan pajak daerah. Untuk mengetahui faktor-faktor yang berpengaruh secara signifikan pada ketidakpatuhan dalam membayar pajak maka diperlukan adanya teknik data mining.

Data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan database besar yang tersimpan dan dalam penyimpanan menggunakan teknik pengenalan pola seperti teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna. Ada beberapa teknik dalam data mining, salah satu dari teknik data mining yaitu pengklasifikasian. Klasifikasi merupakan suatu proses untuk menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia [2].

Dalam pengklasifikasian data ada beberapa metode yang digunakan, salah satu metode yang bisa digunakan yaitu metode Binary Logistic Regression. Binary Logistic Regression adalah salah satu metode regresi yang digunakan untuk menganalisis hubungan antara variabel respon dan beberapa variabel bebas, dengan variabel responnya berupa data kualitatif dikotomi yaitu bernilai 1 untuk menyatakan keberadaan suatu karakteristik dan bernilai 0 untuk menyatakan ketidakberadaan sebuah karakteristik. Binary Logistic Regression tidak mengasumsikan hubungan linear antar variabel bebas dan terikat dikarenakan bentuk variabel terikat yang bersifat kategorik. Dalam Binary Logistic Regression, model yang didapat dengan respon kualitatif pada variabel dependen bertujuan untuk menentukan probabilitas individu dalam keputusan kualitatif [3].

Menurut penelitian sebelumnya yaitu oleh [4] Menjelaskan bahwa Binary Logistic Regression merupakan suatu metode Data Mining dan analisis statistika yang mendeskripsikan hubungan antara perubah respon (dependent variable) yang bersifat kualitatif memiliki dua kategori atau lebih dengan satu atau lebih perubah penjelas (independent variable) berskala kategori atau interval. Faktor kunci untuk mendapatkan hasil duga Binary Logistic Linear yang baik adalah besaran jumlah responden yang representatif dengan keragaman relatif tinggi. Oleh karena itu validasi data menjadi faktor penentu dan krusial dilakukan sebelum analisis data.

Berdasarkan latar belakang yang telah dijelaskan diatas penelitian tentang pengelompokan surat pemberitahuan pajak daerah menggunakan metode binary logistic regression perlu untuk dikaji untuk mencari hubungan antara variabel respon (y) yang bersifat biner dengan variabel prediktor (x) dari 2

kategori yaitu tepat waktu dan tidak yang dinotasikan dengan $y=1$ (patuh) dan $y=0$ (tidak patuh), sehingga Pemerintah Daerah dalam hal ini yaitu Badan Pengelola Keuangan dan Pendapatan Daerah Kota Cirebon dapat mengambil tindakan yang tepat terhadap wajib pajak.

II. METODOLOGI PENELITIAN

Jenis penelitian yang digunakan dalam penelitian ini adalah jenis penelitian kuantitatif. Penelitian kuantitatif salah satu jenis penelitian yang spesifikasinya adalah sistematis, terencana dan terstruktur dengan jelas sejak awal hingga pembuatan desain penelitian. Penelitian ini dilakukan dengan cara menganalisa data pembayaran wajib pajak menggunakan machine learning yang bertujuan untuk mengetahui apakah wajib pajak tersebut patuh atau tidak patuh dalam membayarkan pajaknya. Adapun tahapan metode penelitian dalam penelitian ini terdiri dari pengumpulan data, analisis data, implementasi algoritma, hasil penelitian dan pengujian, dan kesimpulan. Tahapan pengumpulan data merupakan tahap mengumpulkan dataset melalui observasi di tempat penelitian. Tahap analisis data dalam penelitian ini menggunakan software rapidminer sebagai tools dalam pengolahan data miningnya. Tahap implementasi algoritma merupakan tahapan dalam penerapan algoritma dalam data mining. Tahap hasil penelitian dan pengujian merupakan proses pengkajian ulang terhadap validasi hasil penelitian sehingga bisa ditarik kesimpulan dari sebuah penelitian.

Sumber data adalah subjek di mana data bisa diperoleh. Terdapat dua macam sumber data, yakni sumber data primer dan sumber data sekunder. Sumber data primer diperoleh langsung oleh peneliti. Sedangkan sumber data sekunder ialah sumber data yang tidak diperoleh langsung, biasanya sumber data ini didapat dari pihak lain yang terpercaya. Data primer ialah data yang diperoleh dan dikumpulkan langsung dari objek yang diteliti. Data primer adalah data yang didapat atau dikumpulkan langsung di lapangan oleh orang yang melakukan penelitian atau yang bersangkutan memerlukan. Data primer diperoleh dari sumber informan, yaitu individu atau perseorangan seperti hasil wawancara yang dilakukan oleh peneliti, hasil observasi lapangan, dan data mengenai informan. Data Sekunder yaitu data yang didapat dari pihak atau sumber lain yang telah ada. Peneliti tidak mendapat informasi langsung dari narasumber atau objek yang diteliti, melainkan dari data yang telah ada, seperti grafik, tabel, diagram, dan tulisan dari peneliti sebelumnya. Data sekunder digunakan untuk mendukung informasi primer yang telah diperoleh, yakni dari bahan pustaka, literatur, penelitian terdahulu, buku, dan lain sebagainya. Sumber data primer dalam penelitian ini adalah data Surat Pemberitahuan Pajak Daerah (SPTPD) yang diperoleh dari Badan Pengelola Keuangan dan Pendapatan Daerah (BPKPD) Kota Cirebon, melalui Sistem Informasi Pendapatan Daerah atau SIMPENDA Kota Cirebon. Sedangkan data sekundernya adalah data yang berupa dokumentasi seperti foto observasi di tempat penelitian yaitu di Badan Pengelola Keuangan dan Pendapatan Daerah Kota Cirebon. Populasi dalam penelitian ini adalah seluruh wajib pajak yang terdiri dari beberapa jenis pajak yaitu jenis pajak hotel, pajak restoran, pajak hiburan, dan pajak parkir di luar badan jalan yang ada di wilayah Kota Cirebon dengan jumlah 1.402 wajib pajak, dengan pembagian pajak hotel 165 wajib pajak, pajak restoran 1018 wajib pajak,

pajak hiburan 133 wajib pajak, dan pajak parkir 86 wajib pajak. Sampel yang digunakan dalam penelitian ini adalah data pembayaran pajak daerah selama satu tahun berdasarkan jenis pajak hotel, pajak restoran, pajak hiburan, dan pajak parkir di luar badan jalan pada tahun 2021 yang berjumlah 11.046 pembayaran, dengan pembagian, pajak hotel 1.247 pembayaran, pajak restoran 9.240 pembayaran, pajak hiburan 447 pembayaran, dan pajak parkir 472 pembayaran.

Algoritma yang digunakan pada penelitian yaitu algoritma regresi logistik biner. Regresi logistik biner merupakan suatu teknik analisis data yang digunakan untuk menganalisis hubungan antara satu atau dua variabel bebas dengan variabel respon yang bersifat biner atau dikotomik. Variabel bebas pada regresi logistik dapat berupa variabel berskala kategorik maupun variabel yang berskala kontinu sedangkan variabel respon berupa variabel berskala kategorik.

Analisis regresi logistik biner digunakan untuk menganalisis hubungan antara satu variabel respon dan beberapa variabel prediktor, dengan variabel responnya berupa data kualitatif dikotomi yaitu bernilai 1 untuk menyatakan keberadaan sebuah karakteristik dan bernilai 0 untuk menyatakan ketidakberadaan sebuah karakteristik [5]. Model regresi logistik biner digunakan jika variabel responnya menghasilkan dua kategori bernilai 0 dan 1, sehingga mengikuti distribusi Bernoulli sebagai berikut:

$$f(y_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

Dimana:

π_i = peluang kejadian ke-1

y_i = perubahan acak ke-1 yang terdiri dari 0 dan 1

Bentuk model regresi dengan satu variabel prediktor adalah:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$$

Untuk mempermudah menaksir parameter regresi, maka $\pi(x)$ pada persamaan diatas ditransformasikan menghasilkan bentuk logit regresi logistik sebagai berikut:

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x$$

Operasionalisasi Variabel

Berikut merupakan tabel operasionalisasi variabel yang digunakan dalam penelitian ini:

Tabel 1. Operasionalisasi Variabel

Variabel	Nama Variabel	Kode	Keterangan
Dependen	Status Kepatuhan	Y	0 = Tidak Patuh
			1 = Patuh
Independen	Jenis Pajak	X	1 = Hiburan
			2 = Hotel
			3 = Parkir
			4 = Restoran

Variabel Bebas atau Variabel Independen (X): bisa disebut dengan variabel prediksi. Merupakan hasil manipulasi. Atribut atau potensial kasus diberikan pada investigasi hasil penelitian. Adapun variabel bebas atau variabel independen dalam

penelitian ini variabel (X1) pajak hiburan, variabel (X2) pajak hotel, variabel (X3) pajak parkir, variabel (X4) pajak restoran.

Variabel Tidak Bebas atau Variabel Dependen (Y): sering disebut sebagai variabel output, criteria, konsekuen, variabel terikat atau variabel yang dipengaruhi atau yang menjadi sebab akibat. Adapun variabel tidak bebas atau variabel dependen dalam penelitian ini yaitu (Y) patuh dan tidak patuh pembayaran pajak daerah.

III.HASIL DAN PEMBAHASAAN

A.Dataset

Pada penelitian ini data yang digunakan adalah data primer. Data primer sendiri merupakan data yang diperoleh dan dikumpulkan peneliti langsung dari objek yang diteliti. Data primer ini merupakan data pembayaran pajak daerah yang didapatkan dari Sistem Pendapatan Daerah (SIMPENDA) Kota Cirebon, yang berasal dari Badan Pengelola Keuangan Daerah Kota Cirebon, data yang diambil sebanyak 11.406 data. Data tersebut terbagi sesuai dengan jenis pajaknya masing masing yaitu, pajak hiburan 447 data, pajak restoran 9.240 data, pajak hotel 1.247 data, dan pajak parkir 472 data.

ID	TGL PEMBAYARAN	MASA PAJAK	JENIS PAJAK	NO_BUKTI	NAMA	ALAMAT	POK OK	DE NDA	JUMLAH	NPWPD	Keterangan
1	08/01/2021	Januari 2021	HIBURAN	813	SKY GAME SURYA	JL. KARANGGETAS NO.23	700.000	0	700.000	P.1.000994	Patuh
2	08/01/2021	Januari 2021	HIBURAN	847	MASTER PIECE	KOMP CSB MALL	6.566.700	0	6.566.700	P.2.001337	Patuh
3	14/01/2021	Januari 2021	HIBURAN	892	CV WIJAYA GEMILANG / PING ME	JL. DR. CIPTO MIK NO.87	243.000	0	243.000	P.2.001335	Patuh
4	14/01/2021	Januari 2021	HIBURAN	876	NEW PANDA SHIATZU AND REFLEKSI	JL. A YANI NO.39 CIREBON	936.000	0	936.000	P.2.001316	Patuh
5	14/01/2021	Januari 2021	HIBURAN	877	CV PADANG NUSA PERSADA / NEW PANDA	JL. A YANI NO.39 CIREBON	1.274.000	0	1.274.000	P.2.001316	Patuh
6	14/01/2021	Januari 2021	HIBURAN	882	PT PRIMA SARANA MANUNGAL/CIREBON WATERLAND	JL. YOS SUDARSO ADE IRMAS 1	2.217.000	0	2.217.000	P.2.001329	Patuh
7	14/01/2021	Januari 2021	HIBURAN	923	KOKUO	JL. SILIWANGI NO. 80 CIREBON	1.500.000	0	1.500.000	P.2.001304	Patuh
8	14/01/2021	Januari 2021	HIBURAN	1061	CSB XXI STUDIO 1	JL. DR. CIPTO MANGUNKUSUMO No. 26	1351.178	0	1351.178	P.2.000174	Patuh
9	14/01/2021	Januari 2021	HIBURAN	1060	CSB XXI STUDIO 2	JL. DR. CIPTO MANGUNKUSUMO No. 26	6553.178	0	6553.178	P.2.000174	Patuh
10	10/02/2021	Januari 2021	HIBURAN	1059	CSB XXI STUDIO 3	JL. DR. CIPTO MANGUNKUSUMO No. 26	6.553.178	0	6.553.178	P.2.000174	Patuh

Gambar 1. Sampel Dataset

Dari data yang didapatkan penulis, data tersebut memiliki 12 atribut diantaranya yaitu id, tanggal pembayaran, masa pajak, jenis pajak, nomor bukti, nama wajib pajak, alamat, pokok pajak, denda, jumlah pajak terhitung, NPWD, dan keterangan.

B. Seleksi Data

Data selection atau seleksi data merupakan proses menganalisis data-data yang relevan dari database karena sering ditemukan bahwa tidak semua data dibutuhkan untuk di analisis. Dari semua data yang digunakan hanya id, nama, jenis pajak, masa pajak, npwpd, denda, alamat, dan tgl_pembayaran, karena informasi yang terkandung didalam data tersebut sudah mewakili informasi yang dibutuhkan untuk dijadikan indikator penelitian.

ID	Nama	Jenis_Pajak	Masa_Pajak	Keterangan
1	SKY GAME SURYA	Hiburan	Januari 2021	Patuh
2	MASTER PIECE	Hiburan	Januari 2021	Patuh
3	CV WIJAYA GEMILANG / PING ME	Hiburan	Januari 2021	Patuh
4	NEW PANDA SHIATZU AND REFLEKSI	Hiburan	Januari 2021	Patuh
5	CV PADANG NUSA PERSADA / NEW PANDA	Hiburan	Januari 2021	Patuh
6	PT PRIMA SARANA MANUNGAL/CIREBON WATERLAND	Hiburan	Januari 2021	Patuh
7	KOKUO	Hiburan	Januari 2021	Patuh
8	CSB XXI STUDIO 1	Hiburan	Januari 2021	Patuh
9	CSB XXI STUDIO 2	Hiburan	Januari 2021	Patuh
10	CSB XXI STUDIO 3	Hiburan	Januari 2021	Patuh

Gambar 2. Seleksi Data

C. Pembersihan Data

Pembersihan data merupakan proses cleaning data supaya didalam data tidak ada duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data, sehingga data tersebut dapat diolah dan bisa dilakukan proses data mining. Berdasarkan data yang didapatkan peneliti, data yang didapatkan sudah bersih tidak ada data yang null maupun data yang ganda.

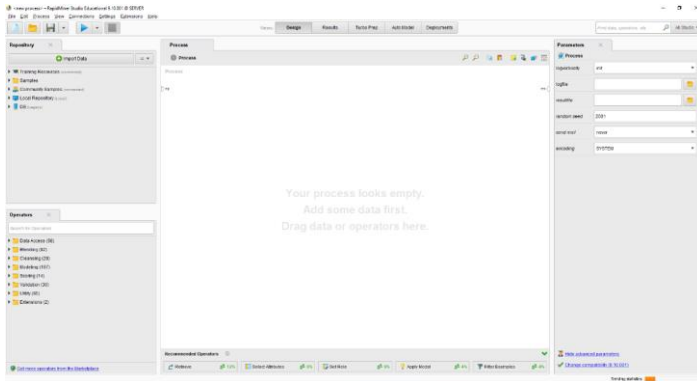
D. Transformasi Data

Transformasi data atau data transformation merupakan proses yang digunakan untuk mengubah data dalam bentuk yang sesuai dalam proses data mining. Dalam penelitian ini teknik transformasi data yang digunakan menggunakan teknik pemilihan atribut, yaitu dengan menambahkan atribut keterangan pada dataset yang dirubah rolenya menjadi atribut label.

E. Pengujian Pada Software RapidMiner

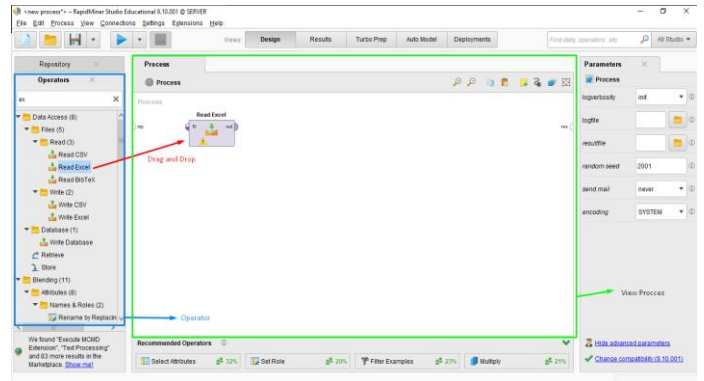
Tujuan utama dalam penelitian ini adalah untuk mengetahui klasifikasi status pembayaran pajak daerah menggunakan metode Binary Logistic Regression. Pengujian untuk mengetahui kepatuhan wajib pajak menggunakan software RapidMiner versi 9.10 dengan menggunakan data training dari data pembayaran pajak daerah.

Setelah Muncul tampilan awal pada RapidMiner maka akan muncul tampilan lembar kerja. Berikut ini tampilan lembar kerja pada RapidMiner.



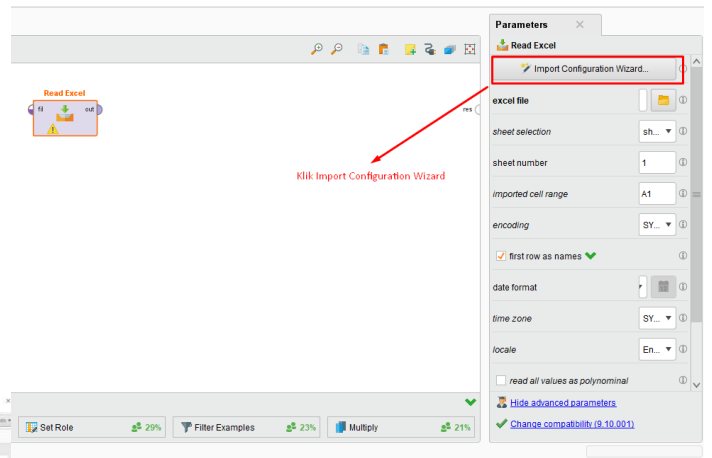
Gambar 3. Tampilan Lembar Kerja RapidMiner

Proses selanjutnya adalah melakukan importing data. Dalam importing data maka dibutuhkan operator, berhubung data yang digunakan peneliti menggunakan format Excel atau (xlsx), maka operator import data yaitu operator read excel dengan cara drag and drop kedalam view process sampai operator read excel tampil dalam view process seperti terlihat pada gambar.



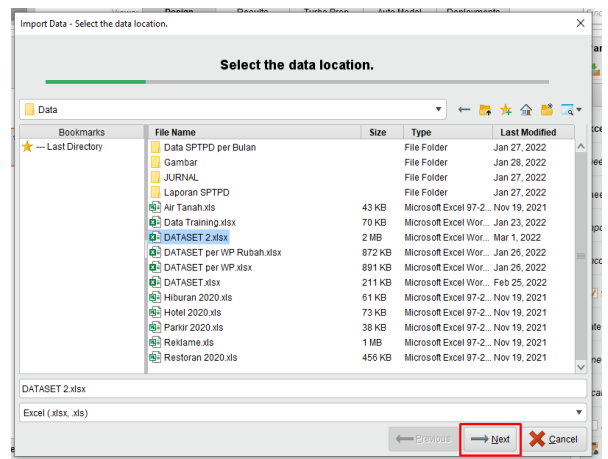
Gambar 4. Tampilan Operator Read Excel

Selanjutnya klik pada operator read excel lalu pada tab parameter klik import configuration wizard untuk mengambil file dataset, seperti pada Gambar 4.



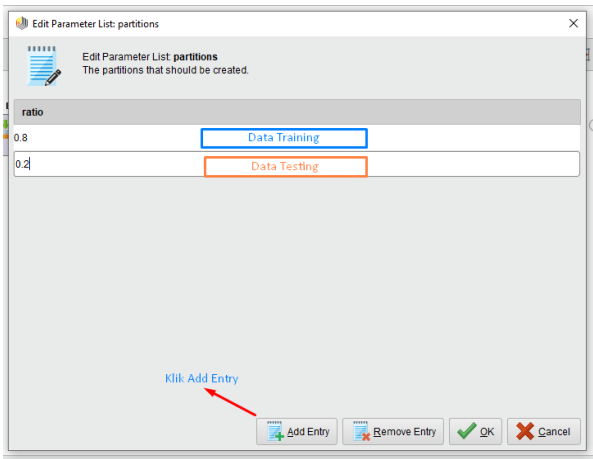
Gambar 5. Import Data

Setelah mengklik import configuration wizard maka akan diarahkan pada form data import wizard tahap 1 untuk memilih lokasi file yang digunakan dalam penelitian ini.



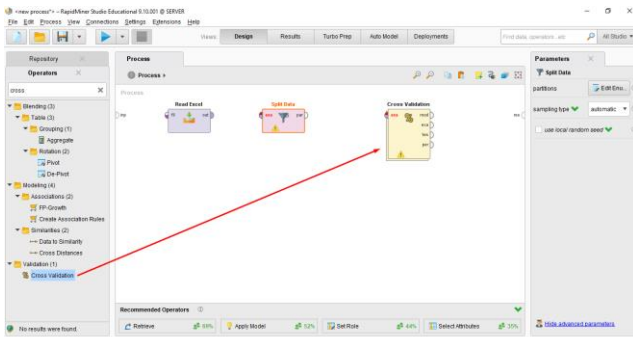
Gambar 6. Import Data

Setelah data berhasil diimport, maka proses selanjutnya yaitu memilih operator split data dengan cara drag and drop operator pada view process. Operator split data berfungsi untuk membagi data menjadi data training dan data testing.



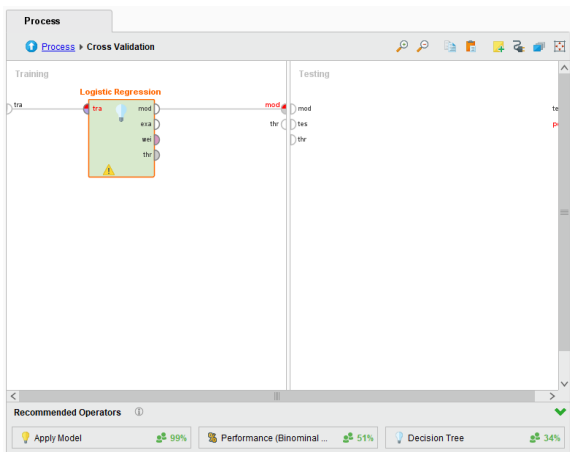
Gambar 7. Proses Split Data

Setelah proses pembagian data dengan operator split data, maka proses selanjutnya memilih operator cross validation. Cross validation berfungsi untuk memvalidasi data dan untuk memaksimalkan nilai akurasi pengolahan data.



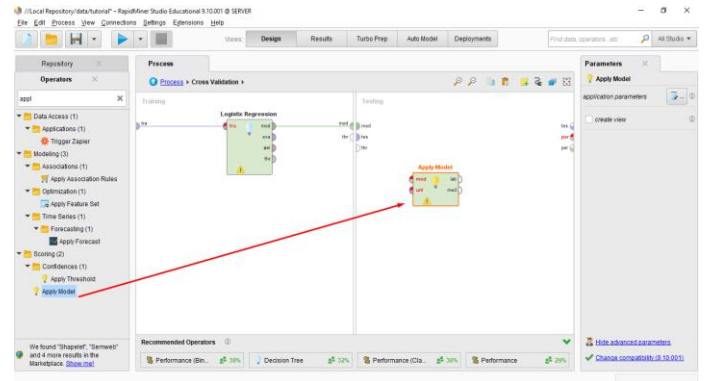
Gambar 8. Operator Cross Validation

Proses selanjutnya memasukkan operator modeling Logistic Regression pada subprocess training dan hubungkan operator ke port training dan port model seperti pada Gambar 9.



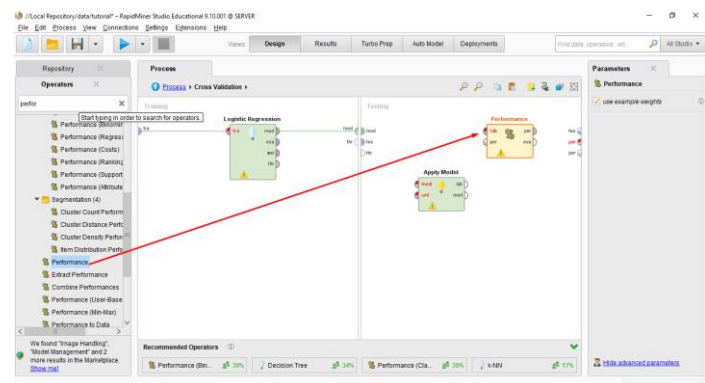
Gambar 9. Operator Logistik Regresi

Selanjutnya masukkan operator Apply Model pada subprocess testing. Operator apply model tersebut digunakan untuk menerapkan model yang telah dilatih sebelumnya menggunakan data training pada unlabeled data pada data testing.



Gambar 10. Apply Model

Setelah memasukkan operator apply model, selanjutnya masukkan operator Performance pada subprocess testing. Operator Performance digunakan untuk mengevaluasi kinerja statistik klasifikasi.



Gambar 11. PerformanceVector

3.2 Pembahasan

Berdasarkan hasil dari pengujian menggunakan rapidminer nilai klasifikasi yang dihasilkan performance vector pada nilai accuracy, precision, dan recall, seperti yang terlihat pada Gambar 12 adalah sebagai berikut:

PerformanceVector (Performance)			
accuracy: 93.97% +/- 0.71% (micro average: 93.97%)			
	true Patuh	true Tidak Patuh	class precision
pred. Patuh	7581	422	94.73%
pred. Tidak Patuh	128	994	88.59%
class recall	98.34%	70.20%	

Gambar 12. Hasil PerformanceVector

1. Accuracy (Ketepatan)

Untuk tingkat *accuracy* atau ketepatan sebesar 93,97 %. *Accuracy* sendiri merupakan jumlah prediksi benar baik itu positif maupun negatif dengan perhitungan seperti dibawah ini:

$$Accuracy = \frac{7581 + 994}{7581 + 422 + 128 + 994} = \frac{8575}{9119} = 0,9397260273 = 93,97\%$$

2. Precision

Precision merupakan jumlah prediksi benar positif dibagi dengan keseluruhan hasil prediksi positif, sehingga didapat berhitung sebagai berikut:

a. Untuk nilai *precision* Patuh

$$precision = \frac{TP}{TP + FP} = \frac{7581}{7581 + 422} = \frac{7581}{8003} = 0,9472697738 = 94,73\%$$

b. Untuk nilai *precision* Tidak Patuh

$$precision = \frac{TN}{TN + FN} = \frac{994}{994 + 179} = \frac{994}{1173} = 0,8859180035 = 88,59\%$$

Sehingga didapatkan hasil bahwa untuk nilai *precision* Patuh sebesar 94,73%, dan untuk nilai *precision* Tidak Patuh sebesar 88,59%.

Recall adalah jumlah prediksi benar positif dibagi dengan keseluruhan data yang benar positif, sehingga didapat perhitungan sebagai berikut:

a. Untuk nilai *recall* Patuh

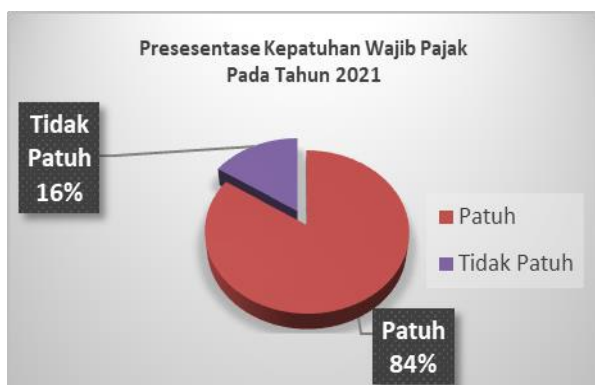
$$recall = \frac{TP}{TP + FN} = \frac{7581}{7581 + 179} = \frac{7581}{7760} = 0,9833960306 = 98,34\%$$

b. Untuk nilai *recall* Tidak Patuh

$$recall = \frac{TN}{TN + FP} = \frac{994}{994 + 177} = \frac{994}{1171} = 0,7019774011 = 70,20\%$$

Sehingga didapatkan hasil bahwa untuk nilai *recall* Patuh sebesar 98,34%, dan untuk nilai *precision* Tidak Patuh sebesar 70,20%.

Berdasarkan analisis yang dilakukan peneliti berikut ini merupakan presentase kepatuhan wajib pajak pada tahun 2021.



Gambar 13. Presentase Kepatuhan Wajib Pajak Tahun 2021

Berdasarkan pada Gambar 4.20 diatas dapat diketahui bahwa presentase kepatuhan wajib pajak dalam membayar pajaknya pada tahun 2021 yaitu wajib pajak patuh sebesar 84% sedangkan wajib pajak yang tidak patuh sebesar 16%. Dari presentase tersebut dapat diketahui bahwa tingkat kepatuhan wajib pajak cukup tinggi karena hanya 16% yang tidak patuh dalam membayar pajaknya.

IV. KESIMPULAN

Berdasarkan hasil dari penelitian yang telah dibahas dan dijelaskan, maka dapat ditarik kesimpulan untuk menjawab pertanyaan penelitian yang ada bahwa, Penelitian ini menggunakan 11.406 data pembayaran pajak daerah di Kota Cirebon yang terbagi menjadi 4 wajib pajak yaitu pajak hiburan, hotel, parkir, dan restoran. Untuk pengolahan menggunakan data mining maka 11.406 data dibagi menjadi 2 bagian yaitu data training sebanyak 9.125 data, dan data testing sebanyak 2.281, dengan presentase 80% data training dan 20% data testing. Pada pengujian menggunakan software rapidminer diperoleh hasil bahwa tingkat akurasi dari performance vector sebesar 93,97% dengan class

precision Patuh sebesar 94,73%, dan Tidak Patuh sebesar 88,59%, dan untuk class recall Patuh sebesar 98,34, dan Tidak Patuh sebesar 70,20%. Dari pemodelan Logistic Regression dan berdasar diagram pie didapatkan hasil tingkat kepatuhan terendah pada jenis pajak restoran dengan presentase 68% seperti yang terlihat pada diagram dibawah ini. Dapat diketahui bahwa presentase tingkat kepatuhan wajib pajak untuk patuh sebesar 84% sedangkan presentase kepatuhan wajib pajak untuk tidak patuh sebesar 16%. Dari presentase tersebut dapat diketahui bahwa tingkat kepatuhan wajib pajak cukup tinggi, namun meskipun tingkat kepatuhan wajib pajak cukup tinggi masih cukup banyak wajib pajak yang tidak patuh dalam membayar pajaknya. Berdasarkan hasil pembahasan dapat diketahui bahwa tingkat tidak patuh untuk jenis pajak restoran sebesar 68%, pajak hotel 19%, pajak parkir 7%, dan pajak hiburan 6%. Jenis pajak restoran memiliki tingkat ketidakpatuhan paling tinggi dengan jumlah 210 wajib pajak tidak patuh. Penerapan data mining sangat penting dalam pengelompokan data dalam hal ini data Surat Pemberitahuan Pajak Daerah, sehingga bisa mempermudah dalam pengklasifikasian wajib pajak yang patuh dan tidak patuh.

V. REFERENCES

- [1] m. Abdurohman, r. Husna, i. Ali, g. Dwilestari, and n. Rahaningsih, "penerapan model klasifikasi dalam tingkat kepuasan layanan publik kelurahan karyamulya dengan menggunakan algoritma decision tree," *inf. Manag. Educ. Prof. J. Inf. Manag.*, vol. 6, no. 1, p. 81, 2022, doi: 10.51211/imbi.v6i1.1678.
- [2] p. Studi, t. Informatika, p. Studi, s. Informasi, p. Studi, and r. Perangkat, "pengelompokan hasil belajar siswa dengan metode clustering k-means saeful anwar 1) , tati suprapti 2) , gifthera dwilestari 3) irfan ali 4)," vol. 4, no. 2, pp. 60–72, 2022.
- [3] f. M. Basysyar, g. Dwilestari, a. Bahtiar, martanto, and d. N. Nuris, "market basketball analysis algorithm for determining products association," *iop conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012040, 2021, doi: 10.1088/1757-899x/1088/1/012040.
- [4] f. M. Basysyar, "clustering data disabilitas menggunakan algoritma k-means di kabupaten cirebon," *jursima (jurnal sist. Inf. Dan ...)*, vol. 9, no. 3, 2021.
- [5] s. Suhari, a. Faqih, and f. M. Basysyar, "sistem informasi kepegawaian mengunakan metode agile development di cv. Angkasa raya," *j. Teknol. Dan inf.*, vol. 12, no. 1, pp. 30–45, 2022, doi: 10.34010/jati.v12i1.6622.
- [6] s. M. A. K-means, . "kata kunci : data mining, kualitas, ujian nasional, algoritma k-means.," vol. 10, no. 1, 2022.
- [7] c. L. Rohmat, i. Ali, t. Suprapti, and u. Aryanti, "aplikasi pemesanan online barbershop berbasis android untuk meningkatkan layanan," vol. 4, no. 2, pp. 37–45, 2021.
- [8] y. A. Wijaya, n. Suarna, iin, r. Hamonangan, and r. Nining, "comparison of machine learning algorithm

- for santander dataset,” *iop conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012032, 2021, doi: 10.1088/1757-899x/1088/1/012032.
- [9] n. Suarna, y. A. Wijaya, mulyawan, t. Hartati, and t. Suprapti, “comparison k-medoids algorithm and k-means algorithm for clustering fish cooking menu from fish dataset,” *iop conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012034, 2021, doi: 10.1088/1757-899x/1088/1/012034.
- [10] s. Turangga and y. A. W, “analisis internet menggunakan parameter quality of service pada alfamart tuparev 70,” vol. 6, no. 1, pp. 392–398, 2022.
- [11] t. Hartati and y. A. Wijaya, “analisis data lalu lintas jaringan di kantor cangehgar cyber operation center menggunakan algoritma k-means network traffic data analysis at cangehgar cyber operation center office using k-means algorithm,” vol. 7, no. 1, pp. 75–84, 2022.
- [12] h. Putri, a. I. Purnamasari, a. R. Dikananda, o. Nurdiawan, and s. Anwar, “penerima manfaat bantuan non tunai kartu keluarga sejahtera menggunakan metode naïve bayes dan knn,” *build. Informatics, technol. Sci.*, vol. 3, no. 3, pp. 331–337, 2021, doi: 10.47065/bits.v3i3.1093.
- [13] a. Z. Zami, o. Nurdiawan, and g. Dwilestari, “klasifikasi kondisi gizi bayi bawah lima tahun pada posyandu melati dengan menggunakan algoritma decision tree,” *j. Sist. Komput. Dan inform.*, vol. 3, pp. 305–310, 2022, doi: 10.30865/json.v3i3.3892.
- [14] e. W. Ramadhona, t. Prasetya, and a. I. Purnamasari, “game edukasi ‘ nihongo kurabu ’ belajar bahasa menggunakan unity 2d berbasis android,” *inf. Manag. Educ. Prof.*, vol. 6, no. 1, pp. 71–80, 2022.
- [15] a. Z. Zami, o. Nurdiawan, and g. Dwilestari, “klasifikasi kondisi gizi bayi bawah lima tahun pada posyandu melati dengan menggunakan algoritma decision tree,” *j. Sist. Komput. Dan inform.*, vol. 3, pp. 305–310, 2022, doi: 10.30865/json.v3i3.3892.
- [16] h. Putri, a. I. Purnamasari, a. R. Dikananda, o. Nurdiawan, and s. Anwar, “penerima manfaat bantuan non tunai kartu keluarga sejahtera menggunakan metode naïve bayes dan knn,” *build. Informatics, technol. Sci.*, vol. 3, no. 3, pp. 331–337, 2021, doi: 10.47065/bits.v3i3.1093.
- [17] d. A. K. Irfan nurdiyanto, odi nurdiawan, nining rahaningsih, ade irfma purnamasari, “penentuan keputusan pemberian pinjaman kredit menggunakan algoritma c.45,” *j. Data sci. Dan inform.*, vol. 1, no. 1, pp. 16–20, 2021.
- [18] a. Faqih, o. Nurdiawan, and a. Setiawan, “ethnomathematics : utilization of crock , ladle , and chopping board for learning material of geometry at the elementary school,” vol. 4, no. 1, pp. 46–55, 2021.
- [19] o. Nurdiawan, f. A. Pratama, d. A. Kurnia, kaslani, and n. Rahaningsih, “optimization of traveling salesman problem on scheduling tour packages using genetic algorithms,” *j. Phys. Conf. Ser.*, vol. 1477, no. 5, 2020, doi: 10.1088/1742-6596/1477/5/052037.
- [20] f. Arie pratama, k. Kaslani, o. Nurdiawan, n. Rahaningsih, and n. Nurhadiansyah, “learning innovation using the zahir application in improving understanding of accounting materials,” *j. Phys. Conf. Ser.*, vol. 1477, no. 3, pp. 0–6, 2020, doi: 10.1088/1742-6596/1477/3/032018.