

Penerapan Data Mining Menggunakan Teknik Classification Untuk Melihat Potensi Kepatuhan Wajib Pajak Badan

Anuqman Fitriadi *, Qamarullah Popalia, Arief Wibowo

Fakultas Teknologi Informasi, Ilmu Komputer, Universitas Budi Luhur, Jakarta, Indonesia
Email: ¹*2311601161@student.budiluhur.ac.id, 2311601104@student.budiluhur.ac.id, arief.wibowo@budiluhur.ac.id
Email Penulis Korespondensi: 2311601161@student.budiluhur.ac.id
Submitted 18-11-2025; Accepted 19-02-2026; Published 28-02-2026

Abstrak

Penerapan data mining dengan teknik classification memiliki potensi besar dalam membantu otoritas pajak mengidentifikasi dan memetakan tingkat kepatuhan wajib pajak badan. Penelitian ini bertujuan untuk membangun model klasifikasi kepatuhan wajib pajak badan menggunakan algoritma Naive Bayes berdasarkan rasio pelaporan Surat Pemberitahuan (SPT) dan rasio pembayaran pajak. Data yang digunakan merupakan data agregat Kantor Pelayanan Pajak (KPP) periode 2022–2024 yang diperoleh dari Direktorat Jenderal Pajak. Tahapan penelitian mengikuti metodologi Knowledge Discovery in Databases (KDD), meliputi seleksi data, preprocessing, transformasi, pemodelan, dan evaluasi. Hasil pengujian menunjukkan bahwa model Naive Bayes mampu mengklasifikasikan tingkat kepatuhan dengan akurasi 100%, precision 1,00, recall 1,00, dan F1-score 1,00. Temuan ini mengindikasikan bahwa rasio pelaporan SPT merupakan faktor dominan dalam menentukan kepatuhan wajib pajak badan. Model yang diusulkan dapat dimanfaatkan sebagai sistem pendukung keputusan untuk membantu otoritas pajak dalam menentukan prioritas pengawasan dan pembinaan wajib pajak badan.

Kata Kunci: Data Mining; Classification; Naive Bayes; Kepatuhan Wajib Pajak; Pajak Badan

Abstract

The application of data mining using classification techniques has significant potential to assist tax authorities in identifying and mapping the compliance levels of corporate taxpayers. This study aims to develop a corporate taxpayer compliance classification model using the Naive Bayes algorithm based on the ratio of Annual Tax Return (SPT) filing and the ratio of tax payments. The data used consist of aggregated data from Tax Service Offices (Kantor Pelayanan Pajak/KPP) for the 2022–2024 period obtained from the Directorate General of Taxes. The research stages follow the Knowledge Discovery in Databases (KDD) methodology, which includes data selection, preprocessing, transformation, modeling, and evaluation. The experimental results indicate that the Naive Bayes model is able to classify compliance levels with an accuracy of 100%, precision of 1.00, recall of 1.00, and an F1-score of 1.00. These findings suggest that the SPT filing ratio is the dominant factor in determining corporate taxpayer compliance. The proposed model can be utilized as a decision support system to assist tax authorities in determining supervision and guidance priorities for corporate taxpayers.

Keywords: Data Mining; Classification; Naive Bayes; Taxpayer Compliance; Corporate Tax

1. PENDAHULUAN

Peran pajak sebagai sumber utama penerimaan negara menjadikan kepatuhan wajib pajak sebagai aspek krusial dalam mendukung stabilitas dan pertumbuhan ekonomi nasional [1]. Wajib pajak badan, sebagai entitas yang memberikan kontribusi signifikan terhadap penerimaan pajak, menjadi salah satu fokus utama pengawasan oleh otoritas pajak. Namun demikian, upaya pemantauan dan peningkatan kepatuhan wajib pajak badan menghadapi tantangan yang semakin kompleks seiring dengan meningkatnya volume data perpajakan serta beragamnya pola perilaku wajib pajak.

Kepatuhan wajib pajak merupakan faktor penting dalam optimalisasi penerimaan pajak suatu negara [2]. Di era digital, ketersediaan data perpajakan dalam jumlah besar membuka peluang untuk melakukan analisis dan prediksi perilaku kepatuhan wajib pajak secara lebih sistematis dengan memanfaatkan teknik data mining [3]. Data mining, sebagai bagian dari Knowledge Discovery in Databases (KDD), menyediakan pendekatan terstruktur untuk mengekstraksi pola dan pengetahuan dari data berukuran besar dan kompleks [4]. Penerapan data mining dalam konteks perpajakan menjadi semakin relevan karena memungkinkan otoritas pajak melakukan analisis prediktif guna mengalokasikan sumber daya pengawasan secara lebih efisien, terutama dengan memfokuskan perhatian pada wajib pajak yang memiliki risiko ketidakpatuhan lebih tinggi [5], [6].

Salah satu teknik data mining yang banyak digunakan dalam analisis kepatuhan wajib pajak adalah classification (klasifikasi). Teknik ini memungkinkan pembangunan model prediktif berbasis data historis untuk mengidentifikasi karakteristik dan pola perilaku wajib pajak, sehingga tingkat kepatuhan dapat diklasifikasikan berdasarkan parameter tertentu yang relevan [7]. Dengan penerapan teknik klasifikasi, otoritas pajak dapat memperoleh insight yang mendukung peningkatan efektivitas pengawasan dan pembinaan wajib pajak.

Sejumlah penelitian terdahulu telah mengkaji penerapan data mining dalam konteks kepatuhan wajib pajak. Beberapa penelitian menggunakan algoritma Naive Bayes untuk mengklasifikasikan potensi kepatuhan wajib pajak kendaraan bermotor maupun Pajak Bumi dan Bangunan (PBB), dengan hasil yang menunjukkan tingkat akurasi yang tinggi dalam mengidentifikasi pola kepatuhan [8], [9]. Selain pendekatan klasifikasi, penelitian lain juga memanfaatkan metode clustering, seperti algoritma K-Means, untuk mengelompokkan wajib pajak berdasarkan tingkat kepatuhannya, sehingga membantu otoritas pajak dalam menyusun strategi pengawasan yang lebih efektif [10].

Meskipun demikian, sebagian besar penelitian sebelumnya masih berfokus pada jenis pajak tertentu, seperti pajak kendaraan bermotor atau PBB, serta menggunakan data pada level individu wajib pajak atau wilayah yang relatif terbatas.

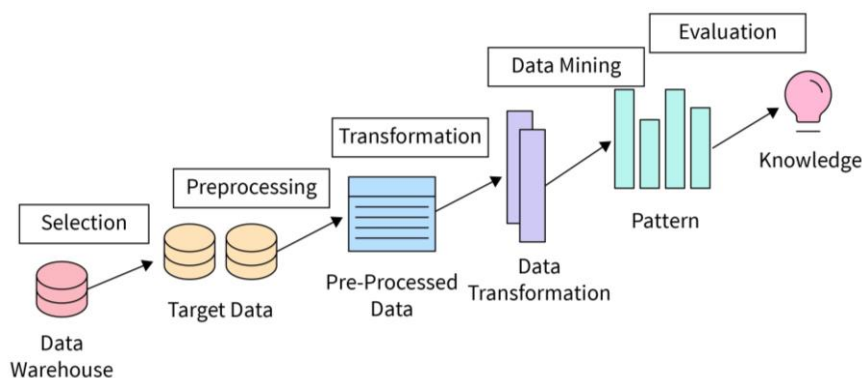
Penelitian yang memanfaatkan data agregat Kantor Pelayanan Pajak (KPP) sebagai unit analisis untuk menilai kepatuhan wajib pajak badan secara lebih luas masih relatif terbatas. Selain itu, penggunaan indikator rasio pelaporan Surat Pemberitahuan (SPT) dan rasio pembayaran pajak secara simultan sebagai dasar klasifikasi kepatuhan pada level KPP belum banyak dikaji.

Berdasarkan kondisi tersebut, terdapat celah penelitian (research gap) berupa belum tersedianya model klasifikasi kepatuhan wajib pajak badan yang memanfaatkan data agregat KPP dengan pendekatan rasio pelaporan dan rasio pembayaran pajak. Penelitian ini bertujuan untuk mengisi celah tersebut dengan menerapkan teknik classification menggunakan algoritma Naive Bayes untuk menganalisis potensi kepatuhan wajib pajak badan pada level KPP.

Kontribusi penelitian ini adalah (1) menyajikan model klasifikasi kepatuhan wajib pajak badan berbasis data agregat KPP, dan (2) mengidentifikasi faktor dominan yang memengaruhi kepatuhan wajib pajak badan, yang diharapkan dapat dimanfaatkan sebagai dasar pengambilan keputusan strategis oleh otoritas pajak dalam menentukan prioritas pengawasan dan pembinaan wajib pajak.

2. METODOLOGI PENELITIAN

Penelitian ini menerapkan metodologi Knowledge Discovery in Databases (KDD) sebagai kerangka kerja utama dalam proses ekstraksi pengetahuan dari data [11], [12]. Metodologi KDD banyak digunakan dalam studi data mining karena menyediakan alur kerja yang terstruktur dan sistematis, mulai dari tahap pemilihan data hingga evaluasi hasil analisis [13], [14]. Melalui penerapan metode KDD, penelitian ini diharapkan mampu menghasilkan pengetahuan yang bersumber dari database yang tersedia, menjawab permasalahan penelitian, serta mendukung pencapaian tujuan penelitian secara lebih optimal. Rangkaian tahapan dalam metodologi KDD disajikan pada Gambar 2.



Gambar 1. Tahapan KDD

Dari Gambar 1, dijelaskan bahwa langkah-langkah dalam tahapan KDD yaitu sebagai berikut:

2.1 Data Selection (Pemilihan Data)

Tahap data selection merupakan proses pemilihan data yang relevan dari sumber data utama untuk membentuk dataset penelitian [15]. Data yang digunakan dalam penelitian ini diperoleh dari basis data Direktorat Jenderal Pajak dan berupa data agregat pada tingkat Kantor Pelayanan Pajak (KPP) periode 2022–2024. Dataset mencakup informasi jumlah Wajib Pajak (WP) Badan terdaftar, jumlah pelaporan Surat Pemberitahuan (SPT) Tahunan WP Badan, serta jumlah pembayaran pajak WP Badan pada masing-masing KPP. Data yang telah dipilih kemudian dipisahkan dari basis data utama dan disusun menjadi dataset yang siap untuk dianalisis lebih lanjut.

2.2 Preprocessing

Tahap preprocessing dilakukan untuk meningkatkan kualitas data sebelum proses pemodelan [16]. Pada tahap ini, dilakukan pemeriksaan terhadap nilai kosong, data duplikat, serta ketidakkonsistenan format data. Nilai yang tidak tersedia (N/A) pada beberapa atribut ditangani dengan melakukan substitusi nilai nol agar dataset tetap dapat diproses. Proses preprocessing ini bertujuan untuk memastikan bahwa data yang digunakan bersih, konsisten, dan layak digunakan dalam proses analisis dan pemodelan.

2.3 Transformasi

Pada tahap transformasi, data yang telah melalui preprocessing diubah ke dalam bentuk yang sesuai dengan kebutuhan algoritma klasifikasi [11]. Transformasi dilakukan dengan menghitung rasio pelaporan SPT Tahunan, yaitu perbandingan antara jumlah pelaporan SPT dengan jumlah WP Badan wajib SPT, serta rasio pembayaran pajak, yaitu perbandingan antara realisasi pembayaran pajak dengan target penerimaan pada masing-masing KPP. Selanjutnya, status kepatuhan pajak diklasifikasikan dan dikonversi ke dalam bentuk numerik, dengan nilai 1 untuk kategori patuh ($\geq 80\%$) dan nilai 0 untuk kategori tidak patuh ($< 80\%$). Transformasi ini bertujuan untuk mempermudah proses pemodelan dan meningkatkan kinerja algoritma klasifikasi.

2.4 Data Mining

Tahap modelling merupakan tahap inti dalam metodologi KDD, yaitu proses pembentukan model untuk menggali pola dan pengetahuan dari data [17]. Pada penelitian ini, algoritma Gaussian Naive Bayes digunakan untuk membangun model klasifikasi kepatuhan wajib pajak badan. Dataset dibagi menjadi dua bagian, yaitu data latih sebesar 80% dan data uji sebesar 20%. Fitur yang digunakan dalam proses pemodelan adalah rasio pelaporan SPT dan rasio pembayaran pajak, sedangkan variabel target adalah status kepatuhan wajib pajak badan. Model yang dihasilkan digunakan untuk mengklasifikasikan KPP ke dalam kategori patuh dan tidak patuh.

2.5 Evaluasi

Tahap interpretasi dan evaluasi bertujuan untuk menilai kinerja model serta memahami makna dari hasil klasifikasi yang diperoleh. Evaluasi model dilakukan menggunakan confusion matrix dan metrik evaluasi, yaitu akurasi, precision, recall, dan F1-score. Hasil evaluasi digunakan untuk mengukur tingkat ketepatan model dalam mengklasifikasikan kepatuhan wajib pajak badan, sekaligus menjadi dasar dalam menarik kesimpulan terkait efektivitas penerapan algoritma Naive Bayes dalam penelitian ini.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penerapan algoritma Naive Bayes dalam mengklasifikasikan tingkat kepatuhan wajib pajak badan pada level Kantor Pelayanan Pajak (KPP). Pembahasan dimulai dari deskripsi data dan pola awal yang terbentuk, dilanjutkan dengan hasil pemodelan dan evaluasi kinerja model, serta diakhiri dengan interpretasi temuan penelitian.

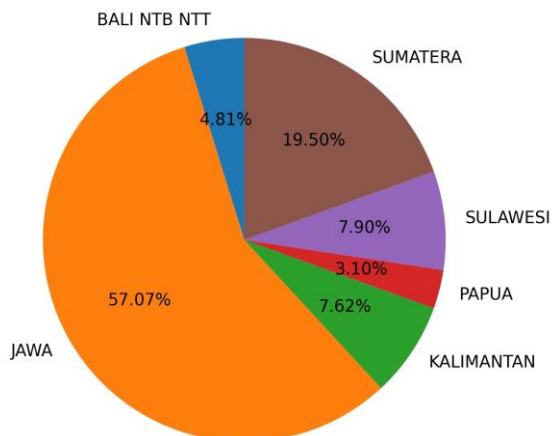
3.1 Deskripsi Data dan Pola Awal

Data yang digunakan dalam penelitian ini merupakan data agregat KPP yang bersumber dari Direktorat Jenderal Pajak, mencakup 346 baris data dengan 14 atribut. Setiap baris data merepresentasikan satu KPP dan menggambarkan kondisi kepatuhan wajib pajak badan periode 2022–2024. Meskipun dataset memuat informasi jumlah WP Badan terdaftar tahun 2021, analisis pemodelan dan klasifikasi difokuskan pada periode 2022–2024 karena ketersediaan data pelaporan dan pembayaran pajak yang lengkap pada periode tersebut, sebagaimana disajikan pada Tabel 1.

Tabel 1. Jumlah Wajib Pajak Terdaftar, Melakukan Pembayaran dan Pelaporan selama Tahun 2021 s.d 2024

NO	KODE & NAMA KPP	DAFTAR20 21	DAFTAR20 22	DAFTAR20 23	...	LAPOR20 24
1	001. KPP Pratama Jakarta Matraman	11,230	11,817	12,330	...	2,570
2	002. KPP Pratama Jakarta Jatinegara	14,935	16,097	17,370	...	3,949
3	003. KPP Pratama Jakarta Pulogadung	18,144	19,066	19,917	...	3,696
4	004. KPP Pratama Jakarta Cakung	5,697	6,685	7,958	...	3,542
5	005. KPP Pratama Jakarta Kramat Jati	16,681	17,681	18,613	...	3,458
6	007. KPP Madya Jakarta Timur	2,370	2,371	2,385	...	1,396
7	008. KPP Pratama Jakarta Duren Sawit	16,629	17,620	18,736	...	3,652
8	009. KPP Pratama Jakarta Pasar Rebo	17,585	18,947	20,508	...	3,964
9	011. KPP Pratama Jakarta Setiabudi Satu	19,912	21,682	23,614	...	8,633
10	012. KPP Pratama Jakarta Kebayoran Baru Satu	16,877	19,040	21,771	...	10,199
...	...	...	...	...	...	...
337	926. KPP Pratama Waingapu	5,330	5,622	5,934	...	1,687
338	941. KPP Pratama Ambon	24,246	25,862	27,593	...	3,194
339	942. KPP Pratama Ternate	15,158	16,598	17,991	...	2,518
340	943. KPP Pratama Tobelo	5,590	6,188	6,759	...	697
341	951. KPP Pratama Sorong	16,982	18,073	19,101	...	3,283
342	952. KPP Pratama Jayapura	28,729	30,455	32,257	...	4,585
343	953. KPP Pratama Timika	8,855	9,573	10,224	...	1,381
344	954. KPP Pratama Biak	9,206	9,817	10,774	...	1,464
345	955. KPP Pratama Manokwari	15,372	16,565	17,649	...	3,490
346	956. KPP Pratama Merauke	8,202	8,749	9,280	...	1,487

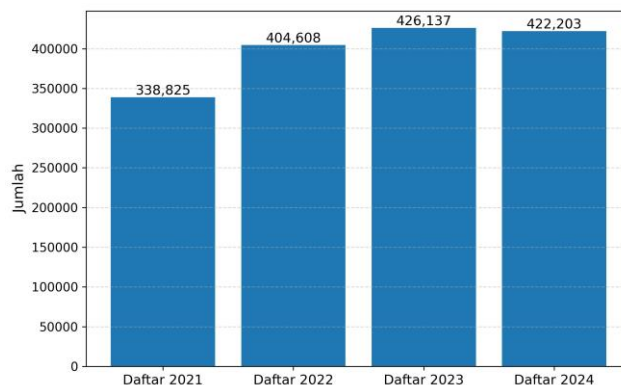
Untuk memperoleh gambaran awal distribusi data, dilakukan visualisasi terhadap jumlah wajib pajak badan berdasarkan wilayah. Gambar 2 disajikan untuk menunjukkan komposisi jumlah WP Badan terdaftar pada masing-masing pulau-pulau besar di Indonesia.



Gambar 2. Komposisi WP Badan Terdaftar

Berdasarkan Gambar 2, terlihat bahwa mayoritas wajib pajak badan terdaftar berada di Pulau Jawa dengan persentase sebesar 57,07%, diikuti oleh Pulau Sumatera (19,50%), Sulawesi (7,90%), Kalimantan (7,62%), Bali dan Nusa Tenggara (4,81%), serta Papua (3,10%). Distribusi ini menunjukkan adanya konsentrasi aktivitas ekonomi dan administrasi perpajakan di wilayah Pulau Jawa.

Mengingat dominasi Pulau Jawa dalam jumlah wajib pajak badan terdaftar, wilayah ini digunakan sebagai representasi analisis wilayah untuk memperoleh gambaran yang lebih mendalam mengenai dinamika pertumbuhan wajib pajak badan. Analisis ini bersifat deskriptif dan tidak menggantikan analisis utama yang mencakup seluruh KPP. Oleh karena itu, Gambar 3 disajikan untuk menggambarkan perkembangan jumlah WP Badan terdaftar di Pulau Jawa pada periode 2021–2024.



Gambar 3. Grafik Pebandingan Jumlah WP Badan di Pulau Jawa

Berdasarkan Gambar 3, terlihat bahwa jumlah wajib pajak badan terdaftar di Pulau Jawa mengalami peningkatan dari tahun ke tahun. Pola ini menunjukkan adanya korelasi positif antara waktu dan pertumbuhan jumlah WP Badan, yang berpotensi pada meningkatnya kompleksitas serta beban pengawasan dan pembinaan oleh Direktorat Jenderal Pajak.

3.2 Pemodelan dan Hasil Klasifikasi Naive Bayes

Berdasarkan tahapan preprocessing dan transformasi data yang telah dijelaskan pada Bab 2, proses pemodelan pada penelitian ini dilakukan menggunakan algoritma Gaussian Naive Bayes.. Pada tahap ini, disajikan contoh dataset berupa jumlah pelaporan SPT dan jumlah pembayaran pajak beserta target per KPP dan per tahun. Data ini sebagai gambaran awal sebelum dilakukan perhitungan rasio dan proses klasifikasi. Contoh dataset jumlah pelaporan SPT Tahunan wajib pajak badan pada level KPP disajikan pada Tabel 2, sedangkan contoh dataset jumlah pembayaran pajak wajib pajak badan ditampilkan pada Tabel 3. Kedua dataset tersebut merepresentasikan data numerik hasil preprocessing yang menjadi dasar perhitungan rasio kepatuhan administrasi dan fiskal.

Tabel 2. Target dan Pelaporan SPT

KANTOR PELAYANAN PAJAK	TARGET_SPT_2022	SPT_2022	% LAPOR_2022	...	% LAPOR_2024
001. KPP Pratama Jakarta Matraman	42,340	36,139	85.35%	...	97.84%
007. KPP Madya Jakarta Timur	1,137	1,138	100.09%	...	100.34%
026. KPP Pratama Jakarta Sawah Besar Satu	16,076	12,580	78.25%	...	104.68%

027. KPP Pratama Jakarta Kemayoran	45,993	32,973	71.69%	...	123.13%
051. KPP Wajib Pajak Besar Tiga	216	222	102.78%	...	127.79%
101. KPP Pratama Banda Aceh	58,193	49,678	85.37%	...	207.02%
112. KPP Pratama Medan Belawan	38,535	33,980	88.18%	...	188.76%
...	...	...	...	...	...
903. KPP Pratama Denpasar Timur	62,758	53,494	85.24%	...	202.63%
907. KPP Pratama Gianyar	102,408	88,857	86.77%	...	180.14%
914. KPP Pratama Mataram Timur	44,735	37,811	84.52%	...	238.97%
941. KPP Pratama Ambon	110,859	92,954	83.85%	...	187.21%
942. KPP Pratama Ternate	53,739	54,400	101.23%	...	279.09%
952. KPP Pratama Jayapura	93,345	80,105	85.82%	...	173.00%
956. KPP Pratama Merauke	18,761	16,957	90.38%	...	286.06%

Tabel 3. Target dan Pembayaran Pajak

KANTOR PELAYANAN PAJAK	TARGET_BAYAR_2 022	BAYAR_2022	%BAYAR20 22	...	%BAYAR20 24
001. KPP Pratama Jakarta Matraman	477,310,544,000	662,581,718,194	138.82%	...	47.37%
007. KPP Madya Jakarta Timur	8,574,702,940,000	9,991,485,783,507	116.52%	...	99.35%
026. KPP Pratama Jakarta Sawah Besar Satu	1,217,118,732,000	1,425,478,083,347	117.12%	...	86.80%
027. KPP Pratama Jakarta Kemayoran	1,163,490,732,000	1,678,969,798,764	144.30%	...	79.06%
		163,477,574,630,4			
051. KPP Wajib Pajak Besar Tiga	141,920,278,425,000	52	115.19%	...	95.21%
101. KPP Pratama Banda Aceh	1,322,740,315,000	1,807,051,814,562	136.61%	...	38.90%
112. KPP Pratama Medan Belawan	770,000,000,000	964,848,292,425	125.30%	...	93.53%
...	...	...	...	...	...
903. KPP Pratama Denpasar Timur	851,806,383,000	1,236,122,081,665	145.12%	...	64.44%
907. KPP Pratama Gianyar	622,857,528,000	766,750,601,804	123.10%	...	57.57%
914. KPP Pratama Mataram Timur	287,441,725,000	428,246,249,829	148.99%	...	45.91%
941. KPP Pratama Ambon	1,425,327,360,000	1,778,835,143,679	124.80%	...	49.25%
942. KPP Pratama Ternate	1,764,983,824,000	2,233,607,806,256	126.55%	...	83.28%
952. KPP Pratama Jayapura	2,155,373,396,000	3,079,334,722,932	142.87%	...	42.97%
956. KPP Pratama Merauke	606,533,011,000	894,669,287,473	147.51%	...	41.47%

Berdasarkan Tabel 2 dan Tabel 3, terlihat adanya variasi jumlah pelaporan SPT dan pembayaran pajak antar KPP. Variasi ini menunjukkan bahwa tingkat kepatuhan setiap KPP tidak bersifat homogen dan memerlukan pengolahan lebih lanjut agar dapat dibandingkan secara proporsional.

Untuk mengatasi perbedaan skala data tersebut, selanjutnya dilakukan perhitungan rasio pelaporan SPT dan rasio pembayaran pajak. Hasil transformasi ini bertujuan untuk menstandarkan nilai kepatuhan sehingga dapat digunakan secara efektif dalam proses pemodelan. Distribusi rasio pelaporan SPT yang telah diproses disajikan pada Tabel 4, sedangkan distribusi rasio pembayaran pajak ditampilkan pada Tabel 5. Informasi ini menunjukkan bahwa sebagian besar KPP memiliki rasio pelaporan di atas ambang batas 80%, yang mengindikasikan tingkat kepatuhan administrasi yang relatif tinggi. Namun demikian, masih terdapat beberapa KPP dengan rasio pelaporan di bawah ambang batas tersebut yang berpotensi diklasifikasikan sebagai tidak patuh.

Pada rasio pembayaran pajak, variasi nilai terlihat lebih besar dibandingkan rasio pelaporan. Kondisi ini menunjukkan bahwa kepatuhan fiskal tidak selalu sejalan dengan kepatuhan administrasi, sehingga kedua variabel tersebut perlu dianalisis secara simultan dalam proses klasifikasi.

Contoh dataset rasio pelaporan SPT dan rasio pembayaran pajak yang telah diproses dan siap digunakan sebagai input model masing-masing disajikan pada Tabel 4 dan Tabel 5. Pada tahap ini, status kepatuhan ditetapkan sebagai variabel target dengan dua kelas, yaitu patuh ($\geq 80\%$) dan tidak patuh ($< 80\%$).

Tabel 4. Rasio dan Status Kepatuhan Pelaporan SPT.

KANTOR PELAYANAN PAJAK	%LAPOR2 022	STATUS2 022	% LAPOR_202 3	STATUS2 023	% LAPOR2024	STATUS20 243
001. KPP Pratama Jakarta Matraman	85.4%	Patuh	180.7%	Patuh	20%	Tidak Patuh
002. KPP Pratama Jakarta Jatinegara	86.1%	Patuh	174.5%	Patuh	21%	Tidak Patuh
003. KPP Pratama Jakarta Pulogadung	99.5%	Patuh	174.2%	Patuh	18%	Tidak Patuh
015. KPP Pratama Jakarta Tebet	85.0%	Patuh	163.5%	Patuh	24%	Tidak Patuh
021. KPP Pratama Jakarta Menteng Satu	86.6%	Patuh	151.6%	Patuh	28%	Tidak Patuh
046. KPP Madya Jakarta Utara	99.9%	Patuh	201.7%	Patuh	140%	Patuh

051. KPP Wajib Pajak Besar Tiga	102.8%	Patuh	205.6%	Patuh	19%	Tidak Patuh
052. KPP Penanaman Modal Asing Satu	106.1%	Patuh	202.9%	Patuh	91%	Patuh
053. KPP Badan dan Orang Asing	110.5%	Patuh	248.9%	Patuh	14%	Tidak Patuh
...	...	...	...	...	...	...

Tabel 5. Rasio dan Status Kepatuhan Pembayaran.

KANTOR PELAYANAN PAJAK	%BAYAR20 22	STATUS20 22	%BAYAR20 23	STATUS20 23	%BAYAR20 24	STATUS20 24
001. KPP Pratama Jakarta Matraman	138.8%	Patuh	122.8%	PATUH	47.37%	Tidak Patuh
007. KPP Madya Jakarta Timur	116.5%	Patuh	102.3%	PATUH	99.35%	Patuh
112. KPP Pratama Medan Belawan	125.3%	Patuh	103.8%	PATUH	93.53%	Patuh
201. KPP Pratama Padang Satu	113.3%	Patuh	102.3%	PATUH	72.00%	Tidak Patuh
215. KPP Pratama Batam Utara	126.5%	Patuh	102.1%	PATUH	76.67%	Tidak Patuh
216. KPP Pratama Pekanbaru Tampan	70.3%	Tidak Patuh	107.1%	PATUH	79.35%	Tidak Patuh
407. KPP Pratama Bekasi Utara	126.8%	Patuh	103.2%	PATUH	81.93%	Patuh
422. KPP Pratama Bandung Tegallega	115.1%	Patuh	104.7%	PATUH	73.85%	Tidak Patuh
451. KPP Pratama Tigaraksa	127.8%	Patuh	109.0%	PATUH	93.64%	Patuh
...	...	...	...	...	...	...

Selanjutnya, dataset dibagi menjadi 80% data latih dan 20% data uji. Model Gaussian Naive Bayes dilatih menggunakan data latih dengan dua variabel input, yaitu rasio pelaporan dan rasio pembayaran pajak. Evaluasi model dilakukan pada data uji menggunakan metrik akurasi, precision, recall, dan F1-score.

Hasil prediksi model Naive Bayes untuk masing-masing tahun dan kategori kepatuhan disajikan pada Tabel 6.

Tabel 6. Hasil Prediksi Naive Bayes

No.	Kategori	Akurasi	Dominan	Keterangan
1	Lapor 2022	100%	PATUH	Mayoritas nilai rasio > 1
2	Lapor 2023	100%	PATUH	Semua KPP memenuhi target pelaporan
3	Lapor 2024	100%	PATUH	Seluruh KPP menunjukkan kepatuhan tinggi
4	Bayar 2022	100%	PATUH	Semua KPP melampaui target pembayaran
5	Bayar 2023	100%	PATUH	Konsisten dengan tren kepatuhan tinggi
6	Bayar 2024	100%	TIDAK PATUH	Terjadi penurunan tingkat kepatuhan

Berdasarkan Tabel 6, model Naive Bayes menunjukkan tingkat akurasi sebesar 100% pada seluruh periode pengamatan. Mayoritas KPP diklasifikasikan sebagai patuh pada aspek pelaporan SPT, sedangkan pada aspek pembayaran pajak tahun 2024 terlihat dominasi KPP yang diklasifikasikan sebagai tidak patuh. Kondisi ini mengindikasikan adanya penurunan tingkat realisasi pembayaran pajak meskipun tingkat kepatuhan administrasi melalui pelaporan SPT tetap tinggi. Secara keseluruhan, hasil ini menunjukkan bahwa algoritma Naive Bayes mampu mengklasifikasikan status kepatuhan KPP secara konsisten berdasarkan kombinasi rasio pelaporan dan rasio pembayaran pajak, serta mampu menangkap perbedaan pola antara kepatuhan administrasi dan kepatuhan fiskal.

3.3 Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan menggunakan confusion matrix dan metrik akurasi, precision, recall, serta F1-score. Tabel 7 menyajikan struktur confusion matrix yang digunakan untuk menghitung metrik evaluasi.

Tabel 7. Confusion Matrix

Aktual / Prediksi	Patuh ($\geq 80\%$)	Tidak Patuh ($< 80\%$)	Total Aktual
Patuh ($\geq 80\%$)	True Positive (TP)	False Negative (FN)	KPP Patuh Aktual
Tidak Patuh ($< 80\%$)	False Positive (FP)	True Negative (TN)	KPP Tidak Patuh Aktual
Total Prediksi	KPP Diprediksi Patuh	KPP Diprediksi Tidak Patuh	Total KPP

Berdasarkan hasil confusion matrix pada masing-masing tahun, tidak ditemukan kesalahan klasifikasi, baik false positive maupun false negative. Seluruh data aktual berhasil diprediksi dengan benar oleh model.

Adapun Hasil Metrik Evaluasi Status Lapor dan Status Bayar menggunakan formula:

$$\text{Akurasi} = (TP + TN) / (TP + FP + FN + TN) \quad (1)$$

$$\text{Precision (Patuh)} = TP / (TP + FP) \quad (2)$$

$$\text{Recall (Patuh)} = TP / (TP + FN) \quad (3)$$

$$F1\text{-Score (Patuh)} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

Karena FP = 0 dan FN = 0 (Asumsi Model Sempurna), semua metrik adalah 1.00

Hasil perhitungan metrik evaluasi untuk status pelaporan dan status pembayaran pajak disajikan pada Tabel 8.

Tabel 8. Metrik Evaluasi Status Laporan dan Status Bayar

Tahun	Akurasi	Precision (Patuh)	Recall (Patuh)	F1-Score (Patuh)
2022	1	1	1	1
2023	1	1	1	1
2024	1	1	1	1.00

Berdasarkan Tabel 8, seluruh metrik evaluasi memiliki nilai 1,00. Nilai ini menunjukkan kesesuaian penuh antara hasil prediksi model dan data aktual pada dataset pengujian. Namun demikian, meskipun metrik evaluasi menunjukkan performa sempurna, perbedaan proporsi antara KPP patuh dan tidak patuh pada masing-masing tahun tetap mencerminkan dinamika kepatuhan yang berbeda-beda di lapangan.

3.4 Pembahasan dan Interpretasi Hasil

Hasil penelitian menunjukkan bahwa rasio pelaporan SPT merupakan faktor yang lebih dominan dibandingkan rasio pembayaran pajak dalam menentukan status kepatuhan wajib pajak badan. KPP dengan rasio pelaporan yang tinggi cenderung diklasifikasikan sebagai patuh, meskipun rasio pembayaran pajaknya bervariasi. Sebaliknya, KPP dengan rasio pelaporan yang rendah tetap terklasifikasi sebagai tidak patuh meskipun memiliki rasio pembayaran yang relatif tinggi.

Contoh Interpretasi :

- KPP Pratama Jakarta Matraman, dengan rasio pelaporan tahun 2022 sebesar 85% dan rasio pembayaran pajak diatas 100%, berhasil diprediksi sebagai patuh, sesuai dengan hasil aktual.
- KPP Madya Jakarta Timur, dengan rasio pelaporan tahun 2022 sebesar 100% dan rasio pembayaran pajak diatas 100%, berhasil diprediksi sebagai patuh, sesuai dengan hasil aktual.
- KPP Pratama Pekanbaru Tampan, dengan rasio pelaporan tahun 2022 sebesar 63% dan rasio pembayaran pajak sebesar 70%, terprediksi sebagai tidak patuh, yang sesuai dengan hasil aktual.
- KPP Pratama Balikpapan Timur, dengan rasio pelaporan tahun 2022 sebesar 77% dan rasio pembayaran pajak sebesar 72%, terprediksi sebagai tidak patuh, yang sesuai dengan hasil aktual.

Temuan ini memperkuat hasil penelitian terdahulu yang menyatakan bahwa kepatuhan administrasi, khususnya pelaporan SPT, merupakan indikator awal kepatuhan fiskal. Dengan kata lain, pelaporan pajak yang konsisten dan tepat waktu menjadi prasyarat utama sebelum kepatuhan pembayaran dapat dinilai secara optimal.

Dibandingkan dengan penelitian sebelumnya yang menerapkan algoritma Naive Bayes pada jenis pajak atau level data yang berbeda, hasil penelitian ini menunjukkan performa model yang sebanding, bahkan lebih stabil pada level agregat KPP. Stabilitas ini didukung oleh penggunaan data agregat dan rasio, yang cenderung memiliki variasi lebih rendah dibandingkan data individu wajib pajak.

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa algoritma Naive Bayes mampu mengklasifikasikan potensi kepatuhan wajib pajak badan secara efektif berdasarkan rasio pelaporan SPT dan rasio pembayaran pajak menggunakan data agregat pada level Kantor Pelayanan Pajak (KPP). Hasil evaluasi menunjukkan kinerja model yang sangat baik, dengan nilai akurasi, precision, recall, dan F1-score sebesar 1,00 pada seluruh periode pengamatan. Temuan utama penelitian ini menunjukkan bahwa rasio pelaporan SPT merupakan faktor yang paling dominan dalam menentukan status kepatuhan wajib pajak badan dibandingkan dengan rasio pembayaran pajak. KPP dengan rasio pelaporan yang rendah cenderung terklasifikasi sebagai tidak patuh meskipun memiliki rasio pembayaran yang relatif tinggi. Hal ini menegaskan bahwa kepatuhan administrasi merupakan indikator awal yang penting dalam menilai kepatuhan fiskal. Model klasifikasi yang diusulkan berpotensi digunakan sebagai sistem pendukung keputusan bagi otoritas pajak dalam menetapkan prioritas pengawasan dan pembinaan wajib pajak badan, khususnya pada KPP yang memiliki rasio pelaporan rendah. Model ini juga dapat diintegrasikan sebagai alat penyaringan awal (early warning system) dalam kerangka pengawasan berbasis risiko di lingkungan Direktorat Jenderal Pajak. Dengan pendekatan berbasis data tersebut, pengelolaan kepatuhan pajak diharapkan dapat dilakukan secara lebih terarah, efektif, dan efisien.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini khususnya kepada Subdirektorat Tata Kelola Data dan Informasi - Direktorat Data dan Informasi Perpajakan - Kantor Pusat Direktorat Jenderal Pajak

REFERENCES

- [1] "THE ROLE OF TAXES IN THE DEVELOPMENT OF THE NATIONAL ECONOMY," International journal of artificial intelligence, vol. 5, no. 02 SE-Articles, pp. 1101–1103, Feb. 2025, [Online]. Available: <https://www.academicpublishers.org/journals/index.php/ijai/article/view/3001>
- [2] H. T. Nguyen and S. N. A. C. Darsono, "The Impacts of Tax Revenue and Investment on the Economic Growth in Southeast Asian Countries," Journal of Accounting and Investment, vol. 23, no. 1 SE-Articles, pp. 128–146, Jan. 2022, doi: 10.18196/jai.v23i1.13270.
- [3] I. A. Maulana and A. P. W. Wibowo, "Analysis of Public Sentiment Text Clustering on Tax Increases using Orange Data Mining on Twitter," Brilliance: Research of Artificial Intelligence, vol. 5, no. 1 SE-Articles, pp. 93–99, Apr. 2025, doi: 10.47709/brilliance.v5i1.5787.
- [4] K. Azmi and A. Evioni, "Application of Data Mining Using Classification for Potential Tax Costs," Journal of Computer Science and Informatics Engineering, vol. 4, no. 1 SE-Articles, pp. 40–48, Jan. 2025, doi: 10.55537/cosie.v4i1.990.
- [5] Y. G. Winata and M. Hadi, "Application of data mining techniques for VAT-Registered Business compliance," Scientax: Jurnal Kajian Ilmiah Perpajakan Indonesia, vol. 4, no. 2 SE-Articles, pp. 243–260, Apr. 2023, doi: 10.52869/st.v4i2.317.
- [6] Y. G. Winata, F. Noor, M. F. Bahar, A. B. Santoso, and E. Sukarno, "Application of data mining to taxpayers issuing fictitious tax invoice using classification techniques," Scientax: Jurnal Kajian Ilmiah Perpajakan Indonesia, vol. 5, no. 1 SE-Articles, pp. 58–73, Oct. 2023, doi: 10.52869/st.v5i1.573.
- [7] M. A. Putri, N. Rahaningsih, F. M. Basysyar, and O. Nurdiawan, "Penerapan Data Mining Menggunakan Metode Clustering Untuk Mengetahui Kelompok Kepatuhan Wajib Pajak Bumi dan Bangunan," vol. 5, no. 2, pp. 145–156, 2022.
- [8] A. Ginanjar and A. S. Wibowo, "Data Mining to Detect Fraud Patterns in a Taxpayer's Financial Statement," Scientax: Jurnal Kajian Ilmiah Perpajakan Indonesia, vol. 6, no. 2 SE-Articles, pp. 135–150, Apr. 2025, doi: 10.52869/st.v6i2.571.
- [9] I. Vol, N. O. Issn, and N. P. Rahmayanti, "Research in Business & Social Science Effect of tax penalties, tax audit, and taxpayers awareness on corporate taxpayers' compliance moderated by compliance intentions," vol. 9, no. 2, pp. 118–124, 2020.
- [10] A. Putera Utama Siahaan, S. Sutiono, S. Pranoto, and R. Sri Mentari, "Analysis of Property Tax Payment Compliance Classification in Tebing Tinggi City Using the C4.5 Decision Tree Algorithm," Journal of Information Technology, computer science and Electrical Engineering (JITCSE), vol. 1, no. 2, pp. 134–138, 2024, doi: 10.61306/jitcse.v1i2.
- [11] B. De La Iglesia, "A Methodology: The KDD Roadmap."
- [12] A. Susanto and ; Meiryani, "Functions, Processes, Stages And Application Of Data Mining," INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH, vol. 8, p. 7, 2019, [Online]. Available: www.ijstr.org
- [13] V. Plotnikova, M. Dumas, and F. Milani, "Adaptations of data mining methodologies: A systematic literature review," PeerJ Comput. Sci., vol. 6, pp. 1–43, 2020, doi: 10.7717/PEERJ-CS.267.
- [14] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases) (© AAAI)," 1996. [Online]. Available: www.ffly.com/
- [15] G. Mariscal, Ó. Marbán, and C. Fernández, "A survey of data mining and knowledge discovery process models and methodologies," Knowl. Eng. Rev., vol. 25, no. 2, pp. 137–166, 2010, doi: DOI: 10.1017/S0269888910000032.
- [16] S. Shahidi, A. Wahid Samadzai, and H. Shahbazi, "Effective Data Preprocessing in Data Science: From Method Selection to Domain-Specific Optimization," Journal of Advanced Computer Knowledge and Algorithms, vol. 2, no. 4, pp. 84–90, Jul. 2025, doi: 10.29103/jacka.v2i4.22886.
- [17] A. Hammad and S. AbouRizk, "Knowledge Discovery in Data: A Case Study," Journal of Computer and Communications, vol. 02, no. 05, pp. 1–28, 2014, doi: 10.4236/jcc.2014.25001.