

Penerapan Data Mining Menggunakan Metode Teknik Klasifikasi Untuk Melihat Potensi Kepatuhan Wajib Pajak Kendaraan

Application of Data Mining Using Classification Technique Methods to See Potential Vehicle Taxpayer Compliance

Rika Dinda Lestari¹, Muhammad Siddik Hasibuan², Sri Wahyuni³

¹²³Program Studi Ilmu Komputer Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara

¹²³Jalan Lap.Golf No.120, Kp.Tengah, Kec. Pancur Batu, Sumatera Utara, Indonesia

E-mail: *¹rkdndstargmail.com, ²muhammadsiddik@uinsu.ac.id,

³sriwahyuni10112018@gmail.com

Abstrak

Tujuan penelitian ini adalah untuk menguji kepatuhan wajib pajak kendaraan bermotor, dengan fokus pada Samsat Medan Selatan dan UPTD BAPENDA Kota Medan. Sumber utama pendanaan pembangunan dan pelayanan negara yang dapat dinikmati seluruh masyarakat adalah pajak, yang menyumbang lebih dari 70% pendapatan negara. Badan Pendapatan Daerah (Bapenda) mengelola Pajak Kendaraan Bermotor (PKB) yang tergolong pajak daerah dan ditangani oleh Kantor Bersama Sistem Administrasi Terpadu Satu Atap (SAMSAT). Kami dapat mengidentifikasi wajib pajak yang patuh dan tidak patuh dengan menggunakan sampel data wajib pajak kendaraan bermotor dari Samsat Medan Selatan, tepatnya pada bulan Mei 2023. Dalam penelitian ini, data mining diterapkan bersamaan dengan Teknik Klasifikasi menggunakan algoritma *Naïve Bayes* untuk mengidentifikasi, dengan akurasi sebesar 70,91%, diantara 110 wajib pajak terdapat yang berpotensi patuh (nilai *prior probability*: 0,564) dan yang berpotensi tidak patuh (nilai *prior probability*: 0,436).

Kata Kunci : PKB, Data Mining, Naïve Bayes, Rapidminer

Abstract

The aim of this research is to examine motor vehicle taxpayer compliance, with a focus on South Medan Samsat and Medan City PEPENDA UPTD. The main source of funding for development and state services that can be enjoyed by the whole community is taxes, which contribute more than 70% of state income. The Regional Revenue Agency (Bapenda) manages Motor Vehicle Tax (PKB) which is classified as a regional tax and is handled by the Joint Office for One-Stop Integrated Administrative Systems (SAMSAT). We can identify compliant and non-compliant taxpayers by using a sample of motor vehicle taxpayer data from the South Medan Samsat, precisely in May 2023. In this research, data mining is applied in conjunction with the Classification Technique using the Naïve Bayes algorithm to identify, with an accuracy of 70.91%, among 110 taxpayers there are those who have the potential to comply (prior probability value: 0.564) and those who have the potential to not comply (prior probability value: 0.436).

Keywords: PKB, Data Mining, Naïve Bayes, Rapidminer

1. PENDAHULUAN

Sumber utama pendanaan pembangunan dan pelayanan negara yang dapat dinikmati seluruh masyarakat adalah pajak, yang menyumbang lebih dari 70% pendapatan negara. Dari sudut hukum perpajakan, pajak adalah suatu kewajiban yang timbul dari suatu peraturan perundang-undangan yang mewajibkan penduduk untuk turut serta dengan cara menyetorkan sejumlah penghasilannya untuk pembangunan negara sesuai dengan pedoman yang tertuang dalam Undang-undang Perpajakan Nomor 16 Tahun 2009.[1]

Sebagai pajak daerah, Pajak Kendaraan Bermotor (PKB) diurus oleh Badan Pendapatan Daerah (Bapenda) dan dikelola oleh Badan Gabungan Sistem Administrasi Terpadu Satu Pintu (SAMSAT). Persyaratan kepatuhan PKB diperkirakan akan meningkat seiring dengan jumlah kendaraan bermotor dari waktu ke waktu. Undang-Undang Nomor 28 Tahun 2009 tentang Pajak Daerah dan Retribusi Daerah mengatur ketentuan mengenai PKB pada Pasal 1 .[2] Jumlah tersebut akan dimasukkan dalam perhitungan pendapatan asli daerah karena PKB termasuk dalam pajak daerah. PKB merupakan pajak yang dipungut terhadap orang yang memiliki atau mengelola kendaraan bermotor. PKB fokus pada orang atau badan tertentu, sebagaimana tercantum dalam Pasal 4 Undang-Undang Nomor 28 Tahun 2009. Apabila wajib pajak berbentuk badan, maka kewajiban perpajakan diwakili oleh pengurus atau kuasa badan tersebut. Pemilik pribadi di PKB dapat meminta orang lain untuk membayar pajak atas namanya agar tidak perlu membayar pajak atas mobilnya, dengan syarat dapat menunjukkan surat identitas, Surat Tanda Nomor Kendaraan (STNK), atau surat kuasa khusus. pengacara. pajak yang mereka bayarkan.[3]

Berdasarkan kajian sebelumnya, mereka yang memiliki banyak kendaraan di PKB akan dikenakan tarif progresif. Apabila Wajib Pajak berdomisili di alamat yang sama, terdaftar di sana, dan tercantum pada kartu keluarga yang sama, maka berlaku tarif progresif. Undang-Undang Nomor 28 Tahun 2009 tentang Pajak Daerah dan Retribusi Daerah mengatur secara rinci mengenai peraturan pajak progresif. Sebelum menetapkan pajak progresif, perlu mempertimbangkan nilai komoditas dan dampak buruk penggunaan mobil, termasuk potensi merugikan infrastruktur transportasi umum. Selain pajak progresif, iuran wajib Dana Kecelakaan Lalu Lintas Jalan dan Bea Balik Nama Kendaraan Bermotor (BBN-KB) merupakan biaya tambahan yang tidak ditanggung oleh PKB.

Seorang wajib pajak dikatakan patuh wajib pajak apabila ia dapat memenuhi hak dan kewajiban perpajakannya dengan tepat serta mentaati semua peraturan perundang-undangan yang berlaku. Seorang wajib pajak dikatakan taat terhadap peraturan perundang-undangan apabila ia mengajukan pajaknya sesuai jadwal, membayar pajaknya tepat waktu, dan menaati seluruh peraturan perundang-undangan yang berlaku. Kepatuhan perpajakan ditentukan oleh beberapa faktor, antara lain ketepatan waktu pembayaran, opini wajar tanpa pengecualian dari akuntan publik setelah tiga tahun penuh menelaah laporan keuangan wajib pajak, tidak ada tunggakan pajak (kecuali diperkenankan menunda atau membayar pajak secara mencicil), dan tidak adanya tuntutan pidana. untuk pelanggaran yang berhubungan dengan perpajakan. Ketidapatuhan perpajakan sering kali disebabkan oleh ketidaktahuan wajib pajak akan pentingnya pajak bagi pertumbuhan nasional. Kita perlu memiliki pemahaman tentang wajib pajak agar dapat sepenuhnya mengapresiasi sikap berbakti terhadap mereka. Karena semakin banyaknya masyarakat yang menyerahkan pajak dan karena UPTD BAPENDA Kota Medan, khususnya di Samsat Medan Selatan, tidak memiliki data iuran wajib pajak, maka mereka tidak dapat menentukan persentase wajib pajak yang patuh dan tidak patuh. Contoh ini memperjelas bahwa mengelola data dalam jumlah besar adalah salah satu manfaat *data mining*. Untuk mengamati kepatuhan wajib pajak, digunakan pendekatan kategorisasi untuk mengetahui kemungkinan kepatuhan wajib pajak (WP).

Proses menciptakan informasi dari sumber terorganisir dan tidak terstruktur dikenal sebagai penemuan pengetahuan dalam *database*. Pengetahuan yang dihasilkan harus direpresentasikan sedemikian rupa sehingga membuat inferensi lebih mudah dan harus dalam

format yang dapat dibaca oleh mesin dan dapat diinterpretasikan oleh mesin. *Knowledge Discovery in Databases (KDD)* adalah kata lain yang digunakan dalam *data mining* yang memiliki arti mirip dengan *data mining*. Tujuan dari *data mining* dan KDD adalah menggunakan data database dan memprosesnya untuk menghasilkan informasi baru yang berharga.[4] Selain itu, banyak ungkapan lain, seperti ekstraksi pengetahuan, arkeologi data, pengerukan data, analisis pola/data, dan penemuan pengetahuan dalam database, memiliki konotasi yang mirip dengan penambangan data. Kesalahpahaman yang umum terjadi adalah bahwa data mining identik dengan kata lain yang banyak digunakan. Proses menciptakan informasi dari sumber terorganisir dan tidak terstruktur dikenal sebagai penemuan pengetahuan dalam database. Beberapa orang memandang *data mining* hanya sebagai salah satu fase penting dalam proses penemuan pengetahuan, namun ada juga yang menginginkan pengetahuan yang dihasilkan berada dalam format yang dapat dibaca mesin dan dapat diinterpretasikan oleh mesin serta mencerminkan penemuan pengetahuan dari data, atau KDD.

Proses penerapan teknik atau metodologi tertentu pada data terpilih untuk menemukan pola atau informasi menarik dikenal dengan istilah *data mining*. Teknik, metode, atau algoritma penambangan data sangat berbeda satu sama lain. Tujuan dan keseluruhan proses KDD pada akhirnya menentukan apakah pendekatan atau algoritma yang terbaik. Dalam istilah teknis, data mining adalah praktik menganalisis database relasional yang sangat besar untuk mengidentifikasi pola atau korelasi pada ratusan atau ribuan variabel.

Klasifikasi adalah proses pengembangan (penggunaan) bentuk-bentuk yang mencirikan dan membedakan kelompok data atau desain yang dimaksudkan untuk memperhitungkan kategori mata pelajaran dengan sebutan kelas yang tidak jelas.[5] Proses pengorganisasian informasi melibatkan dua bagian. Langkah pertama disebut pelatihan, atau pelatihan penyemaian, dan ini melibatkan pengembangan algoritma pengelompokan dan menyajikannya sebagai seperangkat aturan sehingga dapat digunakan untuk menganalisis data pelatihan. Pendekatan kedua, clustering, membuat asumsi tentang ketepatan algoritma klasifikasi berdasarkan data eksperimen.

Arif Rahmat Shaumi, Muhammad Faridz Ali dan M Tsaqif Al Mutawakkil Simbolon telah melakukan penelitian sebelumnya dengan judul "Penerapan Data Mining Menggunakan Metode Teknik *Classification* Untuk Melihat Potensi Kepatuhan Wajib Pajak Bumi Dan Bangunan". Tujuan dari penelitian sebelumnya adalah mengklasifikasikan data wajib pajak yang bayar sangat tepat waktu, tepat waktu dan tidak tepat waktu. dengan teknik pengklasifikasian dan menggunakan metode *naive bayes*. Sebagai pengklasifikasi probabilistik langsung, metode *Naive Bayes* menjumlahkan frekuensi dan kombinasi nilai dari kumpulan data tertentu untuk menentukan kumpulan probabilitas. Dengan menggunakan nilai variabel kelas, algoritma menerapkan teorema Bayes dan mengasumsikan bahwa semua kualitas bersifat independen atau tidak saling bergantung.

Menggunakan kemungkinan konseptual (*Probabilitas*) untuk menyelesaikan kasus Supervised Learning kumpulan data yang berisi Label, Kelas, atau Target sebagai *referensi Naive Bayes Classifier* menggunakan statistik. Pengklasifikasi *Naive Bayes* adalah algoritma klasifikasi multi-fase. Untuk memastikan bahwa pilihan yang dibuat akurat, solusinya terdiri dari Data Pelatihan, Algoritma Pembelajaran, Model, Data Uji, dan Proses Pengujian. Kumpulan data yang besar dapat memperoleh manfaat besar dari kemudahan konstruksi *Naive Model Bayes*.[6]

2. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode deskriptif. Metode deskriptif merupakan pendekatan studi objektif yang bertujuan untuk memberikan gambaran jelas mengenai suatu keadaan dengan mengumpulkan informasi dan menganalisis data. Penelitian deskriptif adalah penelitian yang dilakukan untuk memastikan satu atau lebih variabel bebas, yaitu variabel yang tidak bergantung atau mempunyai hubungan dengan variabel lain.[7]

Teknik statistik untuk melakukan inferensi induktif pada masalah klasifikasi adalah metode *Bayes*. *Teorema Bayes* dapat digunakan untuk menghitung probabilitas dengan sejumlah kondisi sistemik. Kesimpulan yang diambil dari kejadian aktual yang berhubungan dengan satu atau lebih variabel disebut probabilitas bersyarat. Pertama-tama kita membahas ide dasar dan definisi *Teorema Bayes* sebelum menerapkannya pada tugas klasifikasi *Data Mining*. Bentuk umum *teorema Bayes* adalah sebagai berikut:

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}$$

[8]

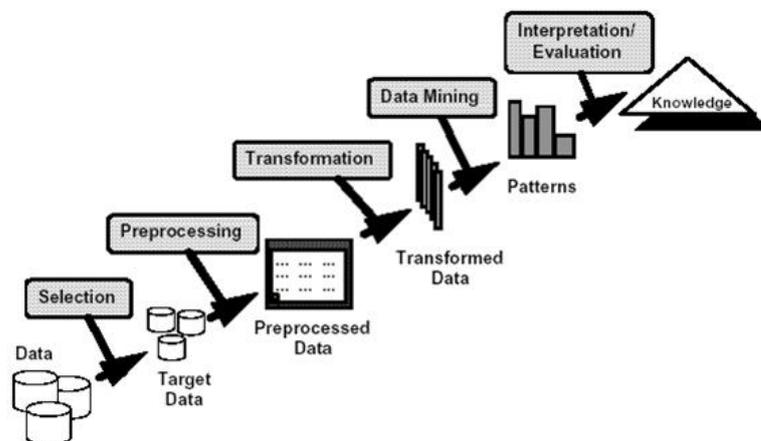
Keterangan :

$P(H|E)$ = Probabilitas

H = Hipotesis

E = Bukti

Pada aplikasi data mining, analisis data dilakukan melalui tahap *Knowledge Discovery in Database* (KDD). Praktik mengambil informasi tersembunyi dari *database* yang sangat besar disebut juga dengan penemuan pengetahuan di *database* (KDD) atau *data mining*. [9] Meski kedua nama ini memiliki arti berbeda, namun keduanya saling berkaitan. Penambangan data juga merupakan langkah dalam proses KDD secara penuh.



Gambar 1. Tahapan KDD

Berdasarkan gambar di atas, proses KDD secara garis besar dinyatakan sebagai berikut:

2.1 Data Selection

Data harus dipilih atau diseleksi dari sekumpulan data operasional sebelum langkah penambangan informasi KDD dapat dimulai. Data yang dipilih akan digunakan dalam proses penambangan data, yang melibatkan identifikasi data yang kami perlukan untuk pemrosesan tambahan dan menyimpannya dalam *file* terpisah dari *database* operasional untuk memfasilitasi penggunaan selanjutnya.

2.2 Pre-processing (Cleaning)

Secara umum, data yang dikumpulkan baik dari eksperimen atau database komersial memiliki entri yang tidak lengkap, seperti ada yang hilang, salah, atau kesalahan sederhana. Selain itu, ada kualitas data unggulan lainnya yang tidak relevan. dihapus karena keberadaannya selanjutnya dapat mempengaruhi presisi atau kualitas hasil data mining.

2.3 Data Transformation

Sebelum digunakan, beberapa teknik data mining memerlukan format data tertentu. Metode umum tertentu, seperti klustering dan analisis asosiasi, misalnya, hanya dapat mengambil data kategorikal. Akibatnya, data numerik kontinu dapat dipisahkan menjadi banyak interval. Istilah umum untuk prosedur ini adalah “*binning*.” Di sini, data yang dibutuhkan untuk pendekatan data mining juga dipilih. Karena beberapa aspek pendekatan *data mining* tertentu bergantung pada tahap ini, maka transformasi dan seleksi data ini juga mempengaruhi kualitas keluaran data mining nantinya.

2.4 Data Mining

Data mining adalah proses penggunaan alat atau strategi tertentu untuk mencari pola atau informasi menarik dalam sekumpulan data. Dalam penambangan data, teknik, metodologi, atau algoritma mungkin berbeda secara signifikan. Tujuan dan keseluruhan proses KDD pada akhirnya menentukan apakah pendekatan atau algoritma yang terbaik. Idealnya, data yang dibutuhkan untuk membuat model yang kuat harus berupa data penelitian yang memadai. Model benchmark yang memiliki lebih banyak data dan lebih sedikit kesalahan adalah model yang unggul.[10]

2.5 Interpretation (Evaluation)

Proses penambangan data menghasilkan pola informasi, yang perlu disajikan dengan cara yang dapat dipahami oleh pihak-pihak yang berkepentingan. Proses KDD memiliki tahapan yang disebut interpretasi. Pada langkah ini perlu dipastikan apakah informasi atau pola yang ditemukan menyimpang dari fakta atau pola yang telah ditetapkan sebelumnya.

Contoh data pembayaran pajak kendaraan bulan Mei 2023 pada Samsat Medan Selatan diperoleh dari UPTD PAPENDA Kota Medan. Hingga 300 WP termasuk di antara sampel data yang dapat diperoleh sebelum prosedur KDD.

3. HASIL DAN PEMBAHASAN

Tahapan *Knowledge Discover in Database* (KDD) digunakan untuk analisis data dalam aplikasi data mining adalah sebagai berikut:

3.1 Data Selection

Dengan menggunakan *rapidminer*, bidang dari sumber data sekarang akan dibagi menjadi bidang yang akan digunakan untuk tahap *Pre-Processing* (pembersihan data) dan bidang yang tidak akan digunakan untuk memperoleh hasil pengetahuan yang diperlukan.

Row No.	No Mesin	Kecamatan	Tgl Jatuh Te...	Tgl Bayar	Estimasi (ha...	Estimasi
104	41538	Mdn Selayang	May 2, 2023	May 27, 2023	-25	Tidak Taat
105	92727	Mdn Selayang	Jun 2, 2023	May 27, 2023	6	Taat
106	05904	Mdn Tuntung...	May 28, 2022	May 27, 2023	-364	Tidak Taat
107	65493	Mdn Selayang	May 29, 2023	May 27, 2023	2	Taat
108	42315	Mdn Kota	Jul 17, 2023	May 29, 2023	49	Taat
109	00064	Mdn Amplas	May 30, 2023	May 29, 2023	1	Taat
110	25242	Mdn Johor	May 30, 2023	May 29, 2023	1	Taat
111	09607	Mdn Johor	May 31, 2023	May 29, 2023	2	Taat
112	75576	Mdn Amplas	May 19, 2022	May 29, 2023	-375	Tidak Taat
113	49392	Mdn Johor	Jul 2, 2023	May 29, 2023	34	Taat
114	27350	Mdn Kota	Apr 29, 2023	May 29, 2023	-30	Tidak Taat
115	09112	?	Jan 13, 2024	May 29, 2023	229	Taat
116	74998	Mdn Johor	Dec 19, 2023	May 29, 2023	204	Taat
117	53036	Mdn Sunggal	Nov 4, 2014	May 29, 2023	-3128	Tidak Taat
118	78409	?	Jun 7, 2023	May 29, 2023	9	Taat

Gambar 2. Hasil Data Selection

Nomor Mesin, Kecamatan, Nomor PKB, Tanggal Jatuh Tempo, dan Tanggal Pembayaran merupakan data-data terkait dalam penelitian ini yang harus disediakan untuk mencapai hasil yang diinginkan dan mengikuti protokol. Selain itu, terdapat dua kolom data lagi: perkiraan dan perkiraan (hari), yang menunjukkan seberapa akurat prediksi kemungkinan kepatuhan. Untuk memilih data yang memenuhi persyaratan, gunakan fungsi seleksi. Sampel data yang tersisa setelah seleksi berjumlah 118 WP.

3.2 Processing(Data Cleaning).

Data yang dipilih pada langkah sebelumnya masih mengandung data yang hilang, duplikat, dan kosong. Selain itu, adanya data yang tidak relevan selanjutnya dapat menurunkan keakuratan hasil data mining, sehingga sebaiknya dihilangkan. Akan ditentukan apakah data tersebut berisi data kosong dan nilai yang hilang dengan mengolah data yang dipilih sebelumnya. Apabila tidak ada data yang kosong atau hilang maka tahap transformasi data akan langsung memprosesnya; namun, bila ada, tahap transformasi data tidak akan tercapai sampai data tersebut dibersihkan.

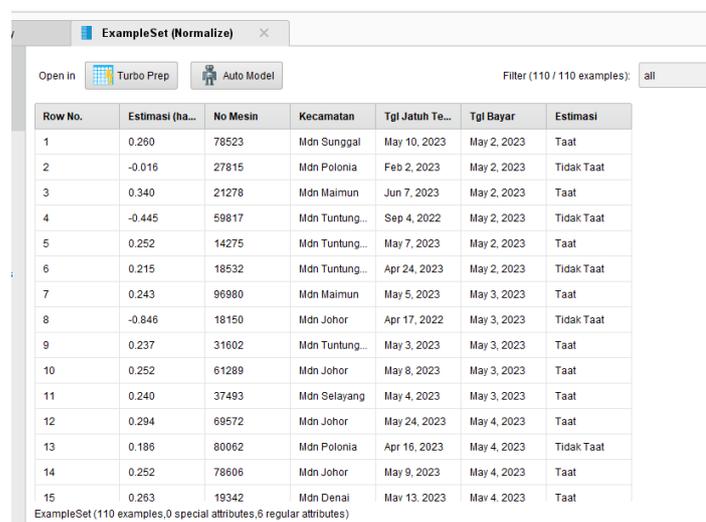
Row No.	No Mesin	Kecamatan	Tgl Jatuh Te...	Tgl Bayar	Estimasi (ha...	Estimasi
96	27452	Mdn Johor	May 25, 2023	May 27, 2023	-2	Tidak Taat
97	50633	Mdn Maimun	Jun 5, 2023	May 27, 2023	9	Taat
98	41538	Mdn Selayang	May 2, 2023	May 27, 2023	-25	Tidak Taat
99	92727	Mdn Selayang	Jun 2, 2023	May 27, 2023	6	Taat
100	05904	Mdn Tuntung...	May 28, 2022	May 27, 2023	-364	Tidak Taat
101	65493	Mdn Selayang	May 29, 2023	May 27, 2023	2	Taat
102	42315	Mdn Kota	Jul 17, 2023	May 29, 2023	49	Taat
103	00064	Mdn Amplas	May 30, 2023	May 29, 2023	1	Taat
104	25242	Mdn Johor	May 30, 2023	May 29, 2023	1	Taat
105	09607	Mdn Johor	May 31, 2023	May 29, 2023	2	Taat
106	75576	Mdn Amplas	May 19, 2022	May 29, 2023	-375	Tidak Taat
107	49392	Mdn Johor	Jul 2, 2023	May 29, 2023	34	Taat
108	27350	Mdn Kota	Apr 29, 2023	May 29, 2023	-30	Tidak Taat
109	74998	Mdn Johor	Dec 19, 2023	May 29, 2023	204	Taat
110	53036	Mdn Sunggal	Nov 4, 2014	May 29, 2023	-3128	Tidak Taat

Gambar 3. Hasil Data Cleansing

Hasil pembersihan data data hilang dan kosong pada data PKB ditunjukkan pada Gambar 3. Data wajib pajak terdapat 110 wajib pajak karena tidak ada nilai yang kosong atau hilang.

3.3 Transformation Data

Setelah dipilih dan dibersihkan, kini data diubah ke dalam format yang sesuai dan sesuai dengan teknik normalisasi data. Operator normalisasi kemudian menggunakan format ini untuk mengubah data asli menjadi data yang sesuai untuk penambangan data. metode penambangan data. Akurasi temuan akan di bawah standar jika data tidak terstandarisasi. Tujuan dari tahapan yang menggunakan metode *Naive Bayes* ini adalah untuk meningkatkan nilai akurasi yang ditemukan pada tahap terakhir.



Row No.	Estimasi (ha...)	No Mesin	Kecamatan	Tgl Jatuh Te...	Tgl Bayar	Estimasi
1	0.260	78523	Mdn Sunggal	May 10, 2023	May 2, 2023	Taat
2	-0.016	27815	Mdn Polonia	Feb 2, 2023	May 2, 2023	Tidak Taat
3	0.340	21278	Mdn Maimun	Jun 7, 2023	May 2, 2023	Taat
4	-0.445	59817	Mdn Tuntung...	Sep 4, 2022	May 2, 2023	Tidak Taat
5	0.252	14275	Mdn Tuntung...	May 7, 2023	May 2, 2023	Taat
6	0.215	18532	Mdn Tuntung...	Apr 24, 2023	May 2, 2023	Tidak Taat
7	0.243	96980	Mdn Maimun	May 5, 2023	May 3, 2023	Taat
8	-0.846	18150	Mdn Johor	Apr 17, 2022	May 3, 2023	Tidak Taat
9	0.237	31602	Mdn Tuntung...	May 3, 2023	May 3, 2023	Taat
10	0.252	61289	Mdn Johor	May 8, 2023	May 3, 2023	Taat
11	0.240	37493	Mdn Selayang	May 4, 2023	May 3, 2023	Taat
12	0.294	69572	Mdn Johor	May 24, 2023	May 4, 2023	Taat
13	0.186	80062	Mdn Polonia	Apr 16, 2023	May 4, 2023	Tidak Taat
14	0.252	78606	Mdn Johor	May 9, 2023	May 4, 2023	Taat
15	0.263	19342	Mdn Denai	May 13, 2023	May 4, 2023	Taat

Gambar 4. Hasil Transformasi Data

Seperti diilustrasikan pada Gambar 4, modifikasi yang dilakukan pada catatan di bidang Nomor Mesin dan Jumlah PKB diubah menjadi format yang sesuai untuk penambangan. Data ini telah diubah sehingga dapat digunakan pada tahap penambangan data berikutnya.

3.4 Data Mining

Saat ini, dua pengujian dijalankan untuk memberikan hasil yang benar: penghitungan melalui *Rapidminer* dan penghitungan yang dilakukan dengan manual. Menggunakan kumpulan data pelatihan 110 catatan, algoritma *Naive Bayes* digunakan untuk menghitung secara manual berikut ini.

Sebelum penghitungan manual dapat dilakukan berdasarkan kumpulan data ini, perkiraan harus diperoleh dengan menyelesaikan penghitungan dan mengembangkan aturan untuk setiap tanggal pembayaran dan tanggal jatuh tempo.

Terdapat 110 data dalam dataset yang dapat digunakan untuk perhitungan manual; beberapa contoh datanya ditunjukkan pada gambar berikut:

No Mesin	Kecamatan	Tgl Jatuh Tempo	Tgl Bayar	Estimasi (hari)
78523	Mdn Sunggal	10/05/2023	02/05/2023	8
27815	Mdn Polonia	02/02/2023	02/05/2023	-89
21278	Mdn Maimun	07/06/2023	02/05/2023	36
59817	Mdn Tuntungan	04/09/2022	02/05/2023	-240
14275	Mdn Tuntungan	07/05/2023	02/05/2023	5
18532	Mdn Tuntungan	24/04/2023	02/05/2023	-8
96980	Mdn Maimun	05/05/2023	03/05/2023	2
18150	Mdn Johor	17/04/2022	03/05/2023	-381
31602	Mdn Tuntungan	03/05/2023	03/05/2023	0
61289	Mdn Johor	08/05/2023	03/05/2023	5
37493	Mdn Selayang	04/05/2023	03/05/2023	1
69572	Mdn Johor	24/05/2023	04/05/2023	20
80062	Mdn Polonia	16/04/2023	04/05/2023	-18
78606	Mdn Johor	09/05/2023	04/05/2023	5
19342	Mdn Denai	13/05/2023	04/05/2023	9
00214	Mdn Maimun	30/05/2023	05/05/2023	25
58542	Mdn Johor	24/05/2023	05/05/2023	19
30171	Mdn Johor	17/05/2023	05/05/2023	12
04094	Mdn Selayang	04/05/2023	05/05/2023	-1
53336	Mdn Kota	04/05/2023	05/05/2023	-1
03102	Mdn Amplas	07/05/2023	06/05/2023	1
48053	Mdn Helvetia	17/04/2023	06/05/2023	-19
66171	Mdn Amplas	09/05/2023	06/05/2023	3
59498	Mdn Tuntungan	16/05/2023	06/05/2023	10
14029	Mdn Denai	17/05/2023	06/05/2023	11
27401	Mdn Tuntungan	08/05/2023	08/05/2023	0
49108	Mdn Johor	07/05/2023	08/05/2023	-1
17397	Mdn Selayang	09/05/2023	08/05/2023	1

Gambar 5. Tampilan Dataset

Menentukan jumlah kelas estimasi dengan membagi seluruh jumlah data dengan jumlah estimasi yang ada, dengan memperhitungkan probabilitas prior dari klasifikasi yang dihasilkan. P1 bagi kelas TAAT dan P2 bagi kelas TIDAK TAAT :

- $P1(\text{Class Estimasi} = \text{"TAAT"}) = \frac{\text{jumlah "TAAT" pada kolom Estimasi}}{110} = \frac{62}{110} = 0,56$
- $P2(\text{Class Estimasi} = \text{"TIDAK TAAT"}) = \frac{\text{jumlah "TIDAK TAAT" pada kolom Estimasi}}{110} = \frac{48}{110} = 0,43$

Selanjutnya berdasarkan masing-masing kecamatan (Mdn Polonia, Mdn Maimun, Mdn Tuntungan, Mdn Johor, Mdn Selayang, Mdn Amplas, Mdn Helvetia, Mdn Simalingkar, Mdn Area, Mdn Petisah, Mdn Sunggal, dan Mdn Kota), hitung jumlah kasusnya. yang sama untuk setiap atribut kelas Estimasi (TAAT, TIDAK TAAT). Kemudian jumlah kasus tersebut dibagi dengan jumlah atribut tiap kecamatan.

- $P(\text{Kecamatan} = \text{"Mdn Polonia"} | \text{Class Estimasi} = \text{"TAAT"}) = 1/5 = 0,2$
- $P(\text{Kecamatan} = \text{"Mdn Polonia"} | \text{Class Estimasi} = \text{"TIDAK TAAT"}) = 4/5 = 0,8$
- $P(\text{Kecamatan} = \text{"Mdn Maimun"} | \text{Class Estimasi} = \text{"TAAT"}) = 5/6 = 0,83$
- $P(\text{Kecamatan} = \text{"Mdn Maimun"} | \text{Class Estimasi} = \text{"TIDAK TAAT"}) = 1/6 = 0,16$
- $P(\text{Kecamatan} = \text{"Mdn Tuntungan"} | \text{Class Estimasi} = \text{"TAAT"}) = 9/18 = 0,5$
- $P(\text{Kecamatan} = \text{"Mdn Tuntungan"} | \text{Class Estimasi} = \text{"TIDAK TAAT"}) = 9/18 = 0,5$
- $P(\text{Kecamatan} = \text{"Mdn Johor"} | \text{Class Estimasi} = \text{"TAAT"}) = 20/27 = 0,74$
- $P(\text{Kecamatan} = \text{"Mdn Johor"} | \text{Class Estimasi} = \text{"TIDAK TAAT"}) = 7/27 = 0,25$
- $P(\text{Kecamatan} = \text{"Mdn Selayang"} | \text{Class Estimasi} = \text{"TAAT"}) = 5/10 = 0,5$
- $P(\text{Kecamatan} = \text{"Mdn Selayang"} | \text{Class Estimasi} = \text{"TIDAK TAAT"}) = 5/10 = 0,5$
- $P(\text{Kecamatan} = \text{"Mdn Denai"} | \text{Class Estimasi} = \text{"TAAT"}) = 5/11 = 0,45$
- $P(\text{Kecamatan} = \text{"Mdn Denai"} | \text{Class Estimasi} = \text{"TIDAK TAAT"}) = 6/11 = 0,54$
- $P(\text{Kecamatan} = \text{"Mdn Amplas"} | \text{Class Estimasi} = \text{"TAAT"}) = 12/19 = 0,63$
- $P(\text{Kecamatan} = \text{"Mdn Amplas"} | \text{Class Estimasi} = \text{"TIDAK TAAT"}) = 7/19 = 0,36$
- $P(\text{Kecamatan} = \text{"Mdn Sunggal"} | \text{Class Estimasi} = \text{"TAAT"}) = 3/5 = 0,6$
- $P(\text{Kecamatan} = \text{"Mdn Sunggal"} | \text{Class Estimasi} = \text{"TIDAK TAAT"}) = 2/5 = 0,4$
- $P(\text{Kecamatan} = \text{"Mdn Kota"} | \text{Class Estimasi} = \text{"TAAT"}) = 2/9 = 0,22$
- $P(\text{Kecamatan} = \text{"Mdn Kota"} | \text{Class Estimasi} = \text{"TIDAK TAAT"}) = 7/9 = 0,77$

Selanjutnya kalikan semua hasil *variable* yang didapat dari hasil $P(\text{Kecamatan} | \text{Class Estimasi})$ dengan sesame kelas estimasinya ($P1$ dan $P2$), adalah sebagai berikut:

- a. Untuk semua atribut *Class Estimasi* = "TAAT"

$$P(E|Class\ Estimasi="TAAT") = 0,2 \times 0,83 \times 0,5 \times 0,74 \times 0,5 \times 0,45 \times 0,63 \times 0,6 \times 0,22 = 0,0011$$
- b. Untuk semua atribut *Class Estimasi* = "TIDAK TAAT"

$$P(E|Class\ Estimasi="TIDAK\ TAAT") = 0,8 \times 0,16 \times 0,5 \times 0,25 \times 0,5 \times 0,54 \times 0,36 \times 0,4 \times 0,77 = 0,0004$$

Selanjutnya kalikan nilai $P(E|Class\ Estimate)$ dengan hasil perkalian seluruh variabel ($P1$ dan $P2$):

- a. *Prior Probabilitas* dikalikan dengan semua kualitas Estimasi Kelas = "TAAT"

$$P1(Class\ Estimasi="TAAT") \times P(E|Class\ Estimasi="TAAT")$$

$$= 0,56 \times 0,0011$$

$$= 0,000616$$
- b. *Prior Probabilitas* dikalikan dengan semua kualitas Estimasi Kelas = "TIDAK TAAT"

$$P2(Class\ Estimasi="TIDAK\ TAAT") \times P(E|Class\ Estimasi="TIDAK\ TAAT")$$

$$= 0,43 \times 0,0004$$

$$= 0,000172$$

$P1(Estimation\ Class="TAAT") > P(E|Estimation\ Class="TAAT")$ adalah perbandingan yang dihasilkan dari hasil komputasi $P(Estimation\ Class) \times P(E|Estimation\ Class)$.

Kesimpulan Kelas Estimasi "TAAT" mempunyai nilai yang lebih tinggi dibandingkan dengan Kelas Estimasi "TIDAK TAAT", berdasarkan perhitungan antara perkalian kedua kelas estimasi tersebut.

Penerapan Algoritma *Naive Bayes* Pada *RapidMiner*

Dalam penggunaan *Rapidminer* hingga algoritma *Naive Bayes* untuk mengidentifikasi kemungkinan kepatuhan wajib pajak. Potensi kepatuhan wajib pajak dan temuan perkiraan *rapidminer* ditunjukkan pada Gambar 6 di bawah ini.

Row No.	Estimasi	No Mesin	Kecamatan	Tgl Jatuh Te...	Tgl Bayar	Estimasi (h...
1	Taat	78523	Midn Sunggal	May 10, 2023	May 2, 2023	8
2	Tidak Taat	27815	Midn Polonia	Feb 2, 2023	May 2, 2023	-89
3	Taat	21278	Midn Maimun	Jun 7, 2023	May 2, 2023	36
4	Tidak Taat	59817	Midn Tuntung...	Sep 4, 2022	May 2, 2023	-240
5	Taat	14275	Midn Tuntung...	May 7, 2023	May 2, 2023	5
6	Tidak Taat	18532	Midn Tuntung...	Apr 24, 2023	May 2, 2023	-8
7	Taat	96980	Midn Maimun	May 5, 2023	May 3, 2023	2
8	Tidak Taat	18150	Midn Johor	Apr 17, 2022	May 3, 2023	-381
9	Taat	31502	Midn Tuntung...	May 3, 2023	May 3, 2023	0
10	Taat	61289	Midn Johor	May 8, 2023	May 3, 2023	5
11	Taat	37493	Midn Selayang	May 4, 2023	May 3, 2023	1
12	Taat	69572	Midn Johor	May 24, 2023	May 4, 2023	20
13	Tidak Taat	80052	Midn Polonia	Apr 16, 2023	May 4, 2023	-18
14	Taat	78506	Midn Johor	May 9, 2023	May 4, 2023	5

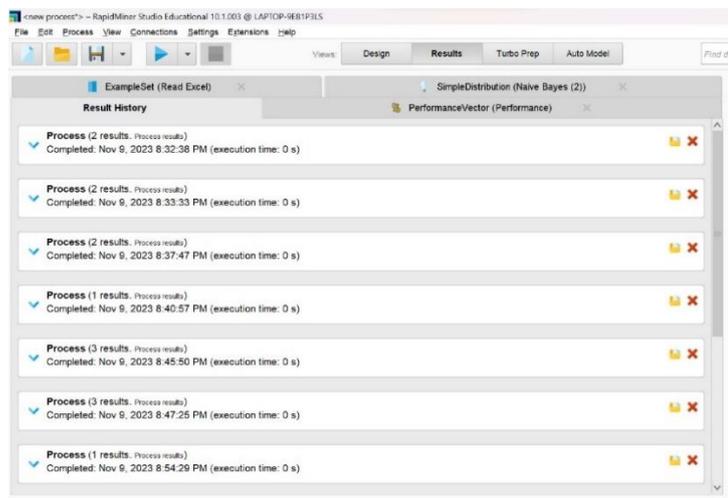
Gambar 6. Hasil Potensi

3.5 Evaluation

Untuk membantu pengguna dalam memahami polanya, pola akhir dari proses penambangan data kini akan ditampilkan secara visual.

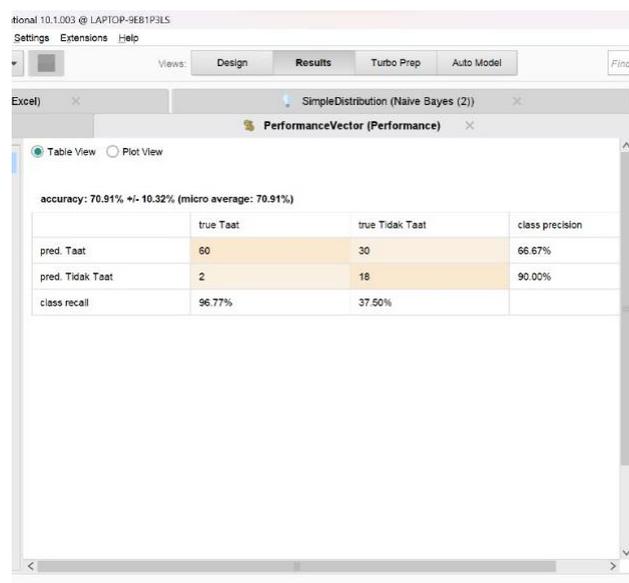
1) Akurasi Prediksi

Metode *Naive Bayes* membutuhkan waktu nol detik (0) *second* untuk dieksekusi dalam pengujian yang dilakukan dengan alat *RapidMiner*. Hasil eksekusi ditampilkan pada Gambar 7 di bawah ini.



Gambar 7. Hasil Eksekusi

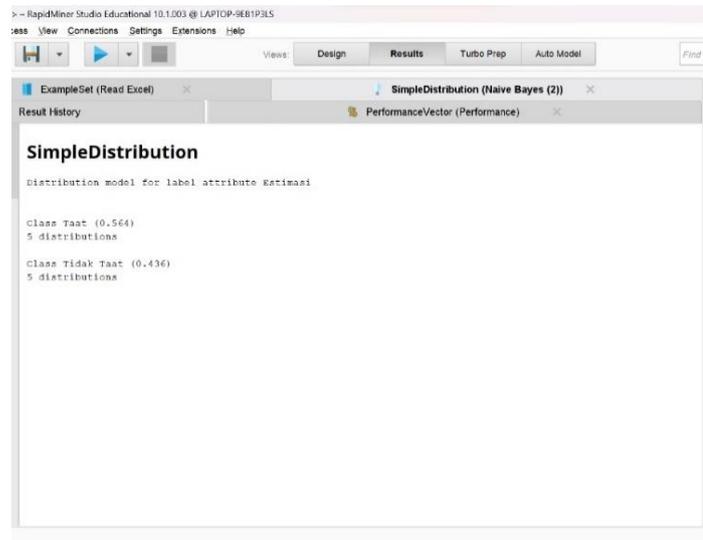
Nilai akurasi hasil model *Naive Bayes* menunjukkan kekuatan nilai akurasi yang dihasilkan oleh algoritma *Naive Bayes*. Model tersebut dapat digunakan untuk mengklasifikasikan potensi kepatuhan wajib pajak, dibuktikan dengan hasil akurasi perhitungan (70,91%). Hasil akurasi ditunjukkan pada Gambar 8. Menyajikan hasil akurasi.

The image shows a screenshot of the 'PerformanceVector (Performance)' window in RapidMiner. The window displays a confusion matrix and accuracy metrics for a 'SimpleDistribution (Naive Bayes (2))' model. The metrics are: accuracy: 70.91% +/- 10.32% (micro average: 70.91%). The confusion matrix is as follows:

	true Taat	true Tidak Taat	class precision
pred. Taat	60	30	66.67%
pred. Tidak Taat	2	18	90.00%
class recall	96.77%	37.50%	

The interface also shows a 'Table View' and 'Plot View' toggle, and a 'Find d...' search box.

Gambar 8. Hasil Akurasi



Gambar 9. Model Distribusi

Gambar 9 menunjukkan bahwa untuk data PKB Samsat Medan Selatan kelas TIDAK TAAT lebih kecil dibandingkan kelas TAAT yang mempunyai nilai sebesar 0,564. Berikut model distribusi label atribut kelas estimasi:

Kelas TAAT : 5 Distribusi
Kelas TIDAK TAAT : 5 Distribusi

2) Hasil pengujian

Penelitian yang dilakukan menghasilkan pengetahuan dan pemahaman baru tentang wajib pajak kecamatan mana yang mungkin tidak patuh. Tahap *Knowledge Discovery Database* (KDD) yang menggunakan data PKB Samsat Medan Selatan per Mei 2023 telah mengidentifikasi mayoritas wajib pajak yang patuh. Metode data mining juga membantu dalam mengidentifikasi wajib pajak mana yang patuh membayar iuran PKB. Tujuan dari penelitian ini adalah untuk menghasilkan data baru mengenai kepatuhan wajib pajak, yang pada akhirnya dapat menimbulkan ketidakakuratan.

Informasi dihasilkan berdasarkan perhitungan data PKB Samsat Medan Selatan bulan Mei 2023 menggunakan algoritma *Naive Bayes* dengan menerapkan teknik klasifikasi dan tingkat akurasi. Hal ini menunjukkan kelas TAAT dengan probabilitas *multiple prior* sebesar 0,56, sedangkan TIDAK TAAT dengan probabilitas prior sebesar 0,43. Berdasarkan perhitungan alat rapidminer kecamatan, yang TIDAK TAAT berada di kecamatan Mdn Tuntungan dari beberapa wajib pajak dengan akurasi sebesar 70,91%. Hal ini dikarenakan adanya kemungkinan ketidaktaatan akibat jumlah PKB yang besar sehingga mengakibatkan jumlah tagihan yang banyak dan menyulitkan masyarakat dalam membayar PKB tahunan.

3) Hasil Analisis

Berdasarkan prosedur penghitungan data mining secara manual dan penggunaan alat rapidminer, diketahui bahwa seluruh wajib pajak Samsat Kecamatan Medan Selatan telah memenuhi kewajiban membayar iuran PKB. Diketahui, 18 kecamatan, termasuk wajib pajak tertentu di Kecamatan Mdn Tuntungan dengan 18 data, diduga tidak patuh terhadap undang-undang. Temuan penelitian ini diharapkan dapat menjadi sumber bagi organisasi pemerintah yang

akan memberi masukan kepada kecamatan-kecamatan tersebut mengenai cara mematuhi persyaratan iuran PKB.

Tabel 1 menampilkan data wajib pajak beserta perhitungan kepatuhan TAAT. Tabel 1 menyajikan jumlah kelas estimasi TAAT dalam dataset yang digunakan.

Tabel 1. Hasil Taat

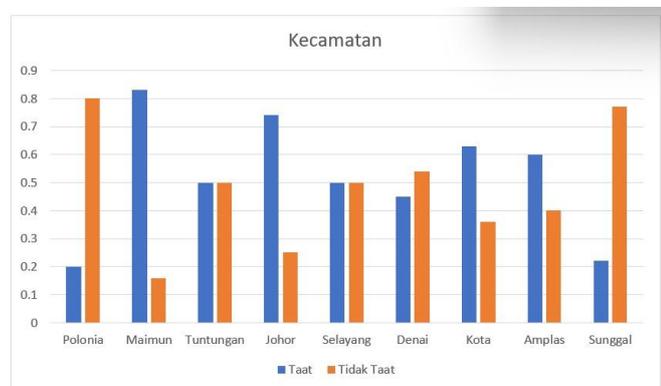
Estimasi	Kecamatan	Jumlah	Prior Probability
TAAT	Mdn Polonia	1	0,2
	Mdn Maimun	5	0,83
	Mdn Tuntungan	9	0,5
	Mdn Johor	20	0,74
	Mdn Selayang	5	0,5
	Mdn Denai	5	0,45
	Mdn Kota	2	0,63
	Mdn Amplas	12	0,6
	Sunggal	3	0,22

Tabel 2. Hasil Tidak Taat

Estimasi	Kecamatan	Jumlah	Prior Probability
TIDAK TAAT	Mdn Polonia	4	0,8
	Mdn Maimun	1	0,16
	Mdn Tuntungan	9	0,5
	Mdn Johor	7	0,25
	Mdn Selayang	5	0,5
	Mdn Denai	6	0,54
	Mdn Kota	7	0,36
	Mdn Amplas	7	0,4
	Sunggal	2	0,77

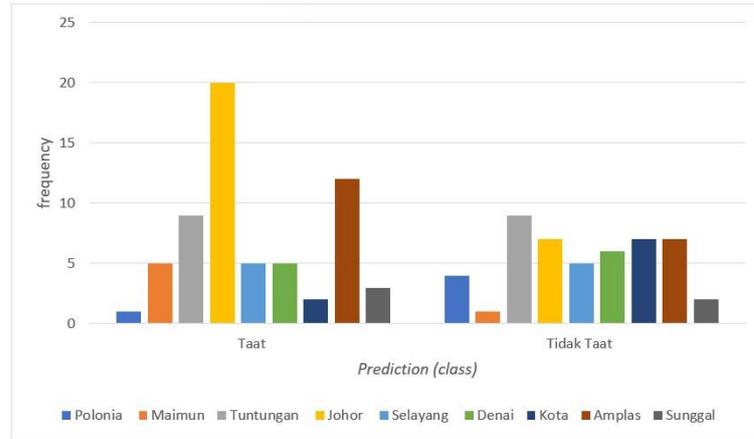
4) Knowledge Representation

Pada titik ini, pengguna disuguhkan pengetahuan yang telah ditemukan melalui berbagai alat bantu visual, antara lain grafik, *histogram*, dan diagram *Pareto*. Anda dapat melihat visualisasinya pada grafik 1, 2, dan 3.



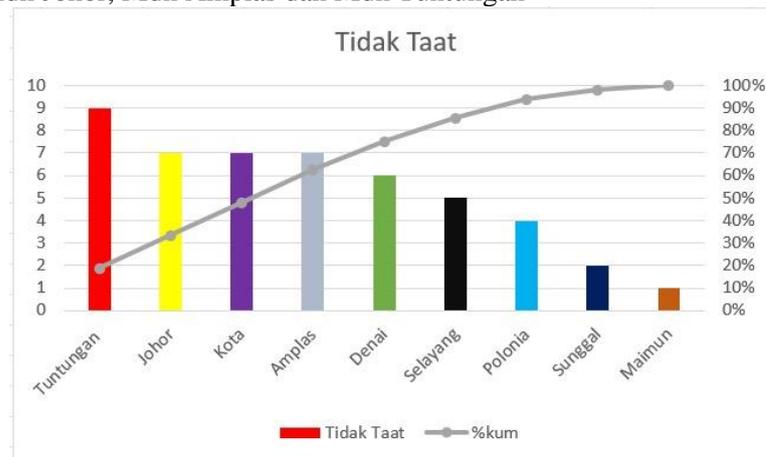
Grafik 1. Chart

Data Wajib Pajak dengan nilai *Prior Probability* TAAT dan TIDAK TAAT ditampilkan pada Grafik 1.



Grafik 2. Histogram

Pada grafik 2 memperlihatkan *Class* Estimasi TAAT dan TIDAK TAAT tertinggi terdapat pada kecamatan Mdn Johor, Mdn Amplas dan Mdn Tuntungan



Grafik 3. Pareto

Pada grafik 3 memperlihatkan *Class* tidak tepat waktu tertinggi terdapat pada kecamatan Mdn Tuntungan.

4. KESIMPULAN DAN SARAN

Tersisa 118 WP setelah diseleksi, berdasarkan perhitungan data mining yang menggunakan metode naïve Bayes dalam pendekatan klasifikasi menggunakan data yang berisi hingga 300 WP. Berdasarkan penelitian dapat disimpulkan bahwa taksiran golongan pembayaran pada wajib pajak kendaraan bermotor yang TAAT lebih besar dibandingkan taksiran golongan pembayaran pada wajib pajak kendaraan bermotor yang TIDAK TAAT. Pembersihan data dilakukan karena masih banyak data yang hilang, dan data akhir digunakan untuk menguji total 110 wajib pajak. Kecamatan Mdn Tuntungan mungkin tidak sesuai dengan data Samsat Medan Selatan, berdasarkan temuan observasi data sampel yang dikumpulkan disana dan metode klasifikasi menggunakan algoritma naïve Bayes sebanyak 0,436, dengan akurasi 70,91%.

DAFTAR PUSTAKA

- [1] Galla Sri Rosiana Bandung and Asmapane Set, "Analisis kepatuhan wajib pajak kendaraan bermotor terhadap penerimaan pajak kendaraan bermotor sebelum dan selama pandemi covid-19 di provinsi Kalimantan Timur," *Jurnal Ekonomi, Manajemen dan Akuntansi*, vol. 25, no. 2, 2023.
- [2] B. Ahmad, B. Romadhoni, M. Adil, F. Ekonomi, and D. Bisnis, "Efektivitas Pemungutan Pajak Kendaraan Bermotor," 2020.
- [3] E. Purwandai, "Implikasi Undang-Undang Nomor 28 Tahun 2009 Tentang Pajak Daerah Dan Retribusi Daerah Terhadap Penerimaan Bphtb Dan Pelayanan Pertanahan," 2014.
- [4] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Islam Indragiri)," 2018.
- [5] H. Derajad Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," *JURNAL INFORMATIKA*, vol. 7, no. 1, 2020, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [6] E. Martantoh and N. Yanih, "Implementasi Metode Naïve Bayes Untuk Klasifikasi Karakteristik Kepribadian Siswa Di Sekolah MTS Darussa'adah Menggunakan PHP MySQL Implementation of Naive Bayes Method for Classification of Student's Personality Characteristics at MTS Darussa'adah School Using PHP Mysql," 2022.
- [7] P. Kepatuhan Wajib Pajak Bumi Dan Bangunan Oleh *et al.*, "Penerapan Data Mining Menggunakan Metode Teknik Classification Untuk Melihat Potensi Kepatuhan Wajib Pajak Bumi Dan Bangunan," 2022. [Online]. Available: www.pajak.go.id
- [8] R. Hisbullah, M. S. Hasibuan, and P. Korespondensi, "Pendekatan Bayes-Hdss Dalam Menentukan Status Pantauan Gizi Balita", doi: 10.25126/jtiik.2023107437.
- [9] M. Imam and D. Udariansyah, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Farmasi Pada Rsud Talang Ubi Kabupaten Pal," *Bina Darma Conference on Computer Science*, [Online]. Available: <https://kbbi.web.id/data>
- [10] M. Rosadi, D. A. Nurhasanah, and M. S. Hasibuan, "Clustering Panjang Ruas Jalan di BBPJN Sumut Menggunakan Algoritma K-Means Clustering The Length Of Road Sections In BBPJN Sumut Using The K-Means Algorithm," *JOURNAL OF COMPUTER SCIENCE AND INFORMATICS ENGINEERING (CoSIE)*, vol. 02, no. 1, p. 2023, 2023, [Online]. Available: <http://creativecommons.org/licenses/by-sa/4.0/>