

## Klasifikasi Masyarakat Wajib Pajak Di Kabupaten Simalungun Menggunakan Metode Naive Bayes

Idris Hanafia Lubis<sup>1</sup>, Poningsih<sup>2</sup>, Ilham Sahputra Saragih<sup>3</sup>

<sup>1</sup>Mahasiswa Program Studi Sistem Informasi STIKOM Tunas Bangsa, Pematangsiantar

<sup>2</sup>AMIK Tunas Bangsa, Pematangsiantar

<sup>3</sup>STIKOM Tunas Bangsa, Pematangsiantar

*Idrislubis95@gmail.com*

### **Abstract**

*Regional tax is a tax set by the regional government with regional regulations, the collection of which is carried out by the regional government and the subsequent results are used to finance the expenditure of regional governments in implementing governance and development in the region. Poor knowledge and understanding of taxes received by the community in paying taxes. The quality of good tax knowledge will greatly affect the smooth running of taxpayers in fulfilling their tax obligations. In this study, the method used in measuring the free or not of taxpayers in Simalungun District uses the method of using Naive Bayes. The parameters used are Modernization of Tax Administration System, Taxation Information Session, Taxation of Awareness, Tax Sanctions, Taxpayer Compliance. The data used in this study provides questionnaires to people who have a type of business in Simalungun District specifically in Java. It is expected that the results of this study can help the government specifically the Regional Revenue Service to understand the understanding and responsibility of the community in fulfilling their tax obligations.*

**Keyword:** Society, Taxpayers, Data Mining, Naive Bayes.

### **Abstrak**

Pajak daerah merupakan pajak yang ditetapkan oleh pemerintah daerah dengan peraturan daerah, yang wewenang pemungutannya dilaksanakan oleh pemerintah daerah dan hasilnya nanti digunakan untuk membiayai pengeluaran pemerintah daerah dalam melaksanakan penyelenggaraan pemerintahan dan pembangunan di daerah. Pengetahuan dan pemahaman yang kurang tentang pajak mengakibatkan kurangnya kesadaran masyarakat dalam membayar pajak. Kualitas pengetahuan pajak yang baik akan sangat mempengaruhi kelancaran wajib pajak dalam memenuhi kewajiban perpajakannya. Dalam penelitian ini, metode yang digunakan dalam mengukur lancar tidaknya masyarakat wajib pajak di Kabupaten Simalungun adalah dengan menggunakan metode *Naive Bayes*. Parameter yang digunakan yaitu Modernisasi Sistem Administrasi Perpajakan, Sosialisasi Perpajakan, Kesadaran Perpajakan, Sanksi Perpajakan, Kepatuhan Wajib Pajak. Data yang digunakan dalam penelitian ini dengan memberikan kuesioner kepada masyarakat yang memiliki jenis usaha di Kabupaten Simalungun khususnya Tanah Jawa. Diharapkan hasil dari penelitian ini dapat membantu pemerintah khususnya Dinas Pendapatan Daerah untuk mengetahui pemahaman dan lancar tidaknya masyarakat dalam memenuhi kewajiban perpajakannya.

**Kata kunci:** Masyarakat, Wajib Pajak, Data Mining, Naive Bayes

## **1. PENDAHULUAN**

Indonesia merupakan negara kesatuan yang menganut asas desentralisasi dalam rangka penyelenggaraan pemerintahan dengan memberikan kesempatan dan keleluasaan kepada Daerah dalam melaksanakan Otonomi Daerah. Otonomi daerah yang berlandaskan Undang-undang Nomor 23 Tahun 2014 tentang Pemerintah Daerah, membahas tentang hak, wewenang dan kewajiban daerah otonom dalam mengatur urusan pemerintahan dan

kepentingan masyarakat setempat sesuai dengan peraturan perundang-undangan. Sumber keuangan daerah salah satu yang dimiliki dan dikelola oleh Pemerintah Daerah adalah Pendapatan Asli Daerah (PAD) yang merupakan pendapatan diperoleh Daerah yang dipungut berdasarkan Peraturan Daerah sesuai dengan peraturan perundang-undangan. Pajak adalah pungutan terhadap masyarakat oleh negara berdasarkan undang-undang yang bersifat memaksa, dan terutang yang wajib dibayar dengan tidak mendapat imbalan secara langsung, yang hasilnya digunakan untuk membiayai pengeluaran negara dalam penyelenggaraan pemerintahan dan pembangunan.

Kepatuhan wajib pajak yaitu dimana wajib pajak memenuhi kewajiban perpajakannya dan melaksanakan hak perpajakan dengan baik dan benar sesuai dengan peraturan dan undang-undang pajak yang berlaku. Pengetahuan dan pemahaman yang kurang tentang pajak mengakibatkan kurangnya kesadaran masyarakat dalam membayar pajak. Masyarakat kurang tertarik akan membayar pajak karena tidak adanya insentif atau timbal balik secara langsung dari negara untuk mereka. Kualitas pengetahuan pajak yang baik akan sangat mempengaruhi kepatuhan wajib pajak dalam memenuhi kewajiban perpajakannya. Semakin tinggi tingkat pengetahuan dan pemahaman wajib pajak, maka semakin mudah pula bagi mereka untuk memahami peraturan perpajakan dan semakin mudah pula untuk memenuhi kewajiban perpajakannya. Salah satunya prasangka buruk masyarakat terhadap petugas pajak harus dirubah menjadi prasangka yang baik, untuk merubah hal tersebut tentu harus menciptakan pelayanan yang memuaskan dan berkualitas. Masyarakat akan membayar pajak dari penghasilan yang diterimanya apabila mereka merasakan pelayanan publik sebanding dengan pembayaran pajaknya, adanya perlakuan yang adil dari pemerintah serta proses perpajakan yang jelas dari pemerintah.

Ada 2 jurnal yang menjadi referensi dalam penelitian ini yaitu pertama, penelitian yang dilakukan [1] yang berjudul “Pengaruh Pemahaman Dan Pengetahuan Wajib Pajak Tentang Peraturan Perpajakan, Kesadaran Wajib Pajak, Kualitas Pelayanan, Dan Sanksi Perpajakan Terhadap Kepatuhan Wajib Pajak Kendaraan Bermotor (Studi Samsat Kota Malang)” membahas tentang kesadaran wajib pajak merupakan sikap wajib pajak yang telah memahami dan mau melaksanakan kewajibannya untuk membayar pajak dan telah melaporkan semua penghasilannya tanpa ada yang disembunyikan sesuai dengan ketentuan yang berlaku. Dan kedua, *Naive Bayes* merupakan suatu teknik prediksi yang mencari probabilitas sederhana berdasarkan pada teorema bayes dengan asumsi independensi yang kuat pada jurnal [1] yang berjudul “Penerapan Metode Naive Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra”. Kesimpulan dari kedua jurnal tersebut adalah metode *Naive Bayes* menggabungkan probabilitas masyarakat wajib pajak dengan kriteria pembayaran wajib pajak untuk mengetahui pemahaman masyarakat dalam memenuhi kewajiban perpajakannya. Semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting. Diharapkan penelitian ini dapat memberikan masukan kepada pemerintahan daerah dalam mengetahui jumlah Masyarakat Wajib Pajak di Kabupaten Simalungun.

## **2. METODOLOGI PENELITIAN**

### **2.1. Data Mining**

*Data Mining* merupakan proses maupun kegiatan mengumpulkan sekumpulan data dalam jumlah yang besar untuk di ekstraksi sehingga menjadi informasi yang dapat digunakan [2][3].

### **2.2. Klasifikasi**

Klasifikasi adalah suatu pekerjaan menentukan objek data yang selanjutnya dimasukan kedalam kelas tertentu berdasarkan sejumlah kelas yang ada. Dua pekerjaan utama yang dilakukan dalam klasifikasi yaitu membangun sebuah model sederhana, melakukan pengenalan /klasifikasi/ prediksi pada suatu objek data lain agar dapat diketahui pada kelas mana objek data tersebut disimpan [4][5].

### 2.3. Pajak

Pajak adalah iuran kepada Negara yang terutang oleh yang wajib membayarnya menurut peraturan - peraturan atau dapat dipaksakan, dengan tidak mendapat prestasi kembali, yang langsung dapat ditunjukkan dan yang gunanya adalah untuk membiayai pengeluaran umum berhubungan dengan tugas Negara [4].

### 2.4. Naive Bayes

*Naive Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan *estimasi* parameter yang diperlukan dalam proses pengklasifikasian. *Naive Bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan [2][6][7]. Untuk menyelesaikan metode *Naive Bayes* dapat dilakukan dengan persamaan-persamaan sebagai berikut:

$$P(H | X) = \frac{P(X | H) * P(H)}{P(X)} \dots\dots\dots(1)$$

Keterangan :

X : Data dengan *class* yang belum diketahui

H : Hipotesis data merupakan suatu *class* spesifik

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

Penjabaran lebih lanjut rumus *Naive Bayes* tersebut dilakukan dengan menjabarkan secara terperinci ( $C|X_1, \dots, X_n$ ) menggunakan aturan perkalian sebagai berikut.

$$\begin{aligned} P(C|x_1, \dots, x_n) &= P(C)P(x_1, \dots, x_n | C) \\ &= P(C)P(X_1|C)P(X_2, \dots, X_n|C, X_1) \\ &= P(C)P(X_1|C)P(X_2|C, X_1)P(X_3, \dots, X_n|C, X_1, X_2)P(X_1|C) \\ &\quad P(X_2|C, X_1)P(X_3|C, X_1, X_2)P(X_4, \dots, X_n|C, X_1, X_2, X_3)P(C) \\ &= P(X_1|C)P(X_2|C, X_1)P(X_3|C, X_1, X_2) \dots \\ &\quad P(X_n|C, X_1, X_2, X_3, \dots, X_{n-1}) \dots\dots\dots(2) \end{aligned}$$

Jika semakin banyak faktor-faktor yang semakin kompleks yang berpengaruh terhadap nilai probabilitas, maka semakin tidak mungkin untuk menghitung nilai tersebut satu persatu. Proses perhitungan akan semakin susah untuk dilakukan, maka disinilah digunakan asumsi independensi yang sangat tinggi, bahwa masing-masing atribut dapat saling bebas. Dengan asumsi tersebut, diperlukan persamaan 3 :

$$P(X_i | X_j) = \frac{P(X_i)P(X_j)}{P(X_j)} = \frac{P(X_i \cap X_j)}{P(X_j)} = P(X_i)$$

Untuk  $i \neq j$ , sehingga

$$P(X_i | C, X_j) = P(X_i | C) \dots\dots\dots(3)$$

Dari persamaan 3 tersebut dapat di ambil kesimpulan bahwa asumsi independensi membuat syarat perhitungan menjadi lebih sederhana. Selanjutnya penjabaran ( $P(C|X_1, \dots, X_n)$ ) dapat disederhanakan menjadi persamaan 4 :

$$P(X_2|C)P(X_3|C) \dots P(C|X_1, \dots, X_n) = P(X_1|C) = \prod_{i=1}^n p(X_i | C) \dots\dots(4)$$

Keterangan :

$\prod_{i=1}^n P(X_i | C)$  = perkalian ranting antar atribut.

Persamaan 5 merupakan teorema *bayes* yang kemudian akan digunakan untuk melakukan perhitungan klasifikasi. Untuk klasifikasi dengan data *continue* atau data angka menggunakan rumus distribusi Gaussian dengan 2 parameter : mean  $\mu$  dan varian  $\sigma$ :

$$P(X_i = X_i | C = c_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} \exp \frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \dots\dots\dots(5)$$

Keterangan :

- P : Peluang
- $X_i$  : Atribut ke i
- $X_j$  : Nilai atribut ke i
- C : Kelas yang dicari
- $C_i$  : Sub kelas Y yang dicari
- $\mu$  : Menyatakan rata-rata dari seluruh atribut
- $\sigma$  : Deviasi standar, menyatakan varian dari seluruh atribut.

Dalam metode *Naive Bayes* diperlukan data latih dan data uji yang ingin diklasifikasikan. Semakin banyak data latih yang yang dilibatkan, semakin baik hasil yang prediksi yang diberikan. Menghitung  $P(C_i)$  yang merupakan probabilitas prior untuk setiap sub kelas C yang akan dihasilkan menggunakan persamaan 6 :

$$P(C_i) = \frac{S_i}{s} \dots\dots\dots(6)$$

Si adalah jumlah data training dari kategori  $C_i$ , dan s adalah jumlah total data training. Menghitung  $P(X_i|C_i)$  yang merupakan probabilitas posterior  $X_i$  dengan syarat C menggunakan persamaan 4.

### 3. HASIL DAN PEMBAHASAN

Hasil dan pembahasan pada bab ini disajikan sesuai penelitian yang dilakukan. Data yang digunakan dalam penelitian ini adalah data bayar pajak berdasarkan pertanyaan kuesioner yang telah diberikan. Perangkat lunak yang digunakan yaitu *rapidminer 5.3* Kumpulan data yang diperoleh digunakan sebagai data masukan dalam membuat model aturan menggunakan algoritma *Naive Bayes* menggunakan *software rapidminer*. Dalam mengetahui lancar tidaknya pembayaran pajak di Kabupaten Simalungun khususnya Tanah Jawa. Penulis akan menampilkan gambaran model aturan dalam mengklasifikasi masyarakat wajib pajak di Kabupaten Simalungun menggunakan algoritma *Naive Bayes*.

#### 3.1. Dataset Penelitian

Metode *Naive Bayes* digunakan dalam pemodelan data yang akan ditampilkan. Data yang akan diolah ada *data training* dan *data testing*. Kriteria data yang telah ditentukan dianalisis menggunakan perangkat lunak *rapidminer 5.3*. Berikut merupakan data yang diolah pada penelitian ini dapat dilihat sebagai berikut:

**Tabel 1.** Data yang dikumpulkan

Responden	Modernisasi Sistem Administrasi Perpajakan	Sosialisasi Perpajakan	Kesadaran Perpajakan	Sanksi Perpajakan	Kepatuhan Wajib Pajak	Hasil
A1	KS	KS	KS	KS	KS	Tidak Lancar
A2	SS	SS	SS	SS	SS	Lancar
A3	SS	KS	KS	S	KS	Tidak Lancar
A4	TS	KS	S	KS	KS	Tidak Lancar
A5	KS	S	SS	S	S	Lancar
A6	SS	S	S	KS	S	Lancar
A7	KS	S	S	S	S	Lancar
A8	KS	S	KS	S	S	Lancar
A9	KS	KS	KS	KS	S	Tidak Lancar
A10	S	KS	S	KS	S	Lancar
A11	KS	S	S	S	S	Lancar
A12	S	S	S	KS	S	Lancar

Responden	Modernisasi Sistem Administrasi Perpajakan	Sosialisasi Perpajakan	Kesadaran Perpajakan	Sanksi Perpajakan	Kepatuhan Wajib Pajak	Hasil
A13	KS	KS	KS	KS	KS	Tidak Lancar
A14	S	S	S	S	S	Lancar
A15	KS	KS	SS	KS	KS	Lancar
A16	SS	KS	SS	S	KS	Lancar
A17	SS	S	KS	KS	S	Lancar
A18	KS	KS	SS	KS	KS	Tidak Lancar
A19	S	KS	SS	TS	S	Lancar
A20	KS	TS	SS	KS	KS	Tidak Lancar

### 3.2. Perhitungan Algoritma Naive Bayes

Perhitungan probabilitas prior kemungkinan Lancar dalam menentukan masyarakat wajib pajak dapat dilihat pada persamaan (6), yaitu :

$$P(H|X) = \frac{P(H)}{P(X)} = \frac{13}{20} = 0,65$$

Sedangkan perhitungan probabilitas tidak Lancar yaitu :

$$P(H|X) = \frac{P(H)}{P(X)} = \frac{7}{20} = 0,35$$

Setelah probabilitas dari masing-masing prior telah diketahui, selanjutnya penulis menghitung masing-masing probabilitas dari setiap kriteria yang digunakan. Kriteria yang digunakan penulis yaitu Modernisasi Sistem Administrasi Perpajakan, Sosialisasi Perpajakan, Kesadaran Perpajakan, Sanksi Perpajakan, Kepatuhan Wajib Pajak. Dalam menentukan probabilitas setiap kriteria, penulis menghitung bagian-bagian yang terdapat pada setiap kriteria, pada penelitian ini bagian-bagian yang terdapat dalam setiap kriteria menggunakan *skala linker 5* yaitu SS (Sangat Setuju), S (Setuju), KS (Kurang Setuju), TS (Tidak Setuju), dan STS (Sangat Tidak Setuju). Sehingga perhitungan probabilitas masing-masing kriteria dapat dilihat pada beberapa tabel-tabel berikut.

Untuk menghitung probabilitas kemungkinan dari kriteria Modernisasi Sistem Administrasi Perpajakan dan Sosialisasi Perpajakan dapat dilihat pada tabel 2:

**Tabel 2.** Probabilitas Kriteria Modernisasi Sistem Administrasi Perpajakan

Modernisasi Sistem Administrasi Perpajakan	Jumlah Kejadian Dipilih		Probabilitas	
	Lancar	Tidak Lancar	Lancar	Tidak Lancar
SS	3	0	0,23	0,00
S	5	1	0,38	0,14
KS	5	5	0,38	0,71
TS	0	1	0,00	0,14
STS	0	0	0,00	0,00
Jumlah	13	7	1,00	1,00

Untuk menghitung probabilitas kemungkinan dari kriteria Modernisasi Sistem Administrasi Perpajakan dan Sosialisasi Perpajakan dapat dilihat pada tabel 3 :

**Tabel 3.** Probabilitas Kriteria Sosialisasi Perpajakan

Sosialisasi Perpajakan	Jumlah Kejadian Dipilih		Probabilitas	
	Lancar	Tidak Lancar	Lancar	Tidak Lancar
SS	1	0	0,08	0,00
S	8	0	0,62	0,00
KS	4	6	0,31	0,86
TS	0	1	0,00	0,14
STS	0	0	0,00	0,00
Jumlah	13	7	1,00	1,00

Untuk menghitung probabilitas kemungkinan dari kriteria Kesadaran Perpajakan dan Sanksi Perpajakan dapat dilihat pada tabel 4:

**Tabel 4.** Probabilitas Kriteria Kesadaran Perpajakan

Kesadaran Perpajakan	Jumlah Kejadian Dipilih		Probabilitas	
	Lancar	Tidak Lancar	Lancar	Tidak Lancar
SS	5	2	0,38	0,29
S	6	1	0,46	0,14
KS	2	4	0,15	0,57
TS	0	0	0,00	0,00
STS	0	0	0,00	0,00
Jumlah	13	7	1,00	1,00

Untuk menghitung probabilitas kemungkinan dari kriteria Kesadaran Perpajakan dan Sanksi Perpajakan dapat dilihat pada tabel 5:

**Tabel 5.** Probabilitas Kriteria Sanksi Perpajakan

Sanksi Perpajakan	Jumlah Kejadian Dipilih		Probabilitas	
	Lancar	Tidak Lancar	Lancar	Tidak Lancar
SS	1	0	0,08	0,00
S	6	1	0,46	0,14
KS	5	6	0,38	0,86
TS	1	0	0,08	0,00
STS	0	0	0,00	0,00
Jumlah	13	7	0,08	1,00

Untuk menghitung probabilitas kemungkinan dari kriteria Kepatuhan Wajib Pajak dapat dilihat pada tabel 6 :

**Tabel 6.** Probabilitas Kriteria Kepatuhan Wajib Pajak

Kepatuhan Wajib Pajak	Jumlah Kejadian Dipilih		Probabilitas	
	Lancar	Tidak Lancar	Lancar	Tidak Lancar
SS	1	0	0,08	0,00
S	10	2	0,77	0,29
KS	2	5	0,15	0,71
TS	0	0	0,00	0,00
STS	0	0	0,00	0,00
Jumlah	13	7	1,00	1,00

Berdasarkan data *training* pada tabel 1 rumus yang digunakan dalam menentukan kelas Lancar dapat dilihat pada persamaan (4). Sehingga untuk menghitung nilai Lancar pada data responden 19 dan 20 adalah sebagai berikut.

$$\begin{aligned}
 P(19 | \text{Lancar}) &= P(\text{Modernisasi Sistem Administrasi Perpajakan} = S | \text{Lancar}) \times \\
 &P(\text{Sosialisasi Perpajakan} = KS | \text{Lancar}) \times P(\text{Kesadaran Perpajakan} = SS | \text{Lancar}) \times \\
 &P(\text{Sanksi Perpajakan} = TS | \text{Lancar}) \times P(\text{Kepatuhan Wajib Pajak} = S | \text{Lancar}) \\
 &= 0,38 \times 0,31 \times 0,38 \times 0,08 \times 0,77 \\
 &= 0,0027
 \end{aligned}$$

$$\begin{aligned}
 P(20 | \text{Lancar}) &= P(\text{Modernisasi Sistem Administrasi Perpajakan} = KS | \text{Lancar}) \times \\
 &P(\text{Sosialisasi Perpajakan} = TS | \text{Lancar}) \times P(\text{Kesadaran Perpajakan} = SS | \text{Lancar}) \times \\
 &P(\text{Sanksi Perpajakan} = KS | \text{Lancar}) \times P(\text{Kepatuhan Wajib Pajak} = KS | \text{Lancar}) \\
 &= 0,38 \times 0,00 \times 0,38 \times 0,38 \times 0,15 \\
 &= 0
 \end{aligned}$$

Sedangkan untuk menghitung nilai Tidak Lancar pada data ke 19 dan 20 rumus yang digunakan sama dengan rumus untuk menentukan nilai Lancar. Sehingga untuk mendapatkan nilai dilakukan sebagai berikut.

$$\begin{aligned}
 P(19 | \text{Tidak Lancar}) &= P(\text{Modernisasi Sistem Administrasi Perpajakan} = S | \text{Lancar}) \times \\
 &P(\text{Sosialisasi Perpajakan} = KS | \text{Lancar}) \times P(\text{Kesadaran Perpajakan} = SS | \text{Lancar}) \times \\
 &P(\text{Sanksi Perpajakan} = TS | \text{Lancar}) \times P(\text{Kepatuhan Wajib Pajak} = S | \text{Lancar}) \\
 &= 0,14 \times 0,86 \times 0,29 \times 0,00 \times 0,29 \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 P(20 | \text{Tidak Lancar}) &= P(\text{Modernisasi Sistem Administrasi Perpajakan} = KS | \\
 &\text{Lancar}) \times P(\text{Sosialisasi Perpajakan} = TS | \text{Lancar}) \times P(\text{Kesadaran Perpajakan} =
 \end{aligned}$$

$$\begin{aligned} \text{SS | Lancar} &\times P(\text{Sanksi Perpajakan} = \text{KS | Lancar}) \times P(\text{Kepatuhan Wajib Pajak} = \\ \text{KS | Lancar}) &= 0,71 \times 0,14 \times 0,29 \times 0,86 \times 0,71 \\ &= 0,0176 \end{aligned}$$

Setelah nilai Lancar dan Tidak Lancar pada data 19 dan 20 telah diketahui. Selanjutnya penulis melakukan perhitungan maksimal masing-masing klasifikasi. Perhitungan data responden 19 untuk menghitung pemaksimalan nilai Lancar yaitu :

$$\begin{aligned} P(\text{Lancar | C}) &= P(\text{Rn | C}) * P(\text{Lancar}) \\ &= P(19 | C) * P(\text{Lancar}) \\ &= 0,0027 \times 0,65 \\ &= 0,0017 \end{aligned}$$

Perhitungan data responden 20 untuk menghitung pemaksimalan nilai Lancar yaitu :

$$\begin{aligned} P(\text{Lancar | C}) &= P(\text{Rn | C}) * P(\text{Lancar}) \\ &= P(20 | C) * P(\text{Lancar}) \\ &= 0 \times 0,35 \\ &= 0 \end{aligned}$$

Sedangkan perhitungan maksimal nilai Tidak Lancar pada data responden 19 yaitu :

$$\begin{aligned} P(\text{Tidak Lancar | C}) &= P(\text{Rn | C}) * P(\text{Tidak Lancar}) \\ &= P(19 | C) * P(\text{Tidak Lancar}) \\ &= 0 \times 0,35 \\ &= 0 \end{aligned}$$

Perhitungan maksimal nilai Tidak Lancar pada data responden 20 yaitu :

$$\begin{aligned} P(\text{Tidak Lancar|C}) &= P(\text{Rn | C}) * P(\text{Tidak Lancar}) \\ &= P(20 | C) * P(\text{Tidak Lancar}) \\ &= 0,0176 \times 0,35 \\ &= 0,0061 \end{aligned}$$

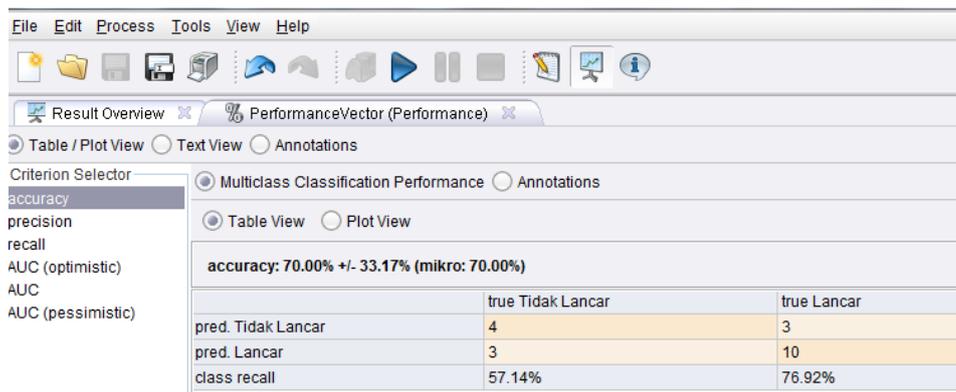
Setelah menghitung pemaksimalan dari nilai Lancar dan Tidak Lancar, selanjutnya penulis membandingkan nilai Lancar dan Tidak Lancar. Sehingga dapat diketahui responden tersebut termasuk kedalam kategori Lancar atau tidak Lancar.

Pada data responden 19 diketahui perhitungan maksimal kelas Lancar = 0,0017 , kelas tidak Lancar = 0. sehingga  $0,0017 > 0$ . Dapat dikatakan data responden 19 data responden yang Lancar dalam membayar Wajib Pajak.

Pada data responden 20 diketahui perhitungan maksimal kelas Lancar = 0 kelas tidak Lancar = 0,0061. sehingga  $0 < 0,0061$ . Dapat dikatakan data responden 20 data responden Tidak Lancar dalam membayar Wajib Pajak.

### 3.3. Validasi Performance

Dalam melakukan validasi *performance* digunakan operator *X-Validation*. Operator *X-validation* melakukan validasi silang untuk memperkirakan kinerja statistik operator pembelajaran (biasanya pada set data yang tak terlihat). Operator ini juga digunakan untuk memperkirakan seberapa akurat suatu model yang akan tampil dalam praktek. Operator *X-Validasi* merupakan operator bersarang yang memiliki dua subproses: *training subprocess* (subproses percobaan) dan *testing subprocess* (subproses pengujian). Subproses percobaan digunakan untuk melatih sebuah model. Model yang terlatih kemudian diterapkan dalam subproses pengujian. Biasanya proses belajar mengoptimalkan parameter model untuk membuat model sesuai dengan data percobaan. Jika kita kemudian mengambil sampel independen dari data pengujian, umumnya model tersebut tidak cocok dengan data percobaan maupun data pengujian. Hal ini disebut dengan istilah '*over-pas*', dan sangat mungkin terjadi ketika ukuran set data training kecil, atau ketika jumlah parameter dalam model besar. Sehingga validasi silang merupakan cara untuk memprediksi kesesuaian model untuk satu set pengujian hipotesis ketika set pengujian eksplisit tidak tersedia. Berikut merupakan hasil pengujian model algoritma Naive Bayes *Classifier* ditunjukkan pada gambar berikut:



	true Tidak Lancar	true Lancar
pred. Tidak Lancar	4	3
pred. Lancar	3	10
class recall	57.14%	76.92%

**Gambar 1.** Nilai Accuracy Performance

Keterangan :

- Jumlah prediksi tidak lancar dan kenyataannya benar tidak lancar adalah 4 *record*.
- Jumlah prediksi lancar dan kenyataannya benar tidak lancar adalah 3 *record*.
- Jumlah prediksi tidak lancar dan kenyataannya benar Lancar adalah 3 *record*.
- Jumlah prediksi lancar dan kenyataannya benar lancar adalah 10 *record*.

Pada gambar Nilai Accuracy Performance dijelaskan bahwa prediksi tidak lancar memiliki nilai dengan *class precision* 57,14%, prediksi lancar memiliki nilai dengan *class precision* 76,92 %.

#### 4. KESIMPULAN DAN SARAN

Berdasarkan hasil perhitungan dan pembahasan Metode *Naive Bayes* dan implementasi *rapidminer* pada bab sebelumnya, penulis dapat menarik kesimpulan sebagai berikut.

- Teknik *Data Mining* menggunakan algoritma *Naive bayes* dapat diterapkan dalam mengklasifikasi masyarakat wajib pajak di Kabupaten Simalungun.
- Data yang sebelumnya diolah dalam perhitungan metode *Naive Bayes* telah diimplementasikan ke dalam *Software rapidminer* yang bertujuan untuk membuktikan bahwa hasilnya sama dengan perhitungan manual pada metode *Naive Bayes*.
- Hasil yang diperoleh dalam mengklasifikasi masyarakat wajib pajak di Kabupaten Simalungun menggunakan *software Rapidminer* adalah sebagai berikut :
  - Jumlah prediksi tidak lancar dan kenyataannya benar tidak lancar adalah 4 *record*.
  - Jumlah prediksi lancar dan kenyataannya benar tidak lancar adalah 3 *record*.
  - Jumlah prediksi tidak lancar dan kenyataannya benar Lancar adalah 3 *record*.
  - Jumlah prediksi lancar dan kenyataannya benar lancar adalah 10 *record*.

Untuk penelitian lebih lanjut ada beberapa saran dan masukan diantaranya,

- Pengujian menggunakan metode/algoritma lainnya agar hasil yang didapatkan lebih baik lagi.
- Diharapkan penelitian ini dapat dijadikan referensi di bidang ilmu pengetahuan dan wawasan untuk penelitian berikutnya, dan dapat dikembangkan untuk memperoleh hasil yang cepat dan akurat.

#### DAFTAR PUSTAKA

- [1] C. Fadlan, S. Ningsih, Dan A. P. Windarto, “Penerapan Metode Naive Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra,” *Jutim*, Vol. 3, No. 1, Hal. 1–8, 2018.

- [2] A. Saleh, “Klasifikasi Metode Naive Bayes Dalam Data Mining Untuk Menentukan Konsentrasi Siswa,” *Ketik*, Hal. 200–208, 2015.
- [3] S. Sundari, I. S. Damanik, A. P. Windarto, Dan H. S. Tambunan, “Analisis K-Medoids Clustering Dalam Pengelompokan Data Imunisasi Campak Balita Di Indonesia,” *Pros. Semin. Nas. Ris. Inf. Sci.*, No. September, Hal. 687–696, 2019.
- [4] H. M. M. Al Laroussi, “Implementasi Algoritma Naive Bayes Sebagai Proses Seleksi Penerima Beasiswa Libyan Embassy Berbasis Web,” 2015.
- [5] Syarli Dan A. A. Muin, “Metode Naive Bayes Untuk Prediksi Kelulusan ( Studi Kasus : Data Mahasiswa Baru Perguruan Tinggi ),” *J. Ilm. Ilmu Komput.*, Vol. 2, No. 1, Hal. 5, 2016.
- [6] C. Fadlan, S. Ningsih, Dan A. P. Windarto, “Penerapan Metode Naive Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra,” *Jutim*, Vol. 3, No. 1, Hal. 1–8, 2018.
- [7] T. Imandasari, E. Irawan, A. P. Windarto, Dan A. Wanto, “Algoritma Naive Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air,” *Pros. Semin. Nas. Ris. Inf. Sci.*, No. November, 2019.