

Taxpayer Awareness Classification Using Decision Tree and Naïve Bayes Methods

Moch Riyadi Maskur A^{1*}, Arief Wibowo^{2*}

* Ilmu Komputer, Universitas Budi Luhur

2211600347@student.budiluhur.ac.id¹, arief.wibowo@budiluhur.ac.id²

Article Info

Article history:

Received 2023-10-12

Revised 2023-11-23

Accepted 2023-12-15

Keyword:

*Decision Tree,
Land and Building Tax,
Awareness,
Taxpayer,
Naïve Bayes.*

ABSTRACT

Land and Building Tax (PBB) has a big influence on a region's PAD. Therefore, regions always strive to increase PBB income as much as possible. Many factors influence the increase in PBB, one of which is public awareness of taxes. Lack of public awareness of taxes causes PBB income to also decrease, and has implications for regional PAD. And conversely, if public awareness of taxes is high, PBB and PAD revenues will also increase. Therefore, a system is needed to measure public awareness of taxes in the region. If public awareness of taxes can be measured, then the relevant agencies can evaluate and map taxpayers in which sub-districts have high or below average awareness. There are several factors that influence taxpayer awareness, including ownership status, tax sector, assessment category, and the number of receivable payments over the past 5 years. By knowing the awareness of taxpayers, the relevant agencies can review the targets for achieving PBB revenue and issue warning letters to taxpayers whose awareness of PBB is lacking. This research uses decision tree and naïve Bayes methods to classify 666,580 datasets obtained from the Cianjur Regency Regional Revenue Management Agency. The stages are carried out by data collection, data preprocessing, training data labeling, classification process, and evaluation. The result of this research is a system that can predict whether taxpayers are aware or not in a sub-district and sub-district or rural area using decision trees and naïve Bayes. The accuracy obtained from the decision tree method was 93.73%, while the accuracy obtained from the naïve Bayes method was 85.61%.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Indonesia merupakan negara yang memegang asas desentralisasi dalam penyelenggaraan pemerintahan dengan memberikan 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 [1].

Pajak Bumi dan Bangunan adalah jenis pajak daerah yang sepenuhnya diatur oleh pemerintah daerah dalam menentukan besar pajaknya, pajak ini sangat penting sekali karena digunakan untuk peningkatan pembangunan dan kesejahteraan rakyat di daerah [2]. Oleh karena itu, salah satu sumber dana untuk pembiayaan pembangunan adalah pajak bumi dan bangunan. Pajak Bumi dan Bangunan juga bisa diartikan sebagai penerimaan negara yang berasal dari rakyat atas kebendaan objek atau bumi, tanah dan atau bangunan

yang sebagian besar hasilnya diserahkan kepada daerah masing-masing untuk meningkatkan pendapatan daerah tersebut yang nilainya di atas NJOPTKP (Nilai Jual Objek Pajak Tidak Kena Pajak) [3].

Maka, untuk meningkatkan pendapatan daerah, biasanya pemerintah daerah menetapkan target penerimaan daerah dari pajak bumi dan bangunan. Ada beberapa faktor yang bisa menyebabkan target penerimaan ini tidak tercapai, yaitu kurangnya kesadaran membayar pajak tepat waktu, lemahnya aparatur pajak terletak pada kesalahan saat melakukan penginputan data, perlawanan aktif yaitu masih ada wajib pajak yang berusaha menghindari, tidak tersampainya surat pemberitahuan pajak terutang (SPPT) karena wajib pajak tidak berada dalam objek pajak, dan masih banyaknya wajib pajak lupa dalam membayar pajak bumi dan bangunan dan perlawanan pasif yaitu minimnya perekonomian dapat mempersulit pembayaran pajak bumi dan bangunan [3].

Dengan secara tidak langsung pendapat daerah sangat bergantung pada kesadaran masyarakat dalam membayar PBB. Wajib Pajak harus memahami alur dan sistem dari Pajak Bumi dan Bangunan agar Wajib Pajak tahu akan kewajibannya dan terhindar dari hambatan dalam perpajakan yang mengakibatkan rendahnya tingkat kesadaran Wajib Pajak dalam membayarkan PBB [2]. Kesadaran bagi wajib pajak atas fungsi perpajakan sebagai pembiayaan negara sangat diperlukan untuk meningkatkan kepatuhan wajib pajak [4].

Kesadaran masyarakat dipengaruhi beberapa faktor yang nantinya akan menjadi *variable* untuk memprediksi sadar atau tidaknya seorang wajib pajak. Faktor kesadaran tersebut diukur dari status kepemilikan, sektor pajak, kategori ketetapan, dan banyaknya piutang selama 5 tahun ke belakang. Bila kesadaran wajib pajak di suatu daerah sudah terukur, maka instansi terkait bisa mengetahui wilayah mana saja yang perlu ditingkatkan kesadarannya dalam membayar pajak. Apalagi target PBB dari tahun ke tahun selalu meningkat, maka perlu dilakukan formulasi untuk mendongkrak pendapatan guna mencapai target.

Berdasarkan penelitian terdahulu yang dilakukan oleh [5] membandingkan algoritma *naïve bayes* dan *decision tree*. Didapatkan hasil akurasi 95,58% untuk *decision tree*, dan 87,69% untuk algoritma *naïve bayes*, dengan begitu algoritma *decision tree* dinilai lebih baik. Dalam penelitian lain, dua algoritma tersebut juga digunakan dan dibandingkan dalam klasifikasi *performance KPI salesman*, didapatkan hasil akurasi *decision tree* yang sempurna yang mencapai angka 100%, sedangkan metode *naïve bayes* hanya 41% [6].

Pada penelitian sebelumnya dilakukan juga dalam klasifikasi diagnosis penyakit covid-19 menggunakan *decision tree*. Dari 125 dataset yang didapat digunakan untuk memprediksi positif atau negatif berdasarkan kelamin, usia, gejala, dan kondisi user. Dari penelitian yang dilakukan didapat akurasi sebesar 90% [7]. Selain itu, dalam penelitian lain digunakan juga untuk pendeteksian *spam* email menggunakan algoritma *naïve bayes* dan *decision tree*. Hasil dari penelitian ini adalah algoritma *decision tree* unggul

dengan akurasi 93,117% dibanding dengan *naïve bayes* yang mendapat akurasi 88,5284% dalam menyortir email *spam* [8]

Maka dalam penelitian ini, metode yang digunakan dalam memprediksi kesadaran wajib pajak ini menggunakan *decision tree* dan *naïve bayes*, dari kedua metode tersebut nantinya akan diukur *performance* nya menggunakan *confusion matrix* agar terlihat mana yang memiliki akurasi paling tinggi. Ada 4 faktor yang dalam pengukuran performa dengan *confusion matrix*, di antaranya yaitu *Accuracy*, *Precision*, *Recall* dan *F1-Score* [8].

II. METODE PENELITIAN

A. PBB (Pajak Bumi dan Bangunan)

Pajak Bumi dan Bangunan adalah pajak yang bersifat kebendaan dalam arti besarnya pajak tentang ditentukan oleh keadaan objek yaitu bumi/tanah dan atau bangunan [3]. Bumi adalah permukaan bumi dan tubuh bumi yang ada di bawahnya, bangunan adalah konstruksi teknik yang ditanam atau dilekatkan secara tetap pada tanah dan atau perairan, termasuk dalam pengertian bangunan adalah jalan lingkungan yang terletak dalam suatu kompleks bangunan seperti hotel, pabrik, dan emplasemenya, jalan tol, kolam renang, pagar mewah, tempat olahraga, galangan kapal, dermaga, taman mewah, tempat penampungan / kilang minyak, air dan gas, pipa minyak, fasilitas lian yang memberikan manfaat.

B. Wajib Pajak

Wajib Pajak atau Subjek Pajak Bumi dan Bangunan (PBB) adalah orang atau badan yang secara nyata mempunyai suatu hak atas bumi atau memperoleh manfaat atas bumi, memiliki, menguasai, atau memperoleh manfaat atas bangunan, dengan demikian tanda pembayaran/pelunasan pajak bukan merupakan bukti pemilikan hak[4]. Yang kemudian Wajib Pajak ini memiliki status kepemilikan tertentu berdasarkan bukti kepemilikan yang ada.

C. Kesadaran Wajib Pajak

Kesadaran Wajib Pajak adalah wajib pajak mau dengan sendirinya melakukan kewajiban perpajakannya seperti mendaftarkan diri, menghitung, membayar dan melaporkan jumlah pajak terutangnya[3] Kesadaran Wajib Pajak dan Sanksi Pajak memiliki pengaruh positif dan signifikan terhadap Kepatuhan Wajib Pajak dalam membayar Pajak Bumi dan Bangunan [2].

Dalam penelitian ini, kesadaran wajib pajak diukur dari beberapa faktor berikut, yaitu

- 1) Status Kepemilikan: status kepemilikan mempengaruhi kesadaran wajib pajak, karena berkaitan dengan tanggungjawab. Orang yang hanya sebagai penyewa tentunya akan berbeda rasa tanggungjawab nya dengan orang yang berstatus sebagai pemilik.
- 2) Sektor Pajak: Sektor pajak terbagi menjadi dua, yaitu sektor perdesaan dan perkotaan. Kesadaran pajak di sektor perdesaan akan berbeda dengan perkotaan, ini

dikarenakan Nilai Jual Objek Pajak dan tingkat penghasilan yang lebih rendah. Faktor penghasilan dapat dijadikan salah satu alasan wajib pajak tidak patuh atau tidak sadar membayar pajak [4].

- 3) Kategori Ketetapan: kategori ketetapan dibedakan menjadi 5 buku, yaitu buku 1 (ketetapan < Rp.100.000), buku 2 (ketetapan Rp.100.001 – Rp.500.000), buku 3 (ketetapan Rp.500.001 – Rp.2.000.000), buku 4 (ketetapan Rp.2.000.001 – Rp.5.000.000), buku 5 (ketetapan > Rp.5.000.000). Dalam penelitian, kategori ketetapan dikelompokkan lagi menjadi 2, yaitu kategori 1,2,3 dan 4,5. Besarnya ketetapan pajak sangat berpengaruh terhadap kesadaran wajib pajak [9].
- 4) Banyak Pembayaran Piutang: ketetapan pajak yang belum atau sudah dibayar akan dilihat selama 5 tahun ke belakang sebagai *history* yang akan menjadi faktor penentu kesadaran wajib pajak.

D. Status Kepemilikan

Status Kepemilikan dalam Pajak Bumi dan Bangunan terbagi menjadi beberapa jenis, yaitu

- 1) Pemilik: orang yang sepenuhnya memiliki atas bumi dan bangunan.
- 2) Penyewa: orang yang menyewa asset berupa tanah atau pun bangunan kepada pemilik.
- 3) Pemakai: orang yang hanya memakai bumi dan bangunan, tetapi tidak memiliki dan menyewa nya.
- 4) Pengelola: orang yang hanya mengelola bumi dan bangunan, dan bukan pemiliknya. Biasanya ini berupa tanah sawah yang dikelolanya, atau bisa juga berbentuk bangunan seperti villa.
- 5) Sengketa: status yang masih belum jelas kepemilikannya, karena surat-surat dan sertifikatnya tidak jelas.

E. Data Mining

Data mining adalah proses mendapatkan atau mencari suatu pengetahuan dalam sebuah database. Beberapa tugas yang dapat dilakukan dengan data mining meliputi: prediksi, klasifikasi, deskripsi, estimasi, *clustering* dan asosiasi.

Beberapa model klasifikasi yang paling sering digunakan diantaranya *decision tree*, *naïve bayes*, *neural network*, *genetic algorithms*, *K- nearest neighbor*, *support vector machine* [5].

F. Klasifikasi

Klasifikasi adalah suatu proses menentukan objek data untuk dimasukkan ke dalam kelas tertentu berdasarkan sejumlah kelas yang ada. Dua hal utama 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[1].

G. Naïve Bayes

Naive Bayes ialah tata cara pengklasifikasian probabilistik simpel. Model ini hendak menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi serta campuran nilai dari dataset yang diberikan [10]. Berikut rumus untuk *naïve bayes* [6]

H. Decision Tree

Model *Decision Tree* adalah berbentuk sebuah pohon yang terdiri dari beberapa bagian yaitu *root node*, *internal node*, dan *terminal node*. Data query menelusuri *root node* dan *internal node* mencapai *terminal node* merupakan proses melakukan klasifikasi pada metode *Decision Tree* ini [11].

Layaknya sebuah pohon, *Decision Tree* mempunyai akar (*root/ node*), batang/cabang (*branch node*), dan daun (*leaf*). Dimana setiap *node* mewakili suatu atribut dan *branch node* mewakili nilai atribut, sedangkan *leaf* digunakan untuk mewakili kelas. *Node* teratas dari pohon keputusan disebut *root/node* [12].

Nilai *Gain* adalah *Information Gain* yang digunakan untuk mencari satu variable/ atribut dari dataset (S) untuk dijadikan *root/ node* dan *branch node*, yaitu satu atribut yang mempunyai nilai *gain* tertinggi. Untuk pencarian nilai *gain* (*information Gain*) dapat dilakukan dengan konsep *Entropy*, *Gini Index* dan *Classification Error*. Nilai *gain* terbesar yang di dapat dari atribut pada dataset (data pembelajaran) pertama kali digunakan untuk mencari atribut yang layak menjadi *root* (akar) pohon keputusan (*decision Tree*). Selanjut proses akan diulang untuk mencari atribut yang menjadi cabang hingga menemukan *leaf* yang merupakan label kelas

Information Gain dengan Nilai *Entropy* adalah formula untuk menghitung homogenitas atribut (A) dari sebuah sample data (S). [13]

Dengan formula:

$$Entropy(S) = \sum_{i=1}^n p_i \log_2 p_i$$

S = Himpunan Kasus dalam dataset

A = Fitur (atribut)

n = jumlah partisi atribut S

Pi = proposi dari Si terhadap S

maka:

$$Entropy(S) = \sum_{i=1}^n p(i|s) \log_2 p(i|s)$$

Maka *Gain(S,A)* adalah *Information Gain* dari attribute A pada koleksi contoh S

$$\text{Gain}(S, A) = \frac{\text{Entrop } y(S)}{\text{Entrop } y(S_i)} \sum_{i=1}^n \frac{|S_i|}{|S|}$$

Information Gain dengan Nilai Gini Index

$$\text{Gini}(t) = 1 - \sum_{i=1}^n [p(i|s)]^2$$

$$\text{Gain}(S, A) = \text{Gini}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \text{Gini}(S_i)$$

Information Gain dengan Classification Error Nilai C. Error diperoleh dari nilai value atribut yang terkecil dari terhadap class label

$$\text{C. Error}(S) = 1 - \max_i [p(i|s)]$$

$$\text{Gain}(S, A) = \text{Error}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \text{Error}(S_i)$$

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

Keterangan :

- X = Data dengan class yang belum diketahui
 H = Hipotesis data merupakan suatu class spesifik
 P(H|X) = Probabilitas hipotesis H berdasar kondisi X (Posteriori Probabilitas)
 P(H) = Probabilitas hipotesis H (Prior Probabilitas)
 P(X|H) = Probabilitas X berdasarkan kondisi hipotesis H
 P(X) = Probabilitas X

I. Missing Value

Missing value adalah hilangnya nilai pada suatu atribut data. Kesalahan input data menjadi penyebab hilangnya nilai tersebut. Hal ini bisa berakibat pada penurunan kualitas data pada proses klasifikasi [14]. Apabila nilai yang hilang bertipe kategori, maka missing value diisi dengan modus data tersebut. Tetapi, bila data tersebut memiliki tipe data numerik, maka akan dilakukan pengecekan distribusi datanya. Jika distribusi datanya normal, maka akan diisi dengan nilai rata-rata atribut tersebut. Sebaliknya missing value akan diisi dengan nilai median dari atribut tersebut [15].

J. Confusion Matrix

Confusion matrix adalah salah satu metode yang digunakan untuk perhitungan akurasi pada konsep data mining. Confidence atau precision adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. Recall atau sensitivity adalah proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar.[7] Terdapat empat istilah sebagai representasi hasil proses klasifikasi pada confusion matrix. Rumus perhitungan accuracy adalah sebagai berikut.

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- 1) True Positive (TP) merupakan data positif yang diprediksi benar.
- 2) True Negative (TN) merupakan data negatif yang diprediksi benar.
- 3) False Positive (FP) merupakan data negatif namun diprediksi sebagai data positif.
- 4) False Negative (FN) merupakan data positif namun diprediksi sebagai negatif.

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Pada penelitian ini diperoleh data sebanyak 666.580 data. Data didapatkan dari Badan Pengelolaan Pendapatan Daerah Kabupaten Cianjur dengan mengekspor data hasil query dari database PBB yang memiliki catatan pembayaran 5 tahun ke belakang dan bukan status kepemilikan yang sengketa. Kemudian dihitung tahun pajak mana saja yang sudah bayar selama 5 tahun sebelumnya. Berikut merupakan tabel-tabel kategori yang digunakan sebagai atribut dalam penelitian ini

TABEL I
DAFTAR NAMA DAN KODE KECAMATAN DI CIANJUR

No	Kecamatan	Kode Kecamatan
1	Agrabinta	3205010
2	Leles	3205011
3	Sindangbarang	3205020
4	Cidaun	3205030
5	Naringgul	3205040
6	Cibinong	3205050
7	Cikadu	3205051
8	Tangeung	3205060
9	Pasirkuda	3205061
10	Pagelaran	3205070
11	Kadupandak	3205080
12	Cijati	3205081
13	Takokak	3205090
14	Sukanagara	3205100
15	Campaka	3205110
16	Campakamulya	3205111
17	Cibeber	3205120
18	Bojongpicung	3205130
19	Haurwangi	3205131

No	Kecamatan	Kode Kecamatan
20	Ciranjang	3205140
21	Sukaluyu	3205141
22	Karangtengah	3205150
23	Cianjur	3205160
24	Cilaku	3205161
25	Warungkondang	3205170
26	Gekbrong	3205171
27	Cugenang	3205180
28	Pacet	3205190
29	Sukaresmi	3205191
30	Cipanas	3205192
31	Mande	3205200
32	Cikalongkulon	3205210

TABEL II
KODE SEKTOR PBB DI CIANJUR

No	Sektor	Kode Sektor
1	Perkotaan	PKT
2	Perdesaan	PDS

TABEL III
DAFTAR KATEGORI KETETAPAN PAJAK

No	Pokok PBB	Kategori
1	Kurang dari sama dengan 2 juta	TAP 1 2 3
2	Lebih dari 2 juta	TAP 4 5

B. Data Preprocessing

Datasets yang sudah terkumpul, kemudian dilakukan *cleansing* dengan pengecekan *missing value*, atau pun redundansi data. Pada penelitian ini ditemukan *missing value* pada data status kepemilikan yang memiliki tipe data *categorical*. Sehingga dilakukan teknik imputasi dengan mengganti data yang *missing value* dengan modus data tersebut, yaitu status pemilik. Setelah dilakukan *preprocessing*, maka data bisa diproses ke tahap berikutnya.

TABEL IV
DATA STATUS KEPEMILIKAN MISSING VALUE

Status Kepemilikan	Jumlah
Pemilik	660.716
Penyewa	186
Pemakai	511
Pengelola	1.608
<Empty>	2.985
Null	574
Total	666.580

Untuk menghilangkan *missing value* tersebut, maka digunakan teknik imputasi, yaitu mengganti nilai yang hilang dengan modus data tersebut.

C. Data Training dan Data Testing

Setelah data sudah *cleansing*, kemudian dilakukan pelabelan manual terhadap data *training* sebanyak 96 data dengan bantuan pejabat pengelola data dan informasi di Badan Pengelolaan Pendapatan Daerah Kabupaten Cianjur.

Berikut sampel data *training* dan data *testing*.

TABEL V
SAMPEL DATA TRAINING

No	Status Kepemilikan	Sektor	Kategori Ketetapan	Pembayaran 5 Tahun Sebelumnya	Kesadaran Wp
1	Pemilik	Perdesaan	Tap 1 2 3	0	Tidak
2	Penyewa	Perdesaan	Tap 1 2 3	0	Tidak
3	Pemakai	Perkotaan	Tap 4 5	0	Tidak
4	Pengelola	Perkotaan	Tap 4 5	0	Tidak
5	Pemilik	Perdesaan	Tap 1 2 3	1	Tidak
6	Penyewa	Perdesaan	Tap 1 2 3	1	Tidak
7	Pemakai	Perkotaan	Tap 4 5	1	Tidak
8	Pengelola	Perkotaan	Tap 4 5	1	Tidak
9	Pemilik	Perdesaan	Tap 1 2 3	2	Tidak
10	Penyewa	Perdesaan	Tap 1 2 3	2	Tidak
11	Pemakai	Perkotaan	Tap 4 5	2	Tidak
12	Pengelola	Perkotaan	Tap 4 5	2	Ya
13	Pemilik	Perdesaan	Tap 1 2 3	3	Tidak
14	Penyewa	Perdesaan	Tap 1 2 3	3	Ya
15	Pemakai	Perkotaan	Tap 4 5	3	Ya
16	Pengelola	Perkotaan	Tap 4 5	3	Ya
17	Pemilik	Perdesaan	Tap 1 2 3	4	Ya
18	Penyewa	Perdesaan	Tap 1 2 3	4	Ya
19	Pemakai	Perkotaan	Tap 4 5	4	Ya
20	Pengelola	Perkotaan	Tap 4 5	4	Ya
21	Pemilik	Perdesaan	Tap 1 2 3	5	Ya
22	Penyewa	Perdesaan	Tap 1 2 3	5	Ya
23	Pemakai	Perkotaan	Tap 4 5	5	Ya
24	Pengelola	Perkotaan	Tap 4 5	5	Ya

Setelah dilakukan pelabelan manual pada data *training*, kemudian data *training* dan Data *testing* di *import* ke dalam *local repository* untuk kemudian dibaca oleh *RapidMiner*. Kemudian dilakukan penentuan *type attribute* dan *role* kolom.

D. Klasifikasi

Data yang sudah diimport kemudian di *apply model* dan *performance* nya dengan menggunakan *decision tree* dan *naïve bayes*. Hasil dari penerapan model dengan dua algoritma ini adalah mengetahui kesadaran Wajib Pajak dalam membayar Pajak Bumi dan Bangunan dengan menggunakan model *decision tree* dan *naïve bayes*.

Pada pemodelan *decision tree* ini, pohon keputusan merupakan hasil proses perhitungan entropi dan *information gain*, setelah dilakukan perhitungan berulang-ulang hingga semua atribut pohon mempunyai kelas dan tidak dapat lagi dilakukan proses perhitungan [16]. Dari hasil pemodelan *decision tree* ini, maka nanti akan terlihat prediksi dari klasifikasi yang telah dilakukan. Dari proses penerapan model *decision tree*, maka didapatkan pohon keputusan seperti pada gambar 1 (lampiran).

Setelah proses pemodelan *decision tree* selesai hingga muncul pohon keputusannya, maka dilakukan pemodelan dengan *naïve bayes* untuk mengetahui *simple distribution* nya. Maka akan dilakukan teknik peramalan dengan dasar probabilitas sederhana yang didasarkan dari penerapan *teorema bayesian* dengan perkiraan independensi kuat [17]. Berikut hasil dari *simple distribution naïve bayes*

SimpleDistribution

Distribution model for label attribute KESADARAN WP

Class TIDAK (0.490)
4 distributions

Class YA (0.510)
4 distributions

Gambar 2. Result Simple Distribution

Dari pohon keputusan terurai sesuai nilai gain yang didapat, mana yang menjadi *root node*, *branch node*, dan *leaf node*. Dan dari pemodelan *naïve bayes* didapat *simple distribution* dengan prediksi class Ya sebesar 51% dan class Tidak sebesar 49%. Kemudian hasil klasifikasi kesadaran wajib pajak akan dibandingkan prediksi nya per kecamatan antara *naïve bayes* dan *decision tree* di tahap evaluasi.

E. Evaluasi

Terdapat beberapa perbedaan prediksi antara dua algoritma *decision tree* dan *naïve bayes*. Hal ini yang mengakibatkan perbedaan akurasi performance yang nantinya akan diukur dengan *confusion matrix*. Berikut perbedaan prediksi Ya dan Tidak dalam klasifikasi kesadaran wajib pajak menggunakan pemodelan *decision tree* dan *naïve bayes*.

TABEL VI
KOMPARASI HASIL PREDIKSI PER KECAMATAN

No	Kecamatan	Decisi on Tree (Ya)	Naïve Bayes (Ya)	Decisio n Tree (Tidak)	Naïve Bayes (Tidak)
1	Agrabinta	9000	9359	16390	16031
2	Leles	12839	13121	4061	3779
3	Sindangbarang	10098	10718	5291	4671
4	Cidaun	10293	10620	4183	3856
5	Naringgul	15384	15747	565	202
6	Cibinong	19340	19739	4143	3744
7	Cikadu	14408	14408	2200	2200
8	Tanggeung	12024	12466	2207	1765
9	Pasirkuda	10388	10388	0	0
10	Pagelaran	19542	20418	3277	2401
11	Kadupandak	13081	13441	2680	2320
12	Cijati	12060	12246	2447	2261
13	Takokak	15381	16074	7267	6574
14	Sukanagara	7452	7708	3421	3165
15	Campaka	14912	16416	5535	4031
16	Campakamulya	7718	7718	262	262
17	Cibeber	17898	20631	21501	18768

No	Kecamatan	Decisi on Tree (Ya)	Naïve Bayes (Ya)	Decisio n Tree (Tidak)	Naïve Bayes (Tidak)
18	Bojongpicung	11696	13119	9281	7858
19	Haurwangi	4535	5371	9916	9080
20	Ciranjang	7423	8285	11902	11040
21	Sukaluyu	11612	13932	12386	10066
22	Karangtengah	12462	17316	27119	22265
23	Cianjur	14438	20993	22359	15804
24	Cilaku	7338	13160	26517	20695
25	Warungkondang	5455	6788	8173	6840
26	Gekbrong	4962	6862	7680	5780
27	Cugenang	4624	5880	11834	10578
28	Pacet	8994	13884	21255	16365
29	Sukaresmi	8354	13018	14385	9721
30	Cipanas	9782	14109	19436	15109
31	Mande	9113	11564	11887	9436
32	Cikalongkulon	7367	9352	17047	15062

Hasil klasifikasi dengan pemodelan *decision tree* dan *naïve bayes* akan dibandingkan *performance* nya, sehingga akan ditemukan akurasi yang paling tinggi di antara dua model tersebut. Akurasi yang didapatkan dari model *decision tree* dan *naïve bayes*, masing-masing mendapatkan 93.73% untuk model *decision tree* dan 85.61% untuk model *naïve bayes*. Sehingga pada penelitian ini, *decision tree* mendapatkan akurasi paling tinggi dalam mengklasifikasi kesadaran Wajib Pajak PBB. Perhitungan akurasi *decision tree*:

$$\frac{349.973 + 274.791}{349.973 + 274.791 + 41.816 + 0} = \frac{624.764}{666.580}$$

$$= 0,937267845 \times 100\% = \mathbf{93,73\%}$$

Apabila disajikan dalam bentuk tabel, maka hasilnya seperti berikut:

TABEL VII
ACCURACY MODEL NAÏVE BAYES

	True Tidak	True Ya	Class Precision
Pred. Tidak	274.791	41.816	86,79%
Pred. Ya	0	349.973	100,00%
Class Recall	100,00%	89,33%	
Accuracy: 93,73%			

Perhitungan akurasi *naïve bayes*:

$$\frac{350.375 + 220.315}{350.375 + 220.315 + 41.414 + 54.476} = \frac{624.764}{666.580}$$

$$= 0,856146299 \times 100\% = \mathbf{85,61\%}$$

Apabila disajikan dalam bentuk tabel, maka hasilnya seperti berikut.

TABEL VIII
ACCURACY MODEL DECISION TREE

	True Tidak	True Ya	Class Precision
Pred. Tidak	220.315	41.414	84,18%
Pred. Ya	54.476	350.375	86,54%
Class Recall	80,18%	89,43%	
Accuracy: 85,61%			

Karena akurasi dari model *decision tree* lebih tinggi dibanding model *naïve bayes*, maka peneliti mencoba memvisualisasikan hasil prediksi kesadaran Wajib Pajak dengan penerapan model *decision tree* dalam bentuk grafik bar, yang dikelompokkan menjadi per kecamatan. Seperti pada gambar 3 untuk hasil klasifikasi yang divisualisasi menjadi gambaran kesadaran Wajib Pajak PBB per kecamatan.

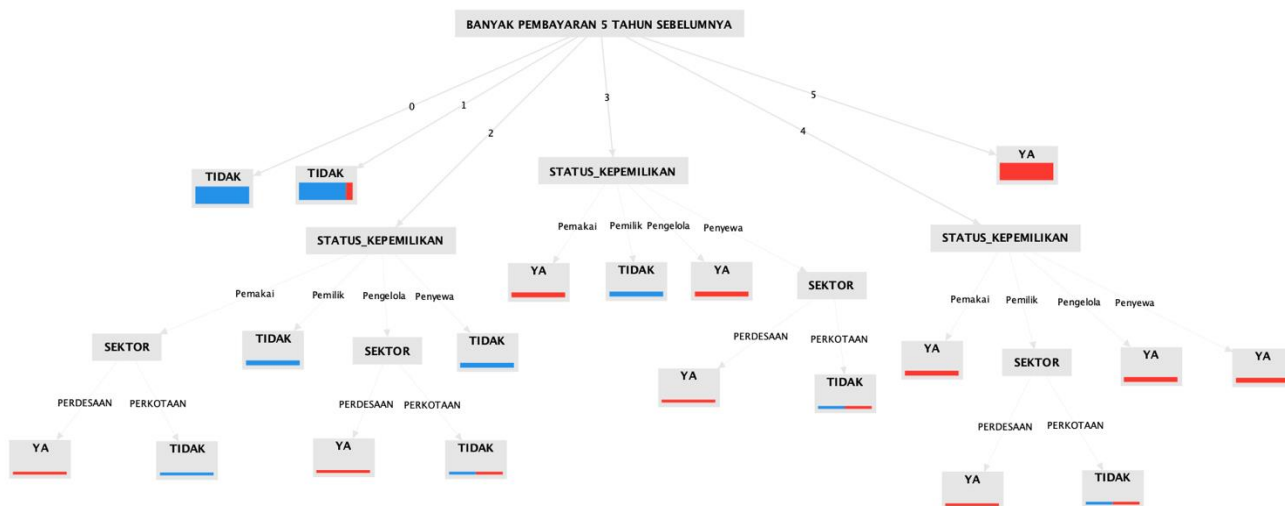
IV. KESIMPULAN

Dari hasil penelitian ini, maka dapat diambil beberapa kesimpulan. Metode *decision tree* memiliki akurasi lebih tinggi, yaitu 93,73%. Sedangkan *naïve bayes* memiliki akurasi sebesar 85,61%. *Decision tree* memiliki hasil yang lebih signifikan dalam pengklasifikasian kesadaran wajib pajak dibanding *naïve bayes* dalam penelitian ini. Dari 666.580 data testing yang telah diprediksi, maka rata-rata 58% memiliki kesadaran sebagai Wajib Pajak PBB dan sekitar 42% belum memiliki kesadaran Wajib Pajak PBB. Kemudian Kecamatan Pagelaran menjadi kecamatan yang memiliki paling banyak Wajib Pajak yang sadar akan Pajak Bumi dan Bangunan bila melihat hasil prediksi dari model *decision tree* dan *naïve bayes*. Kesadaran sebagai Wajib Pajak PBB lebih banyak di sektor perdesaan dibanding sektor perkotaan. Hal ini dikarenakan, pajak PBB di perdesaan lebih murah dibanding perkotaan. Hasil prediksi klasifikasi kesadaran wajib pajak ini dapat digunakan untuk mengevaluasi target penerimaan PBB dan menerbitkan surat peringatan untuk wajib pajak mana saja yang kesadarannya kurang, sehingga di tahun berikutnya bisa didapatkan hasil target yang optimal. Adapun saran untuk penelitian selanjutnya adalah menambahkan atribut umur dan jenjang pendidikan wajib pajak, juga penambahan pengumpulan data menggunakan questioner agar bisa lebih maksimal dalam proses pengklasifikasian kesadaran wajib pajak.

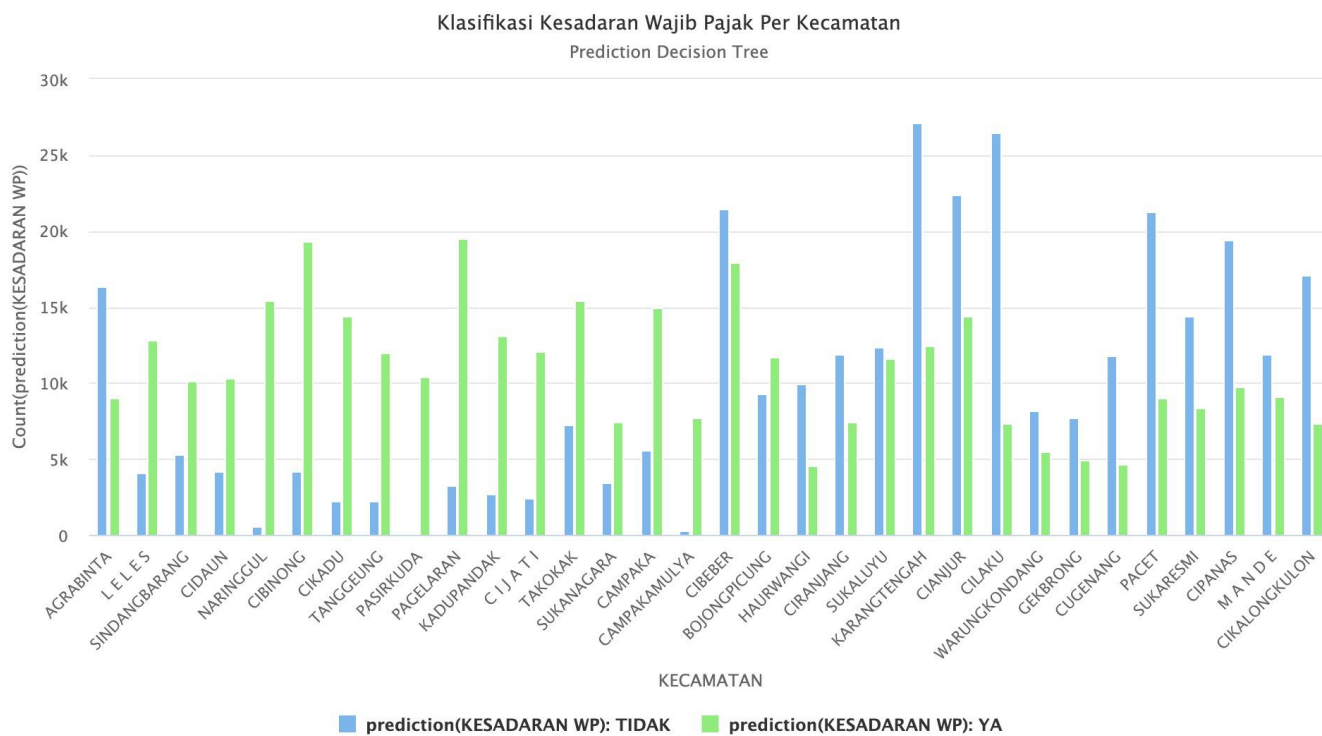
DAFTAR PUSTAKA

- [1] I. Hanafia Lubis and I. Sahputra Saragih, "Klasifikasi Masyarakat Wajib Pajak Di Kabupaten Simalungun Menggunakan Metode Naive Bayes," *BRAHMANA: Jurnal Penerapan Kecerdasan Buatan*, vol. 1, no. 1, pp. 15–23, 2019.
- [2] Y. A. Pravasanti, "Analisis Faktor-Faktor Yang Mempengaruhi Kepatuhan Wajib Pajak Dalam Membayar Pajak Bumi Dan Bangunan," *Jurnal Akuntansi dan Pajak*, vol. 21, no. 01, Jul. 2020, doi: 10.29040/jap.v21i1.1165.
- [3] R. Hidayat and S. R. Wati, "Pengaruh kesadaran Wajib Pajak dan kualitas pelayanan pajak terhadap kepatuhan Wajib Pajak dalam membayar Pajak bumi dan Bangunan di kota Bandung," *Owner*, vol. 6, no. 4, pp. 4009–4020, Oct. 2022, doi: 10.33395/owner.v6i4.1068.
- [4] P. Penghasilan *et al.*, "Effect of Income and Taxpayer Awareness on Compliance in Paying Land and Building Tax (PBB)," 2021. [Online]. Available: <http://e-journal.stie-aub.ac.id>
- [5] B. A. Candra Permana and I. K. Dewi Patwari, "Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naïve Bayes Untuk Prediksi Penyakit Diabetes," *Infotek : Jurnal Informatika dan Teknologi*, vol. 4, no. 1, pp. 63–69, Jan. 2021, doi: 10.29408/jit.v4i1.2994.
- [6] S. Sonia Shabrianti, A. Triayudi, and D. Avrilia Lantana, "Analisis Klasifikasi Performace KPI Salesman Menggunakan Metode Decision Tree Dan Naïve Bayes," *Jurnal Riset Komputer*, vol. 10, no. 1, pp. 2407–389, 2023, doi: 10.30865/jurikom.v10i1.5628.
- [7] R. Ridho and H. Hendra, "Klasifikasi Diagnosis Penyakit Covid-19 Menggunakan Metode Decision Tree," *JUST IT : Jurnal Sistem Informasi, Teknologi Informasi dan Komputer*, vol. 11, no. 3, pp. 69–75, 2022, [Online]. Available: <https://jurnal.umj.ac.id/index.php/just-it/article/view/13594>
- [8] R. S. Lutfiyani and N. Retnowati, "Implementasi Pendeteksian Spam Email Menggunakan Metode Text Mining Dengan Algoritma Naïve Bayes Dan Decision Tree J48," *Jurnal Komputer dan Informatika*, vol. 9, no. 2, pp. 244–252, Oct. 2021, doi: 10.35508/jicon.v9i2.5304.
- [9] S. Komparasi Stimulus Pajak Bumi Dan Bangunan Perdesaan Dan Perkotaan, D. Adhetya Safitra, and A. Hanifah, "Jurnal Anggaran dan Keuangan Negara Indonesia Comparative Study of Stimulus on Land and Building Tax in Rural and Urban Areas," 2022. [Online]. Available: <https://anggaran.e-journal.id/akurasi>
- [10] S. Alim, "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naïve Bayes," 2021.
- [11] A. Tangkelayuk and E. Mailoa, "Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes Dan Decision Tree," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 2, pp. 1109–1119, 2022, [Online]. Available: <http://jurnal.mdp.ac.id>
- [12] N. Rochmawati *et al.*, "Covid Symptom Severity Using Decision Tree," *Proceeding - 2020 3rd International Conference on Vocational Education and Electrical Engineering: Strengthening the framework of Society 5.0 through Innovations in Education, Electrical, Engineering and Informatics Engineering, ICVEE 2020*, pp. 1–5, 2020, doi: 10.1109/ICVEE50212.2020.9243246.
- [13] P. Kasih, "Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara," *Innovation in Research of Informatics (INNOVATICS)*, vol. 1, no. 2, pp. 63–69, 2019, doi: 10.37058/innovatics.v1i2.918.
- [14] M. Lutfi and M. Hasyim, "Penanganan Data Missing Value Pada Kualitas Produksi Jagung Dengan Menggunakan Metode K-Nn Imputation Pada Algoritma C4.5," Online, 2019. [Online]. Available: <http://jurnal.stiki-indonesia.ac.id/index.php/jurnalresistor>
- [15] M. Fajri and A. Primajaya, "Komparasi Teknik Hyperparameter Optimization pada SVM untuk Permasalahan Klasifikasi dengan Menggunakan Grid Search dan Random Search," 2023. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>

- [16] H. Sulistiani and A. A. Aldino, "Decision Tree C4.5 Algorithm for Tuition Aid Grant Program Classification (Case Study: Department of Information System, Universitas Teknokrat Indonesia)," *Edutic - Scientific Journal of Informatics Education*, vol. 7, no. 1, pp. 40–50, 2020, doi: 10.21107/edutic.v7i1.8849.
- [17] A. Wibowo and A. Rohman, "Prediksi Predikat Kelulusan Mahasiswa Menggunakan Naive Bayes dan Decision Tree pada Universitas XYZ," *EXPERT: Jurnal Manajemen Sistem Informasi dan Teknologi*, vol. 12, no. 2, p. 104, Dec. 2022, doi: 10.36448/expert.v12i2.2810.



Gambar 1. Result Decision Tree



Gambar 3. Visualisasi Data Prediksi Kesadaran Wajib Pajak PBB Per Kecamatan