



## Implementasi Algoritma Naïve Bayes untuk Memprediksi Tingkat Kepatuhan Wajib Pajak di Desa Dames Damai

Zur'aini<sup>1</sup>, Yahya<sup>2</sup>, L.M Samsu<sup>3</sup>

<sup>1,2,3</sup>Program Studi Sistem Informasi, Universitas Hamzanwadi

zuraini2761@gmail.com

### Abstrak

Kepatuhan yang tinggi dari wajib pajak adalah suatu keadaan dimana wajib pajak memenuhi semua kewajibannya dalam membayar pajak. Dengan semakin banyaknya wajib pajak dan data yang membayar pajak menyebabkan pihak desa tidak mengetahui seberapa banyak wajib pajak yang patuh dan tidak patuh. Pada penelitian ini menggunakan teknik data mining dengan menerapkan metode klasifikasi menggunakan Algoritma Naïve Bayes dimana algoritma ini terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan jumlah data yang lebih besar. Klasifikasi ini sangat berguna untuk memprediksi wajib pajak mana yang patuh dan tidak patuh dalam pembayaran pajak bumi bangunan (PBB). Dengan menerapkan algoritma naïve bayes dan setelah dilakukan pengujian sebanyak 9 kali validation didapatkan hasil akurasi dan AUC tertinggi yaitu pada K-Fold Validation 8 dengan nilai akurasi sebesar 99.43% dan nilai AUC sebesar 0.991. Dari hasil akurasi tersebut dapat dinyatakan bahwa metode naïve bayes dapat digunakan sehingga kesadaran masyarakat Desa Dames Damai terhadap Pajak Bumi dan Bangunan dapat dikatakan patuh.

**Kata kunci:** Algoritma Naïve bayes, Data Mining, Prediksi, Wajib pajak.

### Abstract

High compliance of the taxpayer is a condition where the taxpayer fulfills all of his obligations in paying taxes. With the increasing number of taxpayers and data paying taxes, the village does not know how many taxpayers are compliant and non-compliant. In this study using data mining techniques by applying the classification method using the Naïve Bayes Algorithm where this algorithm is proven to have high accuracy and speed when applied to a database with a larger amount of data. This classification is very useful for predicting which taxpayers are compliant and non-compliant in paying land and building tax (PBB). By applying the naïve Bayes algorithm and after testing 9 times validation, the highest accuracy and AUC results were obtained, namely in K-Fold Validation 8 with an accuracy value of 99.43% and an AUC value of 0.991. From the results of this accuracy it can be stated that the naïve Bayes method can be used so that the awareness of the people of Dames Damai Village regarding Land and Building Tax can be said to be compliant.

**Keywords:** *Naïve bayes Algorithm, Data Mining, Prediction, Taxpayers.*

DOI : 10.29408/jprinter.v1i2.22005

URL : <https://doi.org/10.29408/jprinter.v1i2.22005>

## 1. Pendahuluan

Pemerintah melaksanakan pembangunan di Indonesia ini tentu memerlukan dana yang cukup besar. Pemasukan kas dari Pajak Bumi dan Bangunan merupakan bagian terpenting untuk pembangunan suatu daerah, dengan hasil yang telah didapatkan pemerintah daerah dapat meningkatkan pembangunan daerah dengan berbagai infrastruktur yang membantu masyarakat dalam melakukan berbagai aktivitas serta menjadikan daerah tersebut menjadi lebih maju. Pajak Bumi dan Bangunan (PBB) adalah pajak negara yang dikenakan terhadap bumi dan bangunan berdasarkan undang-undang nomor 12 tahun 1985.

Kepatuhan yang tinggi dari wajib pajak adalah suatu keadaan dimana wajib pajak memenuhi semua kewajibannya dalam membayar pajak, faktor yang mempengaruhi wajib pajak adalah kesadaran wajib pajak. Kesadaran wajib pajak merupakan kondisi dimana wajib pajak mengerti dan memahami arti, fungsi, maupun tujuan pembayaran kepada negara. Dengan kesadaran wajib pajak yang tinggi, akan memberikan pengaruh untuk meningkatkan kepatuhan pajak yang lebih baik lagi<sup>[1]</sup>.

Dengan melihat kemajuan infrastruktur di Desa Dames Damai yang

semakin berkembang menjadikan pajak bumi dan bangunan pendapatan yang paling dominan. Desa Dames Damai adalah salah satu desa yang berada di Kecamatan Suralaga Kabupaten Lombok Timur Provinsi Nusa Tenggara Barat. Desa Dames Damai terdiri dari 5 dusun yaitu : Dusun Dames, Dusun Burneo, Dusun Buniara, Dusun Sukadamai, dan Dusun Jamar Jenggang.

Namun yang menjadi permasalahan adalah peningkatan jumlah wajib pajak bumi dan bangunan semakin terus mengalami kenaikan, Jumlah wajib pajak untuk tiap tahun nya selalu bertambah, sebanyak 5.826 data wajib pajak yang terbagi dalam waktu 5 tahun terakhir yaitu dari 2018 sampai 2022 pada Kecamatan Suralaga Desa Dames Damai. Dengan semakin meningkatnya jumlah wajib pajak dan penginputan data yang terlalu banyak dan masih manual menyebabkan pihak desa yang mengelola pajak masih belum mengetahui seberapa banyak wajib pajak yang patuh dan seberapa banyak wajib pajak yang tidak patuh dalam membayar iuran pajak bumi dan bangunan tersebut. Dari permasalahan tersebut, peneliti tertarik untuk mengatasi masalah kepatuhan wajib pajak dengan melakukan pengolahan data menggunakan Data Mining, dimana manfaat data mining adalah untuk menangani banyaknya data. Untuk melihat

DOI : 10.29408/jprinter.v1i2.22005

kepatuhan wajib pajak digunakan teknik klasifikasi untuk menentukan potensi kepatuhan wajib pajak dengan menerapkan Algoritma Naïve Bayes.

Klasifikasi adalah proses penemuan model yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya belum di ketahui<sup>[2]</sup>. Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang ditemukan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi sebuah peluang dari suatu objek yang akan digunakan untuk prediksi masa depan. Klasifikasi dalam data mining adalah teknik umum yang membantu dalam mengatur kumpulan data yang rumit dan besar. Naïve Bayes merupakan metode pengklasifikasian paling populer digunakan dengan tingkat keakuratan yang baik. Keuntungan Algoritma Naïve Bayes adalah tingkat nilai error yang didapat lebih rendah walaupun data set berjumlah besar, selain itu akurasi naïve bayes dan kecepatannya lebih tinggi pada saat diaplikasikan ke dalam data set yang jumlahnya lebih besar.

Data Mining adalah serangkaian proses untuk menggali informasi yang belum diketahui dari suatu kumpulan data yang besar sehingga sering disebut

URL : <https://doi.org/10.29408/jprinter.v1i2.22005>

*Knowledge Discovery Database* (KDD).

Data mining adalah salah satu bentuk implementasi yang diterapkan untuk mencari sebuah model dan pola yang mampu melakukan prediksi pada suatu berdasarkan data sebelumnya dalam priode waktu tertentu.

## 2. Tinjauan Pustaka

### 2.1. Penelitian Terkait

- Penelitian yang dilakukan oleh Muhammad Wasil dan Mahfuz dalam jurnal yang berjudul “Analisis Persepsi Masyarakat Terhadap Kinerja Aparatur Desa Dalam Meningkatkan Pelayanan Publik Di Desa Lenek Lauk Menggunakan Metode Naïve Bayes” (2020) dapat disimpulkan bahwa berdasarkan analisa dan implementasi yang telah dilakukan serta proses evaluasi dan validasi menggunakan software rapid miner, maka hasil penelitian dengan metode naïve bayes memiliki tingkat akurasi yang tinggi untuk mengetahui puas atau tidak puasnya masyarakat terhadap kinerja pelayanan public di desa Lenek Lauk. Dari populasi yang berjumlah 5.539 penduduk di desa Lenek lauk, nilai akurasi yang didapatkan adalah sebanyak 92.26% dan nilai AUC pada Kurva ROC adalah sebanyak 0.924<sup>[3]</sup>.

- Penelitian yang dilakukan oleh Baiq Andriska Candra Permana dan Intan Komala Dewi Patwari yang berjudul

DOI : 10.29408/jprinter.v1i2.22005

“Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naïve Bayes Untuk Prediksi Penyakit Diabetes” (2021) dapat disimpulkan bahwa berdasarkan eksperimen yang telah dilakukan terhadap pasien penderita diabetes, untuk klasifikasi decision tree lebih baik dalam prediksi penyakit diabetes dengan nilai akurasi 95,58% dan nilai AUC 0,981 lebih tinggi dibandingkan naïve bayes dengan akurasi 87,69% dan nilai AUC 0,947<sup>[4]</sup>.

- Selanjutnya penelitian yang dilakukan oleh Nurhidayati dan Suhartini yang berjudul “Probabilitas Pembayaran Kredit Tepat Waktu Menggunakan Algoritma Naïve Bayes Pada Koperasi Serba Usaha Daruzzakah Rensing Lombok Timur” (2021) dapat disimpulkan bahwa setelah dilakukannya pengujian sebanyak 9 kali yaitu dari validation 2 sampai 10 maka didapatkan dua hasil akurasi terbaik dari K-Fold Validation 9 dan K-Fold Validation 10. K-Fold Validation 9 memiliki nilai akurasi sebesar 96,43% dan K-Fold Validation 10 memiliki nilai akurasi 96,45%. Dan selisih akurasi keduanya adalah 0,02%. Pada K-Fold Validation 10 dengan akurasi sebesar 96,45% dimana data dibagi menjadi 10 bagian untuk ditraining dan ditesting serta didapatkan juga nilai AUC (Area Under Curve) sebesar 0,942 dimana AUC merupakan parameter yang digunakan dalam analisis klasifikasi dalam

URL : <https://doi.org/10.29408/jprinter.v1i2.22005>

menentukan model terbaik untuk prediksi suatu kelas atau atribut<sup>[5]</sup>.

- Selanjutnya penelitian yang dilakukan oleh Nurhidayati, Yahya, Fathurrahman, L.M Samsu, dan Wajizatul Amnia yang berjudul “Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Penerima Beasiswa (Studi Kasus Universitas Hamzanwadi (2023))” dapat disimpulkan bahwa dari implementasi dan pengujian yang dilakukan peneliti dalam mengklasifikasi kelulusan pemberkasan mahasiswa universitas hamzanwadi yang mendaftarkan diri atau mengusulkan diri untuk mendapatkan beasiswa bidikmisi pada tahun 2021 menggunakan teknik data mining algoritma Naïve Bayes, dengan 10 kali pengujian menggunakan cross validation yaitu k-fold 2 sampai k-fold 10 didapatkan nilai akurasi tertinggi sebesar 91,43% dari k-fold 4 validation. Sedangkan AUC didapatkan 0,996% dengan tingkat diagnose excellent classification. Dari hasil akurasi dan nilai AUC yang didapatkan tersebut, maka dapat disimpulkan bahwa algoritma Naïve Bayes ini lebih akurat dan sangat bagus digunakan dalam menganalisis serta mengklasifikasi penerima beasiswa bidikmisi pada Universitas Hamzanwadi Selong<sup>[6]</sup>.

- Penelitian yang dilakukan oleh Mahpuz, Yahya, dan Muhammad Wasil yang berjudul “Implementasi Algoritma

DOI : 10.29408/jprinter.v1i2.22005

Decision Tree Untuk Mengetahui Faktor Kredit Macet Dan Lancar di Koperasi Serba Usaha Daruzzakah Rensing Lombok Timur” (2020) setelah dilakukannya pengujian sebanyak 9 kali yaitu validation 2 sampai dengan 10 maka didapatkan dua hasil akurasi terbaik dari K-Fold Validation 9 dan K-Fold Validation 10 yang masing-masing memiliki nilai accuracy sebesar 96,43% pada K-Fold Validation 9 dan 96,45% dari K-Fold Validation 10. Dan selisih akurasi keduanya adalah 0.02% yang menunjukkan bahwa terjadi peningkatan 0.02% pada rentang k-fold validation 9 dan 10. Peningkatan setiap uji coba tidaklah selalu sama ini dipengaruhi oleh seberapa banyak pembagian data yang dilakukan. Maka didapatkan nilai accuracy terbaik pada K-Fold Validation 10 dengan akurasi sebesar 96,45% dimana data dibagi menjadi 10 bagian untuk ditraining dan ditesting serta didapatkan juga nilai AUC (Area Under Curve) sebesar 0.942 yang dimana AUC merupakan parameter yang digunakan untuk analisis klasifikasi dalam menentukan prediksi suatu kelas atau atribut, AUC sendiri mempunyai rentang nilai dari 0 sampai 1, yang artinya semakin nilai AUC mendekati angka 1 maka prediksi atau diagnose atributnya semakin bagus. Ini menunjukkan nilai AUC sebesar 0.945 yang didapatkan setelah melakukan pengujian dengan K-Fold Validation 10

URL : <https://doi.org/10.29408/jprinter.v1i2.22005>

sangat baik karena hampir mendekati angka 1<sup>[7]</sup>.

## 2.2. Naïve Bayes

*Naïve Bayes* merupakan suatu bentuk klasifikasi data dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *Teorema Bayes*. Ciri utama dari Naïve Bayes Classifier ini adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi atau kejadian<sup>[5]</sup>.

*Naïve Bayes* termasuk metode klasifikasi yang sangat populer dan masuk dalam sepuluh algoritma terbaik dalam data mining. Naïve bayes merupakan sebuah pengklasifikasian probabilistic sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari data set yang diberikan. Algoritma menggunakan teorema bayes mengasumsikan bahwa semua atribut bersifat independen atau tidak dipengaruhi oleh faktor lain yang diberikan nilai pada variabel kelas. Algoritma Naïve Bayes merupakan salah satu algoritma machine learning yang banyak digunakan untuk tugas klasifikasi. Teorema Bayes memiliki bentuk umum sebagai berikut :

$$P(H|X) = P(H) * P(X|H)/P(X)$$

DOI : 10.29408/jprinter.v1i2.22005

Keterangan :

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik

P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probability)

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

P(X|H) : Probabilitas X berdasar kondisi hipotesis H

P(X) : Nilai probabilitas pada X

### 2.3. Prediksi

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi. Pengertian Prediksi sama dengan ramalan atau perkiraan<sup>[13]</sup>.

### 2.4. K-Fold Cross Validation

Metode k-fold cross validation melakukan generalisasi dengan membagi data ke dalam **K** bagian berukuran sama. Selama proses berlangsung, salah satu dari

URL : <https://doi.org/10.29408/jprinter.v1i2.22005>

partisi dipilih untuk data uji, dan sisanya digunakan untuk data latih (*training*). Metode k-fold cross validation menetapkan  $K = N$ , ukuran dari data set. Pendekatan ini memiliki keuntungan dalam penggunaan data sebanyak mungkin untuk di *training*. Test set secara efektif mencakup keseluruhan data set. K-fold cross validation adalah salah satu teknik untuk mengevaluasi keakuratan model<sup>[16]</sup>.

Cara kerja K-fold cross validation adalah sebagai berikut :

1. Seluruh data dibagi menjadi **K** bagian.
2. Fold ke-1 adalah ketika ke-1 menjadi data uji (*testing*) dan sisanya menjadi data latih (*training*). Selanjutnya, hitung akurasi berdasarkan porsi data tersebut.
3. Fold ke-2 adalah ketika bagian ke-2 menjadi data uji (*testing*) dan sisanya menjadi data latih (*training*). Selanjutnya, hitung akurasi berdasarkan porsi data tersebut.
4. Demikian seterusnya hingga mencapai fold ke-**K**.
5. Hitung rata-rata akurasi dari **N** buah akurasi di atas. Rata-rata akurasi ini menjadi akurasi final.

### 2.5. Confusion Matrix

Salah satu cara untuk menghitung akurasi adalah dengan menggunakan metode *confusion matrix*. *Confusion matrix* terdiri dari baris data uji yang diprediksi

DOI : 10.29408/jprinter.v1i2.22005

URL : <https://doi.org/10.29408/jprinter.v1i2.22005>

benar dan tidak benar suatu data oleh model klasifikasi<sup>[17]</sup>. *Confusion matrix* biasanya digambarkan dengan peta kebenaran yang digunakan untuk memperkirakan antara aktual dan prediksi.

*True Negatif* (TN) = jumlah data negatif yang terdeteksi dengan benar.

*False Positif* (FP) = data negatif namun terdeteksi sebagai data positif.

*False Negatif* (FN) merupakan data positif namun terdeteksi sebagai data negatif.

Metode ini menggunakan tabel matriks seperti terlihat pada tabel 1

## 2.6. Area Under The Curve (AUC)

berikut ini.

Tabel 1. Confussion Matrix

Prediction	Diklasifikasikan sebagai	
	Yes	No
Yes	True Positif (TP)	False Positif (FP)
No	False Negatif (FN)	True Negatif (TN)

Area Under Curve (AUC) adalah daerah yang letaknya di bawah Receiver Operating Characteristic (ROC) curve yang digunakan sebagai indikator pengukuran tingkat akurasi performansi dari masing-masing algoritma klasifikasi. AUC mengukur performansi metode klasifikasi berdasarkan ROC curve. Nilai AUC ditunjukkan dalam skala 0 sampai 1 dimana angka 0 menunjukkan tingkat negatif dan angka 1 menunjukkan tingkat positif. Klasifikasi nilai dari AUC sendiri dibagi menjadi 5 kelompok antara lain<sup>[18]</sup> :

Untuk menghitung akurasi dapat menggunakan persamaan dibawah ini :

1. 0.90 – 1.00 = Excellent Classification
2. 0.80 – 0.90 = Good Classification
3. 0.70 – 0.80 = Fair Classification
4. 0.60 – 0.70 = Poor Classification
5. 0.50 – 0.60 = Failure

- a. Akurasi merupakan presentase jumlah data yang dilakukan pada klasifikasi atau prediksi secara benar oleh algoritma.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{(\text{FP} + \text{FN} + \text{TP} + \text{TN})} \times 100\%$$

## 2.7. RapidMiner

Dimana :

*Rapidminer* merupakan perangkat lunak yang bersifat terbuka (open source). *Rapidminer* adalah sebuah solusi untuk melakukan analisis terhadap data mining,

*True Positif* (TP) = data positif yang terdeteksi benar.

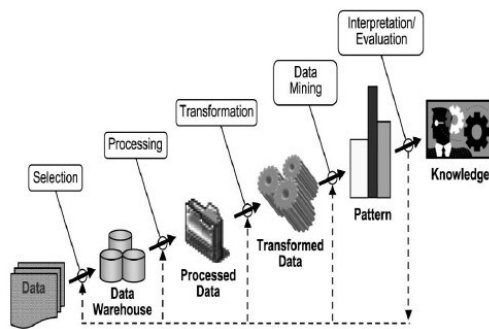
DOI : 10.29408/jprinter.v1i2.22005

URL : <https://doi.org/10.29408/jprinter.v1i2.22005>

text mining, dan analisis prediksi. *Rapidminer* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik.

## 2.8. Tahapan Data Mining

Data Mining merupakan salah satu dari rangkaian *knowledge discovery in database (KDD)*. Adapun serangkaian proses tahapan dalam Data Mining tersebut sebagai berikut<sup>[7]</sup>:



Gambar 1. Tahapan Data Mining

### a) Pembersihan Data (*Data Cleaning*)

Merupakan proses untuk membuang data yang tidak konsisten dan menghilangkan *noise*. Data yang diperoleh baik dari database suatu perusahaan maupun hasil eksperimen memiliki isian yang tidak sempurna seperti data yang hilang, data tidak valid atau bahkan sekedar salah ketik. Data-data yang tidak relevan lebih baik dibuang, pembersihan data sangat penting dilakukan karena akan

mempengaruhi performasi dari teknik data mining.

### b) Integrasi Data (*Data Integration*)

Merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan, dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyedatkan pengambilan aksi nantinya.

### c) Seleksi Data (*Data Selection*)

Data yang ada pada database seringkali tidak semuanya dipakai. Oleh karena itu, hanya data yang sesuai untuk dianalisis yang akan diambil dari database.

### d) Transformasi Data (*Data Transformasi*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan



DOI : 10.29408/jprinter.v1i2.22005

URL : <https://doi.org/10.29408/jprinter.v1i2.22005>

format data yang khusus sebelum bisa di aplikasikan.

e) Proses Mining

Proses mining merupakan proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

f) Evaluasi Pola (*Pattern Evaluation*)

Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi di evaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternative yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining yang lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil diluar dugaan yang mungkin bermanfaat.

### 3. Metode Penelitian

Dalam tahap pengumpulan data, penulis menggunakan metode-metode sebagai berikut :

1. Observasi

Dilakukan dengan mengumpulkan data secara langsung di lokasi

penelitian yaitu Kantor Desa Dames Damai Suralaga Lombok Timur berdasarkan data yang ada.

2. Wawancara

Melakukan tanya jawab dengan kepala desa dan staf desa Dames Damai untuk mengumpulkan keterangan mengenai data pajak bumi dan bangunan.

3. Studi Pustaka

Untuk menunjang metode observasi dan wawancara maka peneliti melakukan pengumpulan informasi dengan mencari referensi-referensi yang berkaitan dengan kasus serupa yang bisa didapatkan melalui buku jurnal, dan internet.

### 4. Hasil dan Pembahasan

Adapun dataset yang akan digunakan setelah melalui proses seleksi dan cleaning sebanyak 348 data yang akan diolah dan terdiri dari 8 atribut. Yang semula jumlah data awal sebanyak 5.826 data dari tahun 2018-2022.

Berikut adalah sampel datasets yang akan digunakan (10 datasets) :

DOI : 10.29408/jprinter.v1i2.22005

URL : <https://doi.org/10.29408/jprinter.v1i2.22005>

Tabel 2. Sampel datasets

NAMA WAJIB PAJAK	LUAS BUNDA MI	LUAS BANGUNAN	PBB YANG HARUS	TOTAL TAGIHAN	SUDAH BAYAR	PERSENTASE	KEPUTUSAN
NASRIE	##	72	####	#####	#####	80%	Patuh
MOH. SUJIRMAN	##	154	####	#####	#####	60%	Tidak Patuh
A BURN	##	30	####	#####	#####	##	Patuh
A SODIK	##	30	####	#####	#####	80%	Patuh
I MASTI	##	30	####	#####	#####	60%	Tidak Patuh
MQ KEL	##	30	####	#####	#####	80%	Patuh
SARBAE	##	30	####	#####	#####	80%	Patuh
I MAHN	##	54	####	#####	#####	80%	Patuh
SAMSUI	##	30	####	#####	#####	##	Patuh

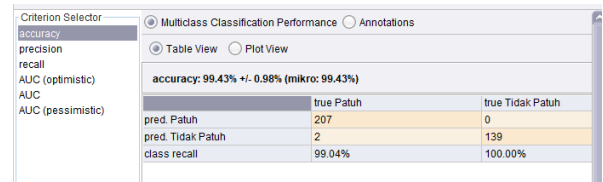
Berikut rincian hasil pengolahan data yang telah dilakukan pada software Rapidminer dari k-fold validation 2 sampai 10 menggunakan algoritma Naïve Bayes.

Tabel 3. Hasil Pengolahan K-Fold Validation 2 sampai 10

cross validation	Accuracy	Precision	Recall	AUC
2	98.28%	96.00%	100.00%	0.98
3	89.37%	83.82%	97.16%	0.96
4	95.11%	96.18%	90.44%	0.99
5	92.27%	89.65%	96.48%	0.97
6	93.68%	90.83%	97.10%	0.97
7	97.99%	96.80%	98.53%	0.98
8	99.43%	98.65%	100.00%	0.99
9	96.57%	94.57%	98.61%	0.98
10	96.29%	94.98%	97.14%	0.97

Berdasarkan hasil uji coba dapat disimpulkan bahwa, nilai *accuracy* tertinggi yaitu pada *K-Fold Validation* 8 dengan nilai *accuracy* sebesar 99.43%, precision 98.65%, recall 100.00%, dan AUC 0.991. Hasil tersebut didapatkan dengan metode *K-Fold Validation* 8 dimana data dibagi

menjadi 8 bagian untuk di *training* dan di *testing*.



Gambar 3. Hasil Akurasi *K-Fold Validation* 8

Keterangan :

1. *True Positif* (TP) = 207 dari 348 datasets hanya 227 data diprediksi sebagai data (Patuh).
2. *False Negatif* (FN) = 2 data diprediksi sebagai data (Patuh) ternyata masuk ke dalam kategori data (Tidak Patuh).
3. *True Negatif* (TN) = 139 data diprediksi sebagai data (Tidak Patuh).
4. *False Positif* (FP) = 0 data diprediksi sebagai data (Tidak Patuh) ternyata masuk ke dalam kategori data (Patuh)

Tabel 4. Hasil nilai algoritma Naïve Bayes model *X-Validation*

Nilai yang dihasilkan algoritma naïve bayes dengan model X-Validation		
<i>True_positive</i>	TP	207
<i>True_negative</i>	TN	139
<i>False_positive</i>	FP	0

DOI : 10.29408/jprinter.v1i2.22005

URL : <https://doi.org/10.29408/jprinter.v1i2.22005>

Nilai yang dihasilkan algoritma naïve bayes dengan model X-Validation

<i>False_negative</i>	FN	2
-----------------------	----	---

Untuk mencari nilai *accuracy*, *precision*, dan *recall* dapat dihitung menggunakan persamaan dibawah ini :

1. Mencari nilai akurasi menggunakan persamaan berikut

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{FP+FN+TP+TN} \\
 &= \frac{207+139}{0+2+207+139} \times 100 \\
 &\quad \% \\
 &= \frac{346}{348} \times 100\% \\
 &= 99.43 \%
 \end{aligned}$$

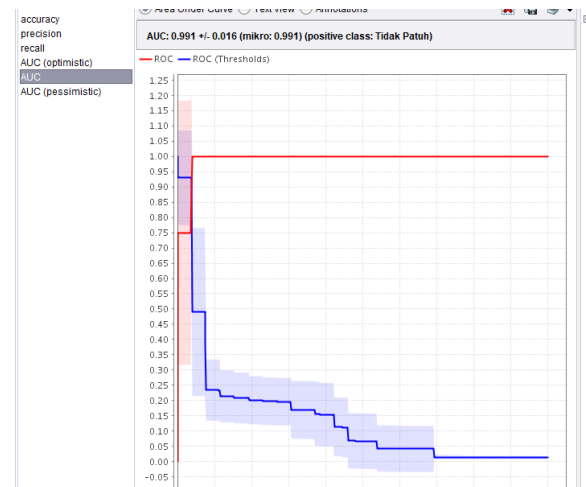
2. Mencari nilai precision menggunakan persamaan berikut

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{FP+TP} \\
 &= \frac{207}{0+207} \times 100 \% \\
 &= \frac{207}{207} \times 100 \% \\
 &= 100\%
 \end{aligned}$$

3. Mencari nilai recall menggunakan persamaan berikut

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{FN+TP} \\
 &= \frac{207}{2+207} \times 100 \% \\
 &= \frac{207}{209} \times 100 \% \\
 &= 99.04 \%
 \end{aligned}$$

Berdasarkan hasil perhitungan yang dilakukan diatas maka diperoleh nilai *accuracy* sebesar 99.43%, *precision* sebesar 100%, *recall* sebesar 99.04%, dan nilai AUC 0.991. adapun tampilan dari nilai AUC dapat dilihat pada gambar 4 di bawah ini.



Gambar 4. Hasil AUC dari *K-Fold Validation* 8

Gambar 4.22 menjelaskan grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.991 dengan tingkat diagnosa *Excellent Classification*.

## 5. Kesimpulan

Berdasarkan hasil pengolahan data yang dilakukan untuk “Memprediksi Tingkat Kepatuhan Wajib Pajak Di Desa Dames Damai”, dapat ditarik kesimpulan yaitu, Penerapan Data Mining pada aplikasi dengan menggunakan algoritma Naïve Bayes untuk memprediksi kepatuhan wajib pajak di desa dames damai menggunakan ketentuan “Patuh” dan “Tidak Patuh”.

DOI : 10.29408/jprinter.v1i2.22005

Setelah dilakukannya pengujian sebanyak 9 kali dari validation 2 sampai dengan 10 maka didapatkan hasil akurasi prediksi kepatuhan wajib pajak sebesar 99.43% dari k-fold validation 8 dengan nilai AUC (*Area Under Curve*) yang hampir mendekati angka 1 yaitu sebesar 0.991 dengan tingkat diagnosa *Excellent Classification*. Hasil ini membuktikan bahwa metode naïve bayes dapat digunakan untuk membantu “Memprediksi Tingkat Kepatuhan Wajib Pajak Di Desa Dames Damai”, karena menghasilkan tingkat keakuratan yang tinggi, sehingga kesadaran masyarakat terhadap pembayaran pajak bumi dan bangunan dapat dikatakan patuh.

#### Daftar Pustaka

- [1] K. rahayu Pustaka, K. Pemikiran, and D. A. N. Hipotesis, “kesadaran wajib pajak untuk meningkatkan kepatuhan pajak,” pp. 7–25, 2017.
- [2] Kamber, “Pengertian Klasifikasi Menurut Para Ahli,” pp. 3–13, 2011.
- [3] M. Wasil, “Analisis Persepsi Masyarakat Terhadap Kinerja Aparatur Desa Dalam Meningkatkan Pelayanan Publik Di Desa Lenek Lauk Menggunakan Metode Naive Bayes,” 2020.
- [4] B. A. Candra Permana and I. K. Dewi Patwari, “Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naïve Bayes Untuk Prediksi Penyakit Diabetes,” *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 63–69, Jan. 2021, doi: 10.29408/jit.v4i1.2994.
- [5] N. Nurhidayati and S. Suhartini, “Probabilitas Pembayaran Kredit
- URL : <https://doi.org/10.29408/jprinter.v1i2.22005>  
Tepat Waktu Menggunakan Algoritma Naïve Bayes Pada Koperasi Serba Usaha Daruzzakah Rensing Lombok Timur,” *Infotek J. Inform. dan Teknol.*, vol. 4, no. 2, pp. 194–204, Jul. 2021, doi: 10.29408/jit.v4i2.3573.
- [6] N. I. Nurhidayati, Y. Yahya, F. Fathurrahman, L. . Samsu, and W. Amnia, “Implementasi Algoritma Naive Bayes Untuk Klasifikasi Penerima Beasiswa (Studi Kasus Universitas Hamzanwadi),” *Infotek J. Inform. dan Teknol.*, vol. 6, no. 1, pp. 177–188, Jan. 2023, doi: 10.29408/jit.v6i1.7529.
- [7] M. Wasil, “Implementasi Algoritma Decision Tree Untuk Mengetahui Faktor Kredit Macet Dan Lancar di Koperasi Serba Usaha Daruzzakah Rensing Lombok Timur,” *J. Inform. dan Teknol.*, vol. 3, no. 2, pp. 93–103, 2020.
- [8] Y. Yahya and H. Bahtiar, “Pengaruh Pertumbuhan Ekonomi Terhadap Tingkat Kesejahteraan Masyarakat Kecamatan Selong Kabupaten Lombok Timur – Nusa Tenggara Barat Menggunakan Algoritma Naive Bayes,” *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 20–28, Jan. 2021, doi: 10.29408/jit.v4i1.2981.
- [9] B. H. Amri Muliawan Nur, “Komparasi Algoritma SVM dan SVM Berbasis PSO Dalam Menganalisa Kinerja Guru SMAN 3 Selong,” *Infotek J. Inform. dan Teknol.*, vol. 2, no. 2, pp. 86–94, 2019.
- [10] M. Saiful and S. Samsuddin, “Implementasi Algoritma Naive Bayes Untuk Memprediksi Predikat Ketuntasan Belajar Siswa Pasca Pandemi Covid 19,” *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 29–38, Jan. 2021, doi: 10.29408/jit.v4i1.2982.
- [11] R. Wariyanto Abdullah, D. Hartanti, H. Permatasari, A. Wicaksono Septyanto,

DOI : 10.29408/jprinter.v1i2.22005

URL : <https://doi.org/10.29408/jprinter.v1i2.22005>

- and Y. Abi Bagaskara, “Penerapan Data Mining untuk Memprediksi Jumlah Produk Terlaris Menggunakan Algoritma Naive Bayes Studi Kasus (Toko Prapti),” *J. Ilm. Inform. Glob.*, vol. 13, no. 2, p. 57154, 2022, doi: 10.36982/jiig.v13i1.2060.
- [12] A. R. Shaumi, M. F. Ali, and M. T. A. M. Simbolon, “Penerapan Data Mining Menggunakan Metode Teknik Classification Untuk Melihat Penerapan Data Mining Menggunakan Metode Teknik Classification Untuk,” *JUKI J. Komput. dan Inform.*, vol. 4, no. 2, pp. 171–181, 2022, [Online]. Available: [www.pajak.go.id](http://www.pajak.go.id)
- [13] Herdianto, “Pengertian Prediksi,” pp. 6–26, 2013.
- [14] M. H. Ma and S. Supatminingsih, “Faktor-Faktor Yang Berpengaruh Terhadap Kepatuhan Wajib Pajak dalam Membayar Pajak Bumi Dan Bangunan,” vol. 20, no. 2, pp. 276–284, 2020.
- [15] B. A. B. Ii, K. Pustaka, and D. A. N. Kerangka, “Pajak Bumi Dan Bangunan Menurut Para Ahli”.
- [16] C. Mustika 2015, “k-fold cross validation dan confusion matrix,” no. 2014, pp. 1–23, 2012.
- [17] B. A. B. Iii, “confusion matrix data mining,” pp. 1–8, 2014.
- [18] B. Kurang, M. Iain, S. Nurjati, and A. Maulana, “Komparasi Algoritma Data Mining Untuk Akurasi Penentuan,” 2012.
- [19] I. W. S. W. Dennis Aprilla C, Donny Aji Baskoro, Lia Ambarwati, *Belajar Data Mining dengan RapidMiner*. 2013.