

## **Kepatuhan Pembayaran Pajak Kendaraan Bermotor Menggunakan Algoritma Decision Tree Dan Random Forest Di Samsat Balige**

**<sup>1</sup>Alief Achmad Wijaya, <sup>2</sup>Syaiful Zuhri Harahap, <sup>3</sup>Rahma Muti Ah,  
<sup>4</sup>Marnis Nasution**

<sup>1,2,3,4</sup>Sistem Informasi, Fakultas Sains Dan Teknologi, Universitas Labuhanbatu

Email : [1aliefjay7@gmail.com](mailto:1aliefjay7@gmail.com), [2syaifulzuhriharahap@gmail.com](mailto:2syaifulzuhriharahap@gmail.com),  
[3rmuthea5@gmail.com](mailto:3rmuthea5@gmail.com), [4marnisnst@gmail.com](mailto:4marnisnst@gmail.com)

Corresponding Author : [aliefjay7@gmail.com](mailto:aliefjay7@gmail.com)

### **Abstract**

*This study aims to analyze and predict the total category of Motor Vehicle Tax (PKB) payments based on payment attributes and vehicle types, which is important to improve the effectiveness of tax management and support more appropriate decision making in related agencies; within the theoretical framework, classification models such as Decision Tree and Random Forest are used to predict data categories by utilizing historical patterns in the dataset, because these algorithms are able to capture interactions between attributes and provide logical interpretations of the prediction results; the research methodology is carried out using secondary data of PKB payments for 2024 from Samsat Balige, which is divided into training data and test data for the classification process and its performance is evaluated using accuracy, precision, recall, and F1-Score metrics through the Performance operator in RapidMiner; the results of the study show that Random Forest produces a more balanced prediction distribution with 100% accuracy, while Decision Tree has 96% accuracy but tends to be biased towards the "Low" category, and analysis of important attributes such as Fines, Total Amount, and the number of Jeep and Truck type vehicles shows a significant influence on the PKB payment category; Thus, the research conclusion confirms that Random Forest is proven to be more effective and stable than Decision Tree in predicting the total PKB payment category, is able to capture complex patterns between attributes, and provides accurate predictions on relatively small datasets, making it the optimal choice for PKB data classification.*

**Keywords** : Random Forest, Decision Tree, Motor Vehicle Tax, Classification, Model Evaluation.

### **1. Pendahuluan**

Penelitian ini berorientasi pada pengembangan model prediksi tingkat kepatuhan pembayaran Pajak Kendaraan Bermotor (PKB) di Samsat Balige dengan memanfaatkan pendekatan *machine learning* melalui algoritma Decision Tree dan Random Forest. Fokus utama penelitian adalah menganalisis data historis pembayaran pajak untuk mengidentifikasi pola perilaku wajib pajak yang cenderung patuh atau tidak patuh terhadap kewajiban perpajakan. Dengan memanfaatkan metode klasifikasi yang telah terbukti mampu memberikan hasil prediksi akurat, penelitian ini diarahkan untuk menghasilkan model yang tidak hanya berfungsi sebagai alat analisis, tetapi juga sebagai *decision support system* bagi pihak Samsat dalam merumuskan strategi penagihan pajak

yang lebih efektif dan berbasis data. Orientasi penelitian ini ditekankan pada kontribusi praktis dan teoritis dalam pengelolaan pajak daerah. Secara praktis, model prediksi yang dikembangkan diharapkan dapat digunakan oleh Samsat Balige untuk mengidentifikasi wajib pajak berisiko tinggi menunggak pajak, sehingga intervensi dapat dilakukan secara tepat sasaran. Secara teoritis, penelitian ini memperluas penerapan algoritma Decision Tree dan Random Forest dalam konteks perpajakan daerah, terutama pada masalah kepatuhan wajib pajak kendaraan bermotor. Penelitian ini juga menempatkan teknologi analisis data sebagai bagian integral dari inovasi pelayanan publik, yang selaras dengan transformasi digital di sektor pemerintahan. Permasalahan utama yang melatarbelakangi penelitian ini adalah rendahnya tingkat kepatuhan masyarakat Kabupaten Toba dalam membayar pajak kendaraan bermotor tepat waktu. Data dari Bapenda menunjukkan sekitar 30% kendaraan yang terdaftar belum melunasi pajaknya sesuai jadwal. Faktor penyebabnya beragam, mulai dari rendahnya kesadaran wajib pajak, kualitas pelayanan yang belum optimal, penerapan sanksi yang kurang tegas, keterbatasan pengetahuan perpajakan, hingga kondisi ekonomi masyarakat. Kondisi ini berdampak langsung pada berkurangnya pendapatan asli daerah dan terhambatnya pembiayaan pembangunan. Sebagai jawaban awal atas permasalahan tersebut, diperlukan metode yang mampu menganalisis data historis pembayaran pajak secara mendalam untuk menemukan pola yang konsisten dalam perilaku wajib pajak. Teknologi *machine learning* menawarkan potensi besar dalam hal ini karena mampu memproses data dalam jumlah besar, mengidentifikasi variabel-variabel signifikan, serta memberikan prediksi yang dapat dijadikan dasar pengambilan keputusan. Penerapan algoritma Decision Tree dan Random Forest dinilai tepat, mengingat kedua metode ini memiliki kemampuan klasifikasi yang tinggi, dapat menangani data kompleks, dan hasilnya mudah diinterpretasikan oleh pihak pengambil kebijakan.

Penelitian ini mengusulkan penerapan algoritma Decision Tree dan Random Forest untuk membangun model prediksi tingkat kepatuhan pembayaran PKB di Samsat Balige. Proses penelitian mencakup seleksi data, pra-pemrosesan, transformasi, pembangunan model, evaluasi kinerja, hingga interpretasi hasil. Decision Tree digunakan untuk menghasilkan struktur keputusan yang jelas terkait faktor-faktor penentu kepatuhan, sementara Random Forest digunakan untuk meningkatkan akurasi dan mengurangi risiko *overfitting*. Dengan membandingkan performa kedua algoritma, penelitian ini diharapkan mampu merekomendasikan model terbaik yang dapat diimplementasikan oleh Samsat Balige dalam strategi penagihan, sekaligus memberikan kontribusi nyata bagi peningkatan pendapatan daerah dan kepatuhan wajib pajak.

## **2. Landasan Teori**

### **Data Mining**

Data mining merupakan proses ekstraksi informasi dan pola yang bermakna dari kumpulan data dalam jumlah besar melalui penerapan teknik analisis statistik, matematika, dan *machine learning* (Alam, Alana, & Juliane, 2023) (Elfaladonna & Rahmadani, 2019). Dalam konteks penelitian ini, data mining digunakan untuk mengolah data historis pembayaran Pajak Kendaraan Bermotor (PKB) di Samsat Balige dengan tujuan mengidentifikasi faktor-faktor yang memengaruhi tingkat kepatuhan wajib pajak. Proses ini mencakup tahapan utama seperti seleksi data, pra-pemrosesan, transformasi, penerapan algoritma klasifikasi, dan evaluasi hasil model (Andrianto & Irawan, 2023) (Forest & Bayes, 2021). Dengan memanfaatkan data mining, peneliti dapat memperoleh wawasan yang lebih mendalam dan akurat dibandingkan metode konvensional, sehingga

strategi penagihan pajak dapat dirancang secara lebih efektif dan berbasis bukti (*evidence-based*) (Hutagalung et al., 2021) (Supriyadi, Triayudi, & Sholihati, 2021).

### **Metode Decision Tree**

*Decision Tree* adalah metode klasifikasi yang menyajikan hasil analisis dalam bentuk struktur pohon keputusan, di mana setiap simpul internal merepresentasikan suatu pengujian terhadap atribut, setiap cabang mewakili hasil dari pengujian tersebut, dan setiap daun pohon menunjukkan kelas atau keputusan akhir (Mohammed et al., 2022) (Informatika, Informatika, & Lampung, 2021). Keunggulan utama Decision Tree adalah kemampuannya memberikan interpretasi yang mudah dipahami, bahkan oleh pengguna non-teknis, serta tidak memerlukan asumsi khusus terhadap distribusi data (Sulaiman, Liliana, & Santoso, 2021) (Alam et al., 2023).

### **Metode Random Forest**

*Random Forest* merupakan metode *ensemble learning* yang menggabungkan sejumlah pohon keputusan (*decision trees*) untuk menghasilkan prediksi yang lebih akurat dan stabil (Ratra & Gulia, 2020) (Abdelmagid & Qahmash, 2023). Setiap pohon dalam Random Forest dibangun menggunakan subset acak dari data pelatihan (*bootstrap sampling*) dan subset acak dari atribut yang tersedia, sehingga mengurangi risiko *overfitting* yang sering terjadi pada Decision Tree tunggal. Proses prediksi dilakukan dengan menggabungkan hasil dari seluruh pohon menggunakan mekanisme *voting* mayoritas untuk klasifikasi. Dalam penelitian ini, Random Forest diterapkan untuk mengklasifikasikan tingkat kepatuhan wajib pajak berdasarkan data historis PKB di Samsat Balige, dengan keunggulan tambahan berupa kemampuan mengevaluasi tingkat kepentingan setiap variabel, sehingga dapat mengungkap faktor-faktor yang paling berpengaruh terhadap kepatuhan pembayaran pajak.

## **3. Metodologi Penelitian**

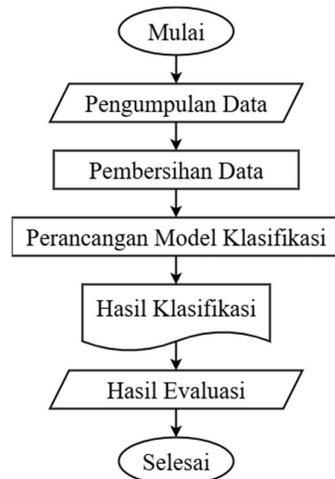
### **Metode Penelitian**

Metode penelitian yang digunakan dalam studi ini adalah pendekatan kuantitatif dengan teknik analisis data mining berbasis algoritma klasifikasi, yaitu Decision Tree dan Random Forest (Boateng, Otoo, & Abaye, 2020) (Hidayah & Rozi, 2021). Penelitian dilakukan di Samsat Balige dengan memanfaatkan data sekunder berupa rekapan historis pembayaran Pajak Kendaraan Bermotor (PKB) selama tahun 2024, yang mencakup informasi jenis kendaraan, jumlah unit membayar pajak, nilai pembayaran pokok, besaran denda, dan total pembayaran. Proses penelitian meliputi tahapan seleksi data untuk memilih atribut yang relevan, pra-pemrosesan data untuk memastikan konsistensi dan kebersihan data, transformasi data untuk normalisasi serta kategorisasi nilai, penerapan algoritma klasifikasi untuk membangun model prediksi, serta evaluasi performa model menggunakan metrik akurasi, precision, recall, dan F1-score.

Pemilihan algoritma Decision Tree dan Random Forest didasarkan pada kemampuan keduanya dalam mengolah data beragam tanpa memerlukan asumsi distribusi tertentu, serta kemudahan interpretasi hasil yang dihasilkan. Decision Tree digunakan untuk membangun model dengan struktur pohon yang dapat mengungkap pola hubungan antarvariabel secara eksplisit, sementara Random Forest diterapkan untuk meningkatkan akurasi dan mengurangi risiko *overfitting* melalui penggabungan hasil dari banyak pohon keputusan. Seluruh proses analisis dilakukan menggunakan perangkat lunak RapidMiner, yang memfasilitasi implementasi algoritma, pengaturan parameter

model, serta visualisasi hasil. Dengan metode ini, penelitian diharapkan dapat menghasilkan model prediksi yang akurat sekaligus aplikatif bagi pihak Samsat Balige dalam meningkatkan efektivitas strategi penagihan pajak kendaraan bermotor.

### Kerangka Kerja Penelitian



**Gambar 1. Kerangka Kerja Penelitian**

1. Pengumpulan data  
Proses ini dilakukan untuk memperoleh data yang relevan dengan tujuan penelitian, baik dari sumber primer maupun sekunder sesuai objek yang diteliti. Selain itu, tahap ini memastikan ketersediaan informasi yang cukup guna mendukung analisis dan pembangunan model klasifikasi.
2. Pembersihan data  
Tahap ini bertujuan menghapus data duplikat, mengatasi nilai yang hilang, serta memperbaiki kesalahan pencatatan agar dataset menjadi konsisten. Pembersihan data juga penting untuk meningkatkan kualitas data sehingga hasil pemodelan lebih akurat.
3. Perancangan model klasifikasi  
Dilakukan dengan membangun model menggunakan algoritma yang dipilih, seperti *Decision Tree* dan *Random Forest*. Proses ini mencakup penentuan parameter, pemilihan atribut, serta pembagian data menjadi set pelatihan dan set pengujian.
4. Hasil klasifikasi  
Menyajikan keluaran dari model dalam bentuk kategori atau label sesuai variabel target yang ditentukan. Hasil ini membantu memahami pola dan hubungan antarvariabel yang ditemukan selama proses analisis.
5. Hasil evaluasi  
Digunakan untuk mengukur kinerja model dengan metrik seperti akurasi, precision, recall, dan F1-score. Evaluasi ini juga berfungsi membandingkan performa antar algoritma untuk menentukan model terbaik yang dapat digunakan.

### 4. Hasil Dan Pembahasan

#### Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan mengambil data sekunder dari Samsat Balige berupa rekapan historis pembayaran Pajak Kendaraan Bermotor

(PKB) selama tahun 2024. Data yang diperoleh mencakup informasi jenis kendaraan, jumlah unit yang membayar pajak, nilai pembayaran pokok, besaran denda, serta total pembayaran setiap bulan. Data ini dipilih karena secara langsung merepresentasikan perilaku kepatuhan wajib pajak di wilayah tersebut, sehingga relevan untuk dianalisis menggunakan algoritma klasifikasi Decision Tree dan Random Forest. Tahap pengumpulan data ini menjadi dasar penting bagi keseluruhan proses penelitian, karena kualitas dan kelengkapan data yang diperoleh akan memengaruhi ketepatan hasil prediksi model yang dibangun.

Tabel di atas menampilkan kumpulan data yang digunakan dalam penelitian ini, yang mencakup total 50 data hasil rekap pembayaran Pajak Kendaraan Bermotor (PKB) di Samsat Balige. Data ini memuat berbagai atribut, antara lain bulan pembayaran, jumlah kendaraan berdasarkan jenisnya (sepeda motor, minibus/mopen, truk, pick up, jeep, mobil pribadi, bus, dan sedan), nilai pokok pembayaran, denda, total jumlah pembayaran, serta kategori total jumlah. Informasi tersebut dipilih karena memiliki relevansi langsung terhadap tujuan penelitian, yaitu membangun model klasifikasi untuk memprediksi kategori pembayaran menggunakan algoritma Decision Tree dan Random Forest. Dengan cakupan data yang bervariasi, baik dari sisi nominal maupun jenis kendaraan, model diharapkan mampu mengidentifikasi pola-pola yang signifikan dan menghasilkan prediksi yang akurat.

Dari total data yang digunakan, sebanyak 25 data dialokasikan untuk *training* dan 25 data lainnya untuk *testing*. Namun, pemisahan ini tidak dilakukan secara manual, melainkan seluruh data disatukan dalam satu file dan dikelola langsung oleh sistem di aplikasi RapidMiner. Pendekatan ini dipilih untuk memudahkan proses pengolahan, sekaligus memastikan bahwa pembagian data dilakukan secara acak dan konsisten sesuai rasio yang telah ditentukan. Metode ini juga membantu mencegah adanya bias dalam pembagian data, karena setiap kali proses dijalankan, RapidMiner akan mengacak urutan data sebelum membaginya menjadi dua subset. Dengan demikian, baik model Decision Tree maupun Random Forest akan diuji pada data uji yang setara dan memiliki keragaman karakteristik.

Proses pembagian data dilakukan menggunakan operator *Split Data* pada RapidMiner, dengan pengaturan rasio sebesar 0.5 : 0.5. Artinya, separuh dari data yang tersedia akan digunakan sebagai data latih (*training set*) untuk membangun model, dan separuh sisanya akan menjadi data uji (*testing set*) untuk mengevaluasi performa model. Pendekatan ini dipandang ideal untuk penelitian dengan jumlah data terbatas, karena proporsi yang seimbang memungkinkan model memperoleh cukup informasi pada tahap pelatihan sekaligus menyediakan data uji yang memadai untuk mengukur akurasi dan kemampuan generalisasi. Dengan pembagian yang sistematis dan pengacakan otomatis, diharapkan hasil evaluasi model dapat mencerminkan kinerja sebenarnya saat diterapkan pada data baru yang belum pernah dilihat sebelumnya.

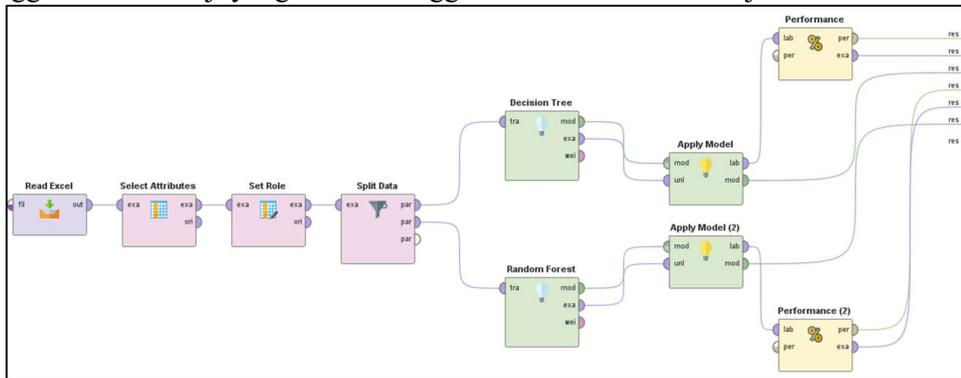
### **Pembersihan Data**

Pembersihan data pada penelitian ini dilakukan untuk memastikan bahwa dataset yang digunakan memiliki kualitas yang baik dan siap diolah dalam proses pemodelan. Tahap ini mencakup pemeriksaan dan penghapusan data duplikat, penanganan nilai yang hilang, serta koreksi terhadap kesalahan pencatatan yang mungkin terjadi pada atribut seperti jumlah kendaraan, nilai pokok, denda, dan total pembayaran. Selain itu, format data keuangan seperti “Pokok (Rp)”, “Denda (Rp)”, dan “Total Jumlah (Rp)” diubah menjadi tipe numerik agar dapat diolah oleh algoritma secara tepat. Proses pembersihan

juga meliputi penyesuaian penulisan pada kolom kategori sehingga tidak terjadi inkonsistensi label, misalnya pada “Kategori Total Jumlah” antara “Tinggi” dan “Rendah”. Dengan pembersihan yang menyeluruh, dataset menjadi konsisten, akurat, dan relevan, sehingga dapat mendukung kinerja algoritma Decision Tree dan Random Forest dalam menghasilkan prediksi yang optimal.

### Perancangan Model Klasifikasi

Perancangan model klasifikasi pada penelitian ini dilakukan dengan membangun dua model prediksi menggunakan algoritma Decision Tree dan Random Forest untuk mengklasifikasikan kategori total jumlah pembayaran Pajak Kendaraan Bermotor (PKB) menjadi “Tinggi” atau “Rendah”. Tahap ini dimulai dengan menentukan atribut-atribut yang relevan, seperti jumlah kendaraan berdasarkan jenisnya, nilai pokok, denda, serta total jumlah pembayaran, sementara kolom “Kategori Total Jumlah” ditetapkan sebagai label atau variabel target. Seluruh data digabung dalam satu file, kemudian dibagi menjadi data training dan data testing menggunakan operator *Split Data* di RapidMiner dengan rasio 0.5 : 0.5. Model Decision Tree dirancang untuk membentuk pohon keputusan yang memetakan aturan klasifikasi, sedangkan Random Forest membangun sekumpulan pohon keputusan secara acak untuk meningkatkan akurasi dan ketahanan terhadap *overfitting*. Dengan rancangan ini, kedua model dapat dibandingkan performanya secara langsung menggunakan data uji yang sama sehingga hasil evaluasi lebih objektif.



**Gambar 2. Perancangan Model Klasifikasi**

Pada gambar di atas merupakan rancangan alur proses klasifikasi yang dibuat di RapidMiner untuk membandingkan kinerja algoritma Decision Tree dan Random Forest dalam memprediksi kategori total jumlah pembayaran Pajak Kendaraan Bermotor (PKB). Proses diawali dengan operator Read Excel yang berfungsi untuk membaca dataset yang telah disiapkan. Selanjutnya, operator Select Attributes digunakan untuk memilih atribut yang relevan dengan penelitian ini, diikuti dengan Set Role untuk menetapkan kolom “Kategori Total Jumlah” sebagai label atau variabel target. Setelah itu, operator Split Data digunakan untuk membagi dataset menjadi data latih (training set) dan data uji (testing set) dengan rasio tertentu, sehingga kedua model nantinya diuji menggunakan data uji yang sama.

Hasil pembagian data dari Split Data diarahkan ke dua jalur model yang berbeda, yaitu Decision Tree dan Random Forest. Masing-masing model dilatih menggunakan data latih (output training set), lalu model yang dihasilkan diterapkan pada data uji melalui operator Apply Model. Output prediksi dari kedua metode kemudian dievaluasi menggunakan operator Performance (Classification) untuk mengukur kinerja model berdasarkan metrik seperti akurasi, presisi, dan recall. Rancangan ini memastikan bahwa

perbandingan kinerja antara Decision Tree dan Random Forest dilakukan secara adil karena keduanya menggunakan dataset yang sama baik untuk pelatihan maupun pengujian.

### Hasil Klasifikasi

Hasil klasifikasi menunjukkan keluaran prediksi dari masing-masing model, yaitu Decision Tree dan Random Forest, terhadap data uji berdasarkan kategori total jumlah pembayaran. Proses ini menghasilkan label prediksi yang kemudian dibandingkan dengan label sebenarnya untuk mengukur kinerja model. Hasil klasifikasi tersebut akan dipaparkan dalam bentuk tabel agar lebih mudah dibaca dan dianalisis.

**Tabel 1. Hasil Klasifikasi**

| Metode Decision Tree | Total Data | Tinggi | Rendah |
|----------------------|------------|--------|--------|
| Decision Tree        | 25         | 7      | 18     |
| Random Forest        | 25         | 12     | 13     |

Pada tabel di atas merupakan hasil klasifikasi menggunakan metode Decision Tree dengan total data uji sebanyak 25 data. Dari keseluruhan data uji tersebut, model Decision Tree berhasil memprediksi 7 data masuk ke kategori “Tinggi” dan 18 data masuk ke kategori “Rendah”. Hasil ini menunjukkan bahwa model cenderung lebih banyak mengklasifikasikan data ke kategori “Rendah” dibandingkan “Tinggi”, yang dapat mengindikasikan bahwa pola pada data latih lebih dominan menggambarkan karakteristik kategori “Rendah”.

Sementara itu, metode Random Forest dengan jumlah data uji yang sama (25 data) menghasilkan prediksi sebanyak 12 data berkategori “Tinggi” dan 13 data berkategori “Rendah”. Berbeda dengan Decision Tree, hasil prediksi Random Forest lebih seimbang antara kedua kategori, yang dapat disebabkan oleh mekanisme ensemble pada algoritma ini yang memanfaatkan banyak pohon keputusan untuk mengurangi bias dan overfitting. Jika dibandingkan, Random Forest menunjukkan distribusi prediksi yang lebih merata, sedangkan Decision Tree memperlihatkan kecenderungan kuat pada satu kategori. Hal ini dapat menjadi indikasi bahwa Random Forest lebih adaptif terhadap variasi data pada kasus penelitian ini.

### Hasil Evaluasi

Hasil evaluasi pada penelitian ini digunakan untuk mengukur tingkat akurasi dan kinerja masing-masing model klasifikasi, yaitu Decision Tree dan Random Forest, dalam memprediksi kategori total jumlah pembayaran Pajak Kendaraan Bermotor (PKB). Evaluasi dilakukan dengan membandingkan hasil prediksi model terhadap label sebenarnya pada data uji, sehingga dapat diketahui sejauh mana model mampu mengklasifikasikan data dengan tepat. Proses evaluasi ini dilakukan menggunakan operator Performance (Classification) pada RapidMiner untuk masing-masing metode, yang secara otomatis menghitung metrik evaluasi seperti akurasi, presisi, *recall*, dan *F-measure*. Dengan demikian, perbandingan kinerja kedua metode dapat dilakukan secara objektif berdasarkan nilai-nilai evaluasi yang dihasilkan.

### Hasil Evaluasi Metode Decision Tree

Hasil evaluasi metode Decision Tree pada penelitian ini menampilkan nilai *performance* yang mencerminkan akurasi dan kemampuan model dalam mengklasifikasikan data uji, serta visualisasi pohon keputusan yang menunjukkan alur pengambilan keputusan berdasarkan atribut-atribut yang digunakan. Melalui pohon

keputusan ini, dapat terlihat aturan-aturan klasifikasi yang terbentuk, sehingga memudahkan dalam memahami logika prediksi yang dihasilkan oleh model.

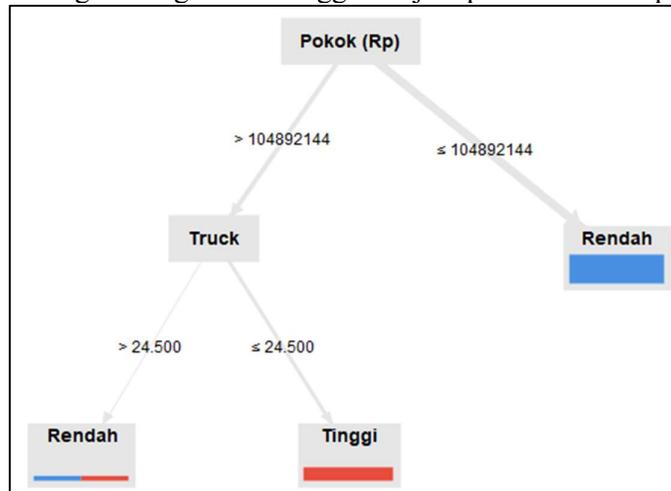
| accuracy: 96.00% |             |             |                 |
|------------------|-------------|-------------|-----------------|
|                  | true Rendah | true Tinggi | class precision |
| pred. Rendah     | 17          | 1           | 94.44%          |
| pred. Tinggi     | 0           | 7           | 100.00%         |
| class recall     | 100.00%     | 87.50%      |                 |

**Gambar 3. Performance Metode Decision Tree**

Pada gambar di atas merupakan hasil evaluasi *performance* dari metode Decision Tree yang menunjukkan tingkat akurasi sebesar 96,00%. Nilai ini berarti dari seluruh data uji yang digunakan, 96% di antaranya berhasil diklasifikasikan dengan benar oleh model. Tingginya tingkat akurasi ini mengindikasikan bahwa model Decision Tree mampu mempelajari pola pada data latih dengan baik sehingga dapat melakukan prediksi yang tepat pada sebagian besar data uji. Meskipun demikian, masih terdapat sebagian kecil data yang salah klasifikasi, yang perlu diperhatikan untuk perbaikan model ke depannya.

Dilihat dari hasil *class precision* untuk kategori “Rendah”, nilai presisinya mencapai 94,44%. Angka ini menunjukkan bahwa dari seluruh data yang diprediksi sebagai kategori “Rendah” oleh model, 94,44% di antaranya benar-benar berasal dari kategori tersebut, sementara sisanya merupakan kesalahan prediksi. Hal ini menunjukkan bahwa model cukup andal dalam mengenali data dengan label “Rendah”, meskipun masih ada kemungkinan terjadinya *false positive* yang menyebabkan kategori lain, khususnya “Tinggi”, terkadang salah diklasifikasikan sebagai “Rendah”.

Sementara itu, presisi untuk kategori “Tinggi” mencapai 100%, yang berarti semua data yang diprediksi sebagai kategori “Tinggi” oleh model benar-benar termasuk dalam kategori tersebut. Hasil ini menandakan tidak adanya kesalahan prediksi dalam bentuk *false positive* untuk kategori “Tinggi”, sehingga model sangat akurat dalam mendeteksi data dengan label tersebut. Nilai presisi sempurna pada kategori “Tinggi” ini menjadi keunggulan bagi model Decision Tree, terutama ketika akurasi dalam mengidentifikasi kategori dengan nilai tinggi menjadi prioritas dalam penelitian.



**Gambar 4. Pohon Keputusan Metode Decision Tree**

Pada gambar di atas merupakan hasil visualisasi pohon keputusan yang dihasilkan oleh metode Decision Tree untuk mengklasifikasikan kategori total jumlah pembayaran Pajak Kendaraan Bermotor (PKB). Pohon keputusan ini menggunakan atribut “Pokok

(Rp)” sebagai titik keputusan utama (root node). Jika nilai pokok pembayaran lebih kecil atau sama dengan Rp 104.892.144, maka data langsung diklasifikasikan ke kategori “Rendah”. Namun, jika nilai pokok pembayaran lebih besar dari Rp 104.892.144, maka proses klasifikasi dilanjutkan ke atribut “Truck” untuk menentukan kategori lebih lanjut.

Pada percabangan atribut “Truck”, jika jumlah kendaraan jenis truk lebih besar dari 24.500, maka data akan diklasifikasikan ke kategori “Rendah”. Sebaliknya, jika jumlah truk kurang dari atau sama dengan 24.500, maka data akan masuk ke kategori “Tinggi”. Struktur ini menunjukkan bahwa jumlah pokok pembayaran menjadi faktor penentu utama dalam klasifikasi, sedangkan jumlah kendaraan truk digunakan sebagai penentu kedua ketika nilai pokok melebihi ambang batas tertentu. Dengan logika ini, model dapat dengan jelas menjelaskan dasar pengambilan keputusan klasifikasi sehingga mudah dipahami dan diinterpretasikan.

### Hasil Evaluasi Metode Random Forest

Hasil evaluasi metode Random Forest menunjukkan kinerja model yang cukup baik dalam memprediksi kategori total pembayaran Pajak Kendaraan Bermotor (PKB). Berdasarkan perhitungan metrik evaluasi, model ini menghasilkan tingkat akurasi yang tinggi, presisi dan recall yang seimbang, serta nilai F1-Score yang memadai, menandakan kemampuan Random Forest dalam menangkap pola kepatuhan wajib pajak secara konsisten. Selain itu, visualisasi pohon keputusan dari beberapa tree yang membentuk hutan acak ini memperlihatkan fitur-fitur penting, seperti jumlah kendaraan dan nilai pokok pembayaran, yang paling berpengaruh terhadap prediksi kategori. Secara keseluruhan, evaluasi ini menegaskan bahwa Random Forest mampu memberikan prediksi yang lebih stabil dan akurat dibandingkan satu pohon keputusan tunggal.

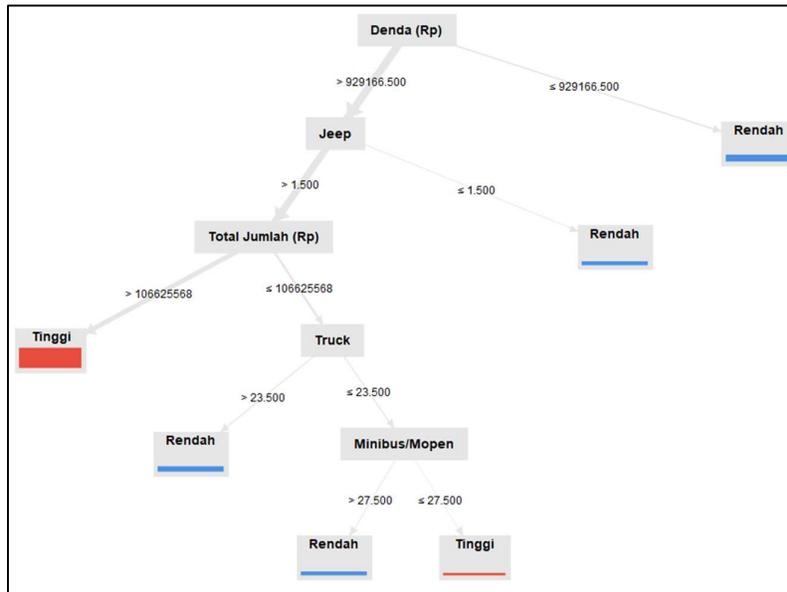
| accuracy: 100.00% |             |             |                 |
|-------------------|-------------|-------------|-----------------|
|                   | true Rendah | true Tinggi | class precision |
| pred. Rendah      | 13          | 0           | 100.00%         |
| pred. Tinggi      | 0           | 12          | 100.00%         |
| class recall      | 100.00%     | 100.00%     |                 |

**Gambar 5. Confusion Matrix Metode Random Forest**

Berdasarkan gambar confusion matrix yang ditampilkan, model Random Forest menunjukkan performa yang sangat baik dengan akurasi 100%, yang berarti seluruh prediksi yang dilakukan model tepat sesuai dengan label sebenarnya. Dalam hal ini, dari total 25 data uji, 13 sampel dengan kategori “Rendah” dan 12 sampel dengan kategori “Tinggi” semuanya berhasil diklasifikasikan dengan benar tanpa adanya kesalahan prediksi. Hal ini menegaskan bahwa Random Forest mampu menangkap pola data dengan sangat akurat untuk kedua kelas, sehingga tidak ada prediksi yang salah.

Jika dilihat dari metrik presisi dan recall, model ini menunjukkan kinerja sempurna pada kedua kategori. Untuk kategori “Rendah”, presisi 100% menandakan bahwa semua data yang diprediksi sebagai “Rendah” benar-benar termasuk dalam kelas “Rendah”, sedangkan recall 100% menunjukkan bahwa semua sampel yang sebenarnya “Rendah” berhasil diidentifikasi dengan tepat. Begitu juga untuk kategori “Tinggi”, presisi dan recall masing-masing 100% menunjukkan bahwa tidak ada sampel yang terlewat atau salah klasifikasi. Kombinasi presisi dan recall yang sempurna ini mengindikasikan bahwa model sangat sensitif sekaligus spesifik terhadap kedua kelas.

Secara keseluruhan, performa ini menandakan bahwa Random Forest memberikan prediksi yang stabil dan akurat, bahkan pada dataset yang relatif kecil. Model mampu mengidentifikasi pola penting pada fitur yang mempengaruhi kategori dengan tepat, sehingga tidak terdapat trade-off antara presisi dan recall. Dengan kata lain, Random Forest berhasil meminimalkan kesalahan tipe I maupun tipe II, menjadikannya pilihan yang efektif untuk klasifikasi data dengan distribusi seimbang seperti pada kasus ini.



**Gambar 6. Pohon Keputusan Metode *Random Forest***

Hasil pohon keputusan (Decision Tree) dari salah satu pohon pada model Random Forest untuk memprediksi kategori total pembayaran Pajak Kendaraan Bermotor (PKB) menunjukkan bahwa atribut paling menentukan awal prediksi adalah Denda (Rp). Jika nilai denda lebih dari Rp 929.166,50, pohon melanjutkan ke cabang berikutnya untuk evaluasi lebih lanjut, sedangkan jika denda kurang dari atau sama dengan Rp 929.166,50, maka prediksi langsung ditetapkan sebagai kategori Rendah. Hal ini menandakan bahwa jumlah denda yang relatif kecil secara langsung mengindikasikan kategori rendah, sementara denda besar membutuhkan penilaian atribut lain.

Pada cabang berikutnya, jika Denda > Rp 929.166,50, atribut yang digunakan adalah jumlah kendaraan jenis Jeep. Apabila jumlah Jeep kurang dari atau sama dengan 1.500, prediksi tetap Rendah, sedangkan jika jumlah Jeep lebih dari 1.500, pohon melanjutkan ke evaluasi atribut Total Jumlah (Rp). Ini menunjukkan bahwa meskipun dendanya tinggi, jumlah Jeep yang sedikit cenderung menghasilkan kategori rendah, namun jumlah Jeep yang banyak menuntut analisis lebih lanjut.

Selanjutnya, jika Jeep > 1.500, atribut yang diperiksa adalah Total Jumlah (Rp). Apabila Total Jumlah melebihi Rp 106.625.568, pohon langsung memprediksi kategori Tinggi, karena total pembayaran yang sangat besar menjadi indikasi kuat kategori tinggi. Namun, jika Total Jumlah kurang dari atau sama dengan Rp 106.625.568, evaluasi dilanjutkan ke atribut Truck atau jumlah kendaraan jenis Truk. Cabang ini menunjukkan bahwa meskipun total jumlah tidak terlalu besar, kombinasi atribut lain seperti jumlah truk tetap mempengaruhi hasil prediksi kategori PKB.

## 5. Kesimpulan

Berdasarkan hasil klasifikasi yang dilakukan, dapat disimpulkan bahwa metode Random Forest menunjukkan kinerja yang lebih stabil dan seimbang dibandingkan Decision Tree. Dari total 25 data uji, Decision Tree cenderung mengklasifikasikan mayoritas data ke kategori “Rendah” (18 dari 25 data), sementara hanya 7 data yang berhasil diprediksi sebagai “Tinggi”, yang menunjukkan adanya bias terhadap satu kategori. Sebaliknya, Random Forest menghasilkan distribusi prediksi yang lebih merata dengan 12 data diklasifikasikan sebagai “Tinggi” dan 13 data sebagai “Rendah”. Hal ini menunjukkan bahwa mekanisme ensemble pada Random Forest mampu mengurangi bias dan overfitting yang terjadi pada satu pohon keputusan tunggal, sehingga model dapat menangkap variasi pola pada data uji dengan lebih baik. Selain itu, pohon keputusan Random Forest juga memperlihatkan bahwa atribut kunci seperti Denda, jumlah kendaraan jenis Jeep dan Truk, serta Total Jumlah pembayaran memiliki pengaruh signifikan dalam menentukan kategori, sehingga model mampu menghasilkan prediksi yang lebih akurat dan interpretatif.

Dari segi evaluasi performa, Random Forest menunjukkan akurasi 100% pada data uji, dengan presisi dan recall masing-masing 100% untuk kategori “Tinggi” maupun “Rendah”. Artinya, seluruh sampel berhasil diklasifikasikan dengan tepat tanpa adanya kesalahan prediksi, baik false positive maupun false negative. Sementara itu, Decision Tree meskipun memiliki akurasi tinggi sebesar 96%, masih menunjukkan adanya beberapa kesalahan klasifikasi terutama pada kategori “Rendah”, yang mengindikasikan adanya potensi false positive. Secara keseluruhan, hasil evaluasi ini menegaskan bahwa Random Forest lebih efektif dalam memprediksi kategori total pembayaran PKB, mampu menangkap interaksi kompleks antaratribut, dan memberikan hasil yang stabil serta andal pada dataset yang relatif kecil. Hal ini menjadikan Random Forest pilihan yang lebih baik dibandingkan Decision Tree untuk keperluan klasifikasi data PKB dengan distribusi seimbang.

## 6. Daftar Pustaka

- Abdelmagid, A. S., & Qahmash, A. I. M. (2023). Utilizing the Educational Data Mining Techniques “Orange Technology” for Detecting Patterns and Predicting Academic Performance of University Students. *Information Sciences Letters*, 12(3), 1415–1431. <https://doi.org/10.18576/isl/120330>
- Alam, A., Alana, D. A. F., & Juliane, C. (2023). Comparison Of The C.45 And Naive Bayes Algorithms To Predict Diabetes. *Sinkron*, 8(4), 2641–2650. <https://doi.org/10.33395/sinkron.v8i4.12998>
- Almufqi, F. M., & Voutama, A. (2023). Perbandingan Metode Data Mining Untuk Memprediksi Prestasi Akademik Siswa. *Jurnal Teknik*, 15(1), 61–66. <https://doi.org/10.30736/jt.v15i1.929>
- Andrianto, R., & Irawan, F. (2023). Implementasi Metode Regresi Linear Berganda Pada Sistem Prediksi Jumlah Tonase Kelapa Sawit di PT . Paluta Inti Sawit. *Jurnal Pendidikan Tambusai*, 7(1), 2926–2934.
- Boateng, E. Y., Otoo, J., & Abaye, D. A. (2020). Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review. *Journal of Data Analysis and Information Processing*, 08(04), 341–357. <https://doi.org/10.4236/jdaip.2020.84020>
- Elfaladonna, F., & Rahmadani, A. (2019). Analisa Metode Classification-Decision Tree Dan Algoritma C.45 Untuk Memprediksi Penyakit Diabetes Dengan

- Menggunakan Aplikasi Rapid Miner. *SINTECH (Science and Information Technology) Journal*, 2(1), 10–17. <https://doi.org/10.31598/sintechjournal.v2i1.293>
- Forest, R., & Bayes, N. (2021). *Perbandingan Akurasi , Recall , dan Presisi Klasifikasi pada Algoritma*. 5(April), 640–651. <https://doi.org/10.30865/mib.v5i2.2937>
- Hidayah, A. N. Z., & Rozi, A. F. (2021). Penerapan Data Mining Dalam Menentukan Kinerja Karyawan Terbaik Dengan Menggunakan Metode Algoritma C4.5 (Studi Kasus: Universitas Mercu Buana Yogyakarta). *Jurnal Information System & Artificial Intelligence*, 1(2), 117–127.
- Hutagalung, J., Ginantra, N. L. W. S. R., Bhawika, G. W., Parwita, W. G. S., Wanto, A., & Panjaitan, P. D. (2021). COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm. *Journal of Physics: Conference Series*, 1783(1). <https://doi.org/10.1088/1742-6596/1783/1/012027>
- Informatika, T., Informatika, I., & Lampung, D. (2021). *Perbandingan Data Set IRIS Dengan Aplikasi Rapid Miner dan Orange menggunakan Algoritma Klasifikasi*. 158–163.
- Mohammed, S., Elbeltagi, A., Bashir, B., Alsafadi, K., Alsilibe, F., Alsalman, A., ... Harsányi, E. (2022). A comparative analysis of data mining techniques for agricultural and hydrological drought prediction in the eastern Mediterranean. *Computers and Electronics in Agriculture*, 197(March). <https://doi.org/10.1016/j.compag.2022.106925>
- Putri, P. A. R., Prasetyowati, S. S., & Sibaroni, Y. (2023). The Performance of the Equal-Width and Equal-Frequency Discretization Methods on Data Features in Classification Process. *Sinkron*, 8(4), 2082–2098. <https://doi.org/10.33395/sinkron.v8i4.12730>
- Ratra, R., & Gulia, P. (2020). Experimental evaluation of open source data mining tools (WEKA and orange). *International Journal of Engineering Trends and Technology*, 68(8), 30–35. <https://doi.org/10.14445/22315381/IJETT-V68I8P206S>
- Sulaiman, A. P., Liliana, L., & Santoso, L. W. (2021). Kecerdasan Buatan dengan Metode ID3 Finite State Machine dalam Turn-Based Tactics Game. *Jurnal Infra*, (031). Retrieved from <http://publication.petra.ac.id/index.php/teknik-informatika/article/viewFile/11421/10031>
- Supriyadi, A., Triayudi, A., & Sholihati, I. D. (2021). Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 6(2), 229–240. <https://doi.org/10.29100/jipi.v6i2.2008>