*Sistemasi: Jurnal Sistem Informasi* ISSN:2302-8149
*Volume 14, Nomor 5, 2025: 2450-2463* e-ISSN:2540-9719

# Klasifikasi Status Pembayaran *Invoice* Bank Menggunakan Regresi Logistik dan *Random Forest*

## *Payment Status Classification Invoice Bank Using Logistic Regression and Random Forest*

**[1]Farah Anindia Putri*, [2] Mujiati Dwi Kartikasari**
[1,2]Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia
Jalan Kaliurang Km 14,5 Sleman, Yogyakarta, Indonesia
*e-mail: 21611162@students.uii.ac.id

(***received*:** 26 June 2025, ***revised*:** 19 July 2025, ***accepted*:** 20 July 2025)

**Abstrak**

Manajemen pembayaran merupakan aspek penting dalam operasional keuangan bank, khususnya dalam memastikan kelancaran transaksi pengadaan barang dan jasa. *Invoice* sebagai dokumen resmi memiliki peran dalam menentukan apakah suatu transaksi dapat segera diproses atau mengalami penundaan. Meskipun peran *invoice* sangat sentral, kajian empiris mengenai faktor-faktor yang memengaruhi status pembayarannya masih terbatas, terutama dalam konteks institusi perbankan. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi status pembayaran *invoice* berdasarkan jenis perusahaan, jenis pengadaan, dan nilai *invoice*. Metode yang digunakan meliputi regresi logistik dan *random forest* untuk membandingkan performa klasifikasi kedua pendekatan. Hasil analisis menunjukkan bahwa jenis pengadaan dan nilai *invoice* berpengaruh signifikan terhadap status pembayaran, dengan nilai *invoice* sebagai variabel paling dominan berdasarkan p-*value*. Pada model *random forest*, nilai *invoice* juga menunjukkan tingkat kepentingan tertinggi. Dari sisi akurasi, *random forest* memberikan hasil yang lebih unggul dengan akurasi sebesar 94,47%, dibandingkan regresi logistik sebesar 59,30%. Meskipun keduanya memiliki nilai presisi yang hampir sama (sekitar 97%), *random forest* mencatat recall sebesar 97,41% dan F1-score yang lebih tinggi dibandingkan regresi logistik (recall 69,19%). Temuan ini menunjukkan bahwa *random forest* merupakan metode yang lebih efektif dalam memprediksi status pembayaran dan berpotensi mendukung pengambilan keputusan berbasis data dalam sistem manajemen pembayaran perbankan.

**Kata kunci:** *invoice*, pembayaran, *random forest*, regresi logistik

***Abstract***

*Payment management is an essential aspect of a bank's financial operations, particularly in ensuring the smooth execution of procurement transactions for goods and services. The invoice, as an official document, plays a role in determining whether a transaction can be processed promptly or experiences a delay. Despite its central role, empirical research exploring the factors influencing invoice payment status remains limited, especially within the context of banking institutions. This study aims to analyze the factors that affect invoice payment status based on company type, procurement type, and invoice value. The methods employed include logistic regression and random forest to compare the classification performance of both approaches. The analysis reveals that procurement type and invoice value significantly influence payment status, with invoice value emerging as the most dominant variable based on the smallest p-value. In the random forest model, invoice value also ranks highest in terms of variable importance. In terms of accuracy, the random forest model outperforms logistic regression, achieving an accuracy of 94.47% compared to 59.30%. Although both methods yield similar precision (approximately 97%), random forest demonstrates a substantially higher recall (97.41%) and F1-score, whereas logistic regression records a recall of only 69.19%. These findings suggest that random forest is a more effective method for predicting payment status and holds strong potential for supporting data-driven decision-making in bank payment management systems.*

***Keywords:*** *invoice, logistic regression, payment, random forest*

 ISSN:2302-8149
e-ISSN:2540-9719

# 1 Pendahuluan

Bank merupakan lembaga keuangan yang berperan dalam menghimpun dana dari masyarakat melalui kredit, pembiayaan, atau bentuk lainnya, dengan tujuan utama untuk meningkatkan kesejahteraan masyarakat [1]. Secara umum, fungsi utama bank mencakup tiga aspek, yaitu penghimpunan dana, penyaluran dana, serta pelaksanaan aktivitas dalam sistem pembayaran [2]. Dalam konteks ini, pengelolaan sistem pembayaran menjadi elemen krusial yang mendukung kelancaran operasional, termasuk di sektor perbankan [3], [4]. Untuk menjamin efektivitas proses tersebut, dibutuhkan sistem yang terstruktur dan efisien. Salah satu komponen penting dalam proses pembayaran adalah *invoice*, yaitu dokumen resmi yang digunakan sebagai alat penagihan. Menurut Kamus Besar Bahasa Indonesia (KBBI), *invoice* merupakan daftar barang kiriman yang memuat rincian seperti nama barang, jumlah, dan harga yang harus dibayarkan oleh pembeli [5].

Ketepatan dan kelancaran proses pembayaran memegang peranan penting dalam menjaga kredibilitas perusahaan serta mengurangi risiko finansial akibat keterlambatan pembayaran. Status pembayaran, seperti “terbayar” dan “tidak terbayar”, menjadi indikator utama dalam mengevaluasi efektivitas sistem keuangan perusahaan. Namun, pada praktiknya, pengelolaan data pembayaran yang belum optimal sering kali menimbulkan berbagai kendala, seperti inkonsistensi informasi, keterlambatan pencairan dana, hingga kesalahan dalam pengambilan keputusan strategis. Meskipun beberapa sistem pelaporan telah diterapkan, analisis terhadap keterkaitan antara berbagai faktor yang memengaruhi status pembayaran masih jarang dilakukan secara mendalam. Sebagian besar analisis yang tersedia masih bersifat deskriptif dan dilakukan secara manual, sehingga belum memberikan kontribusi signifikan dalam mendukung pengambilan keputusan berbasis data.

Temuan awal yang diperoleh melalui wawancara dengan pegawai di salah satu bank di Indonesia menunjukkan bahwa status pembayaran suatu *invoice*—apakah “terbayar” atau “tidak terbayar”—dipengaruhi oleh sejumlah faktor, seperti jenis perusahaan, jenis pengadaan, dan nilai *invoice*. Praktik di lapangan turut memperkuat temuan ini, mengindikasikan bahwa karakteristik-karakteristik tersebut memiliki peran dalam menentukan kelancaran proses pembayaran. Meskipun telah tersedia sistem pelaporan internal, hubungan antara faktor-faktor tersebut dengan status pembayaran belum dianalisis secara kuantitatif dan sistematis.

Berdasarkan permasalahan tersebut, penelitian ini mengadopsi pendekatan statistika berupa regresi logistik untuk mengevaluasi hubungan antara variabel independen dan variabel dependen, serta membandingkannya dengan metode *random forest* guna menentukan pendekatan prediktif yang paling efektif dalam memetakan status pembayaran. Regresi logistik merupakan metode analisis yang digunakan untuk menggambarkan hubungan antara variabel dependen biner (dua kategori) dengan satu atau lebih variabel independen [6]. Melalui regresi logistik, diharapkan penelitian ini dapat memberikan wawasan yang lebih dalam mengenai faktor-faktor yang memengaruhi status pembayaran, sekaligus memperkuat manajemen keuangan dan sistem pembayaran yang berbasis data.

Selain regresi logistik, metode *random forest* turut digunakan sebagai teknik klasifikasi alternatif untuk memprediksi status pembayaran berdasarkan variabel-variabel independen. *Random forest* dipilih karena kemampuannya dalam menangani data yang kompleks serta keunggulannya dalam mengatasi masalah *overfitting* [7]. Metode ini juga dikenal memiliki stabilitas dan akurasi yang tinggi, karena bekerja dengan membentuk kumpulan pohon keputusan (*decision trees*) yang berkolaborasi dalam proses klasifikasi [8].

Dengan demikian, penelitian ini tidak hanya bertujuan mengidentifikasi faktor-faktor yang memengaruhi status pembayaran, tetapi juga mengevaluasi dan membandingkan performa regresi logistik dan *random forest* sebagai dua pendekatan klasifikasi. Hasilnya diharapkan dapat menjadi dasar bagi pengembangan sistem prediksi yang akurat dan andal dalam mendukung pengambilan keputusan, khususnya dalam konteks manajemen keuangan dan sistem pembayaran di sektor perbankan. Penelitian ini menjadi relevan untuk mendorong transformasi institusi keuangan dari pendekatan manual menuju sistem analitik prediktif berbasis data, dalam rangka meningkatkan efisiensi operasional serta mengurangi risiko finansial akibat keterlambatan pembayaran.

 ISSN:2302-8149
e-ISSN:2540-9719

## 2 Tinjauan Literatur

Penelitian terkait klasifikasi dengan metode regresi logistik dan *random forest* pada data tidak seimbang sudah banyak dilakukan, namun topik pembayaran *invoice* pada suatu bank di Indonesia masih jarang diteliti terutama jika datanya tidak *balanced.* Penelitian terhadulu oleh Agustia et al. 2025 [9] yang melakukan komparasi algoritma *naive bayes*, *random forest*, dan regresi logistik dalam analisis sentimen terhadap isu judi *online*. Penelitian ini menunjukkan bahwa *random forest* memiliki performa terbaik dengan akurasi tertinggi. Namun, penelitian tersebut hanya berfokus pada hasil klasifikasi tanpa membahas pengaruh masing-masing variabel independen terhadap variabel dependennya.

Sementara itu Andriani and Susilanungrum 2023 [10] mengaplikasikan regresi logistik dengan pendekatan SMOTE dalam klasifikasi waktu tunggu pilot di pelabuhan. Penelitian ini memberikan gambaran tentang pentingnya keseimbangan data dalam meningkatkan kemampuan prediksi model terhadap kelas minoritas, namun tidak membandingkan performa regresi logistik dengan metode lain seperti *random forest* atau mengevaluasi validitas model secara keseluruhan melalui uji *likelihood*, uji *wald*, dan uji kesesuaian model dan interpretasi variabel dalam klasifikasi.

Penelitian yang dilakukan oleh Fasilkom 2025 [11] berfokus pada peningkatan akurasi klasifikasi tutupan lahan dengan menggunakan algoritma *random forest* terhadap citra satelit Sentinel-2 di Provinsi Jambi. Model yang dibangun menunjukkan akurasi tinggi dalam mengklasifikasikan jenis tutupan lahan dengan pendekatan utama berupa optimasi parameter dan validasi spasial. Meskipun sama-sama menggunakan *random forest*, pendekatan dan konteks penelitian ini berbeda dengan penelitian peneliti yang berfokus pada klasifikasi status pembayaran berbasis data keuangan. Dalam penelitian ini, *random forest* tidak hanya dievalusasi dari sisi akurasi, tetapi juga dari kemampuannya mengenali kelas minoritas melalui penyesuaian bobot kelas (`class_weight=balanced`), serta mengenai kontribusi setiap variabel melalui *importance* variabel.

Ketiga penelitian tersebut menunjukkan bahwa metode regresi logistik dan *random forest* pada data *imbalanced* sudah pernah dilakukan. Dapat disimpulkan bahwa *random forest* lebih unggul daripada regresi logistik. Lalu teknik *balanced* data seperti SMOTE dan `class_weight=balanced` mampu menangani data minoritas. Pada ketiga penelitian tersebut fokus hanya pada akurasi model, tanpa mengecek variabel independen mana yang berpengaruh terhadap variabel dependennya. Penelitian peneliti bertujuan untuk membandingkan kedua metode berdasarkan hasil akurasinya lalu meninjau pengaruh variabel (jenis pengadaan, jenis perusahaan, dan nilai *invoice*) terhadap variabel dependennya yaitu status pembayaran. Selain itu, peneliti mengecek validitas model secara statistik dan efektivitas pendekatan *balancing* data (SMOTE dan `class_weight=balanced`) pada data pembayaran di suatu bank.

## 3 Metode Penelitian

Sebagai alat penelitian, *microsoft excel* dan *python* digunakan untuk membantu dalam pengorganisasian, pengolahan, dan analisis data. *Microsoft excel* dan *python* memungkinkan peneliti untuk menyusun data secara sistematis, melakukan klasifikasi variabel, serta mempersiapkan data untuk analisis regresi logistik dan *random forest* yang akan dilakukan.

Data yang diambil merupakan data transaksi pembayaran dan dilakukan pengambilan data di salah satu Bank di Indonesia. Data yang diambil mencakup transaksi pembayaran yang terjadi pada periode Januari-Juni 2024 yang berjumlah 991 data dengan variabel yang digunakan adalah status pembayaran, jenis perusahaan, jenis pengadaan, dan nilai *invoice*, sesuai dengan kebutuhan peneliti dan ketersediaan data. Penjelasan lebih lanjut mengenai variabel yang digunakan baik variabel independen maupun dependen akan dijelaskan pada **Tabel 1**.

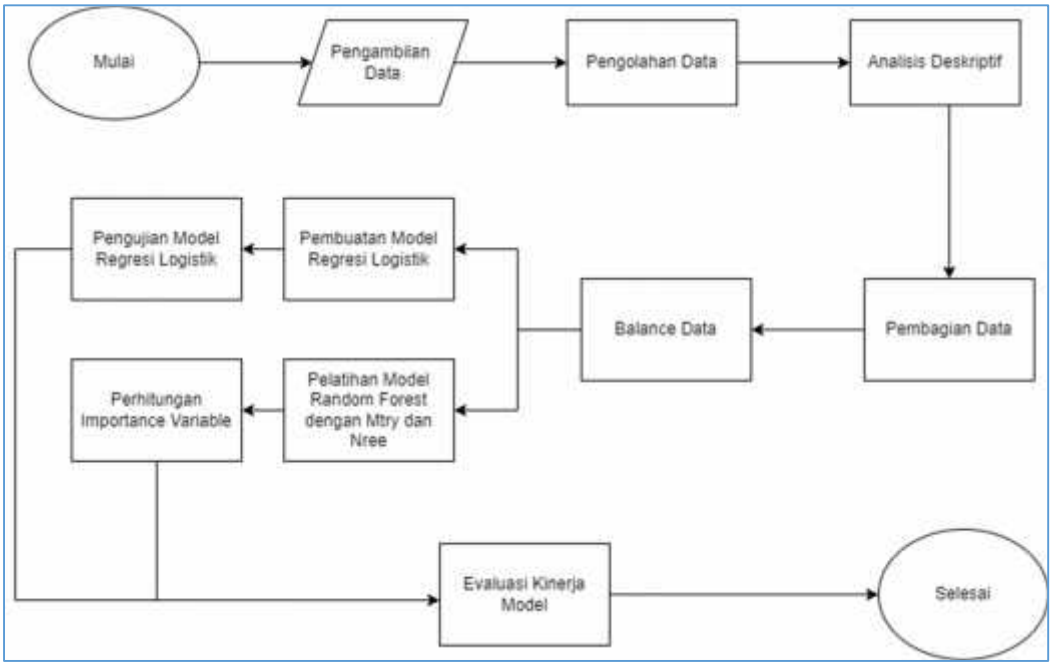
**Tabel 1 Definisi operasional variabel**

| Variabel | Definisi Operasional | Satuan | Skala Data |
|---|---|---|---|
| Status Pembayaran (Y) | Kondisi yang menunjukkan suatu pembayaran telah diselesaikan sesuai dengan ketentuan atau masih dalam proses penyelesaian. | - | Nominal<br>0: Tidak Terbayar<br>1: Terbayar |

 ISSN:2302-8149
e-ISSN:2540-9719

| Variabel | Definisi Operasional | Satuan | Skala Data |
|---|---|---|---|
| Jenis Perusahaan | Klasifikasi organisasi bisnis berdasarkan bidang usaha atau aktivitas utama yang dijalankan. | - | Nominal<br>0: Bank *Group*<br>1: Bank *Non Group* |
| Jenis Pengadaan | Klasifikasi pembelian barang atau jasa yang dilakukan oleh perusahaan atau organisasi untuk mendukung operasionalnya. | - | Nominal<br>0: IT<br>1: *Non* IT |
| Nilai *Invoice* | Nilai yang tertera pada dokumen *invoice* yang menggambarkan total biaya yang harus dibayarkan oleh perusahan kepada penyedia barang atau jasa. | Rupiah | Rasio |

Selanjutnya metode penelitian ini juga mencakup tahapan mulai dari pengambilan data, pengolahan data, hingga evaluasi kinerja model berdasarkan metode dan variabel yang digunakan. Seluruh alur penelitian disajikan dalam bentuk *flowchart* pada **Gambar 1** untuk memudahkan pemahaman.




**Gambar 1 Tahapan penelitian**

### 3.1 Pembagian Data

Klasifikasi adalah proses membangun model atau fungsi yang dapat mengidentifikasi serta membedakan konsep dengan tujuan memperkirakan kelas suatu objek yang belum diketahui labelnya [12]. Dalam pengklasifikasian data terdapat dua proses, pada tahap ini, data dibagi menjadi dua bagian yaitu data *training* (data pelatihan) dan data *testing* (data pengujian). Data *training* digunakan untuk membangun dan melatih model, sedangkan data *testing* digunakan untuk mengevaluasi kinerja model [13].

### 3.2 *Imbalance* Data

*Imbalance data* menunjukkan kondisi di mana distribusi data antara kategori dalam suatu dataset tidak merata. Ketidakseimbangan ini dapat menyebabkan model klasifikasi kesulitan mengenali pola pada kelas minoritas [14]. Maka dari itu, peneliti melakukan *balancing* sebelum pelatihan model pada data latih menggunakan algoritma SMOTE sehingga jumlah kelas untuk status pembayaran dengan klasifikasi terbayar dan tidak terbayar menjadi seimbang pada regresi logistik [8]. Pada SMOTE peneliti

 ISSN:2302-8149
e-ISSN:2540-9719

mereplikasikan data pada kelas minor untuk mencari tetangga terdekat untuk menggunakan jarak *euclidean*. Lalu menghitung *syntethic* data pada kelas minor dengan rumus [10]:

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_p - y_p)^2} \quad (1)$$

$$x_s = x_i + (x_k - x_i)\tau \quad (2)$$

Pada r*andom forest* menggunakan parameter `class_weight='balanced'` selama pelatihan model sehingga data yang digunakan adalah data asli [11]. Pada *random forest* tidak mengubah data tetapi mengatur bobot kelas saat training model.

Teknik *balancing* hanya mengubah distribusi variabel dependen karena masalah utama yang ingin diatasi adalah ketidakseimbangan kelas dari variabel dependen, bukan di distribusi variabel independen.

### 3.3 Regresi Logistik

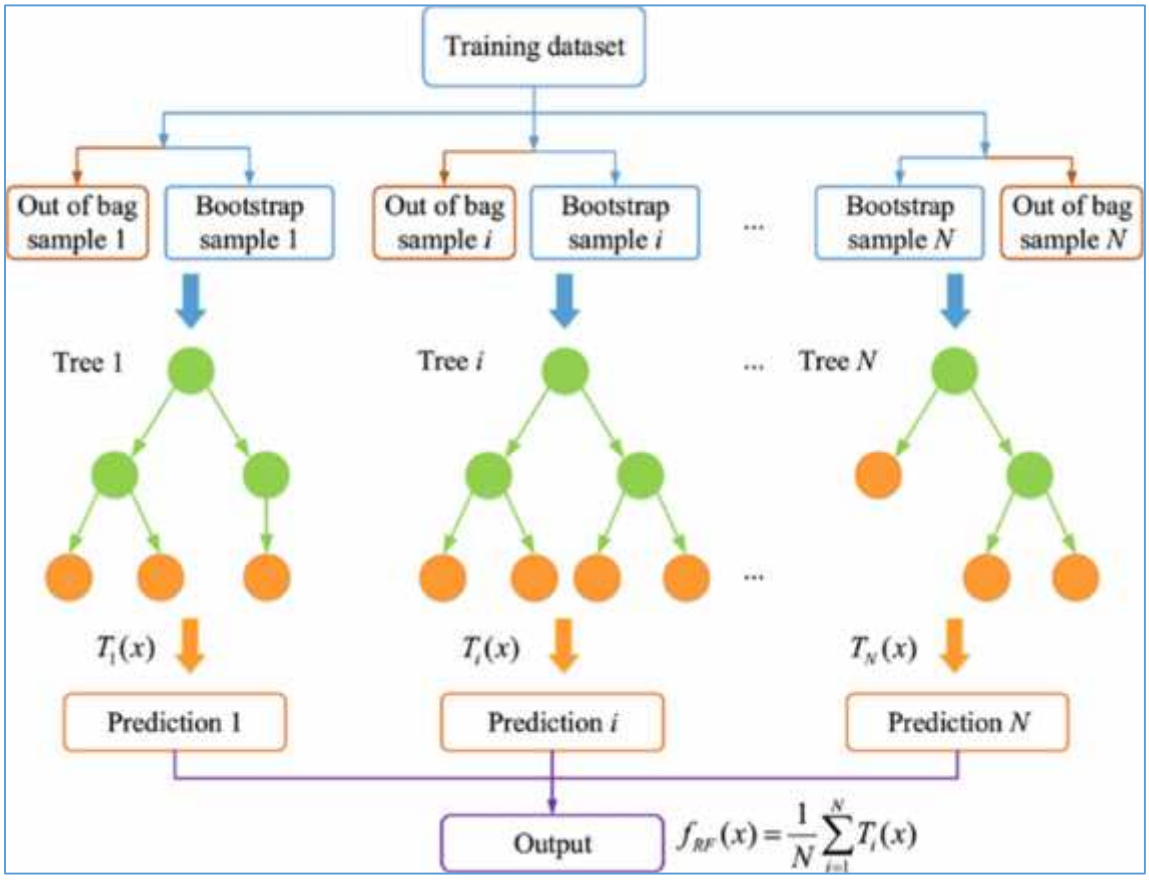
Membentuk persamaan regresi dari variabel yang telah ditentukan. Pada tahap ini mencari model regresi yang semua variabelnya signifikan dengan data yang sudah seimbang. Jika terdapat variabel yang tidak signifikan, maka dilakukan eliminasi variabel dengan menghapus variabel yang mempunyai p-*value* paling besar. Lalu didapatkan model regresi logistiknya [15]:

$$\pi(x) = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_p X_p}} \quad (3)$$

Terdapat beberapa uji yang dilakukan yaitu uji rasio *likelihood* uji *wald*, dan uji kesesuaian model. Uji *likelihood* digunakan untuk mengetahui koefisien $\beta$ terhadap variabel dependen secara serentak, uji *wald* digunakan untuk menguji signifikansi variabel independen terhadap dependen, dan uji kesesuaian model bertujuan untuk menilai seberapa baik model cocok dengan data [16].

### 3.4 *Random Forest*

Klasifikasi *random forest* memanfaatkan data *training* yang sudah *balance* dengan menentukan nilai *mtry* atau peubah penjelas dan *ntree* atau jumlah pohon agar mendapatkan model yang optimal dan nilai *error*OOB yang bernilai kecil [17]. Setelah didapatkan *mtry* dan *ntree*, dapat dibentuk pohon sebagai berikut [18]:




**Gambar 2 Model struktur *random forest***

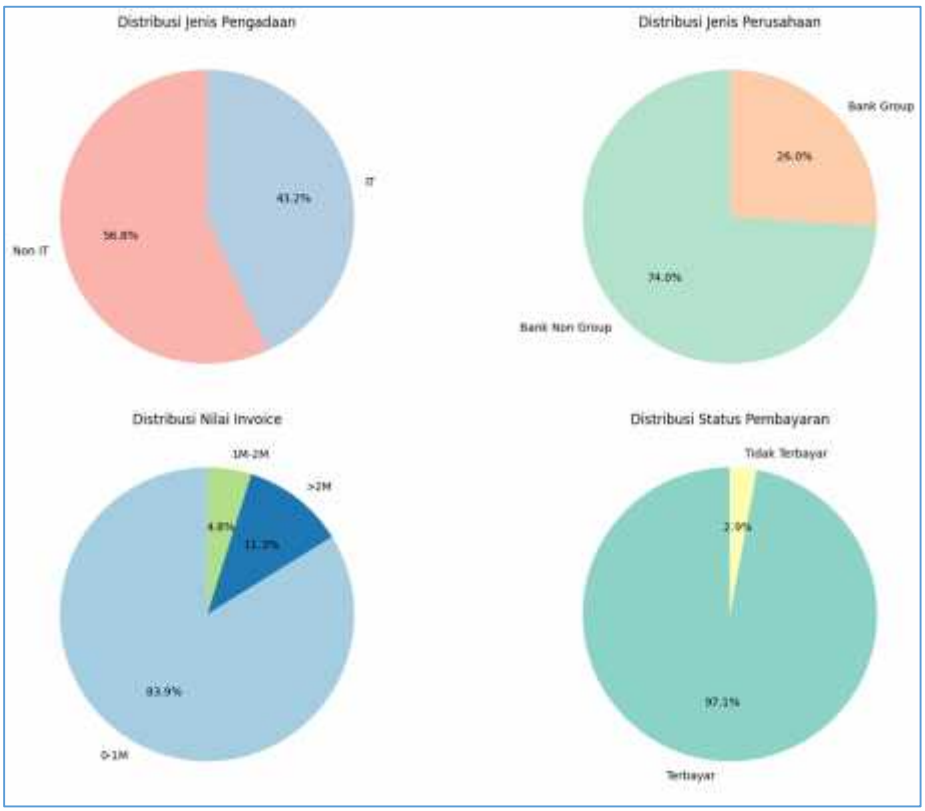
*Importance variable* mengindentikasikan sejauh mana masing-masing variabel berperan dalam mempengaruhi hasil analisis atau prediksi. Nilai *importance* yang baik adalah nilai yang tinggi, karena menunjukkan bahwa variabel tersebut banyak digunakan dalam membagi data secara efektif [11].

### 3.5 *Confusion Matrix*

Membuat tabel klasifikasi untuk melihat efektifikasi pemodelan dan menghitung nilai akurasi, presisi, *recall*, dan *F1-Score* berdasarkan penerapan model regresi logistik dan *random forest*. Hal ini sangat penting karena tingkat akurasi model prediksi akan menentukan kualitas prediksi yang dihasilkan [9], [19].

## 4 Hasil dan Pembahasan

Tahap awal analisis dilakukan dengan melihat karakteristik masing-masing variabel yang digunakan dalam penelitian, yaitu status pembayaran, jenis perusahaan, jenis pengadaan, dan nilai *invoice*. Untuk mempermudah pemahaman terhadap distribusi data, analisis deskriptif digunakan untuk mengubah sekumpulan data mentah menjadi bentuk yang lebih mudah dipahami yang berbentuk informasi yang lebih ringkas [20]. Peneliti menggunakan *piechart* untuk memahami data yang digunakan, yang menggambarkan proporsi masing-masing kategori.




**Gambar 3 Analisis deskriptif**

Dari hasil visualisasi pada **Gambar 3** untuk variabel jenis pengadaan diperoleh sebesar 57% dari total transaksi merupakan pengadaan Non IT, sedangkan sisanya yaitu 43% termasuk dalam kategori pengadaan IT. Lalu, untuk jenis perusahaan dapat diketahui bahwa mayoritas perusahaan yang terlibat dalam transaksi pembayaran di periode Januari-Juni 2024 adalah perusahaan yang tidak tergabung dalam Bank *Group* dengan proporsi sebesar 74%, sedangkan perusahaan yang termasuk dalam *Bank Group* hanya sebesar 26% dari total transaksi. Selanjutnya yaitu untuk besar *invoice* berada pada rentang 0-1 Miliar yaitu sebesar 83%, *invoice* dengan nilai 1-2 miliar hanya mencakup 4,8% dan *invoice* dengan nilai lebih dari 2 miliar sebesar 11,3% dari total data. Hasil dari *pie chart* menunjukkan bahwa mayoritas transaksi pembayaran pada Bulan Januari-Juni 2024 bernilai kecil hingga menengah dengan pengadaan Non IT di Bank Non Group, sedangkan transaksi dengan nilai besar relatif lebih sedikit.

Berdasarkan *pie chart* yang dibuat dari data pembayaran *invoice* bank di Indonesia periode Januari-Juni 2024, diketahui bahwa bahwa lebih dari 50% *invoice* telah terbayar yaitu sebesar 97%. Sedangkan yang tidak terbayar sebesar 3%. Hasil ini menunjukkan bahwa secara umum proses pembayaran di Bank berjalan dengan baik dan mayoritas kewajiban pembayaran berhasil diselesaikan.

### 4.1 Regresi Logistik

Analisis regresi logistik dilakukan untuk melihat pengaruh variabel independen terhadap status pembayaran. Langkah pertama dalam analisis ini adalah menguji signifikansi masing-masing variabel independen terhadap variabel dependen menggunakan uji *wald*. Hal ini digunakan untuk mengetahui apakah variabel tersebut berkontribusi secara signifikan dalam model. Hasil dari uji wald akan disajikan dalam **Tabel 2**.

**Tabel 2 Uji *wald* pertama**

| Variabel | Koefisien | p-*value* | *Wald* |
|---|---|---|---|
| *Intercept* | $-7{,}2557x\,10^{-0}$ | $1{,}8614x\,10^{-0}$ | 27,1718 |
| Jenis Pengadaan | $2{,}6711x\,10^{0}$ | $1{,}8226x\,10^{-7}$ | 319,541 |
| Jenis Perusahaan | $-1{,}8227\,x\,10^{-1}$ | $2{,}2228\,x\,10^{-1}$ | 1,489 |
| Nilai *Invoice* | $1{,}2823\,x\,10^{-1}$ | $4{,}8518\,x\,10^{-5}$ | 16,505 |

Pada **Tabel 2** dengan tingkat signifikansi sebesar $\alpha = 5\%$ dan kriteria pengambilan keputusan tolak $H_0$ jika $p - v < \alpha$, maka berdasarkan hasil pengujian dapat disimpulkan bahwa variabel jenis perusahaan tidak berpengaruh terhadap status pembayaran karena nilai $p - v > 0{,}05$. Sebaliknya, variabel jenis pengadaan dan nilai *invoice* memiliki $p - v < 0{,}05$, sehingga keduanya berpengaruh secara signifikan terhadap status pembayaran pada tingkat kepercayaan 95%.

Berdasarkan hasil uji *wald* **Tabel 2**, variabel jenis perusahaan tidak berpengaruh signifikan terhadap status pembayaran. Oleh karena itu,variabel tersebut dihapus dari model regresi logistik untuk memperoleh model yang lebih efisien dan hanya terdiri dari variabel yang relevan. Hasil analisis regresi logistik setelah penghapusan variabel jenis perusahaan disajikan pada **Tabel 3**.

**Tabel 3 Uji *wald* kedua**

| Variabel | Koefisien | p-*value* | *Wald* |
|---|---|---|---|
| *Intercept* | $-8{,}727x\,10^{-0}$ | $9{,}47\,x\,10^{-3}$ | $1{,}51\,x\,10^{2}$ |
| Jenis Pengadaan | $-2{,}6899x\,10^{0}$ | $5{,}33\,x\,10^{-7}$ | $3{,}27\,x\,10^{2}$ |
| Nilai *Invoice* | $1{,}265\,x\,10^{-1}$ | $4{,}71\,x\,10^{-5}$ | $1{,}66\,x\,10^{1}$ |

Berdasarkan hasil uji *wald* setelah proses eliminasi variabel secara *backward*, dilakukan pengujian signifikansi masing-masing variabel terhadap status pembayaran dengan $\alpha = 5\%$. Diperoleh bahwa seluruh variabel yang ada dalam model memiliki $p - v < 0{,}05$, sehingga dapat disimpulkan bahwa semua variabel tersebut berpengaruh signifikan terhadap status pembayaran, dan tidak ada variabel yang perlu dieliminasi lebih lanjut dari model.

Dari kedua variabel tersebut, nilai *invoice* menunjukkan p-*value* yang lebih kecil $4{,}71\,x\,10^{-5}$ dibandingkan jenis pengadaan $5{,}33\,x\,10^{-7}$ , namun jika dilihat dari konteks positif-negatif koefisien dan nilai *wald*, nilai *invoice* memiliki koefisien positif yang menandakan bahwa semakin besar nilai *invoice*, semakin tinggi peluang *invoice* tersebut untuk terbayar. Dengan demikian, nilai *invoice* dapat disimpulkan sebagai variabel yang paling berpengaruh dan paling relevan dalam memprediksi status pembayaran pada regresi logistik.

Setelah melakukan uji signifikansi parsial dengan uji *wald*, langkah selanjutnya adalah melakukan uji *likelihood* untuk menguji signifikansi model secara keseluruhan. Uji ini bertujuan untuk mengetahui apakah model regresi logistik yang dibangun lebih baik model tanpa independen.

**Tabel 4 Uji *likelihood***

| G | df | p-*value* |
|---|---|---|
| 486,4959 | 2 | $2{,}28x\,10^{-1}$ |

Berdasarkan **Tabel 4** dengan tingkat signifikansi $\alpha = 5\%$, jika $p - v \quad < 0{,}05$ maka $H_0$ ditolak. Berdasarkan hasil uji *likelihood*, diperoleh nilai $p - v \quad < 0{,}05$, sehingga dapat disimpulkan bahwa model regresi logistik secara keseluruhan signifikansi dan layak digunakan untuk memprediksi status pembayaran.

Setelah dilakukan uji *likelihood* untuk menilai signifikansi model secara keseluruhan, langkah selanjutnya adalah melakukan uji kesesuaian model untuk mengevaluasi seberapa baik model regresi yang dibangun mampu menggambarkan data yang diamati.

**Tabel 5 Uji kesesuaian model**

| | *Chi-Square* | df | p-*value* |
|---|---|---|---|
| *Pearson Chi-Square* | 176,9587 | 8 | $4{,}48x\ 10^{-3}$ |

Berdasarkan **Tabel 5** dengan tingkat signifikansi $\alpha = 5\%$, jika $p - v \quad < 0{,}05$ maka $H_0$ ditolak. Berdasarkan hasil uji kesesuaian model, diperoleh nilai $p - v \quad < 0{,}05$, sehingga dapat disimpulkan bahwa $H_0$ ditolak. Dengan demikian, dapat disimpulkan bahwa terdapat perbedaan yang signifikan antara nilai observasi dan nilai prediksi model, yang menunjukkan bahwa model tidak sesuai dengan data.

Setelah itu, pembentukan model akhir regresi logistik dilakukan apabila semua variabel independen yang diuji sudah berpengaruh signifikan terhadap variabel dependen. Dari hasil uji *wald* setelah penghapusan variabel jenis perusahaan, dapat dinyatakan bahwa semua variabel yang diteliti berpengaruh signifikan terhadap status pembayaran. Maka dapat dinyatakan dengan persamaan regresi logistik, sebagai berikut:

$$(x) = \frac{\exp(-8{,}727x\ 10^{-0} + 2{,}6899_{J(} + 1{,}265\ x\ 10^{-1}{}_{N} )}{1 + \exp(-8{,}727x\ 10^{-0} + 2{,}6899_{J} + 1{,}265\ x\ 10^{-1}{}_{N} )} \quad (4)$$

$$l(\quad(\bar{\pi}) = \ln\left(\frac{\bar{\pi}}{1-\bar{\pi}}\right) \quad (5)$$

$$= -8{,}727x\ 10^{-0} + 2{,}6899_{J}$$

$$+ 1{,}265\ x\ 10^{-1}{}_{N}$$

Setelah membangun model regresi logistik, langkah selanjutnya adalah mengevaluasi kinerja model menggunakan data uji. Evaluasi dilakukan dengan menghitung metrik-metrik penting dalam klasifikasi, yaitu akurasi, presisi, *recall*, dan F1-*Score*, yang diperoleh melalui *confusion matrix*. Metrik ini memberikan gambaran seberapa baik model dalam memprediksi status pembayaran, khususnya dalam mengenali kelas mayoritas maupun minoritas. Evaluasi ini penting untuk menilai efektivitas model dalam konteks data yang tidak seimbang.

**Tabel 6 *Confusion matrix* regresi logistik**

| Aktual | Prediksi | | Total |
|---|---|---|---|
| | Tidak Terbayar | Terbayar | |
| Tidak Terbayar | 2 | 4 | 6 |
| Terbayar | 77 | 116 | 193 |
| Total | 79 | 120 | 199 |

Hasil klasifikasi menggunakan regresi logistik dapat dilihat pada **Tabel 6** terdapat 193 observasi yang termasuk dalam data berkategori terbayar. Sedangkan untuk kategori tidak terbayar ada sebanyak 6 observasi. Namun hasil pada pengujian memprediksikan ada sebanyak 120 observasi pada kategori terbayar dan 79 pada kategori tidak terbayar. Dari tabel *confusion matrix* didapatkan hasil pengukuran kinerja algoritma klasifikasi sebagai berikut:

1. Akurasi

 ISSN:2302-8149
e-ISSN:2540-9719

$$a = \frac{T + T}{T + F + F + T} x\,100\% \quad (6)$$
$$= \frac{1 + 2}{1 + 7 + 4 + 2}$$
$$= \frac{1}{1} = 0{,}5930x100\% = 59{,}30\%$$

2. Presisi

$$p = \frac{T}{T + F} x\,100\% \quad (7)$$
$$= \frac{1}{1 + 4}$$
$$= 0{,}9667x100\% = 96{,}67\%$$

3. *Recall/Sensitivity*

$$r = \frac{T}{T + F} x\,100\% \quad (8)$$
$$= \frac{1}{1 + 7}$$
$$= 0{,}6010x100\% = 60{,}10\%$$

4. F1-*Score*

$$F1 - S = \frac{2 x p \quad x r}{p \quad + r} \quad (9)$$
$$= \frac{2 x 0{,}9 \quad x 0{,}6}{0{,}9 \quad + 0{,}6}$$
$$= \frac{1{,}1}{1{,}5} = 0{,}7414x\,100\% = 74{,}14\%$$

Dapat dilihat pada hasil perhitungan, nilai *sensitivity* atau pengukuran proporsi *true positive* (terbayar) yang diidentifikasi dengan benar sebesar 60,10%. Hasil akurasi data *testing* sebesar 59,30% menunjukkan bahwa metode regresi logistik untuk melakukan prediksi dari data *testing* atau untuk data baru dengan hasil prediksi tepat sebesar atau bisa dikatakan bisa memprediksi dengan cukup baik.

### 4.2 *Random Forest*

Sebelum melakukan metode *random forest* akan dilakukan penentuan nilai *mtry* atau peubah penjelas dan *ntree* atau jumlah pohon agar mendapatkan model yang optimal dan nilai *error*OOB yang bernilai kecil. Penentuan *mtry* agar didapatkan nilai optimal dapat dilakukan dengan tiga cara, yaitu [21], [17]:

$$m = \frac{1}{2}\sqrt{p} = \frac{\sqrt{3}}{2} = \frac{1{,}7}{2} = 0{,}865 = 1 \quad (10)$$

$$m = \left|\sqrt{p}\right| = \left|\sqrt{3}\right| = 1{,}73 = 2 \quad (11)$$

$$m = 2\,x\,\left|\sqrt{p}\right| = 2\,x\,\left|\sqrt{3}\right| = 2 \text{ x } 1{,}73 = 3{,}46 = 3 \quad (12)$$

Setelah menentukan *mtry*, selanjutnya yaitu mencoba *mtry* yang telah didapatkan untuk melakukan klasifikasi, maka didapatkan nilai *error*OOB sebagai berikut:

**Tabel 7 Pengujian nilai *error*OOB *mtry***

| *Mtry* | *error*OOB |
|---|---|
| 1 | 0,06313 |
| 2 | 0,06691 |
| 3 | 0,05808 |

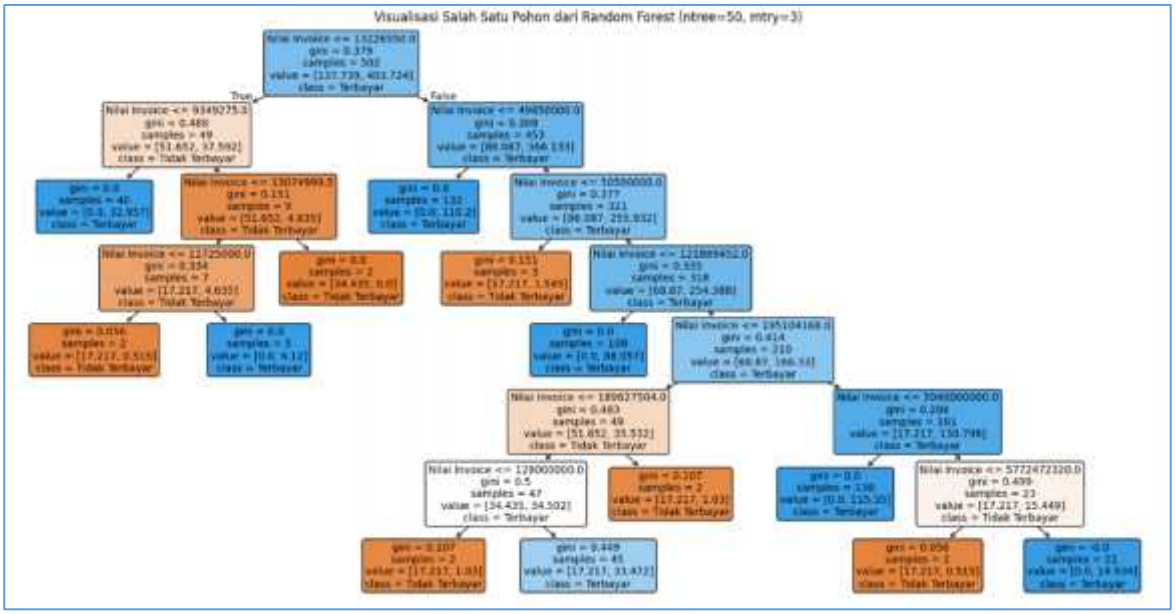
Didapatkan nilai *error*OOB terendah yaitu 0,05808 pada saat nilai *mtry* sebesar 3. Maka peneliti mengambil keputusan untuk menggunakan nilai *mtry* = 3. Langkah selanjutnya yaitu menentukan jumlah pohon (*ntree*) yang akan digunakan menggunakan nilai *mtry* yang didapatkan sebelumnya. Jumlah pohon (*ntree*) yang akan dicoba oleh peneliti dimulai dari 25, 50, 100, 250, 300, 500, dan 1000.

 *ISSN:2302-8149*
*e-ISSN:2540-9719*

**Tabel 8 Pengujian nilai *error*OOB *ntree***

| *Ntree* | *Error*OOB |
|---|---|
| 25 | 0,0606 |
| 50 | 0,0555 |
| 100 | 0,0580 |
| 250 | 0,0631 |
| 300 | 0,0618 |
| 500 | 0,0707 |
| 1000 | 0,0656 |

Pada **Tabel 8** adalah nilai *error* yang dihasilkan oleh setiap jumlah pohon dengan nilai *mtry* sebesar 3. Peneliti melakukan pengujian dengan jumlah pohon (*mtry*) berbeda, didapatkan nilai *error*OOB terendah pada saat jumlah pohon (*ntree*) sebesar 50 dan dihasilkan nilai *error*OOB terendah sebesar 0,0555. Setelah mendapatkan nilai *mtry*=3 dan *ntree* optimum=50, maka kemudian nilai tersebut digunakan untuk membuat pohon seperti gambar berikut:




**Gambar 4 Pohon *random forest***

Visualisasi salah satu pohon keputusan dari model *random forest* (dengan *ntree* = 50 dan *mtry* = 3) menunjukkan bahwa variabel nilai *invoice* menjadi pemisah utama dalam menentukan status pembayaran, yang mengindikasikan bahwa variabel tersebut memiliki kontribusi paling dominan dalam membedakan status pembayaran. Hal ini juga dapat disebabkan oleh karakteristik data, di mana variabel lain seperti jenis pengadaan dan jenis perusahaan tidak memberikan informasi yang cukup kuat untuk meningkatkan akurasi pemisahan dalam struktur pohon, sehingga variabel tersebut tidak muncul dalam pohon.

Cabang-cabang awal pada pohon banyak membagi data berdasarkan batas nilai tertentu dari *invoice*, yang mengindikasikan bahwa besarnya nominal tagihan sangat berpengaruh terhadap kemungkinan *invoice* dibayar atau tidak. Misalnya, *invoice* dengan nilai lebih besar dari Rp132 juta cenderung diklasifikasikan sebagai *terbayar*, sedangkan *invoice* dengan nilai jauh lebih kecil memiliki kemungkinan lebih besar untuk *tidak terbayar*. Selain itu, tampak bahwa pohon menghasilkan simpul-simpul akhir dengan nilai gini yang kecil, menandakan pemisahan kelas yang cukup baik dalam pohon tersebut. Hal ini mengindikasikan bahwa model mampu menangkap pola klasifikasi yang cukup akurat berdasarkan nilai *invoice*. Setelah itu, menentukan prediksi dan kemudian akan di uji tingkat akurasinya, *recall*, presisi, dan *F1-Score* terhadap data testing atau data baru.

**Tabel 9 *Confusion matrix random forest***

| Aktual | Prediksi | | Total |
|---|---|---|---|
| | Tidak Terbayar | Terbayar | |
| Tidak Terbayar | 0 | 6 | 6 |
| Terbayar | 5 | 188 | 193 |
| Total | 5 | 194 | 199 |

Hasil klasifikasi menggunakan *random forest* dapat dilihat pada **Tabel 9** terdapat 193 observasi yang termasuk dalam data berkategori terbayar, sedangkan untuk kategori tidak terbayar ada sebanyak 6 observasi. Namun, hasil pada pengujian memprediksikan ada sebanyak 194 pada kategori terbayar dan 5 pada kategori tidak terbayar. Berdasarkan tabel *confusion matrix* didapatkan hasil pengukuran kinerja algoritma klasifikasi sebagai berikut:

1. Akurasi

$$a = \frac{T \ +T}{T \ +F \ +F \ +T} x\ 100\% \quad (13)$$

$$= \frac{1 \ +1}{1 \ +0+6+5}$$

$$= \frac{1}{1} = 0{,}9447x100\% = 94{,}47\%$$

2. Presisi

$$p = \frac{T}{T \ +F} x\ 100\% \quad (14)$$

$$= \frac{1}{1 \ +6}$$

$$= 0{,}9691x100\% = 96{,}91\%$$

3. *Recall/Sensitivity*

$$r = \frac{T}{T \ +F} x\ 100\% \quad (15)$$

$$= \frac{1}{1 \ +5}$$

$$= 0{,}9741x100\% = 97{,}41\%$$

4. F1-*Score*

$$F1 - S = 2\frac{p \quad x\, r}{p \quad +r} \quad (16)$$

$$= \frac{2\,x\,0{,}9 \quad x\,0{,}9}{0{,}9 \quad +0{,}9}$$

$$= 0{,}9708\,x\,100\% = 97{,}08\%$$

Dapat dilihat pada hasil perhitungan, nilai *sensitivity* atau pengukuran proporsi *true positive* (terbayar) yang diidentifikasi dengan benar sebesar 97,41%. Hasil akurasi data *testing* sebesar 94,47% menunjukkan bahwa metode *random forest* untuk melakukan prediksi dari data testing atau untuk data baru dengan hasil prediksi tepat sebesar 94,47% atau bisa dikatakan bisa memprediksi dengan sangat baik.

Pada analisis *random forest*, model dibangun menggunakan data asli tanpa teknik penyeimbangan. Oleh karena itu, hasil *importance variable* yang diperoleh mencerminkan tingkat kepentingan masing-masing variabel berdasarkan data asli tersebut.

**Tabel 10 *Importance variable random forest***

| Variabel | *Mean Decrease Accuracy* |
|---|---|
| Nilai *Invoice* | 0,8694 |
| Jenis Pengadaan | 0,0742 |
| Jenis Perusahaan | 0,0562 |

**Tabel 10** menggambarkan tingkat *importance variable* pada model klasifikasi *random forest* yang telah dilakukan. *Importance variable* mengindikasikan sejauh mana masing-masing variabel berperan dalam mempengaruhi hasil analisis atau prediksi. Semakin besar nilai *importance variable*, semakin besar kontribusi variabel tersebut dalam hasil analisis. Berdasarkan hasil ini, nilai *invoice* merupakan variabel dengan kontribusi terbesar, diikuti oleh jenis pengadaan dan jenis perusahaan. Nilai

*importance* tertinggi ditemukan pada nilai *invoice* dengan angka 0,8694 yang menunjukkan bahwa variabel ini memiliki pengaruh signifikan terhadap hasil prediksi. Sebaliknya, jenis pengadaan dan jenis perusahaan memiliki nilai *importance* yang lebih rendah (0,0742 dan 0,0562) yang mengindikasikan bahwa meskipun keduanya berkontribusi, perannya dalam mempengaruhi hasil prediksi lebih kecil dibandingkan dengan nilai *invoice*.

Hal ini juga diperkuat oleh hasil visualisasi pada **Gambar 4** di mana seluruh percabangan dalam pohon tersebut hanya menggunakan variabel nilai *invoice*. Tidak munculnya variabel jenis pengadaan dan jenis perusahaan dalam visualisasi tersebut menunjukkan bahwa kedua tidak terpilih sebagai pemisah terbaik dalam pohon yang divisualisasikan. Namun, hal ini tidak berarti kedua variabel tersebut tidak berkontribusi sama sekali dalam model. Dengan demikian, dapat disimpulkan bahwa nilai *invoice* merupakan variabel yang paling berpengaruh dalam menentukan status pembayaran pada model *random forest*.

### 4.3 Perbandingan

Setelah dilakukan analisis klasifikasi menggunakan regresi logistik dan *random forest*, didapatkanlah nilai akurasi, presisi, recall dan F1-*Score* dari masing-masing metode. Berikut adalah hasil perbandingannya dari kedua metode.

**Tabel 11 Perbandingan hasil**

| | Regresi Logistik | *Random Forest* |
|---|---|---|
| Akurasi | 59,30% | 94,47% |
| Presisi | 96,67% | 96,91% |
| *Recall* | 69,19% | 97,41% |
| F1-*Score* | 74,14% | 97,08% |

**Tabel 11** menunjukkan perbandingan performa antara dua metode klasifikasi yaitu regresi logistik dan *random forest* dalam memprediksi status pembayaran. Masing-masing metrik evaluasi digunakan untuk menilai kualitas model, baik secara keseluruhan maupun spesifik terhadap kelas terbayar. *Random forest* memiliki akurasi 94,47% lebih tinggi dibandingkan regresi logistik yaitu 59,30%, menunjukkan bahwa secara keseluruhan *random forest* lebih andal dalam mengklasifikasikan data dengan benar. Lalu presisi untuk kedua model memiliki presisi tinggi yaitu diatas 96% artinya model cukup baik dalam memprediksi kelas mayoritas. *Recall* juga sangat baik pada *random forest* 97,41% dibandingkan regresi logistik 69,19% artinya *random forest* hampir tidak melewatkan data yang seharusnya diklasifikasikan sebagai (terbayar). *F1-Score* sangat tinggi pada *random forest* 97,08% dibandingkan regresi logistik 74,14%, mengindikasikan keseimbangan yang baik antara presisi dan *recall*.

Performa *random forest* yang lebih unggul disebabkan oleh cara kerjanya yang lebih fleksibel dalam mengenali pola-pola rumit pada data. Berbeda dengan regresi logistik yang hanya mampu menangkap hubungan linier antar variabel, *random forest* dapat memahami hubungan yang tidak beraturan atau lebih kompleks. *Random forest* bekerja dengan membangun banyak pohon keputusan dan menggabungkan hasilnya untuk membuat prediksi yang lebih akurat. Selain itu, metode ini mampu bekerja dengan baik pada data yang mengandung campuran variabel dengan skala data nominal dan rasio. Inilah yang membuat *random forest* lebih stabil dan efektif dalam memprediksi klasifikasi status pembayaran dibandingkan regresi logistik.

## 5 Kesimpulan

Berdasarkan hasil analisis regresi logistik, ditemukan bahwa variabel jenis pengadaan dan nilai *invoice* berpengaruh signifikan terhadap status pembayaran *invoice* bank, dengan nilai *invoice* sebagai variabel paling dominan berdasarkan nilai p-*value* terkecil pada uji *wald*. Sementara itu, pada metode *random forest*, nilai *invoice* juga menunjukkan tingkat kepentingan tertinggi dengan nilai *importance* sebesar 0,8694. Hal ini mengindikasikan bahwa besarnya nilai *invoice* secara konsisten menjadi faktor yang paling memengaruhi kemungkinan suatu *invoice* akan terbayar atau tidak. Dalam hal performa klasifikasi, *random forest* secara signifikan lebih unggul dibanding regresi logistik. *Random forest* menghasilkan akurasi sebesar 94,47%**,** jauh lebih tinggi dibanding regresi logistik yang hanya 59,30%. Selain itu, meskipun presisi kedua metode hampir sama yaitu mendekati 97%, *random forest* memiliki

nilai *recall* (97,41%) dan F1-*Score* (97,08%) yang jauh lebih tinggi dibanding regresi logistik (*recall* 69,19% dan F1-*Score* 74,14%). Ini menunjukkan bahwa bank dapat menggunakan model *random forest* sebagai alat bantu untuk mengklasifikasikan status pembayaran secara lebih akurat, serta menjadikan nilai *invoice* sebagai indikator penting dalam proses manajemen pembayaran. Pemanfaatan metode *machine learning* seperti *random forest* dapat membantu bank dalam mengurangi faktor risiko keterlambatan atau kesalahan pembayaran, serta meningkatkan efisiensi pengambilan keputusan berbasis data. Sebagai saran untuk penelitian selanjutnya, disarankan untuk menambah variabel baru, seperti lamanya waktu pemrosesan *invoice* atau status verifikasi dokumen. Selain itu, perluasan rentang data hingga Desember 2024 juga dapat meningkatkan ketepatan model dan hasil prediksi.

## Referensi

[1] A. Firda, Kurniati, A. Rahman, and M. Tabran, "Perbandingan Kinerja Bank Syariah dan Bank Konvensional dalam Melaksanakan Transaksi," *Al-Ubudiyah J. Pendidik. dan Stud. Islam*, vol. 4, no. 2, pp. 20–29, 2023, doi: 10.55623/au.v4i2.216.

[2] O. J. Keuangan, "Bank Umum," Jakarta, 2024. [Online]. Available: https://www.ojk.go.id/id/kanal/perbankan/Pages/Bank-Umum.aspx

[3] C. D. Pratama and S. Gischa, "Sistem Pembayaran: Definisi dan Perannya dalam Perekonomian," *Kompas.com*, Jakarta, Nov. 23, 2020. [Online]. Available: https://www.kompas.com/skola/read/2020/11/23/175246869/sistem-pembayaran-definisi-dan-perannya-dalam-perekonomian

[4] Y. S. Atmaja and D. H. Paulus, "Partisipasi Bank Indonesia Dalam Pengaturan Digitalisasi Sistem Pembayaran Indonesia," *Masal. Huk.*, vol. 51, no. 3, pp. 271–286, 2022, doi: 10.14710/mmh.51.3.2022.271-286.

[5] C. D. Ribkauli, "Prosedur Pembuatan Dokumen Invoice pada PT Visi Insan Pratama," Politeknik Negeri Jakarta, 2022. [Online]. Available: https://repository.pnj.ac.id/id/eprint/7181/5/JudulPendahuluanDanPenutup.pdf

[6] S. Haridanti, R. Adawiyah, G. S. Ariadne, F. A. Y. Putri, and Z. Ulfa, "Analisis Regresi Non Linear Model Logistik," *Seniati*, vol. 3, no. 2, pp. 62–67, 2020.

[7] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 677–690, 2022, doi: 10.30812/matrik.v21i3.1726.

[8] N. Sharfina and N. G. Ramadhan, "Analisis SMOTE Pada Klasifikasi Hepatitis C Berbasis Random Forest dan Naïve Bayes," *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 8, no. 1, p. 33, 2023, doi: 10.31328/jointecs.v8i1.4456.

[9] D. N. Agustia, R. R. Suryono, U. T. Indonesia, L. Ratu, and K. B. Lampung, "Comparison Of Naïve Bayes , Random Forest , And Logistic Regression Algorithms For Sentiment Analysis Online Gambling Komparasi Algoritma Naïve Bayes , Random Forest , Dan Logistic Regresion Untuk Analisis," vol. 10, no. 1, pp. 284–295, 2025.

[10] C. M. F. Andriani and D. Susilaningrum, "Klasifikasi Waiting Time for Pilot di Pelabuhan Tanjung Perak Menggunakan Metode Regresi Logistik - Synthetic Minority Oversampling Technique (SMOTE)," *J. Sains dan Seni ITS*, vol. 12, no. 1, 2023, doi: 10.12962/j23373520.v12i1.109844.

[11] J. Fasilkom, "Peningkatan Akurasi Klasifikasi Tutupan Lahan Menggunakan Random Forest pada Data Sentinel-2 di Jambi Author : Akhiyar Waladi," vol. 15, no. 1, pp. 17–24, 2025.

[12] S. U. Panjaitan *et al.*, "Uji Metode Naive Bayes Classifier Dalam," vol. 7, pp. 450–462, 2022.

[13] B. N. Azmi, A. Hermawan, and D. Avianto, "Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 4, no. 4, pp. 281–290, 2023, doi: 10.35746/jtim.v4i4.298.

[14] E. Sutoyo and M. A. Fadlurrahman, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 379, 2020, doi: 10.26418/jp.v6i3.42896.

[15] D. Kartikasari, "Analisis Faktor-Faktor Yang Mempengaruhi Level Polusi Udara Dengan Metode Regresi Logistik Biner," *MATHunesa J. Ilm. Mat.*, vol. 8, no. 1, pp. 55–59, 2020, doi: 10.26740/mathunesa.v8n1.p55-59.
[16] M. A. Suhendra, D. Ispriyanti, and S. Sudarno, "Ketepatan Klasifikasi Pemberian Kartu Keluarga Sejahtera Di Kota Semarang Menggunakan Metode Regresi Logistik Biner Dan Metode Chaid," *J. Gaussian*, vol. 9, no. 1, pp. 64–74, 2020, doi: 10.14710/j.gauss.v9i1.27524.
[17] M. Rosada and O. Mukhti, "Classification of Dental Caries in RSGM Baiturrahmah Using the Random Forest Method," *UNP J. Stat. Data Sci.*, vol. 2, no. 2, pp. 130–136, 2024.
[18] M. Zhu, Y. Yang, X. Feng, Z. Du, and J. Yang, "Robust Modeling Method for Thermal Error of CNC Machine Tools Based on Random Forest Algorithm," *J. Intell. Manuf.*, vol. IV, pp. 2013–2026, 2023.
[19] H. Junianto, R. E. Saputro, B. A. Kusuma, D. Intan, and S. Saputra, "Comparison Of Logistic Regression And Random Forest In Sentiment Analysis Of Disdukcapil Application Reviews Komparasi Logistic Regression Dan Random Forest Dalam," vol. 5, no. 6, pp. 1539–1547, 2024.
[20] L. D. Martias, "Statistika Deskriptif Sebagai Kumpulan Informasi," *Fihris J. Ilmu Perpust. dan Inf.*, vol. 16, no. 1, p. 40, 2021, doi: 10.14421/fhrs.2021.161.40-59.
[21] C. Yulianto, "Model Penilaian Tanah Massal Berbasis Parcel-Based Mass Valuation Using Random Forest in Surakarta City," pp. 26–39, 2024.