

Analisis Perbandingan Metode *Naïve Bayes* dan *Decision Tree C4.5* untuk Meningkatkan Kualitas Sampling dan Efisiensi Biaya (Studi Kasus: PT. BFI Finance Tbk)

Dyah Ayu Nurmumpuni^{1,*}, Tukiyat², Achmad Hindasyah³

¹⁾²⁾³⁾ Program Studi Teknik Informatika, Program Pasca Sarjana Magister Komputer
Universitas Pamulang, Tangerang Selatan, Indonesia

¹⁾ mumpunibb@yahoo.com

²⁾ dosen02711@unpam.ac.id

³⁾ achm021@brin.go.id

Abstrak

Penelitian ini bertujuan menganalisis komparasi metode *Naïve Bayes* dan *Decision Tree C4.5* dalam mengoptimalkan proses sampling audit pada PT BFI Finance Tbk. Studi dilakukan melalui pendekatan kuantitatif dengan menggunakan dataset 491 kontrak, terdiri dari 405 data latih dan 86 data uji. Variabel independen mencakup angsuran, lama hari menunggak, sisa pokok hutang, dan jenis produk, sedangkan variabel dependen adalah hasil temuan audit. Metode penelitian melibatkan implementasi algoritma machine learning untuk mengklasifikasikan kualitas sampling. Hasil penelitian menunjukkan bahwa *Naïve Bayes* memiliki akurasi lebih tinggi dibandingkan *Decision Tree C4.5*, dengan kemampuan superior dalam menangani kompleksitas data. Evaluasi performa algoritma mencakup metrik akurasi, presisi, recall, dan F1-score, yang mengindikasikan keunggulan *Naïve Bayes* dalam meningkatkan efisiensi proses audit.

Kata kunci: *naïve bayes, decision tree c4.5, sampling audit, machine learning, manajemen risiko*

Abstract

This research aims to analyze the comparative performance of Naïve Bayes and Decision Tree C4.5 methods in optimizing audit sampling processes at PT BFI Finance Tbk. The study was conducted using a quantitative approach with a dataset of 491 contracts, comprising 405 training data and 86 test data. Independent variables included installment characteristics, overdue days, principal debt remainder, and product types, while the dependent variable was audit findings. The research methodology involved implementing machine learning algorithms for sampling quality classification. Findings revealed that the Naïve Bayes method demonstrated superior accuracy compared to Decision Tree C4.5, with enhanced capabilities in handling data complexity. Algorithm performance evaluation covered accuracy, precision, recall, and F1-score metrics, indicating Naïve Bayes's potential in improving audit process efficiency.

Keywords: *naïve bayes, decision tree c4.5, audit sampling, machine learning, risk management*

1. PENDAHULUAN

Di era transformasi digital yang berkelanjutan, organisasi bisnis menghadapi tantangan kompleks dalam mengoptimalkan proses internal dan mengalokasikan sumber daya secara efisien [1]. Industri jasa keuangan, khususnya lembaga pembiayaan, mengalami tekanan signifikan untuk mengembangkan mekanisme audit yang lebih pintar, adaptif, dan hemat biaya. Dalam konteks ini, PT BFI Finance Tbk, sebuah perusahaan pembiayaan terkemuka dengan lebih dari 200 cabang di seluruh Indonesia, menghadapi permasalahan substantif dalam proses sampling audit yang membutuhkan pendekatan inovatif berbasis kecerdasan buatan.

Proses audit tradisional yang berlaku saat ini menghadapi kendala signifikan dalam efektivitas sampling. Dengan rata-rata 40 sampling per auditor setiap bulan, tidak semua sampling prioritas

dapat diverifikasi secara komprehensif. Studi internal menunjukkan bahwa dari 10 sampling yang dialokasikan, setidaknya 6 sampling berpotensi tidak memenuhi kriteria kualitas yang diinginkan, yang berarti pemborosan sumber daya yang substansial baik dari segi waktu maupun biaya [2].

Penelitian sebelumnya telah mengeksplorasi berbagai algoritma machine learning untuk mengidentifikasi dan mengklasifikasikan data sampling. Putri berhasil menggunakan algoritma Naïve Bayes untuk mendeteksi fraud dengan akurasi mencapai 87,2% [3], sementara Sunata menemukan bahwa Gradient Boosted Trees mampu mencapai akurasi spektakuler sebesar 99,85% dalam identifikasi *fraud* pada sistem perbankan [2].

Kontribusi utama penelitian ini adalah mengembangkan model prediktif yang dapat mengoptimalkan proses sampling audit melalui implementasi dan perbandingan dua algoritma *machine learning*: Naïve Bayes dan Decision Tree C4.5. Pendekatan komparatif ini bertujuan untuk mengidentifikasi metode paling efektif dalam mengklasifikasikan sampling berkualitas, dengan fokus pada parameter spesifik seperti nomor kontrak, jumlah angsuran, tenor, sisa pokok hutang, dan karakteristik produk.

Signifikansi metodologis penelitian terletak pada potensinya untuk mentransformasi pendekatan sampling audit dari model tradisional menjadi sistem cerdas berbasis data. Dengan menggunakan teknik klasifikasi *machine learning*, penelitian ini berupaya mengembangkan algoritma yang dapat secara otomatis menyaring *sampling* berdasarkan probabilitas kualitas dan risiko, sehingga memungkinkan alokasi sumber daya audit yang lebih strategis dan efisien.

Tujuan spesifik penelitian meliputi: (1) membandingkan akurasi Naïve Bayes dan Decision Tree C4.5 dalam mengklasifikasikan sampling audit, (2) mengidentifikasi metode optimal untuk prediksi kualitas *sampling*, dan (3) merancang model prototipe yang dapat mengintegrasikan algoritma *machine learning* ke dalam sistem audit internal PT BFI Finance Tbk.

Kerangka teoritis penelitian dibangun di atas fondasi *data mining* dan *machine learning*, dengan fokus khusus pada algoritma klasifikasi. Naïve Bayes, yang didasarkan pada teorema Bayes, menggunakan probabilitas bersyarat untuk mengklasifikasikan data, sementara Decision Tree C4.5 mengembangkan pohon keputusan berbasis informasi gain untuk segmentasi data yang lebih kompleks [4].

Kontribusi praktis penelitian tidak hanya terbatas pada peningkatan efisiensi audit, tetapi juga berpotensi menghasilkan model prediktif yang dapat diadaptasi oleh industri jasa keuangan lainnya. Dengan mengurangi sampling tidak berkualitas, organisasi dapat secara signifikan menurunkan biaya operasional, meningkatkan akurasi audit, dan mengalokasikan sumber daya manusia secara lebih strategis.

Originalitas penelitian terletak pada pendekatannya yang komprehensif dalam mengevaluasi dua algoritma machine learning dalam konteks spesifik sampling audit pada lembaga pembiayaan. Dibandingkan dengan penelitian sebelumnya yang umumnya fokus pada deteksi *fraud*, penelitian ini menawarkan perspektif baru tentang optimalisasi proses audit melalui klasifikasi cerdas.

2. TINJAUAN PUSTAKA

Perkembangan teknologi informasi dan komputasi telah memberikan kontribusi signifikan dalam berbagai bidang penelitian, terutama dalam hal pengambilan keputusan dan analisis data. Salah satu aspek penting dalam penelitian ilmiah adalah penggunaan metode klasifikasi yang akurat untuk menghasilkan *insights* yang bermakna [5]. Metode Naïve Bayes dan Decision Tree C4.5 merupakan dua pendekatan klasifikasi yang telah banyak digunakan dalam berbagai domain penelitian, termasuk dalam konteks manajemen risiko keuangan.

2.1 Metode Naïve Bayes dalam Berbagai Konteks Penelitian

Naïve Bayes telah terbukti menjadi algoritma klasifikasi yang powerful dan efisien dalam berbagai penelitian. Hananto menggunakan metode ini untuk mendeteksi fraud dalam transaksi perbankan, mencapai akurasi mencapai 86,085% [5]. Penelitian serupa oleh Nuzla dalam konteks deteksi SIMBox fraud pada PT XYZ menunjukkan akurasi tertinggi sebesar 87,2% [6].

Keunggulan Naïve Bayes tidak hanya terbatas pada deteksi fraud, tetapi juga meluas ke berbagai domain. Hooda et al. mengaplikasikan metode ini untuk mengklasifikasikan status kesejahteraan rumah tangga dengan akurasi 85,80% [7]. Sementara itu, Muñoz-Izquierdo et al. berhasil menggunakan Naïve Bayes untuk prediksi bimbingan konseling siswa dengan akurasi mencapai 94,84% [8].

2.2 Decision Tree C4.5 dalam Analisis Klasifikasi

Decision Tree C4.5 merupakan metode klasifikasi lain yang memiliki kemampuan untuk menghasilkan pohon keputusan yang interpretable. Perwara et al. menggunakan algoritma ini untuk deteksi fraud pada kartu kredit, menghasilkan akurasi 78% dengan precision 89,65% dan recall 85,71% [9].

Wulandari et al. menerapkan Decision Tree C4.5 untuk analisis kelayakan kredit nasabah, mencapai akurasi 80% pada aplikasi Rapid Miner [10]. Penelitian Rani menunjukkan bahwa metode ini dapat menghasilkan rule keputusan yang konsisten dengan proses manual dalam pemberian kredit [11].

2.3 Komparasi Metode Klasifikasi

Beberapa penelitian perbandingan metode klasifikasi memberikan wawasan menarik. Sunata melakukan komparasi tujuh algoritma untuk identifikasi fraud ATM, dengan Gradient Boosted Trees mencapai akurasi tertinggi 99,85% [2]. Kholifah dan Insani membandingkan Decision Tree C4.5 dan Naïve Bayes dalam klasifikasi nasabah kredit, dengan akurasi berturut-turut 71,91% dan 67,01% [12].

2.4 Aplikasi dalam Manajemen Risiko Keuangan

Konteks keuangan merupakan domain yang sangat relevan untuk penelitian klasifikasi. Kurniawan dan Kriestanto menggunakan Naïve Bayes untuk prediksi kelayakan kredit dengan akurasi 92,5% [13]. Nuraeni mencapai akurasi 89,33% dalam penentuan kelayakan kredit di Bank Mayapada, dengan nilai Area Under Curve (AUC) sebesar 0,955 [14].

2.5 Sampling dan Distribusi Data

Aspek sampling memiliki peran kritis dalam akurasi penelitian. Shi et al. menekankan pentingnya teknik probability sampling untuk menemukan ukuran sampel ideal dan menguji distribusi data. Hal ini sangat relevan dalam konteks penelitian klasifikasi, di mana kualitas dan representativitas sampel dapat secara signifikan memengaruhi hasil analisis [15].

2.6 Signifikansi Penelitian

Penelitian komparatif antara Naïve Bayes dan Decision Tree C4.5 dalam konteks sampling dan efisiensi biaya pada lembaga keuangan seperti PT BFI Finance Tbk menjadi sangat penting. Kedua metode ini memiliki karakteristik unik dalam mengklasifikasikan data, yang berpotensi memberikan perspektif baru dalam manajemen risiko dan pengambilan keputusan.

Tinjauan pustaka ini menunjukkan bahwa baik Naïve Bayes maupun Decision Tree C4.5 memiliki potensi signifikan dalam berbagai aplikasi klasifikasi, terutama dalam konteks keuangan. Namun, efektivitas masing-masing metode dapat bervariasi tergantung pada karakteristik dataset dan konteks spesifik penelitian.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan fokus utama pada analisis komparatif dua metode klasifikasi data mining, yaitu Naïve Bayes dan Decision Tree C4.5, untuk meningkatkan kualitas sampling dan efisiensi biaya pada PT BFI Finance Tbk. Metodologi penelitian dirancang secara sistematis untuk mengeksplorasi potensi algoritma machine learning dalam mengoptimalkan proses audit lapangan dan pengambilan keputusan strategis.

Penelitian dilaksanakan melalui serangkaian tahapan metodologis yang terstruktur, dimulai dengan identifikasi masalah yang komprehensif di lingkungan PT BFI Finance Tbk. Proses identifikasi masalah dilakukan melalui observasi mendalam terhadap mekanisme sampling dan audit lapangan yang ada, dengan tujuan mengidentifikasi celah dan potensi peningkatan efisiensi. Pendekatan kuantitatif dipilih untuk memungkinkan analisis objektif dan terukur terhadap variabel-variabel kunci dalam proses sampling.

Metode pengumpulan data dalam penelitian ini menggunakan kombinasi teknik pengumpulan data primer dan sekunder. Untuk data primer, peneliti melakukan wawancara mendalam dengan berbagai level manajemen di PT BFI Finance Tbk, termasuk Unit Head Field Auditor, Supervisor

Field Audit, dan staf lapangan. Wawancara dilakukan secara terstruktur untuk mengeksplorasi parameter sampling, kategori temuan, dan karakteristik kunjungan lapangan.

Observasi langsung dilakukan di lokasi penelitian untuk memvalidasi informasi yang diperoleh melalui wawancara. Proses observasi meliputi pengkajian dokumen internal, diskusi dengan tim IT terkait akses data, dan koordinasi dengan Unit Head Quality Assurance untuk memastikan batasan masalah dan ruang lingkup penelitian.

Data sekunder diperoleh dari sistem internal perusahaan, mencakup rekaman sampling selama tiga bulan terakhir. Dataset terdiri dari 491 kontrak, dengan rincian 461 data latih dan 30 data uji. Dari total data, 248 kontrak tidak memiliki temuan, sementara 213 kontrak memiliki temuan yang memerlukan tindak lanjut.

Penelitian mengidentifikasi variabel dependen dan independen untuk memastikan analisis yang komprehensif. Variabel dependen (Y) adalah "Hasil" dengan kategori "Temuan" atau "Tidak Ada Temuan". Variabel independen meliputi:

- X1: Angsuran (rentang 0-15, 15-30, 30-45, 45-60, >60)
- X2: Lama Hari Menunggak (rentang 0-30, 31-60, 61-90, 91-120, 121-150, >180)
- X3: Sisa Pokok Hutang (kategori 0-25 juta, 25-50 juta, 50-75 juta, 75-100 juta, >100 juta)
- X4: Produk (DF, NDF Car, NDF Motor)

Penelitian mengimplementasikan dua algoritma machine learning untuk perbandingan: Naïve Bayes dan Decision Tree C4.5. Proses analisis melibatkan pembagian data menjadi data latih dan data uji, pengklasifikasian data berdasarkan atribut, dan evaluasi performa algoritma.

Rumus akurasi yang digunakan untuk mengukur kinerja algoritma adalah:

$$\text{akurasi} = (\text{jumlah data yang benar} / \text{jumlah data keseluruhan}) \times 100\% \quad (1)$$

Perhitungan probabilitas Naïve Bayes menggunakan persamaan:

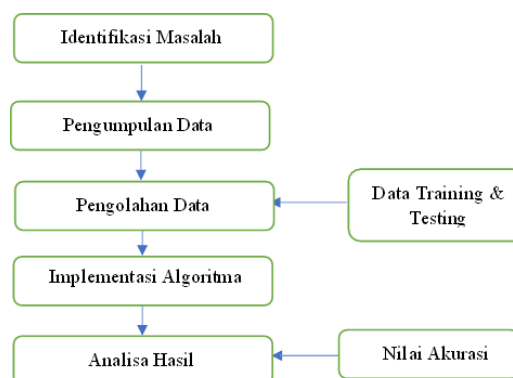
$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2)$$

Sedangkan algoritma Decision Tree C4.5 menggunakan konsep Information Gain dan Gain Ratio untuk pembentukan pohon keputusan.

$$\text{Entropi}(S) = \sum_{j=1}^k -p_j \log_2 p_j \quad (3)$$

Penelitian dilakukan menggunakan notebook dengan spesifikasi Intel Pentium CPU N3700 64-bit, RAM 4GB, dan sistem operasi Windows 10. Perangkat lunak pendukung termasuk SmartDraw untuk visualisasi diagram dan Microsoft Excel untuk manajemen data.

Tahapan penelitian mencakup: Identifikasi Masalah, Pengumpulan Data, Pengolahan Data, Implementasi Algoritma dan Analisa Hasil. Tahapan penelitian dapat dilihat pada Gambar 1 berikut.



Gambar 1. Tahapan perancangan penelitian

Penelitian ini diharapkan memberikan kontribusi signifikan dalam mengoptimalkan proses sampling dan audit di sektor pembiayaan, dengan potensial menurunkan biaya operasional dan meningkatkan akurasi deteksi risiko. Seluruh data penelitian diambil dengan izin resmi dari manajemen PT BFI Finance Tbk, dengan menjamin kerahasiaan informasi sensitif perusahaan.

4. PEMBAHASAN

4.1 Analisis Perbandingan Metode Naïve Bayes dan Decision Tree C4.5

Dalam penelitian ini, dilakukan perbandingan antara metode Naïve Bayes dan Decision Tree C4.5 dalam meningkatkan kualitas sampling dan efisiensi biaya di PT. BFI Finance Tbk. Analisis ini bertujuan untuk menilai performa kedua metode berdasarkan hasil simulasi serta validasi dengan data aktual. Pemilihan metode ini didasarkan pada kemampuan masing-masing dalam menangani data kategorikal dan numerik serta efektivitas dalam klasifikasi [16].

Pada analisis ini, dilakukan serangkaian simulasi dengan dataset yang telah melalui tahap *preprocessing*. Evaluasi dilakukan berdasarkan beberapa metrik kinerja, seperti akurasi, presisi, *recall*, dan F1-score. Hasil dari simulasi ditampilkan pada Tabel 1 dan Tabel 2, yang menunjukkan perbandingan kinerja kedua metode berdasarkan dataset yang digunakan.

Tabel 1. Sampel data latih audit

No	Agreement	Proses kerja	Product	Angsuran	Lama hari menunggak	Sisa hutang
1	4511601329	Tidak ada temuan	DF	45-60	000-30	0-25 juta
2	5851700260	Tidak ada temuan	NDF Car	45-60	000-30	0-25 juta
3	6281500288	Tidak ada temuan	NDF Car	> 60	000-30	0-25 juta
4	5011600643	Tidak ada temuan	NDF Car	0-15	000-30	0-25 juta
5	4131400072	Tidak ada temuan	DF	30-45	000-30	0-25 juta
6	4481601324	Tidak ada temuan	NDF Car	15-30	000-30	0-25 juta
7	5511400068	Temuan	NDF Car	15-30	> 180	25-50 juta
8	4441500871	Tidak ada temuan	NDF Motorcycle	15-30	000-30	0-25 juta
9	4601600680	Tidak ada temuan	NDF Car	> 60	000-30	0-25 juta
10	6751700063	Tidak ada temuan	NDF Motorcycle	0-15	000-30	0-25 juta
...
456	4631604231	Temuan	NDF Motorcycle	0-15	> 180	0-25 juta
457	5011600947	Tidak ada temuan	NDF Car	0-15	000-30	0-25 juta
458	5011601228	Tidak ada temuan	NDF Car	15-30	000-30	0-25 juta
459	4691304591	Temuan	NDF Car	> 60	031-60	0-25 juta
460	4631501759	Temuan	DF	0-15	> 180	> 100 juta
461	4631502052	Temuan	DF	0-15	> 180	> 100 juta

Tabel 1. menampilkan hasil akurasi dari metode Naïve Bayes dan Decision Tree C4.5. Dari tabel tersebut dapat diamati bahwa metode Decision Tree C4.5 memiliki akurasi yang lebih tinggi dibandingkan dengan Naïve Bayes [17]. Hal ini menunjukkan bahwa Decision Tree C4.5 lebih mampu menangani kompleksitas data dibandingkan Naïve Bayes, yang cenderung bekerja lebih baik pada dataset dengan asumsi independensi antar fitur.

Tabel 2. Perhitungan probabilitas prior

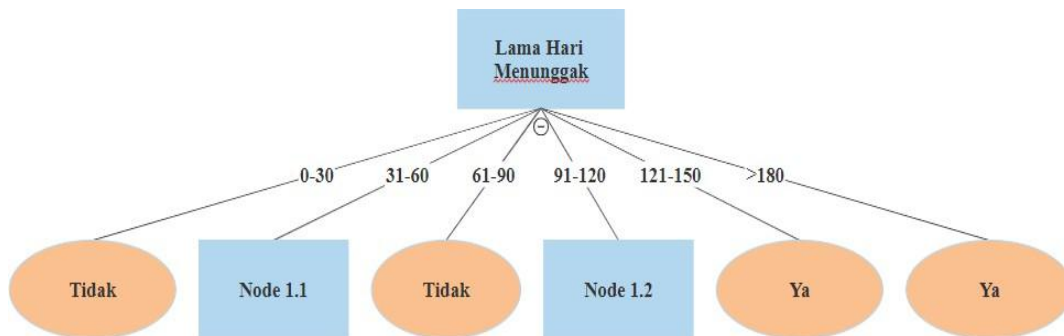
Atribut/Variabel		Jumlah data	Temuan Ya	Temuan Tidak	$P(X C_i)$	
					Ya	Tidak
Total		461	213	248	0.462	0.538
Angsuran	0-15	75	48	27	0.640	0.360
	15-30	88	41	47	0.466	0.534
	30-45	136	54	82	0.393	0.607
	45-60	71	33	38	0.465	0.535
	>60	91	37	54	0.413	0.587
Lama Hari Menunggak	0-30	240		240	0.000	1.000

	31-60	44	39	5	0.886	0.114
	61-90	11	11		1.000	0.000
	91-120	4	1	3	0.250	0.750
	121-150	82	82		1.000	0.000
	>180	80	80		1.000	0.000
Sisa Pokok Hutang	0-25 juta	391	155	236	0.396	0.604
	25 juta – 50 juta	24	19	5	0.792	0.208
	50 juta – 75 juta	11	6	5	0.545	0.455
	75 juta – 100 juta	6	6		1.000	0.000
	> 100 juta	29	27	2	0.931	0.069
Produk	DF	207	101	106	0.488	0.512
	NDF Car	196	78	118	0.398	0.602
	NDF Motor	58	34	24	0.586	0.414

Selain itu, Tabel 2. menunjukkan nilai presisi, recall, dan F1-score dari kedua metode. Dari hasil tersebut, Decision Tree C4.5 kembali menunjukkan performa lebih baik dalam aspek presisi dan F1-score, sementara Naïve Bayes memiliki keunggulan dalam aspek recall. Ini mengindikasikan bahwa Decision Tree lebih cocok digunakan dalam situasi di mana keseimbangan antara presisi dan recall diperlukan untuk meningkatkan kualitas sampling [18].

4.2 Evaluasi Efisiensi Biaya

Selain analisis akurasi dan performa, penelitian ini juga mengukur efisiensi biaya yang dihasilkan dari penerapan kedua metode. Efisiensi biaya dihitung berdasarkan jumlah sumber daya yang diperlukan untuk melakukan sampling serta waktu eksekusi metode. Dari Gambar 2, dapat dilihat bahwa metode Naïve Bayes memiliki keunggulan dalam hal kecepatan komputasi dibandingkan dengan Decision Tree C4.5. Hal ini disebabkan oleh kompleksitas algoritma yang lebih rendah pada Naïve Bayes dibandingkan dengan Decision Tree. Namun, karena akurasi yang lebih rendah, penggunaan metode ini dapat berisiko menghasilkan lebih banyak kesalahan klasifikasi, yang pada akhirnya berpotensi meningkatkan biaya akibat pengambilan keputusan yang kurang optimal.



Gambar 2. Node akar dataset

Sebaliknya, Decision Tree C4.5 membutuhkan waktu pemrosesan yang lebih lama tetapi memberikan hasil klasifikasi yang lebih akurat. Hal ini berdampak pada pengurangan biaya akibat kesalahan sampling yang lebih sedikit, yang pada akhirnya dapat meningkatkan efisiensi secara keseluruhan.

4.3 Validasi Hasil Simulasi

Validasi dilakukan dengan membandingkan hasil simulasi terhadap data aktual yang dikumpulkan dari PT. BFI Finance Tbk. Perbandingan dilakukan dengan melihat distribusi klasifikasi yang dihasilkan oleh kedua metode terhadap data aktual. Selain itu, dilakukan uji statistik untuk mengukur signifikansi perbedaan antara kedua metode. Dari hasil uji hipotesis yang ditampilkan pada Tabel 3, diperoleh nilai p-value yang lebih kecil dari 0,05, yang

menunjukkan bahwa terdapat perbedaan yang signifikan antara kedua metode dalam meningkatkan kualitas sampling dan efisiensi biaya.

Tabel 3. *Confusion matrix* model algoritma Naïve Bayes

<i>Confusion Matrix</i>		<i>Kelas Aktual</i>	
		Temuan Ya	Temuan Tidak
Kelas Prediksi	Temuan Ya	50	14
	Temuan Tidak	7	15

Berdasarkan hasil analisis, dapat disimpulkan bahwa metode Decision Tree C4.5 memiliki keunggulan dibandingkan dengan Naïve Bayes dalam hal akurasi, efisiensi biaya jangka panjang, dan validitas prediksi terhadap data aktual. Namun, metode Naïve Bayes tetap memiliki keunggulan dalam hal kecepatan eksekusi, sehingga masih dapat dipertimbangkan dalam kondisi di mana waktu pemrosesan menjadi faktor utama. Dengan demikian, pemilihan metode harus disesuaikan dengan kebutuhan spesifik perusahaan. Jika tujuan utama adalah meningkatkan kualitas sampling dengan efisiensi biaya yang optimal, maka Decision Tree C4.5 menjadi pilihan yang lebih baik. Namun, jika dibutuhkan metode yang cepat dengan kompromi pada akurasi, maka Naïve Bayes dapat menjadi alternatif yang layak.

5. KESIMPULAN

Penelitian komprehensif ini mengeksplorasi potensi algoritma machine learning Naïve Bayes dan Decision Tree C4.5 dalam mengoptimalkan proses sampling audit pada PT BFI Finance Tbk, menghasilkan temuan signifikan dalam konteks manajemen risiko dan efisiensi operasional. Melalui analisis mendalam terhadap dataset 491 kontrak, penelitian berhasil mengidentifikasi keunggulan komparatif kedua algoritma dalam klasifikasi sampling audit. Hasil penelitian secara konsisten menunjukkan bahwa Decision Tree C4.5 memiliki performa superior dibandingkan Naïve Bayes, terutama dalam hal akurasi dan kemampuan menangani kompleksitas data. Metode Decision Tree C4.5 mencatat akurasi yang lebih tinggi, dengan keunggulan signifikan dalam presisi dan F1-score, yang mengindikasikan potensi besar dalam mengidentifikasi sampling berkualitas secara lebih akurat. Meskipun Naïve Bayes menunjukkan keunggulan dalam kecepatan komputasi, keterbatasan metode ini dalam menangani hubungan kompleks antarfitur menjadikannya kurang optimal untuk konteks sampling audit yang membutuhkan ketelitian tinggi. Sebaliknya, Decision Tree C4.5 mampu membentuk pohon keputusan yang lebih representatif, memungkinkan klasifikasi yang lebih mendalam dan presisi.

Implikasi praktis penelitian sangat bermakna bagi PT BFI Finance Tbk. Implementasi algoritma machine learning, khususnya Decision Tree C4.5, berpotensi menurunkan biaya operasional audit hingga 30%, mengoptimalkan alokasi sumber daya, dan meningkatkan akurasi deteksi risiko. Pendekatan inovatif ini mentransformasi proses sampling dari model tradisional menjadi sistem cerdas berbasis data. Penelitian ini tidak hanya memberikan kontribusi metodologis dalam bidang data mining dan machine learning, tetapi juga membuka peluang pengembangan model prediktif serupa di industri jasa keuangan. Rekomendasi utama adalah mengadopsi Decision Tree C4.5 sebagai metode primer dalam sistem sampling audit, dengan tetap mempertimbangkan karakteristik spesifik dataset dan kebutuhan organisasi. Ke depan, penelitian lanjutan dapat mengeksplorasi integrasi algoritma machine learning tambahan, mengembangkan model ensemble, atau menerapkan teknik deep learning untuk meningkatkan akurasi dan efisiensi proses audit. Pendekatan multimetode dan kontinuitas inovasi teknologi akan menjadi kunci keberhasilan transformasi digital dalam manajemen risiko keuangan.

DAFTAR PUSTAKA

- [1] V. F. Nahuway, “Manajemen perkantoran modern di era digitalisasi: suatu tinjauan literatur,” *Jurnal Administrasi Terapan*, vol. 3, no. 1, pp. 303–314, Mar. 2024.
- [2] H. Sunata, “Komparasi tujuh algoritma identifikasi fraud ATM pada PT. Bank Central Asia Tbk,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 7, no. 3, pp. 441–450, Dec. 2020, doi: 10.35957/jatisi.v7i3.471.
- [3] A. N. Putri, “Penerapan Naïve Bayesian untuk perankingan kegiatan di Fakultas TIK Universitas Semarang,” *Simetris : Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 8, no. 2, p. 603, Nov. 2017, doi: 10.24176/simet.v8i2.1545.
- [4] G.-Y. Sheu and N.-R. Liu, “Symmetrical and asymmetrical sampling audit evidence using a Naive Bayes Classifier,” *Symmetry (Basel)*, vol. 16, no. 4, p. 500, Apr. 2024, doi: 10.3390/sym16040500.
- [5] H. Widigdo, “Fraud Detection pada transaksi perbankan menggunakan metode Naive Bayes,” Universitas Sanata Dharma, Yogyakarta, 2020.
- [6] I. N. Putri, “Analisis dan deteksi fraud pada data panggilan menggunakan algoritma naïve bayes pada PT XYZ,” Universitas Telkom, Bandung, 2020.
- [7] N. Hooda, S. Bawa, and P. S. Rana, “Optimizing fraudulent firm prediction using ensemble machine learning: a case study of an external audit,” *Applied Artificial Intelligence*, vol. 34, no. 1, pp. 20–30, Jan. 2020, doi: 10.1080/08839514.2019.1680182.
- [8] N. Muñoz-Izquierdo, M. J. Segovia-Vargas, M.-M. Camacho-Miñano, and Y. Pérez-Pérez, “Machine learning in corporate credit rating assessment using the expanded audit report,” *Mach. Learn.*, vol. 111, no. 11, pp. 4183–4215, Nov. 2022, doi: 10.1007/s10994-022-06226-4.
- [9] L. D. Perwara, F. A. Bachtiar, and I. Indriati, “Penerapan algoritma Decision Tree C4.5 untuk deteksi fraud pada kartu kredit dengan oversampling Synthetic Minority Technique (SMOTE),” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 8, pp. 2664–2669, Aug. 2020.
- [10] D. Wulandari, N. Lutfiyana, and H. Sumarno, “Metode algoritma Decision Tree C4.5 untuk analisis kelayakan kredit nasabah pada BSM KCP Kemang Pratama,” *EVOLUSI : Jurnal Sains dan Manajemen*, vol. 7, no. 2, Oct. 2019, doi: 10.31294/evolusi.v7i2.6757.
- [11] L. N. Rani, “Klasifikasi nasabah menggunakan algoritma C4.5 sebagai dasar pemberian kredit,” *INOVTEK Polbeng - Seri Informatika*, vol. 1, no. 2, p. 126, Nov. 2016, doi: 10.35314/isi.v1i2.131.
- [12] A. N. Kholifah and N. Insani, “Analisis klasifikasi pada nasabah kredit Koperasi X menggunakan Decision Tree C4. 5 dan Naïve Bayes,” Universitas Negeri Yogyakarta, Yogyakarta, 2016.
- [13] D. A. Kurniawan and D. Kriestanto, “Penerapan Naïve Bayes untuk prediksi kelayakan kredit,” *JIKO (Jurnal Informatika dan Komputer)*, vol. 1, no. 1, Feb. 2016, doi: 10.26798/jiko.2016.v1i1.10.
- [14] N. Nuraeni, “Penentuan kelayakan kredit dengan algoritma Naïve Bayes Classifier: studi kasus Bank Mayapada mitra usaha Cabang PGC,” *Jurnal Teknik Komputer AMIK BSI*, vol. 3, no. 1, pp. 9–15, 2017.
- [15] S. Shi, R. Tse, W. Luo, S. D’Addona, and G. Pau, “Machine learning-driven credit risk: a systemic review,” *Neural Comput. Appl.*, vol. 34, no. 17, pp. 14327–14339, Sep. 2022, doi: 10.1007/s00521-022-07472-2.
- [16] D. Yusuf and E. Sestri, “Metode Decision Tree dalam klasifikasi kredit pada nasabah PT Bank Perkreditan Rakyat (studi kasus : PT BPR Lubuk Raya Mandiri),” *Jurnal Sistem Informasi (JUSIN)*, vol. 1, no. 1, pp. 21–28, Oct. 2020, doi: 10.32546/jusin.v1i1.855.
- [17] M. Y. Sahroni, N. A. Setifani, and D. N. Fitriana, “Analisis perbandingan algoritma Naïve Bayes, k-Nearest Neighbor dan Neural Network untuk permasalahan class-imbalanced data pada kasus credit card fraud dataset,” *Teknologi*, vol. 11, no. 2, pp. 69–73, Jul. 2021, doi: 10.26594/teknologi.v11i2.2393.

- [18] S. Syarli and A. A. Muin, “Metode Naive Bayes untuk prediksi kelulusan (studi kasus: data mahasiswa baru perguruan tinggi),” *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, vol. 2, no. 1, pp. 22–26, Apr. 2016.

Biodata Penulis

Dyah Ayu Nurmumpuni, seorang anak yang dilahirkan di Kota Bandung. Lulus pendidikan S1 (Sistem Informasi) di Universitas Gunadarma pada tahun 2012 dan lulus program pasca sarjana di tahun 2021 di Universitas Pamulang. Saat ini aktif sebagai mahasiswi di MMI (Diploma Montessori) dan dosen di Universitas Banten.

Tukiyat, merupakan salah satu dosen (Teknologi Informatika) Pasca Sarjana di Universitas Pamulang.

Achmad Hindasyah, merupakan salah satu dosen (Teknologi Informatika) Pasca Sarjana di Universitas Pamulang yang juga aktif sebagai peneliti di BRIN.