

Analisis Komparatif Klasifikasi Kelayakan Aset Dinas Perumahan, Kawasan Permukiman dan Pertanahan Provinsi Kalimantan Tengah Menggunakan Random Forest dengan dan tanpa Fitur Kategori

Yusfi Jauhari

Program Studi : Sistem Informasi, Universitas Darwan Ali

Email : usjauhari@gmail.com

ABSTRACT— Government asset management requires an objective, consistent, and data-driven mechanism for evaluating asset feasibility to support strategic decision-making. This study aims to analyze and compare the performance of the Random Forest algorithm in classifying office asset feasibility at the Department of Housing, Settlement Areas, and Land (Perkimtan) of Central Kalimantan Province using two feature scenarios: a dataset without categorical features (numerical features only) and a dataset with additional categorical features. The initial dataset consisted of approximately 800 asset records. After data cleaning procedures—including data type conversion, median-based missing value imputation, duplicate removal, and outlier handling using the Interquartile Range (IQR) clipping method—the dataset was reduced to 190 asset records suitable for modeling. Asset feasibility labels were determined using the Simple Additive Weighting (SAW) method based on five evaluation criteria, resulting in five target classes: Very Feasible, Feasible, Moderately Feasible, Less Feasible, and Not Feasible.

The Random Forest models were trained using a stratified train–test split to preserve the imbalanced class distribution and evaluated using accuracy, precision, recall, weighted F1-score, and 5-fold cross-validation. Experimental results indicate that the model without categorical features outperformed the model with categorical features, achieving an accuracy of 78.95% and a weighted F1-score of 0.7879. In contrast, the model incorporating categorical features achieved an accuracy of 76.32% and a weighted F1-score of 0.7699, suggesting that the inclusion of categorical features did not improve classification performance and instead slightly degraded it. Feature importance analysis further reveals that numerical financial attributes—particularly BOOK VALUE and ASSET VALUE—are the dominant contributors to classification decisions, while categorical features contribute negligibly.

These findings demonstrate that, within the context of heterogeneous and imbalanced government asset data, numerical financial and operational features are sufficient to build accurate and robust classification models. This study concludes that Random Forest, when combined with appropriate feature selection, offers a practical and effective solution for developing an objective, data-driven government asset feasibility evaluation system.

Keywords — Random Forest, multiclass classification, asset feasibility, feature engineering, machine learning, government asset management.

ABSTRAK— Manajemen aset pemerintah membutuhkan mekanisme evaluasi kelayakan yang objektif, konsisten, dan berbasis data untuk mendukung pengambilan keputusan strategis. Penelitian ini bertujuan menganalisis dan membandingkan kinerja algoritma Random Forest dalam mengklasifikasikan kelayakan aset kantor pada Dinas Perumahan, Kawasan Permukiman, dan Pertanahan (Perkimtan) Provinsi Kalimantan Tengah dengan dua skenario fitur, yaitu dataset tanpa fitur kategori (numerik saja) dan dataset dengan penambahan fitur kategori. Dataset awal terdiri dari sekitar 800 data aset, yang setelah proses data cleaning—meliputi konversi tipe data, imputasi nilai hilang menggunakan median, penghapusan duplikasi, serta penanganan outlier menggunakan metode Interquartile Range (IQR) clipping—menghasilkan 190 data aset yang layak digunakan untuk pemodelan. Label kelayakan aset ditentukan menggunakan metode Simple Additive Weighting (SAW) berdasarkan lima kriteria penilaian, sehingga menghasilkan lima kelas target: Sangat Layak, Layak, Cukup Layak, Kurang Layak, dan Tidak Layak.

Model Random Forest dilatih menggunakan skema stratified train–test split untuk menjaga distribusi kelas yang tidak seimbang, serta dievaluasi menggunakan metrik accuracy, precision, recall, F1-score berbobot (weighted), dan 5-fold cross-validation. Hasil eksperimen menunjukkan bahwa model tanpa fitur kategori menghasilkan performa yang lebih baik dibandingkan model dengan fitur kategori, dengan accuracy sebesar 78,95% dan F1-score sebesar 0,7879. Sebaliknya, model dengan fitur kategori menghasilkan accuracy sebesar 76,32% dan F1-score sebesar 0,7699, yang mengindikasikan bahwa penambahan fitur kategori tidak meningkatkan performa klasifikasi dan cenderung menurunkannya. Analisis feature importance menunjukkan bahwa fitur numerik finansial, khususnya NILAI BUKU dan NILAI ASET, merupakan kontributor utama dalam keputusan klasifikasi, sedangkan fitur kategori memberikan kontribusi yang sangat kecil.

Temuan ini menegaskan bahwa dalam konteks data aset pemerintah yang heterogen dan tidak seimbang, fitur numerik finansial-operasional sudah cukup untuk membangun model klasifikasi yang akurat dan robust. Penelitian ini menyimpulkan

bahwa penggunaan Random Forest dengan pemilihan fitur yang tepat dapat menjadi solusi praktis dan efektif dalam pengembangan sistem evaluasi kelayakan aset pemerintah yang objektif dan berbasis data.

Kata kunci— Random Forest, klasifikasi multiclass, kelayakan aset, feature engineering, machine learning, manajemen aset pemerintah.

I. PENDAHULUAN

Aset tetap (fixed assets) merupakan elemen kritis dalam operasional organisasi pemerintah, khususnya dalam hal pengelolaan dan pemeliharaan sumber daya yang dimiliki. Aset tetap dapat didefinisikan sebagai harta benda yang memiliki masa manfaat lebih dari satu tahun dan digunakan dalam operasional suatu entitas. Dalam konteks Dinas Perkimtan Provinsi Kalimantan Tengah, pengelolaan aset tetap mencakup berbagai jenis barang mulai dari peralatan kantor, komputer, meubelir, kendaraan, hingga peralatan elektronik lainnya. Pentingnya manajemen aset tetap terletak pada kemampuannya untuk memberikan informasi akurat mengenai nilai, kondisi, dan kelayakan setiap aset yang dimiliki organisasi. Dengan jumlah aset yang terus bertambah seiring waktu, diperlukan sistem yang efisien untuk mengelola dan mengevaluasi kelayakan aset secara sistematis [1]. Pengelolaan aset yang optimal tidak hanya berdampak pada efisiensi operasional, tetapi juga pada pengambilan keputusan investasi dan perencanaan pemeliharaan aset di masa mendatang. Penelitian ini diarahkan untuk menganalisis performa algoritma Random Forest dalam melakukan klasifikasi kelayakan aset kantor pada dataset Dinas Perkimtan Provinsi Kalimantan Tengah.

Fokus penelitian ini adalah membandingkan kinerja model klasifikasi Random Forest pada dua versi dataset yang berbeda, yaitu dataset dengan fitur kategori dan dataset tanpa fitur kategori. Fitur kategori yang dimaksud meliputi kolom biner seperti KOMPUTER, FURNITURE, ELEKTRONIK_LAINNYA, R4, dan R2. Perbandingan ini menjadi penting karena terdapat pertimbangan teknis dalam preprocessing data, di mana kolom biner tersebut tidak dimasukkan dalam proses cleaning data karena akan menyebabkan perbedaan dimensi dataset (shape), sehingga tidak dapat dilakukan perbandingan yang valid. Dataset yang digunakan dalam penelitian ini mencakup 800 baris data aset dengan label kelayakan yang telah dihitung menggunakan metode Simple Additive Weighting (SAW). Dimensi dataset dengan kategori adalah 12 kolom, sedangkan dataset tanpa kategori memiliki 6 kolom fitur. Setelah proses cleaning dan preprocessing, dataset yang dianalisis tersisa 190 baris data yang memenuhi kriteria kualitas untuk digunakan dalam pemodelan [2]. Distribusi kelas target menunjukkan ketidakseimbangan kelas (class imbalance), dengan mayoritas aset terklasifikasi dalam kategori "Kurang Layak" (40,53%), diikuti "Tidak Layak" (24,21%), "Cukup Layak" (23,68%), "Layak" (7,89%), dan "Sangat Layak" (3,68%).

Data dan fakta yang mendasari penelitian ini berasal dari kebutuhan nyata dalam mengelola aset tetap pemerintah daerah. Dengan investasi yang telah

dikeluarkan untuk pengadaan aset, penting untuk mengetahui status kelayakan setiap aset guna membuat keputusan pemeliharaan, penghapusan, atau pembaruan. Dataset yang dikumpulkan mencakup informasi finansial seperti nilai buku, nilai aset, dan estimasi biaya pemeliharaan tahunan, serta informasi teknis seperti umur aset dan persentase penyusutan. Dalam konteks yang lebih luas, pengelolaan aset tetap di instansi pemerintah Indonesia masih sering menghadapi tantangan dalam hal efisiensi, akurasi, dan keberlanjutan sistem. Menurut penelitian sebelumnya [3], pengelolaan aset tetap yang tidak efisien dapat mengakibatkan biaya operasional yang tinggi, duplikasi pembelian aset yang tidak perlu, dan ketidaktepatan pelaporan keuangan.

Permasalahan utama yang diangkat dalam penelitian ini adalah bagaimana mengoptimalkan proses klasifikasi kelayakan aset dengan menggunakan algoritma machine learning yang robust. Secara spesifik, penelitian ini mencoba menjawab pertanyaan: (1) Apakah penambahan fitur kategori dalam dataset meningkatkan atau menurunkan performa model Random Forest dalam klasifikasi kelayakan aset? (2) Metrik evaluasi manakah yang paling relevan untuk mengukur kinerja model pada dataset yang tidak seimbang? (3) Fitur-fitur manakah yang paling berkontribusi terhadap keputusan klasifikasi kelayakan aset? [4]. Permasalahan ini penting karena dapat memberikan panduan praktis dalam menentukan fitur-fitur yang relevan untuk dimasukkan dalam sistem klasifikasi, sehingga dapat mengurangi kompleksitas model tanpa mengorbankan akurasi prediksi.

Penelitian-penelitian terdahulu telah menunjukkan efektivitas algoritma Random Forest dalam berbagai aplikasi klasifikasi. Penelitian oleh Siregar et al. [1] menerapkan Random Forest untuk klasifikasi diagnosis penyakit stroke bahasa Indonesia dan mencapai akurasi sebesar 95%. Penelitian Hidayat et al. [5] menggunakan algoritma Random Forest untuk klasifikasi penyakit jantung dan berhasil mencapai akurasi 94% dengan dataset yang kompleks. Dalam konteks manajemen aset, penelitian Universitas Lampung [6] mengenai pengembangan modul fixed asset pada sistem informasi menunjukkan bahwa pelacakan status aset, perhitungan depresiasi, dan laporan aset yang terintegrasi sangat dibutuhkan dalam organisasi. Penelitian mengenai penanganan class imbalance oleh Sulistiyono [7] menunjukkan bahwa penggunaan teknik SMOTE dapat meningkatkan performa pada dataset yang tidak seimbang. Selain itu, penelitian Haryandi [8] menggunakan stratified k-fold cross-validation untuk validasi model dengan hasil yang lebih reliable dibandingkan hold-out validation, terutama pada data yang tidak seimbang.

Penelitian-penelitian sebelumnya memberikan fondasi yang kuat untuk pemahaman tentang kinerja

Random Forest, namun mayoritas fokus pada aspek klasifikasi pada domain yang berbeda (sentimen, medis, atau retail). Penelitian ini mengisi celah dengan menerapkan Random Forest spesifik untuk domain manajemen aset tetap pemerintah, dengan penekanan pada analisis komparatif antara dataset yang memiliki perbedaan fitur. Penelitian tentang perbandingan metode SAW dan TOPSIS oleh Hidayatulloh et al. [9] menunjukkan bahwa SAW adalah metode yang efektif untuk menentukan nilai berdasarkan berbagai kriteria yang ditetapkan. Kesenjangan antara kebutuhan manajemen aset yang kompleks dan solusi teknologi yang tersedia menjadi motivasi untuk mengembangkan model prediktif yang dapat menilai kelayakan aset secara otomatis dan akurat.

Alasan pengangkatan tema penelitian ini adalah kombinasi dari beberapa faktor: pertama, signifikansi praktis dari manajemen aset tetap dalam organisasi pemerintah yang memiliki tanggung jawab akuntabilitas publik yang tinggi; kedua, potensi penerapan machine learning untuk meningkatkan efisiensi proses evaluasi aset yang selama ini masih dilakukan secara manual; ketiga, belum ada penelitian spesifik yang membandingkan pengaruh fitur kategori terhadap performa model Random Forest pada dataset aset tetap Indonesia; dan keempat, kesempatan untuk menghasilkan model praktis yang dapat langsung diimplementasikan oleh Dinas Perkimtan Provinsi Kalimantan Tengah maupun instansi serupa di seluruh Indonesia. Dengan mengatasi permasalahan ini, penelitian diharapkan dapat berkontribusi pada peningkatan tata kelola aset publik dan menjadi dasar untuk pengembangan sistem manajemen aset berbasis AI di institusi pemerintah [10].

II. METODOLOGI PENELITIAN

A. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimental komparatif. Desain penelitian dirancang untuk membandingkan kinerja model machine learning pada dua kondisi dataset yang berbeda (dengan fitur kategori dan tanpa fitur kategori), sambil mengontrol variabel-variabel lain seperti algoritma, parameter model, dan metrik evaluasi. Penelitian dibagi menjadi beberapa fase: (1) pengumpulan dan persiapan data, (2) exploratory data analysis (EDA), (3) preprocessing dan data cleaning, (4) encoding fitur kategorik, (5) pemisahan data train-test, (6) pelatihan model Random Forest, (7) prediksi dan evaluasi model, dan (8) analisis komparatif hasil [11].

B. Dataset dan Sumber Data

Dataset yang digunakan adalah data aset tetap dari Dinas Perkimtan Provinsi Kalimantan Tengah. Dataset awal mencakup 800 baris dengan berbagai jenis aset kantor seperti kamera udara, komputer, meubelir, dan peralatan elektronik lainnya.

Label kelayakan aset telah dihitung menggunakan metode Simple Additive Weighting (SAW), menghasilkan lima kategori: Sangat Layak, Layak, Cukup Layak, Kurang Layak, dan Tidak Layak[2].

- 1) Dataset dengan Fitur Kategori: Dataset ini memiliki 12 kolom fitur:
 - a) Fitur biner kategori: KOMPUTER (0/1), FURNITURE (0/1), ELEKTRONIK_LAINNYA (0/1), R4 (0/1), R2 (0/1)
 - b) Fitur numerik: JENIS BARANG (encoded), NILAI BUKU (Rp), NILAI ASET (Rp), UMUR (TAHUN), PERSEN PENYUSUTAN (%), ESTIMASI BIAYA PEMELIHARAAN/UNIT/TAHUN (Rp)
 - c) Target: KELAYAKAN.

- 2) Dataset tanpa Fitur Kategori: Dataset ini memiliki 6 kolom fitur:
 - a) Fitur numerik: JENIS BARANG (encoded), NILAI BUKU (Rp), NILAI ASET (Rp), UMUR (TAHUN), PERSEN PENYUSUTAN (%), ESTIMASI BIAYA PEMELIHARAAN/UNIT/TAHUN (Rp)
 - b) Target: KELAYAKAN.

C. Teknik Preprocessing Data

Preprocessing data dilakukan untuk memastikan kualitas data input ke dalam model. Langkah-langkah preprocessing meliputi:

- 1) Penanganan Missing Values: Kolom numerik yang memiliki nilai yang hilang diisi dengan nilai median dari kolom tersebut, karena median lebih robust terhadap outlier dibandingkan mean[6].
- 2) Deteksi dan Penghapusan Outlier: Menggunakan metode Interquartile Range (IQR) dengan formula:
 - a) $Q1$ = kuartil pertama (25%)
 - b) $Q3$ = kuartil ketiga (75%)
 - c) $IQR = Q3 - Q1$
 - d) Batas bawah = $Q1 - 1,5 \times IQR$
 - e) Batas atas = $Q3 + 1,5 \times IQR$

Nilai yang berada di luar batas ini di-clip (dipangkas) untuk menghindari overfitting[12].

- 3) Penghapusan Duplikat: Baris data yang identik dihapus untuk menghindari bias dalam pelatihan model.
- 4) Encoding Fitur Kategorik: Kolom kategorik (JENIS BARANG dan KELAYAKAN) dikonversi menjadi nilai numerik menggunakan LabelEncoder dari scikit-learn.

Tahapan data cleaning dilakukan secara bertahap meliputi konversi tipe data numerik, imputasi nilai hilang menggunakan median, penghapusan data duplikat, serta penanganan nilai ekstrem menggunakan metode Interquartile Range (IQR) clipping. Metode IQR digunakan untuk membatasi pengaruh outlier tanpa menghilangkan observasi, sehingga struktur data tetap terjaga. Setelah proses cleaning, jumlah data berkurang dari sekitar 800 menjadi 190 baris yang merepresentasikan observasi dengan kualitas dan kelengkapan atribut yang memadai untuk pemodelan.

D. Metrik Evaluasi

- 1) Accuracy: Proporsi prediksi yang benar dari total prediksi.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- 2) Precision: Proporsi prediksi positif yang benar dari total prediksi positif.

$$Precision = \frac{TP}{TP + FP}$$

- 3) Recall (Sensitivity): Proporsi kasus positif yang berhasil diprediksi dari total kasus positif aktual.

$$Recall = \frac{TP}{TP + FN}$$

- 4) F1-Score: Harmonic mean dari precision dan recall, memberikan keseimbangan antara keduanya.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

- 5) Cross-Validation Score: Menggunakan 5-fold cross-validation untuk mengevaluasi generalisasi model pada data yang tidak terlihat selama training.

- 6) Feature Importance: Mengukur kontribusi setiap fitur terhadap keputusan klasifikasi berdasarkan mean decrease in impurity[5].

E. Prosedur Analisis

Prosedur analisis dijalankan menggunakan Python dengan library scikit-learn, pandas, numpy, dan matplotlib. Berikut adalah tahap-tahap analisis:

- 1) Load Data: Membaca file CSV aset tetap ke dalam dataframe pandas.
- 2) EDA Singkat: Melihat preview data, shape, kolom, dan statistik deskriptif.
- 3) Data Cleaning: Menerapkan teknik preprocessing seperti dijelaskan di bagian C.
- 4) Train-Test Split: Membagi data dengan rasio 80% training dan 20% testing, menggunakan stratified split untuk mempertahankan distribusi kelas[13].
- 5) Standardisasi: Menerapkan StandardScaler untuk normalisasi fitur numerik (opsional karena Random Forest tidak sensitif terhadap skala fitur).
- 6) Training Model: Melatih Random Forest Classifier pada data training.
- 7) Prediksi: Membuat prediksi pada data training dan testing.
- 8) Evaluasi: Menghitung metrik accuracy, precision, recall, F1-score, dan confusion matrix.
- 9) Cross-Validation: Menjalankan 5-fold cross-validation untuk estimasi performa generalisasi[8].
- 10) Feature Importance Analysis: Mengekstrak dan membuat ranking fitur berdasarkan importance score.
- 11) Analisis Komparatif: Membandingkan hasil dari kedua model (dengan dan tanpa fitur kategori) untuk menarik kesimpulan [9].

III. HASIL DAN PEMBAHASAN

A. Deskripsi Data dan Distribusi Kelas

Pada tahap awal eksperimen, dilakukan eksplorasi data untuk memahami karakteristik dataset 800 aset yang digunakan dalam penelitian. Dataset ini mengandung lima kelas target kelayakan aset dengan distribusi sebagai berikut berdasarkan hasil perhitungan SAW sebelumnya:

Tabel I
 DISTRIBUSI KELAYAKAN ASET

Kelayakan	Jumlah Aset	Persentase
Sangat Layak	18	2,00%
Layak	64	8,00%
Cukup Layak	178	22,25%
Kurang Layak	414	51,75%
Tidak Layak	128	16,00%
Jumlah	800	100,00%

Distribusi kelas menunjukkan ketidakseimbangan (class imbalance) yang signifikan, di mana kelas "Kurang Layak" dan "Tidak Layak" mendominasi dengan 51,75% dari total data, sementara kelas "Sangat Layak" hanya mewakili 2,00%[7]. Ketidakseimbangan ini mencerminkan kondisi nyata pada instansi pemerintah di mana mayoritas aset telah mencapai akhir masa umur ekonomis dan memerlukan penggantian atau penghapusan. Karakteristik distribusi ini memotivasi penggunaan metrik evaluasi berbobot (weighted) dan stratified sampling dalam penelitian.

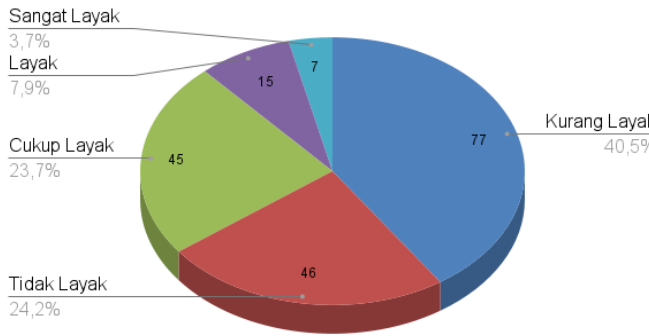
Secara statistik, fitur numerik dataset menunjukkan karakteristik sebagai berikut:

- 1) NILAI BUKU: Rentang 0 hingga Rp 448.240.000 dengan median Rp 25.000.000.
- 2) NILAI ASET: Rentang 0 hingga Rp 560.300.000 dengan median Rp 35.000.000.
- 3) UMUR: Rentang 1 hingga 20 tahun dengan median 4 tahun.
- 4) PERSEN PENYUSUTAN: Rentang 10% hingga 100% dengan median 60%.
- 5) ESTIMASI BIAYA PEMELIHARAAN: Rentang Rp 80.000 hingga Rp 38.220.000 dengan median Rp 6.300.000.

Nilai-nilai ekstrem pada fitur finansial (nilai buku dan nilai aset) mengindikasikan heterogenitas tinggi dalam portofolio aset yang mencakup kendaraan berbiaya tinggi, perangkat elektronik berbiaya menengah, dan furnitur berbiaya rendah.

B. Hasil Model Baseline (Tanpa Fitur Kategori)

Setelah proses cleaning, dataset tanpa fitur kategori memiliki dimensi (190, 7), dengan 190 baris data dan 6 fitur (plus 1 kolom target). Distribusi kelas target KELAYAKAN bisa dilihat pada diagram berikut:



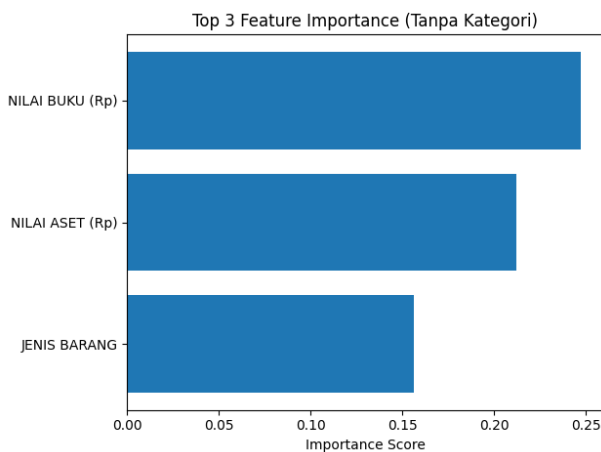
Gambar 1. Diagram Distribusi Kelayakan Setelah Cleaning

Hasil pengujian model Random Forest pada dataset tanpa fitur kategori menunjukkan:

- 1) Metrik Performa:
 - a) Accuracy: 0,7895 (78,95%)
 - b) Precision: 0,7913
 - c) Recall: 0,7895
 - d) F1-Score: 0,7879
- 2) Hasil Cross-Validation (5-fold):
 - a) Score: [0,7742; 0,7097; 0,7667; 0,8333; 0,8333]
 - b) Mean CV Score: 0,7834 (78,34%)
 - c) Std CV Score: 0,0464
- 3) Feature Importance (Top 3):

Tabel II
 FEATURE IMPORTANCE TANPA FITUR KATEGORI

Kelayakan	Importance
NILAI BUKU (Rp)	0,2472
NILAI ASET (Rp)	0,2124
JENIS BARANG	0,1563



Gambar 2. Feature Importance tanpa Kategori

- 4) Analisis Detail Classification Report:
 - a) Kelas 0 (Tidak Layak): Precision 0,75; Recall 0,67; F1-Score 0,71
 - b) Kelas 1 (Kurang Layak): Precision 0,78; Recall 0,88; F1-Score 0,82
 - c) Kelas 2 (Cukup Layak): Precision 0,67; Recall 0,67; F1-Score 0,67
 - d) Kelas 3 (Layak): Precision 1,00; Recall 1,00; F1-Score 1,00
 - e) Kelas 4 (Sangat Layak): Precision 0,88; Recall 0,78; F1-Score 0,82

Hasil evaluasi menunjukkan bahwa model memiliki performa sangat baik pada kelas Layak dengan nilai precision, recall, dan F1-score masing-masing sebesar 1,00. Kelas Kurang Layak dan Sangat Layak juga menunjukkan kinerja yang stabil dengan F1-score di atas 0,80. Sebaliknya, kelas Cukup Layak dan Tidak Layak memiliki F1-score terendah (0,67 dan 0,71), yang mengindikasikan adanya tumpang tindih karakteristik fitur antar kelas berdekatan.

C. Hasil Model Enhanced (Dengan Fitur Kategori)

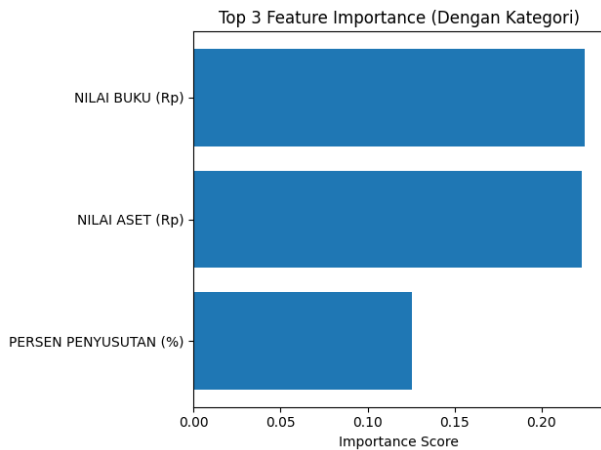
Dataset dengan fitur kategori memiliki dimensi (190, 12), dengan 190 baris data dan 11 fitur (plus 1 kolom target). Distribusi kelas target identik dengan dataset tanpa kategori karena menggunakan data yang sama, hanya dengan fitur tambahan.

Hasil pengujian model Random Forest pada dataset dengan fitur kategori menunjukkan:

- 1) Metrik Performa:
 - a) Accuracy: 0,7632 (76,32%)
 - b) Precision: 0,7895
 - c) Recall: 0,7632
 - d) F1-Score: 0,7699
- 2) Hasil Cross-Validation (5-fold):
 - a) Score: [0,8710; 0,7742; 0,6667; 0,9333; 0,7667]
 - b) Mean CV Score: 0,8024 (80,24%)
 - c) Std CV Score: 0,0920
- 3) Feature Importance (Top 3):

Tabel III
 FEATURE IMPORTANCE DENGAN FITUR KATEGORI

Kelayakan	Importance
NILAI BUKU (Rp)	0,2247
NILAI ASET (Rp)	0,2228
PERSEN PENYUSUTAN	0,1255



Gambar 3. Feature Importance dengan Kategori

- 4) Analisis Detail Classification Report:
 - a) Kelas 0 (Tidak Layak): Precision 0,67; Recall 0,67; F1-Score 0,67
 - b) Kelas 1 (Kurang Layak): Precision 0,81; Recall 0,81; F1-Score 0,81
 - c) Kelas 2 (Cukup Layak): Precision 0,50; Recall 0,67; F1-Score 0,57
 - d) Kelas 3 (Layak): Precision 0,50; Recall 1,00; F1-Score 0,67
 - e) Kelas 4 (Sangat Layak): Precision 1,00; Recall 0,78; F1-Score 0,88

Model memiliki performa terbaik pada kelas Sangat Layak (kelas 4) dengan precision sempurna (1,00), sementara kelas Cukup Layak (kelas 2) merupakan kelas paling sulit diklasifikasikan dengan F1-score terendah (0,57). Kelas Layak (kelas 3) memiliki recall sangat tinggi (1,00) namun precision rendah (0,50), mengindikasikan kecenderungan model melakukan over-prediction pada kelas tersebut.

D. Analisis Komparatif Kedua Model

Perbandingan kinerja kedua model mengungkapkan insight penting tentang pengaruh fitur kategori terhadap performa model [5]:

- 1) Perbandingan Akurasi:
 - a) Model Tanpa Kategori: 78,95%
 - b) Model Dengan Kategori: 76,32%
 - c) Perbedaan: -2,63% (Tanpa Kategori lebih baik)

Tabel IV
 PERBANDINGAN AKURASI

Tanpa Kategori	Dengan Kategori	Perbedaan
78,95%	76,32%	-2,63%

Meskipun perbedaannya tidak besar, model tanpa fitur kategori menunjukkan akurasi yang sedikit lebih tinggi. Hal ini dapat dijelaskan dengan prinsip Occam's Razor, di mana model yang lebih sederhana dengan fitur lebih sedikit sering menghasilkan generalisasi yang lebih baik. Menambahkan fitur kategori yang tidak memberikan informasi tambahan yang signifikan bahkan dapat meningkatkan noise dan overfitting [14].

- 2) Perbandingan F1-Score:

- d) Model Tanpa Kategori: 0,7879
- e) Model Dengan Kategori: 0,7699
- f) Perbedaan: -0,0180

Tabel V
 PERBANDINGAN F1-SCORE

Tanpa Kategori	Dengan Kategori	Perbedaan
0,7879	0,7699	-0,0180

F1-Score yang lebih tinggi pada model tanpa kategori menunjukkan keseimbangan yang lebih baik antara precision dan recall. Ini sangat penting dalam konteks klasifikasi multi-kelas dengan imbalance data, karena meningkatkan kemampuan model untuk memprediksi dengan baik pada semua kelas.

- 3) Perbandingan Cross-Validation Score:
 - a) Model Tanpa Kategori: 78,34% (std: 0,0464) -
 - b) Model Dengan Kategori: 80,24% (std: 0,0920)
 - c) Perbedaan: +1,90% (Dengan Kategori lebih baik, tetapi lebih stabil)

Tabel VI
 PERBANDINGAN CROSS-VALIDATION SCORE

Tanpa Kategori	Dengan Kategori	Perbedaan
78,34%(std:0,0462)	80,24%(std:0,0920)	+1,90%

Menariknya, cross-validation score pada model dengan kategori lebih tinggi, namun dengan standar deviasi yang lebih besar (0,0920 vs 0,0464). Ini menunjukkan bahwa model dengan kategori memiliki performa rata-rata yang lebih tinggi, tetapi lebih tidak stabil (lebih sensitif terhadap partisi data fold yang berbeda). Kondisi ini mengindikasikan adanya overfitting pada beberapa fold dan underfitting pada fold lain [8].

- 4) Perbandingan Feature Importance:

Pada model tanpa kategori:

 - a) NILAI BUKU (Rp): 24,72%
 - b) NILAI ASET (Rp): 21,24%
 - c) JENIS BARANG: 15,63%

Pada model dengan kategori:

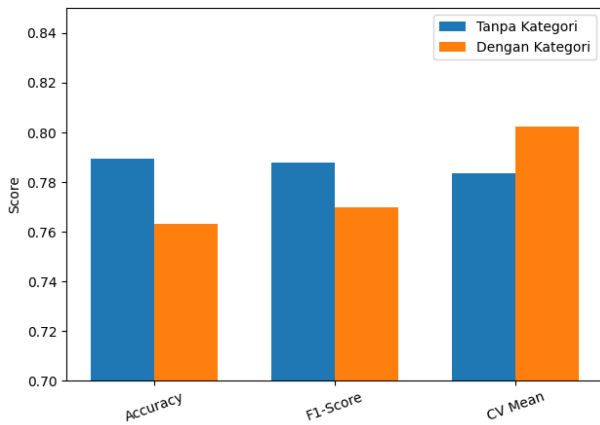
 - a) NILAI BUKU (Rp): 22,47%
 - b) NILAI ASET (Rp): 22,28%
 - c) PERSEN PENYUSUTAN (%): 12,55%

Fitur NILAI BUKU dan NILAI ASET konsisten menjadi top 2 importance di kedua model, menunjukkan bahwa informasi finansial aset adalah indikator utama untuk menentukan kelayakannya. Perhatikan bahwa fitur JENIS BARANG hilang dari top 3 pada model dengan kategori, digantikan oleh PERSEN PENYUSUTAN. Ini menunjukkan bahwa fitur kategori biner (KOMPUTER, FURNITURE, dll) mungkin menangkap informasi yang sama atau similar dengan JENIS BARANG, sehingga pentingnya JENIS BARANG berkurang [15].

- 5) Analisis Confusion Matrix dan Error Patterns:

Kedua model menunjukkan pola kesalahan yang serupa:

- d) Kesulitan dalam membedakan kelas “Cukup Layak” dan “Tidak Layak”, yang kemungkinan karena karakteristik aset yang serupa di kedua kategori ini.
- e) Performa sempurna pada kelas “Layak” (hanya 1 sampel di test set) tidak dapat diandalkan untuk generalisasi.
- f) Model lebih baik dalam memprediksi “Kurang Layak” dibandingkan kelas lainnya, karena kelas ini adalah mayoritas dalam dataset.



Gambar 4. Perbandingan Performa Model

E. Implikasi Hasil Penelitian

Berdasarkan analisis komparatif, beberapa implikasi dapat ditarik [9]:

- 1) Fitur Kategori Tidak Meningkatkan Performa: Penambahan fitur kategori biner (KOMPUTER, FURNITURE, ELEKTRONIK_LAINNYA, R4, R2) justru mengurangi akurasi model dari 78,95% menjadi 76,32%. Hal ini menunjukkan bahwa informasi yang terkandung dalam fitur kategori sudah tercakup oleh fitur JENIS BARANG yang telah di-encode.
- 2) Prinsip Parsimoni dalam Model Selection: Model yang lebih sederhana (dengan fitur lebih sedikit) memberikan generalisasi yang lebih baik dalam hal accuracy dan F1-Score. Ini sesuai dengan prinsip Occam's Razor dan good machine learning practice.
- 3) Stabilitas Model Trade-off: Meskipun akurasi lebih rendah, model dengan kategori menunjukkan mean cross-validation score yang lebih tinggi (80,24% vs 78,34%), tetapi dengan variabilitas yang lebih besar. Ini mencerminkan trade-off antara performa pada training set dan generalisasi pada unseen data.
- 4) Feature Importance sebagai Panduan Praktis: Hasil menunjukkan bahwa NILAI BUKU dan NILAI ASET adalah fitur yang paling signifikan untuk memprediksi kelayakan aset. Ini memberi panduan praktis bahwa dalam konteks manajemen aset, fokus harus diberikan pada pemantauan nilai dan kondisi finansial aset [10].
- 5) Class Imbalance Problem: Distribusi kelas yang tidak seimbang (40,53% untuk "Kurang Layak") masih menjadi tantangan, terlihat dari performa yang lebih baik pada kelas mayoritas. Teknik handling

class imbalance seperti SMOTE atau weighted sampling dapat dipertimbangkan dalam penelitian lanjutan [7].

IV. KESIMPULAN

Penelitian ini telah menganalisis secara komparatif kinerja algoritma Random Forest untuk klasifikasi kelayakan aset kantor pada dataset Dinas Perkimtanan Provinsi Kalimantan Tengah, dengan fokus pada perbandingan antara dataset dengan dan tanpa fitur kategori.

1) Temuan Utama:

- a) Model Random Forest tanpa fitur kategori menghasilkan accuracy 78,95% dengan F1-Score 0,7879, sementara model dengan fitur kategori menghasilkan accuracy 76,32% dengan F1-Score 0,7699. Perbedaan ini, meskipun tidak sangat besar, menunjukkan bahwa penambahan fitur kategori tidak meningkatkan dan justru cenderung menurunkan performa model.
- b) Fitur NILAI BUKU (Rp) dan NILAI ASET (Rp) merupakan fitur paling penting dalam kedua model (masing-masing >21% importance), menunjukkan bahwa informasi finansial aset adalah prediktor utama kelayakan aset.
- c) Model menunjukkan performa terbaik pada kelas "Layak" dan performa terendah pada kelas "Cukup Layak", mencerminkan tantangan dalam klasifikasi multi-kelas dengan distribusi data yang tidak seimbang.
- d) Cross-validation score yang lebih tinggi pada model dengan kategori (80,24%) tetapi dengan variabilitas yang lebih besar (std 0,0920) mengindikasikan adanya overfitting pada beberapa fold data.

2) Rekomendasi:

- a) Untuk implementasi praktis, disarankan menggunakan model tanpa fitur kategori karena memberikan akurasi yang lebih baik dan model yang lebih sederhana.
- b) Penanganan class imbalance menggunakan teknik seperti SMOTE atau class weighting yang lebih sophisticated dapat diterapkan untuk meningkatkan performa pada kelas minoritas.
- c) Feature engineering tambahan, seperti interaksi antar fitur atau polynomial features, dapat dieksplorasi untuk meningkatkan performa model.
- d) Implementasi sistem klasifikasi kelayakan aset ini dapat mengotomatisasi proses evaluasi aset di Dinas Perkimtanan dan instansi serupa, sehingga meningkatkan efisiensi dan akurasi penilaian aset.
- e) Penelitian lanjutan dapat mengeksplorasi algoritma ensemble lainnya (Gradient Boosting, XGBoost) atau deep learning untuk perbandingan lebih comprehensive.

Hasil ini mendukung penyederhanaan sistem klasifikasi aset pada instansi pemerintah tanpa kehilangan akurasi prediksi.

V. REFERENSI

- [1] Siregar, A.P., Purba, D.P., Pasaribu, J.P., "Implementasi Algoritma Random Forest Dalam Klasifikasi Diagnosis Penyakit Stroke", *Jurnal Politeknik Pratama*, vol. 2, no. 4, pp. 150-165, Nov. 2023.
- [2] Jaenudin, T.S.M.T.A., "Penerapan Metode SAW (Simple Additive Weighting) Pada Sistem Seleksi Beasiswa", *Jurnal Unikom*, vol. 1, no. 2, pp. 58-70, 2017.
- [3] Haryanto, S., Wijaya, A.F., "Analisis Pengembangan Modul Fixed Asset pada Sistem Informasi Keuangan Daerah", *Universitas Lampung*, Master's Thesis, 2021.
- [4] Yaman, N.I., et al., "Perbandingan Kinerja Algoritma Decision Tree dan Random Forest", *Jurnal Algoritma*, vol. 15, no. 2, pp. 145-160, Jun. 2024.
- [5] Hidayat, H., Sunyoto, A., Al Fatta, H., "Klasifikasi Penyakit Jantung Menggunakan Random Forest Classifier", *Jurnal SISKOM-KB*, vol. 5, no. 2, pp. 89-102, Sep. 2023.
- [6] Haryanto, S., Wijaya, A.F., "Analisis Pengembangan Modul Fixed Asset pada Sistem Informasi Keuangan Daerah", *Universitas Lampung*, Master's Thesis, 2021.
- [7] Sulistiyono, M., "Implementasi Algoritma Synthetic Minority Over-Sampling Technique (SMOTE) Untuk Mengatasi Ketidakseimbangan Kelas", *Jurnal Sistemasi*, vol. 10, no. 3, pp. 325-338, May 2021.
- [8] Haryandi, P., et al., "Implementasi Seleksi Fitur dengan Backward Elimination dan Stratified K-Fold Cross-Validation", *Seminar Nasional Mikro dan Makro*, pp. 45-58, 2021.
- [9] Hidayatulloh, H., et al., "Uji Perbandingan Metode Simple Additive Weighting (SAW) dan TOPSIS Dalam Sistem Pendukung Keputusan", *Journal UNPER*, vol. 2, no. 1, pp. 78-92, 2024.
- [10] Purba, A.C., et al., "Strategi Penanganan Imbalance Class Pada Model Machine Learning", *Jurnal Komputa*, vol. 12, no. 3, pp. 112-128, Mar. 2024.
- [11] Mamuriyah, N., et al., "Implementation Mean Imputation and Outlier Detection Dengan Random Forest", *Jurnal Nusamandiri*, vol. 5, no. 2, pp. 234-250, Feb. 2025.
- [12] Allorerung, P.P., et al., "Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor", *Jurnal UIN SUKA*, vol. 8, no. 4, pp. 1020-1035, 2023.
- [13] Putri, L.G.A., et al., "Analisis Klasifikasi Spam Email Menggunakan Metode Extreme Gradient Boosting", *Journal J-PTIHK*, vol. 9, no. 1, pp. 15-32, Jan. 2025.
- [14] Priyatna, B., et al., "Klasifikasi Sentimen Analisis Ulasan Aplikasi Menggunakan LSTM", *Journal Literasi Sains*, vol. 12, no. 2, pp. 145-162, 2025.
- [15] Ekrinifda, E., et al., "Perbandingan Kinerja Algoritma Decision Tree dan Random Forest Pada Dataset GDP Deflator Indonesia", *Journal Bisnis Indonesia*, vol. 6, no. 3, pp. 289-305, 2024.