Jurnal Manajemen dan Teknologi
https://journal.stmiki.ac.id

E-ISSN: 3047-8251 | P-ISSN: 3047-826X
Vol. 3 No. 2 (2026) | November

OPEN ACCESS

https://doi.org/ 10.63447/jmt.v3i2.2195

Klasifikasi Kualitas Karet Menggunakan Metode *Random Forest* Pada Perusahaan XYZ

Ongki Juliandra [1], Alsella Meiriza [2*], Hardini Novianti [3], Mira Afrina [4], Ken Ditha Tania [5]

[1,2*,3,4,5] Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya, Kota Palembang, Provinsi Sumatera Selatan, Indonesia.

Corresponding Email: ongkijuliandra2@gmail.com [1]

Histori Artikel:
Dikirim 16 Oktober 2026; *Diterima dalam bentuk revisi* 20 Oktober 2026; *Diterima* 30 Oktober 2026; *Diterbitkan* 25 November 2026. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Perusahaan XYZ, pengolah crumb rubber di Sumatra, Indonesia, menetapkan kategori kualitas Low, Medium, atau High pada setiap lot produksi melalui pengujian manual di laboratorium, sementara parameter proses yang dicatat belum dihubungkan dengan kategori tersebut. Penelitian ini mengukur seberapa baik Random Forest meniru kategori itu dari enam parameter (suhu oven, viskositas, ukuran partikel, kadar abu, kekerasan, dan elastisitas) pada 1.031 lot periode Februari hingga Juni 2026, dengan hanya 13,5% lot berkategori Low. Ketidakseimbangan kelas ditangani dengan bobot kelas seimbang. Random Forest dibandingkan dengan Support Vector Machine (SVM), Decision Tree, dan dua pembanding naif melalui validasi hold-out, validasi silang berulang, dan validasi silang bersarang. Pada validasi silang, Random Forest hasil penyetelan mencapai macro-F1 0,409, lebih tinggi daripada Decision Tree (0,365; p = 0,028) dan tidak berbeda nyata dari SVM (0,385; p = 0,13). Namun, nilai itu hanya sedikit di atas tebakan acak berstrata (0,330), dan tidak ada model yang melampaui akurasi tebakan kelas High saja (0,507). Kelas Low paling sulit dikenali (F1 sekitar 0,23). Keenam parameter hampir tidak berbeda antarkelas, sehingga keterbatasan tampaknya terletak pada informasi dalam data. Model hanya layak menjadi alat bantu penyaringan awal untuk menandai lot yang perlu diuji ulang di laboratorium, dan tidak boleh dipakai untuk meloloskan atau menolak lot.

Kata Kunci: Random Forest; Kualitas Karet; Klasifikasi Multikelas; Ketidakseimbangan Kelas; Validasi Silang; Parameter Proses.

Abstract

Company XYZ, a crumb rubber processor in Sumatra, Indonesia, assigns each production lot a Low, Medium, or High quality category by laboratory testing, but recorded parameters are not linked to them. This study tests how well Random Forest reproduces the categories from six parameters in 1,031 lots, only 13.5% of which are Low. Random Forest was compared with Support Vector Machine (SVM), Decision Tree, and two naive baselines using hold-out, repeated, and nested cross-validation. Cross-validated macro-F1 of the tuned Random Forest was 0.409, higher than Decision Tree (0.365; p = 0.028) and not significantly different from SVM (0.385; p = 0.13), yet only slightly above random guessing (0.330). No model exceeded the accuracy of always guessing High (0.507), and the Low class fared worst (F1 about 0.23). The six parameters barely differ between classes, so the data carry little class information. The model suits only flagging lots for laboratory re-testing.

Keyword: Random Forest; Rubber Quality; Multiclass Classification; Class Imbalance; Cross-Validation; Process Parameters.

Jurnal Manajemen dan Teknologi
https://journal.stmiki.ac.id

E-ISSN: 3047-8251 | P-ISSN: 3047-826X
Vol. 3 No. 2 (2026) | November

OPEN ACCESS

https://doi.org/ 10.63447/jmt.v3i2.2195

1. Pendahuluan

Asia Tenggara mendominasi produksi karet alam, dan Indonesia memiliki salah satu area perkebunan terluas di kawasan ini (Littidej et al., 2024). Karet blok seperti STR 20 dibuat dari koagulum kebun seperti *cup lump* (Puttipipatkajorn & Puttipipatkajorn, 2021), dan kategori kualitas dari laboratorium menentukan penggolongan dan harga suatu lot. Pembelajaran mesin pada karet sejauh ini terutama diterapkan di hulu: memperkirakan hasil panen dari variabel spasial (Littidej et al., 2024), kadar karet kering dari data cuaca historis (Areewattana et al., 2025), dan kadar karet kering *cup lump* dari spektrum (Puttipipatkajorn & Puttipipatkajorn, 2021). Di sisi pabrik, Putri et al. (2024) memantau viskositas Mooney dan *plasticity retention index* (PRI) pada *crumb rubber* SIR 3CV 60 dengan peta kendali dan menemukan Cpk di bawah 1 untuk keduanya, tetapi pendekatan itu mendeteksi penyimpangan parameter uji dan tidak memprediksi kategori kualitas suatu lot. Di manufaktur yang lebih luas, *machine learning* dipakai antara lain untuk pengendalian kualitas (Rosihan et al., 2026). *Random Forest* pernah memprediksi jumlah *chip* tertolak dari data mesin (Supriyadi et al., 2025), dan pada data tabular berukuran sedang model berbasis pohon mengungguli jaringan saraf dalam (Grinsztajn et al., 2022). *Random Forest*, *Support Vector Machine* (SVM), dan *Decision Tree* termasuk pengklasifikasi dasar yang banyak dipakai (Alnuaimi & Albaldawi, 2024) dan pernah dibandingkan pada data kualitas berkelas tidak seimbang, misalnya kualitas anggur (Zaza et al., 2023). Dua masalah evaluasi muncul saat kualitas diklasifikasikan dari data proses. Pertama, kategori kualitas jarang seimbang, dan akurasi lalu memberi nilai tinggi pada model yang condong ke kelas mayoritas (Ghosh et al., 2024); metrik rata-rata makro memberi bobot sama pada tiap kelas tetapi punya kelemahan sendiri (Opitz, 2024). Kedua, satu kali pembagian data latih dan uji dengan beberapa ratus kasus memberi estimasi yang tidak stabil, dan estimasi varians yang lazim pada validasi silang terlalu kecil karena antarfold berbagi data latih (Bates et al., 2024). Bila hiperparameter disetel, penyetelan harus berada di dalam putaran validasi agar skor tidak terlalu optimistis (Bischl et al., 2023).

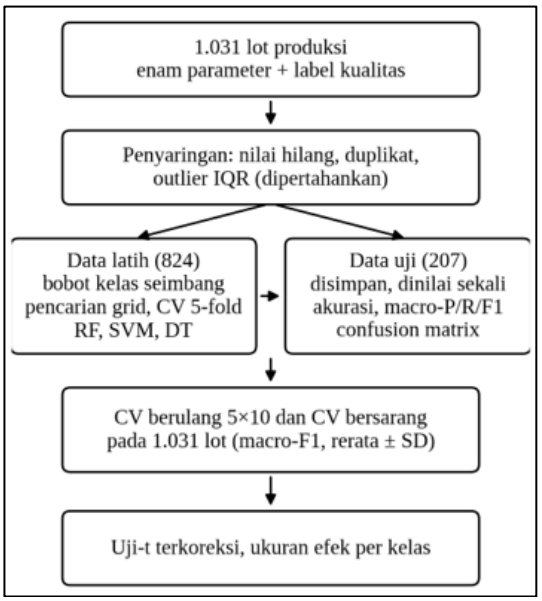
Penelitian ini berbeda dari studi-studi di atas pada objek, target, dan evaluasinya. Objeknya kategori kualitas yang ditetapkan laboratorium pabrik pada produk olahan, bukan hasil panen atau mutu bahan baku; targetnya tiga kelas tidak seimbang, sedangkan studi hulu memprediksi besaran kontinu; dan evaluasinya memakai tiga skema validasi dengan uji-t terkoreksi untuk resampling berulang. Di antara studi yang ditinjau di sini, belum ada yang menguji apakah parameter yang dicatat pabrik karet remah cukup untuk memprediksi kategori kualitas buatan laboratoriumnya, dan itulah kesenjangan yang diisi penelitian ini. *Random Forest* dipilih sebagai model utama karena datanya tabular dan berukuran sedang (1.031 lot, enam parameter numerik berskala sangat berbeda), kondisi ketika model berbasis pohon cenderung unggul (Grinsztajn et al., 2022). Model ini tidak memerlukan standardisasi, menangkap hubungan non-linear dan interaksi tanpa asumsi bentuk fungsi, lebih tahan derau daripada *boosting* , menekan *overfitting* dibanding pohon tunggal (Salman et al., 2024), menerima pembobotan kelas (Akbar & Sanjaya, 2023), dan menyediakan ukuran kepentingan variabel. SVM (berbasis margin) dan *Decision Tree* (pohon tunggal) dipilih sebagai pembanding karena mewakili pendekatan yang berbeda. Perusahaan XYZ, pengolah karet remah di Sumatra, Indonesia, mencatat enam parameter untuk setiap lot, dan laboratoriumnya menetapkan kategori *Low*, *Medium*, atau *High*; belum ada model yang menghubungkan keduanya. Penelitian ini (i) membangun *Random Forest* yang memprediksi kategori dari keenam parameter, (ii) membandingkannya dengan SVM, *Decision Tree*, dan dua pembanding naif melalui validasi *hold-out*, validasi silang berulang, dan validasi silang bersarang, serta (iii) memeriksa apakah keenam parameter memuat informasi kelas yang cukup bagi pengklasifikasi apa pun. Kontribusinya adalah kasus terdokumentasi saat alur kerja standar yang divalidasi cermat hanya menghasilkan kinerja sedang, disertai bukti penyebabnya.

Jurnal Manajemen dan Teknologi
https://journal.stmiki.ac.id

E-ISSN: 3047-8251 | P-ISSN: 3047-826X
Vol. 3 No. 2 (2026) | November
OPEN ACCESS
https://doi.org/ 10.63447/jmt.v3i2.2195

2. Metode Penelitian

Gambar 1 merangkum alur kerja. Seluruh komputasi memakai Python dengan scikit-learn , pandas, dan SciPy pada satu inti CPU, dan *random seed* yang dipakai adalah 42.



Gambar 1. Alur Kerja Penelitian

2.1 Data dan Penyaringan

Data berupa catatan produksi Perusahaan XYZ periode 1 Februari hingga 2 Juni 2026 yang diterima secara elektronik pada Agustus 2026, tanpa identitas perusahaan dan kode lot. Setiap dari 1.031 catatan adalah satu lot dengan enam parameter numerik (suhu oven dalam °C, viskositas dalam cP, ukuran partikel dalam mesh, kadar abu dalam %, kekerasan dalam Shore A, dan elastisitas dalam MPa) serta kategori kualitas yang ditetapkan laboratorium setelah pengujian manual. Tiga kolom administratif (nomor catatan, tanggal, kode lot) tidak dipakai. Kategori *High*, *Medium*, dan *Low* masing-masing mencakup 523 (50,7%), 369 (35,8%), dan 139 (13,5%) lot, dengan rasio 3,8 berbanding 1 antara kelas terbesar dan terkecil. Penetapan kategori mutu di pabrik karet remah lazim mengacu pada standar karet spesifikasi teknis Indonesia (SNI 1903), yang menggolongkan mutu menurut batas parameter uji seperti plastisitas awal (Po), PRI, viskositas Mooney, kadar abu, kadar kotoran, dan kadar zat menguap (Achmad et al., 2023). Kriteria rinci kategori *Low*, *Medium*, dan *High* di Perusahaan XYZ merupakan penilaian internal laboratorium, sedangkan dataset hanya memuat keenam parameter dan kategori akhirnya. Kategori karena itu diperlakukan sebagai label acuan, dan seluruh kinerja model diukur terhadapnya. Bila kadar abu dan viskositas ikut menentukan kategori, hubungannya dengan kategori seharusnya lebih kuat daripada yang teramati pada Tabel 1. Keenam parameter adalah parameter numerik yang tersedia pada catatan produksi tiap lot, sehingga pemilihannya mengikuti ketersediaan data. Kaitan tiap parameter dengan mutu karet remah didukung literatur. Viskositas Mooney dan kadar abu termasuk parameter mutu SIR 20 (Achmad et al., 2023), dan Putri et al. (2024) menempatkan viskositas Mooney sebagai satu dari dua parameter yang harus dijaga pada *crumb rubber*. Suhu oven dan ukuran partikel adalah kondisi pengeringan; suhu pengering yang sesuai dan ukuran butir yang seragam menentukan efektivitas pengeringan dan mutu akhir SIR 20 (Prastia et al., 2025). Kekerasan dan elastisitas adalah sifat fisik vulkanisat, dan sifat semacam ini (kekerasan, kuat tarik, perpanjangan putus) dipakai untuk membandingkan jenis karet alam, termasuk karet standar Indonesia. Data disaring terhadap nilai hilang, baris duplikat, dan kode lot duplikat. Nilai di luar $Q1 - 1{,}5 \times IQR$ dan $Q3 + 1{,}5 \times IQR$ ditandai lalu dibandingkan dengan distribusi frekuensi tiap parameter: nilai yang berulang pada

Jurnal Manajemen dan Teknologi
https://journal.stmiki.ac.id
E-ISSN: 3047-8251 | P-ISSN: 3047-826X
Vol. 3 No. 2 (2026) | November
OPEN ACCESS
https://doi.org/ 10.63447/jmt.v3i2.2195

banyak lot dianggap pengaturan sah, dan nilai yang muncul sekali dianggap kemungkinan salah catat. Tidak ada nilai yang dihapus, karena sebagian besar nilai yang ditandai ternyata pengaturan sah. Keputusan ini sesuai untuk model berbasis pohon, yang menentukan pembelahan dari urutan nilai. Pada SVM yang peka terhadap skala, pengaruh tiga nilai yang diduga salah catat diperkirakan kecil (tiga dari 1.031 lot); analisis sensitivitas tanpa ketiganya belum dilakukan.

2.2 Model dan Penyetelan

Ketidakseimbangan kelas ditangani dengan pembobotan kelas pada fungsi kerugian, bukan dengan *resampling*. Untuk n lot latih, K kelas, dan n_c lot pada kelas c, bobot seimbang kelas c adalah:

$$w_c = \frac{n}{Kn_c} \quad (1)$$

Sehingga satu lot *Low* dihitung sekitar 3,8 kali lebih berat daripada satu lot *High*. SMOTE tidak dipakai karena membentuk titik sintetis melalui interpolasi, dan titik-titik itu belum tentu mengikuti distribusi asli kelas minoritas (Elreedy et al., 2024). Dengan hanya 139 lot Low, titik semacam itu dapat menyimpangkan gambaran kelas. *Random Forest* adalah model utama (alasannya pada Bagian 1), SVM menjadi pembanding berbasis margin, dan *Decision Tree* pohon tunggal. *Random Forest* memakai bobot kelas seimbang (Akbar & Sanjaya, 2023) dan 100 pohon, nilai bawaan scikit-learn. SVM memakai kernel *radial basis function* dan fitur terstandardisasi karena bergantung pada jarak sedangkan rentang parameter sangat berbeda (viskositas 1.500 hingga 4.000 cP, kadar abu 0,1 hingga 0,9%). *Decision Tree* memakai pengaturan bawaan, dan ketiga model berbobot kelas seimbang. Dua pembanding naif ditambahkan: selalu memprediksi kelas mayoritas (*High*), dan menarik label acak sesuai frekuensi kelas pada data latih. *Random Forest* disetel melalui pencarian menyeluruh atas 108 kombinasi jumlah pohon (100, 200, 300), kedalaman maksimum (tanpa batas, 10, 20, 30), jumlah sampel minimum untuk membelah simpul (2, 5, 10), dan jumlah sampel minimum per daun (1, 2, 4). Setiap kombinasi dinilai dengan validasi silang 5-fold pada data latih (Bischl et al., 2023), dengan kriteria macro-F1 agar kelas *Low* yang kecil dihitung sama pentingnya dengan kelas *High*.

2.3 Evaluasi

Tiga skema evaluasi dipakai. Pada skema *hold-out*, pembagian berstrata 80:20 menghasilkan 824 lot latih dan 207 lot uji; setiap model dilatih pada data latih dan dinilai satu kali pada data uji. Pada skema berulang, validasi silang berstrata 5-fold diulang 10 kali (50 fold uji) untuk pembanding naif, algoritma tanpa penyetelan, dan *Random Forest* dengan parameter hasil penyetelan yang dibuat tetap. Karena parameter tersebut dipilih pada data yang beririsan dengan fold validasi silang, baris itu mungkin sedikit optimistis. Skema bersarang menghilangkan bias ini: pada setiap dari lima fold luar, pencarian grid penuh dijalankan pada bagian latih luar, dan model pemenang dinilai pada bagian uji luar. Validasi bersarang hanya dijalankan satu kali pengulangan (lima fold luar dengan masing-masing 540 pelatihan model) karena keterbatasan waktu komputasi. Akurasi serta presisi, *recall*, dan F1 rata-rata makro dilaporkan. Macro-F1 adalah rata-rata skor F1 per kelas:

$$F1_{\text{macro}} = \frac{1}{K}\sum_{k=1}^{K}\frac{2P_kR_k}{P_k+R_k} \quad (2)$$

Dengan P_k dan R_k adalah presisi dan *recall* kelas k serta K = 3. Macro-F1 memberi kelas *Low* bobot yang sama dengan kelas *High*, hal yang tidak dilakukan akurasi. Kedua metrik dilaporkan bersama nilai per kelas. Model dibandingkan melalui selisih antarfold d_j pada suatu metrik. Karena antarfold beririsan pada data latih, uji-t biasa meremehkan varians (Bates et al., 2024), sehingga dipakai uji-t terkoreksi untuk resampling berulang:

Jurnal Manajemen dan Teknologi
https://journal.stmiki.ac.id

E-ISSN: 3047-8251 | P-ISSN: 3047-826X
Vol. 3 No. 2 (2026) | November
OPEN ACCESS
https://doi.org/ 10.63447/jmt.v3i2.2195

$$t = \frac{\bar{d}}{\sqrt{\left(\frac{1}{J}+\frac{n_{te}}{n_{tr}}\right)s_d^2}} \qquad (3)$$

Dengan J = 50 adalah jumlah fold, $\bar{d}$ dan s_d^2 adalah rerata dan varians d_j, serta $n_{te}/n_{tr} = 1/4$ adalah rasio ukuran data uji terhadap data latih. Nilai-p diperoleh dari distribusi t dengan $J - 1$ derajat bebas. Keenam parameter dibandingkan antarkelas dengan uji Kruskal-Wallis, memakai $\varepsilon^2 = H(n + 1)/(n^2 - 1)$ sebagai ukuran efek. Ketiga skema mengacak lot dari satu periode (1 Februari hingga 2 Juni 2026); pengujian pada periode lain belum dilakukan.

3. Hasil dan Pembahasan

3.1 Hasil

3.1.1 Karakteristik Data

Tidak ada nilai hilang maupun duplikat. Tabel 1 memuat rerata dan simpangan baku tiap parameter per kelas, dan reratanya berdekatan pada keenam parameter. Suhu oven, viskositas, dan elastisitas tidak berbeda antarkelas (p = 0,94; 0,16; dan 0,90). Ukuran partikel, kadar abu, dan kekerasan berbeda (p = 0,031; 0,013; dan 0,004), tetapi ukuran efeknya sangat kecil (ε^2 0,007 hingga 0,011) dan hanya kekerasan yang tetap di bawah ambang Bonferroni 0,0083.

Tabel 1. Rerata±SD keenam parameter per kelas dan ukuran efek Kruskal-Wallis (ε^2)

Parameter	*Low*	*Medium*	*High*	ε^2
Oven (°C)	122,2±12,6	122,6±11,4	122,3±11,4	0,000
Viskositas (cP)	2.496±784	2.610±751	2.524±704	0,004
Ukuran partikel (mesh)	96,1±56,4	102,9±52,5	107,1±52,2	0,007
Kadar abu (%)	0,46±0,23	0,51±0,22	0,52±0,22	0,009
Kekerasan (Shore A)	67,6±11,8	69,0±11,9	70,9±10,0	0,011
Elastisitas (MPa)	3,17±1,78	3,12±1,64	3,07±1,38	0,000

Penyaringan IQR menandai 94 nilai suhu oven di bawah 105 °C, 328 nilai ukuran partikel di luar 50 hingga 130 mesh, dan satu nilai kekerasan sebesar 20 Shore A. Sebagian besar adalah pengaturan wajar yang berada di luar batas karena kebanyakan lot memakai pengaturan lain: 93 nilai suhu oven tepat 100 °C, serta 121 nilai ukuran partikel berupa 20 atau 40 mesh dan 206 berupa 200 mesh. Tiga nilai yang tersisa (10 °C, 250 mesh, dan 20 Shore A) hanya muncul sekali dan mungkin salah catat; ketiganya dipertahankan karena tiga lot dari 1.031 kecil kemungkinannya memengaruhi analisis berbasis pohon. Dari sisi kontribusi pemisahan kelas, kekerasan teratas (ε^2 = 0,011), disusul kadar abu (0,009) dan ukuran partikel (0,007), sedangkan suhu oven dan elastisitas praktis nol. Urutan ini univariat dan semua efeknya sangat kecil, jadi tidak ada parameter yang sendirian memuat informasi kelas memadai; kepentingan variabel di dalam model *Random Forest* belum dihitung.

3.1.2 Hasil Hold-out

Tabel 2 menunjukkan hasil *hold-out*. Selalu memprediksi *High* berakurasi 0,507 pada 207 lot uji, tetapi macro-F1-nya hanya 0,224 karena tidak pernah mengenali lot *Low* atau *Medium*. *Random Forest* bawaan punya akurasi tertinggi di antara model yang belajar (0,493), tetap di bawah pembanding itu. Macro-F1 ketiga algoritma hampir sama (0,380 untuk *Random Forest*, 0,387 untuk SVM, 0,372 untuk *Decision Tree*), sehingga pembagian ini tidak memisahkan mereka. Galat baku akurasi sekitar 0,5 pada 207 lot adalah sekitar 0,035, sama besar dengan selisih pada Tabel 2.

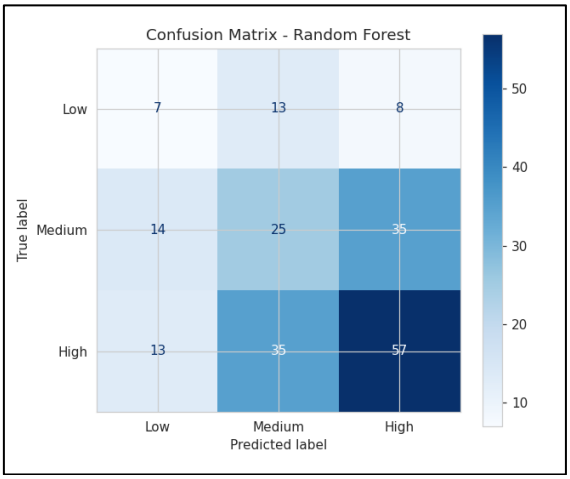
Tabel 2. Hasil hold-out pada 207 lot uji

Model	Akurasi	Macro-P	Macro-R	Macro-F1
Mayoritas	0,507	0,169	0,333	0,224

Jurnal Manajemen dan Teknologi
https://journal.stmiki.ac.id
E-ISSN: 3047-8251 | P-ISSN: 3047-826X
Vol. 3 No. 2 (2026) | November
OPEN ACCESS
https://doi.org/ 10.63447/jmt.v3i2.2195

RF, bawaan	0,493	0,395	0,383	0,380
SVM	0,435	0,392	0,402	0,387
Decision Tree	0,440	0,373	0,372	0,372
RF, disetel	0,430	0,373	0,377	0,374

Penyetelan memilih 100 pohon, kedalaman maksimum 10, minimum 10 sampel untuk membelah, dan minimum 4 sampel per daun. Pada pembagian ini model hasil penyetelan lebih rendah daripada yang bawaan (akurasi 0,430 dibanding 0,493; macro-F1 0,374 dibanding 0,380), tetapi selisih 0,006 itu jauh lebih kecil daripada derau yang mungkin terjadi, dan validasi silang (Bagian 3.3) menunjukkan arah sebaliknya. Gambar 2 menampilkan *confusion matrix* model hasil penyetelan. Lot *High* dikenali paling baik (*recall* 0,543; presisi 0,570), disusul *Medium* (0,338 dan 0,343), dan *Low* paling buruk (0,250 dan 0,206). Kesalahan terbesar adalah pertukaran antara *Medium* dan *High*: 35 lot *Medium* diprediksi sebagai *High* dan 35 lot *High* sebagai *Medium*. Hanya 7 dari 28 lot *Low* yang ditemukan, dan 34 lot diprediksi sebagai *Low*, sehingga 27 dari prediksi tersebut keliru.



Gambar 2. *Confusion matrix Random Forest* hasil penyetelan pada data uji hold-out (jumlah dan recall per baris)

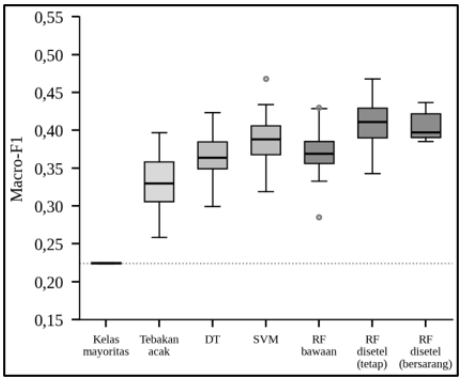
3.1.3 Hasil Validasi Silang

Tabel 3 memuat hasil validasi silang berulang dan bersarang, dan Gambar 3 menunjukkan sebaran macro-F1 antarfold. Pada akurasi, tidak ada model yang melampaui kelas mayoritas (0,507). *Random Forest* bawaan paling mendekati (0,502; $p = 0,71$ terhadap pembanding), dan SVM jelas di bawahnya (0,439; $p < 0,001$). Pada macro-F1 gambarnya berbalik, dan kelas mayoritas tertinggal jauh dari semua model (0,224 dibanding 0,365 hingga 0,409). *Random Forest* bawaan 0,146 di atasnya ($p < 0,001$) tetapi hanya 0,041 di atas tebakan acak sesuai frekuensi kelas (0,330; $p = 0,054$).

Tabel 3. Akurasi dan macro-F1 hasil validasi silang (rerata±SD antarfold)

Model	Akurasi	Macro-F1
Mayoritas	0,507±0,002	0,224±0,001
Tebakan acak	0,391±0,030	0,330±0,032
Decision Tree	0,439±0,030	0,365±0,029
SVM	0,439±0,031	0,385±0,029
RF, bawaan	0,502±0,030	0,370±0,028
RF, disetel (tetap)	0,478±0,030	0,409±0,029
RF, disetel (bersarang)	0,481±0,027	0,406±0,022

Jurnal Manajemen dan Teknologi
https://journal.stmiki.ac.id
E - ISSN: 3047-8251 | P - ISSN: 3047-826X
Vol. 3 No. 2 (2026) | November
OPEN ACCESS
https://doi.org/ 10.63447/jmt.v3i2.2195



Gambar 3. Macro-F1 pada fold-fold validasi silang (garis titik-titik: pembanding kelas mayoritas)

Penyetelan lebih berpengaruh di sini daripada pada *hold-out*: dengan parameter dibuat tetap, *Random Forest* mencapai macro-F1 0,409 (perbandingan lengkap pada Tabel 5). Baris parameter tetap sedikit optimistis, tetapi validasi bersarang memberi nilai hampir sama (0,406±0,022, 0,004 lebih rendah pada lima fold yang sama). Penyetelan melepas sebagian akurasi (0,478 dibanding 0,502; tidak signifikan, p = 0,12). Hiperparameter terpilih berbeda antarkelima fold luar, sedangkan macro-F1 putaran dalam para pemenang hanya 0,37 hingga 0,42, jadi permukaan pencariannya datar.

Tabel 4. F1 per kelas (rerata antarfold)

Model	*Low*	*Medium*	*High*
Tebakan acak	0,145	0,353	0,491
Decision Tree	0,149	0,395	0,551
SVM	0,231	0,350	0,575
RF, bawaan	0,089	0,390	0,633
RF, disetel (tetap)	0,231	0,397	0,599
RF, disetel (bersarang)	0,218	0,400	0,602

Tabel 4 menunjukkan di mana perubahan terjadi. *Random Forest* bawaan hampir mengabaikan kelas *Low* (F1 0,089; *recall* 0,062) walau berbobot seimbang; dugaan yang masuk akal, tetapi tidak diuji, adalah pohon yang tumbuh penuh menghafal sedikit lot *Low* pada tiap sampel *bootstrap*. Penyetelan menaikkan F1 *Low* menjadi 0,231, sama dengan SVM (*recall* 0,245), sementara *recall High* turun dari 0,71 menjadi 0,60 dan *Medium* berubah sedikit. *Low* tetap kelas tersulit pada semua model, dengan F1 terbaik di bawah 0,25. Tabel 5 merangkum posisi *Random Forest* hasil penyetelan terhadap model lain beserta asal selisihnya per kelas.

Tabel 5. Selisih Random Forest hasil penyetelan terhadap model lain pada validasi silang berulang

Pembanding	Macro-F1	Selisih	Nilai-p	Δ *Low*	Δ *Medium*	Δ *High*
RF, bawaan	0,370	+0,039	0,005	+0,142	+0,007	−0,034
Decision Tree	0,365	+0,044	0,028	+0,082	+0,002	+0,048
SVM	0,385	+0,024	0,13	0,000	+0,047	+0,024

Selisih macro-F1 berasal dari kelas yang berbeda untuk tiap pembanding. Keunggulan atas *Random Forest* bawaan seluruhnya datang dari kelas *Low* (+0,142), dengan harga penurunan 0,034 pada *High*; atas *Decision Tree* terutama dari *Low* (+0,082) dan *High* (+0,048); dan atas SVM dari *Medium* (+0,047) dan *High* (+0,024), sedangkan kelas *Low* sama persis (0,231). Jadi tidak ada model yang mengenali *Low* lebih baik daripada SVM dan *Random Forest* hasil penyetelan, dan dengan p = 0,13 data ini tidak mendukung klaim bahwa *Random Forest* lebih baik daripada SVM.

Jurnal Manajemen dan Teknologi
https://journal.stmiki.ac.id
E-ISSN: 3047-8251 | P-ISSN: 3047-826X
Vol. 3 No. 2 (2026) | November
OPEN ACCESS
https://doi.org/ 10.63447/jmt.v3i2.2195

3.1.4 Mengapa Kinerja Tetap Sedang

Bukti yang ada menunjuk ke arah data. Tiga keluarga algoritma berakhir dalam selisih 0,044 pada macro-F1 validasi silang (0,365 hingga 0,409), yang terbaik hanya 0,08 di atas tebakan acak, dan keenam parameter hampir tidak berbeda antarkelas (Tabel 1, $\varepsilon^2 \leq 0{,}011$). Karena pembobotan kelas, penyetelan, dan tiga keluarga algoritma sama-sama berhenti di batas yang sama, batas itu lebih mungkin berada pada informasi dalam keenam parameter daripada pada pilihan metode. Beberapa penjelasan mungkin untuk kinerja sedang, dan tidak satu pun diuji langsung di sini. Tabel 6 merinci tiap penjelasan, indikator yang seharusnya teramati, dan langkah pengujiannya. Derau label menurunkan kinerja generalisasi secara serius (Song et al., 2023), meski kajian itu menyoroti jaringan saraf dalam dan belum tentu berlaku sama untuk *Random Forest*.

Tabel 6. Penjelasan kinerja sedang, indikator yang dapat diamati, dan langkah pengujian atau perbaikannya

Penjelasan	Yang seharusnya teramati	Pengujian atau perbaikan
Derau pada label laboratorium	Kinerja terbatas pada model apa pun; kesalahan menumpuk dekat batas kategori	Uji ulang sampel lot oleh teknisi lain dan ukur kesepakatannya (misalnya kappa)
Bahan baku, jenis koagulan, dan lama penyimpanan tidak tercatat	Kinerja naik bila variabel itu ditambahkan	Catat untuk lot baru, lalu ulangi tiga skema validasi
Kategori berurutan dianggap label terpisah	Kesalahan terutama antarkelas bertetangga	Bandingkan dengan pengklasifikasi ordinal
Kategori ditentukan uji lain di luar keenam parameter	Parameter hampir tidak berbeda antarkelas (Tabel 1)	Telusuri parameter uji penentu kategori di laboratorium

Data yang ada hanya menyentuh penjelasan ketiga: pada *hold-out*, pertukaran terbesar terjadi antara *Medium* dan *High* (masing-masing 35 lot), tetapi keduanya juga kelas terbesar sehingga pola ini belum membedakan penjelasan ordinal dari pengaruh ukuran kelas. Studi pada konteks terkait melaporkan hasil lebih kuat, tetapi targetnya berbeda. *Random Forest* memprediksi *chip* tertolak dengan galat absolut rata-rata 0,64 (Supriyadi et al., 2025), dan model berbasis hutan untuk hasil panen karet melaporkan AUC 0,73 hingga 0,85 (Littidej et al., 2024). Keduanya memprediksi besaran yang berkaitan langsung dengan masukannya, yaitu data mesin terhadap *chip* tertolak dan faktor spasial terhadap hasil panen, sehingga perbandingan dengan penelitian ini hanya indikatif. Alasan menambah variabel bahan baku dan penyimpanan bertumpu pada hal yang sama: Putri et al. (2024) mencatat kualitas bahan baku, metode produksi, mesin, faktor manusia, dan lingkungan sebagai penyebab menyimpangnya viskositas Mooney dan PRI, dan Achmad et al. (2023) menunjukkan jenis koagulan menghasilkan karakteristik karet SIR 20 yang berbeda; dataset ini tidak memuat pemasok, jenis koagulan, maupun lama penyimpanan.

Bagi Perusahaan XYZ, model ini paling jauh berfungsi sebagai alat bantu penyaringan awal. Strategi selalu memprediksi *High* berakurasi 0,507 tetapi tidak pernah menemukan lot *Low* atau *Medium* (macro-F1 0,224), sedangkan model berbobot seimbang menukar sebagian akurasi dengan pengenalan kelas kecil (*recall High* turun dari 0,71 menjadi 0,60, F1 *Low* naik dari 0,089 menjadi 0,231). Pada data uji *hold-out*, model menandai 34 lot sebagai *Low* dan 7 benar (presisi 0,206). Karena *Low* hanya 13,5% data uji (28 dari 207 lot), lot yang ditandai sekitar 1,5 kali lebih mungkin berkategori *Low* daripada lot acak, 21 dari 28 lot *Low* (75%) lolos tanpa tanda, dan sekitar empat dari lima tanda keliru. Sinyal ini terlalu lemah untuk mengurangi pengujian laboratorium, sehingga model tidak boleh dipakai untuk meloloskan atau menolak lot. Pemakaian yang masuk akal adalah menentukan urutan prioritas uji atau uji ulang ketika kapasitas laboratorium terbatas, dengan keputusan tetap pada hasil laboratorium. Penelitian ini terbatas pada satu pabrik, satu periode empat bulan, dan label dari satu laboratorium, dengan hanya 139 lot *Low*. Semua skema evaluasi mengacak lot dari periode yang sama; pengujian pada periode produksi lain belum dilakukan karena data hanya

Jurnal Manajemen dan Teknologi
https://journal.stmiki.ac.id
E-ISSN: 3047-8251 | P-ISSN: 3047-826X
Vol. 3 No. 2 (2026) | November
OPEN ACCESS
https://doi.org/ 10.63447/jmt.v3i2.2195

sampai 2 Juni 2026, dan kinerja pada lot berikutnya bisa lebih rendah bila proses atau bahan baku bergeser. Analisis sensitivitas tanpa tiga nilai yang diduga salah catat belum dilakukan. Validasi bersarang hanya memakai lima fold luar sehingga estimasinya kurang stabil, grid penyetelan tergolong kecil, dan keluarga model lain seperti *gradient boosting* atau pengklasifikasi ordinal tidak dicoba.

3.2 Pembahasan

Hasil penelitian ini menunjukkan bahwa Random Forest hasil penyetelan mencapai macro-F1 0,409 pada validasi silang berulang dan 0,406 pada validasi silang bersarang. Nilai ini lebih tinggi daripada Decision Tree (0,365) dan tidak berbeda nyata dari SVM (0,385). Namun, angka tersebut hanya sedikit di atas tebakan acak berstrata (0,330), dan tidak ada model yang melampaui akurasi strategi yang selalu memprediksi kelas High (0,507). Temuan ini sejalan dengan kecenderungan umum bahwa model berbasis pohon unggul pada data tabular berukuran sedang (Grinsztajn et al., 2022), tetapi keunggulan tersebut belum tampak di sini. Keunggulan Random Forest atas Decision Tree pun bersumber terutama dari kelas Low, sehingga peningkatannya bersifat terbatas. Pola ini menunjukkan bahwa pilihan algoritma bukan faktor penentu utama. Ketiga keluarga algoritma berhenti di batas performa yang hampir sama, yang lebih mungkin berasal dari informasi dalam data itu sendiri. Pemilihan metrik evaluasi turut memengaruhi kesimpulan. Pada data dengan kelas tidak seimbang, akurasi dapat memberi nilai tinggi pada model yang sekadar mengikuti kelas mayoritas (Ghosh et al., 2024). Dalam penelitian ini, strategi selalu memprediksi High memperoleh akurasi 0,507, padahal macro-F1-nya hanya 0,224. Metrik macro-F1 memberi bobot sama pada setiap kelas, tetapi metrik ini juga memiliki kelemahan tersendiri sehingga perlu dilaporkan bersama nilai per kelas (Opitz, 2024). Pembobotan kelas seimbang yang digunakan pada Random Forest sejalan dengan pendekatan yang dikaji Akbar dan Sanjaya (2023) untuk data tidak seimbang. Keputusan untuk tidak memakai SMOTE juga didukung oleh Elreedy et al. (2024), yang menunjukkan bahwa titik sintetis hasil interpolasi belum tentu mengikuti distribusi asli kelas minoritas. Hal ini penting karena kelas Low hanya berjumlah 139 lot. Kesimpulan penelitian ini juga bergantung pada cara validasi. Estimasi dari satu kali pembagian data latih dan uji dengan beberapa ratus kasus cenderung tidak stabil (Bates et al., 2024), sehingga penelitian ini memakai tiga skema validasi dan uji-t terkoreksi untuk resampling berulang. Penyetelan hiperparameter dilakukan di dalam putaran validasi agar skor tidak terlalu optimistis (Bischl et al., 2023). Selisih macro-F1 antara Random Forest hasil penyetelan dan SVM ($p = 0,13$) menunjukkan bahwa data ini belum cukup kuat untuk menyatakan bahwa Random Forest lebih baik dari SVM.

Kesimpulan yang lebih tepat adalah bahwa ketiga algoritma memiliki kinerja yang setara dan sama-sama terbatas. Karena itu, penelitian ini tidak mengklaim keunggulan metode tertentu, melainkan menekankan keterbatasan informasi pada parameter yang tersedia. Keterbatasan informasi tampak dari hasil analisis univariat. Keenam parameter hampir tidak berbeda antarkelas, dengan ukuran efek ε^2 maksimum 0,011. Kondisi ini mengingatkan pada pola yang ditemukan dalam studi sebelumnya. Putri et al. (2024) mencatat bahwa penyimpangan viskositas Mooney dan PRI pada crumb rubber SIR 3CV 60 dapat berasal dari kualitas bahan baku, metode produksi, mesin, faktor manusia, dan lingkungan. Dataset penelitian ini tidak memuat sebagian faktor tersebut. Achmad et al. (2023) juga menunjukkan bahwa jenis koagulan menghasilkan karakteristik karet SIR 20 yang berbeda, sedangkan jenis koagulan tidak tercatat dalam data ini. Penelitian pada konteks terkait juga melaporkan hasil yang lebih kuat, misalnya Supriyadi et al. (2025) yang memprediksi jumlah chip tertolak dengan galat absolut rata-rata 0,64, serta Littidej et al. (2024) yang memperoleh AUC 0,73 hingga 0,85 pada prediksi hasil panen karet. Namun, kedua studi tersebut memprediksi besaran yang berkaitan langsung dengan masukannya, sehingga perbandingannya hanya bersifat indikatif. Ketidakpastian pada label juga perlu diperhitungkan. Kategori Low, Medium, dan High ditetapkan laboratorium melalui pengujian manual, dan kriteria rincinya merupakan penilaian internal yang tidak sepenuhnya tercatat dalam data. Derau pada label dapat menurunkan kemampuan generalisasi model secara serius (Song et al., 2023), meskipun kajian tersebut berfokus pada jaringan saraf dalam. Pola kesalahan pada hold-out menunjukkan bahwa

E-ISSN: 3047-8251 | P-ISSN: 3047-826X
Vol. 3 No. 2 (2026) | November
OPEN ACCESS
https://doi.org/ 10.63447/jmt.v3i2.2195

sebagian besar kesalahan terjadi antara Medium dan High, masing-masing 35 lot. Pola ini dapat mengindikasikan batas kategori yang kabur, tetapi pola tersebut juga bisa terjadi karena kedua kelas tersebut memang yang terbesar. Dengan demikian, penjelasan ini belum dapat dipastikan. Audit konsistensi pelabelan, misalnya dengan uji ulang sampel oleh teknisi berbeda, menjadi langkah awal yang penting. Dari sisi praktis, temuan ini memperkuat posisi model sebagai alat bantu penyaringan awal, bukan alat keputusan. Model berbobot seimbang menukar sebagian akurasi dengan pengenalan kelas kecil. Recall kelas High turun dari 0,71 menjadi 0,60, sementara F1 kelas Low naik dari 0,089 menjadi 0,231. Pada data uji hold-out, model menandai 34 lot sebagai Low, tetapi hanya 7 yang benar, dan 21 dari 28 lot Low lolos tanpa tanda. Dengan tingkat presisi seperti ini, model belum layak mengurangi pengujian laboratorium. Penggunaan yang masuk akal adalah menentukan urutan prioritas uji ulang ketika kapasitas laboratorium terbatas, dengan keputusan akhir tetap berada pada hasil laboratorium. Arah ini juga konsisten dengan upaya pengendalian kualitas berbasis data yang ada, seperti pemantauan dengan peta kendali oleh Putri et al. (2024) dan penerapan machine learning untuk pengendalian kualitas oleh Rosihan et al. (2026), yang menempatkan model sebagai pendukung keputusan, bukan pengganti penilaian ahli.

4. Kesimpulan

Random Forest yang dilatih dengan enam parameter proses pada 1.031 lot Perusahaan XYZ hanya mencapai kinerja sedang dalam mengklasifikasikan kategori *Low*, *Medium*, dan *High* (*macro-F1* validasi silang 0,409; 0,406 pada validasi bersarang). Model ini lebih baik daripada *Decision Tree* dan *Random Forest* bawaan, tidak berbeda nyata dari SVM, tetapi hanya 0,079 di atas tebakan acak berstrata dan tidak melampaui akurasi 0,507 dari strategi selalu memprediksi *High*. Kelas *Low* hanya ditemukan sekitar seperempatnya. Keenam parameter nyaris tidak berbeda antarkelas ($\varepsilon^2 \leq 0{,}011$), sehingga batas kinerja tampaknya ada pada informasi dalam data. Model hanya layak menjadi alat bantu penyaringan awal dan tidak boleh menentukan kelulusan lot. Penelitian lanjutan perlu (i) mengaudit konsistensi pelabelan laboratorium, (ii) menguji pengklasifikasi ordinal, (iii) menambahkan data bahan baku, jenis koagulan, dan lama penyimpanan, serta (iv) memvalidasi model pada lot dari periode produksi lain.

5. Daftar Pustaka

Achmad, F., Deviany, D., Nuranisa, A., Antika, R., Suhartono, S., & Suharto, S. (2023). Fisibilitas pemanfaatan koagulan alami terhadap karakteristik karet pada produksi SIR 20. *Jurnal Penelitian Karet*, *41*(2), 153–168. https://doi.org/10.22302/ppk.jpk.v41i2.876.

Akbar, H., & Sanjaya, W. K. (2023). Kajian performa metode class weight random forest pada klasifikasi imbalance data kelas curah hujan. *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, *3*(1). https://doi.org/10.20885/snati.v3i1.30.

Alnuaimi, A. F. A. H., & Albaldawi, T. H. K. (2024). An overview of machine learning classification techniques. *BIO Web of Conferences*, *97*. https://doi.org/10.1051/bioconf/20249700133.

Areewattana, C., Pramotkul, N., Chanjatunat, T., & Deepaisarn, S. (2025). Cup-lump rubber quality estimation from historical weather data using machine learning regressors. *Science & Technology Asia*, *30*(3), 51–59. https://doi.org/10.14456/scitechasia.2025.46.

Bates, S., Hastie, T., & Tibshirani, R. (2024). Cross-validation: What does it estimate and how well does it do it? *Journal of the American Statistical Association*, *119*(546), 1434–1445. https://doi.org/10.1080/01621459.2023.2197686.

E-ISSN: 3047-8251 | P-ISSN: 3047-826X
Vol. 3 No. 2 (2026) | November
OPEN ACCESS
https://doi.org/ 10.63447/jmt.v3i2.2195

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A.-L., Deng, D., & Lindauer, M. (2023). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *WIREs Data Mining and Knowledge Discovery*, *13*(2), Article e1484. https://doi.org/10.1002/widm.1484.

Elreedy, D., Atiya, A. F., & Kamalov, F. (2024). A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. *Machine Learning*, *113*(7), 4903–4923. https://doi.org/10.1007/s10994-022-06296-4.

Ghosh, K., Bellinger, C., Corizzo, R., Branco, P., Krawczyk, B., & Japkowicz, N. (2024). The class imbalance problem in deep learning. *Machine Learning*, *113*(7), 4845–4901. https://doi.org/10.1007/s10994-022-06268-8.

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). *Why do tree-based models still outperform deep learning on tabular data?* arXiv. https://doi.org/10.48550/arXiv.2207.08815.

Littidej, P., Kromkratoke, W., Pumhirunroj, B., Buasri, N., Prasertsri, N., Sangpradid, S., & Slack, D. (2024). Enhanced rubber yield prediction in high-density plantation areas using a GIS and machine learning-based forest classification and regression model. *Forests*, *15*(9), Article 1535. https://doi.org/10.3390/f15091535.

Opitz, J. (2024). A closer look at classification evaluation metrics and a critical reflection of common evaluation practice. *Transactions of the Association for Computational Linguistics*, *12*, 820–836. https://doi.org/10.1162/tacl_a_00675.

Prastia, B., Bursamin, B., Fernandes, R., & Hidayad, S. (2025). Analysis of dryer control system for non-conformity of the crumb rubber products in PT. Bukit Angkasa Makmur Bengkulu. *Jurnal Sains Agro*, *10*(2).

Putri, G. R., Nadiyah, K., Rishelin, N., Futeri, R., & Susanto, R. A. (2024). Pengendalian kualitas Mooney viscosity (MV) dan plasticity retention index (PRI) dalam crumb rubber SIR 3CV 60 dengan metode statistical quality control (SQC) di perusahaan karet. *Jurnal PASTI*, *18*(3), 315–328. https://doi.org/10.22441/pasti.2024.v18i3.003.

Puttipipatkajorn, A., & Puttipipatkajorn, A. (2021). Spectroscopic measurement approaches in evaluation of dry rubber content of cup lump rubber using machine learning techniques. *International Journal of Agricultural and Biological Engineering*, *14*(3), 207–213. https://doi.org/10.25165/j.ijabe.20211403.6298.

Rosihan, R. I., Kumalasari, R., Widyantoro, M., Saputra, Y., & Haekal, J. (2026). Prediksi kualitas kain pada proses inspeksi tekstil menggunakan machine learning. *Jurnal REKAVASI*, *14*(1), 13–19.

Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, *2024*, 69–79. https://doi.org/10.58496/BJML/2024/007.

Song, H., Kim, M., Park, D., Shin, Y., & Lee, J.-G. (2023). Learning from noisy labels with deep neural networks: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, *34*(11), 8135–8153. https://doi.org/10.1109/TNNLS.2022.3152527.

Supriyadi, A., Sunge, A. S., & Tedi, N. (2025). Analysis of predicting the number of rejected chips using random forest at PT. Wahyu Kartumasindo Internasional. *Journal of Computer Networks,*

E-ISSN: 3047-8251 | P-ISSN: 3047-826X
Vol. 3 No. 2 (2026) | November
OPEN ACCESS
https://doi.org/ 10.63447/jmt.v3i2.2195

Architecture and High Performance Computing, *7*(4), 1128–1139. https://doi.org/10.47709/cnahpc.v7i4.7028.

Zaza, S., Atemkeng, M., & Hamlomo, S. (2023). *Wine feature importance and quality prediction: A comparative study of machine learning algorithms with unbalanced data.* arXiv. https://doi.org/10.48550/arXiv.2310.01584.