KUADRAN
Jurnal Analitika Data dan Ilmu Komputer

Published by PT Yetti Afrida Center Bengkulu
e-ISSN: 3165-3677 | p-ISSN: 3165-3723
Vol. 1 No. 1 Oktober 2026 | Pages 43-56
https://yettiafridacenter.com/kuadran
This Article is licensed under a CC BY-SA 4.0 International License

Prediksi Aktivitas Antioksidan Senyawa Bioaktif Menggunakan Pendekatan Machine Learning Berbasis Deskriptor Molekul

Qomariah Hasanah*
Universitas Islam Negeri Fatmawati Sukarno Bengkulu, Indonesia
Email: qomariah.hasanah@mail.uinfasbengkulu.ac.id

ABSTRACT

Purpose: *This study builds machine learning models that predict the DPPH radical-scavenging activity of phenolic acids from simple, mechanistically designed molecular descriptors, and examines which structural features drive the activity.*

Design/Methodology: *Activity data (pTEAC) for 21 phenolic acids measured under a single protocol were taken from Jing et al. (2012), keeping the original split of 15 training and 6 test compounds. Fifteen two-dimensional descriptors were computed, including the number and position of hydroxyl and methoxy groups, electron-donating groups ortho/para to the phenolic OH, conjugation, TPSA, and Hammett constants. Multiple linear regression, partial least squares, k-nearest neighbours, support vector regression, and random forest were compared using leave-one-out cross-validation, an external test set, 100-fold Y-randomization, permutation importance, and a leverage-based applicability domain.*

Findings: *Random forest performed best ($Q^2LOO = 0.892$; external $R^2pred = 0.767$) and passed Y-randomization, while k-NN and SVR overfitted. The number of electron-donating groups ortho/para to the phenolic OH and TPSA dominated the model. A two-descriptor linear equation was interpretable but statistically weak ($Q^2LOO = 0.409$). The 2D models remained below published 3D CoMFA/CoMSIA models ($R^2pred \approx 0.98$); most of the test error came from one compound flagged as outside the applicability domain.*

Practical Implications: *The workflow can be used to rank phenolic compounds from natural extracts before laboratory testing, provided that predictions are restricted to the applicability domain.*

Originality/Value: *The study combines transparent mechanism-based descriptors with a complete OECD-oriented validation on real experimental data and reports honestly where the 2D approach falls short of 3D models.*

Keywords: *antioxidant activity; phenolic acids; QSAR; machine learning; random forest; molecular descriptors.*

Received: 17/09/2026 Accepted: 24/09/2026 Published: 12/10/2026

A. INTRODUCTION (PENDAHULUAN)

Latar Belakang

Stres oksidatif terjadi ketika produksi spesies oksigen reaktif melampaui kemampuan sistem pertahanan antioksidan tubuh. Kondisi ini berkaitan dengan kerusakan lipid, protein, dan DNA, serta dihubungkan dengan berbagai penyakit degeneratif (Halliwell & Gutteridge, 2015). Senyawa fenolik alami, termasuk asam fenolat dan flavonoid, banyak diteliti karena mampu menetralkan radikal bebas melalui donasi atom hidrogen atau elektron. Asam fenolat seperti asam kafeat, asam ferulat, dan asam galat terdapat luas dalam sayuran, buah, biji-bijian, kopi, dan teh, sehingga relevan bagi pengembangan pangan fungsional dan bahan alam berbasis kekayaan hayati Indonesia.

Penentuan aktivitas antioksidan secara eksperimen, misalnya dengan uji DPPH, ABTS, atau FRAP, membutuhkan bahan kimia, instrumen, dan waktu. Ketika jumlah kandidat senyawa banyak, pengujian satu per satu menjadi mahal. Karena itu, berkembang pendekatan hubungan kuantitatif struktur–aktivitas (Quantitative Structure–Activity Relationship, QSAR) yang memprediksi aktivitas dari struktur molekul (Cherkasov et al., 2014). Pendekatan ini dapat membantu menyaring kandidat sebelum pengujian laboratorium dan sekaligus menjelaskan gugus fungsi yang paling berperan.

Pembelajaran mesin (machine learning) memperluas QSAR klasik dengan algoritma yang mampu menangkap hubungan nonlinear, seperti Random Forest dan Support Vector Regression

This Article is licensed under a CC BY-SA 4.0 International License

(Breiman, 2001; Smola & Schölkopf, 2004). Namun, model yang kompleks rentan mengalami overfitting, terutama ketika jumlah data sedikit. Data antioksidan yang diukur dengan protokol seragam pada umumnya hanya mencakup puluhan senyawa. Pada kondisi seperti ini, validasi yang ketat, uji pengacakan respons, dan penentuan domain aplikabilitas menjadi syarat agar model dapat dipercaya (Tropsha, 2010; OECD, 2007).

Rumusan Masalah dan Tujuan

Penelitian sebelumnya telah membangun model CoMFA dan CoMSIA untuk aktivitas penangkapan radikal DPPH asam fenolat pangan (Jing et al., 2012). Model tiga dimensi tersebut memerlukan penyejajaran struktur dan perhitungan medan molekul yang tidak sederhana. Pertanyaan yang diajukan dalam penelitian ini adalah apakah deskriptor molekul dua dimensi yang mudah dihitung dan mudah ditafsirkan, dikombinasikan dengan algoritma pembelajaran mesin, dapat memberikan prediksi yang sebanding sambil tetap menjelaskan mekanisme kimianya.

Tujuan penelitian ini adalah (1) menghitung deskriptor molekul yang relevan secara mekanistik untuk asam fenolat; (2) membangun dan membandingkan lima model regresi, yaitu regresi linear berganda, Partial Least Squares, k-Nearest Neighbors, Support Vector Regression, dan Random Forest; (3) memvalidasi model dengan validasi silang leave-one-out, set uji eksternal, dan uji pengacakan respons; (4) mengidentifikasi deskriptor yang paling menentukan aktivitas; dan (5) menerapkan model untuk memprediksi aktivitas beberapa asam fenolat yang tidak termasuk dalam data.

B. LITERATURE REVIEW (KAJIAN PUSTAKA)

Mekanisme Antioksidan Senyawa Fenolik

Senyawa fenolik menetralkan radikal bebas melalui beberapa mekanisme. Pada mekanisme transfer atom hidrogen (HAT), gugus hidroksil fenolik mendonorkan atom hidrogen kepada radikal, sehingga kemudahan reaksi ditentukan oleh entalpi disosiasi ikatan O–H. Mekanisme lain adalah transfer elektron tunggal yang diikuti transfer proton (SET-PT) dan kehilangan proton berurutan yang diikuti transfer elektron (SPLET), yang lebih dominan pada pelarut polar (Leopoldini et al., 2011). Pada semua mekanisme tersebut, kestabilan radikal fenoksil yang terbentuk menjadi faktor kunci.

Kajian struktur–aktivitas klasik menunjukkan bahwa gugus orto-dihidroksi (katekol) dan tri-hidroksi (pirogalol) meningkatkan aktivitas secara tajam. Radikal fenoksil pada katekol distabilkan oleh ikatan hidrogen intramolekul dan dapat teroksidasi lebih lanjut menjadi kuinon (Rice-Evans et al., 1996). Gugus metoksi pada posisi orto terhadap hidroksil juga menstabilkan radikal melalui efek donor elektron, walaupun pengaruhnya lebih lemah daripada gugus hidroksil kedua (Cuvelier et al., 1992). Pada asam fenolat, rantai samping propenoat seperti pada asam sinamat memperluas konjugasi sehingga radikal lebih terdelokalisasi dibandingkan pada turunan asam benzoat.

Uji DPPH dan Satuan TEAC

Uji DPPH mengukur penurunan absorbansi radikal stabil 2,2-difenil-1-pikrilhidrazil ketika bereaksi dengan antioksidan (Brand-Williams et al., 1995). Hasil uji dapat dinyatakan sebagai IC50 atau sebagai Trolox Equivalent Antioxidant Capacity (TEAC), yaitu konsentrasi Trolox yang memberikan efek setara dengan konsentrasi tertentu senyawa uji. Untuk keperluan pemodelan, nilai TEAC sering diubah ke skala logaritma negatif, pTEAC = −log TEAC,

This Article is licensed under a CC BY-SA 4.0 International License

sehingga sebarannya lebih simetris. Pada skala ini, nilai pTEAC yang lebih kecil menunjukkan aktivitas yang lebih kuat.

Deskriptor Molekul dan QSAR

Deskriptor molekul adalah representasi numerik dari sifat struktur, misalnya jumlah atom atau gugus fungsi, sifat topologi, sifat elektronik, dan sifat fisikokimia (Todeschini & Consonni, 2009). Deskriptor dapat dihitung dengan perangkat lunak seperti RDKit atau PaDEL, tetapi deskriptor yang dirancang khusus berdasarkan mekanisme sering lebih mudah ditafsirkan. Parameter Hammett σ, misalnya, merangkum efek elektronik substituen pada cincin aromatik dan telah lama dipakai dalam kimia fisik organik (Hansch et al., 1991). Luas permukaan polar topologis (TPSA) dapat dihitung dari kontribusi fragmen tanpa geometri tiga dimensi (Ertl et al., 2000).

Pedoman OECD untuk validasi model QSAR menyebutkan lima prinsip: titik akhir yang terdefinisi, algoritma yang jelas, domain aplikabilitas yang terdefinisi, ukuran kecocokan dan daya prediksi yang memadai, serta interpretasi mekanistik bila memungkinkan (OECD, 2007). Prinsip ini menegaskan bahwa akurasi saja tidak cukup. Model juga harus dapat dijelaskan dan digunakan secara bertanggung jawab hanya pada senyawa yang mirip dengan data latih.

Machine Learning untuk Prediksi Aktivitas

Regresi linear berganda (MLR) menghasilkan persamaan yang mudah ditafsirkan, tetapi hanya mampu menangkap hubungan linear. Partial Least Squares (PLS) cocok bila deskriptor saling berkorelasi tinggi, karena data diproyeksikan ke beberapa komponen laten (Wold et al., 2001). k-Nearest Neighbors memprediksi aktivitas berdasarkan senyawa yang paling mirip, sejalan dengan prinsip kemiripan struktur. Support Vector Regression memetakan data ke ruang berdimensi tinggi melalui fungsi kernel (Smola & Schölkopf, 2004). Random Forest menggabungkan banyak pohon keputusan dan menyediakan ukuran kepentingan variabel (Breiman, 2001). Pada data kecil, Random Forest relatif tahan terhadap overfitting, tetapi kemampuannya mengekstrapolasi di luar rentang data latih terbatas.

Validasi pada data kecil menuntut kehati-hatian. Validasi silang leave-one-out (LOO) menggunakan hampir seluruh data untuk pelatihan pada setiap iterasi, tetapi dapat memberikan estimasi yang terlalu optimistis. Karena itu, validasi silang perlu dilengkapi set uji eksternal dan uji pengacakan respons (Y-randomization). Pada uji pengacakan, nilai aktivitas diacak lalu model dibangun ulang. Model yang baik harus jauh lebih unggul daripada model yang dilatih pada data acak (Tropsha, 2010). Perangkat lunak terbuka scikit-learn menyediakan implementasi seluruh algoritma dan prosedur validasi tersebut (Pedregosa et al., 2011).

C. METHOD (METODE)

Desain Penelitian dan Sumber Data

Penelitian ini merupakan studi komputasi kuantitatif dengan pendekatan QSAR berbasis pembelajaran mesin. Data aktivitas diambil dari Jing et al. (2012), yang mengukur aktivitas penangkapan radikal DPPH sejumlah asam fenolat pangan dalam satu protokol yang seragam. Reaksi menggunakan 0,2 mmol/L DPPH, absorbansi diukur pada 515 nm selama 40 menit, dan hasilnya dinyatakan sebagai pTEAC. Penggunaan data dari satu laboratorium dengan protokol yang sama mengurangi variasi antarpengukuran yang sering menjadi masalah dalam penggabungan data dari banyak sumber.

Dari 24 senyawa pada data asli, tiga senyawa dikeluarkan. Dua senyawa (nomor 19 dan 21) memiliki gugus rantai samping yang hanya disajikan sebagai gambar sehingga strukturnya tidak

This Article is licensed under a CC BY-SA 4.0 International License

dapat dipastikan, dan satu senyawa (nomor 5) memiliki notasi substituen yang ambigu. Dengan demikian, data akhir terdiri atas 21 senyawa. Pembagian data latih dan uji mengikuti pembagian yang ditetapkan oleh penulis asli, sehingga hasilnya dapat dibandingkan langsung dengan model CoMFA dan CoMSIA mereka.

Perhitungan Deskriptor Molekul

Deskriptor dihitung dengan program Python yang disusun peneliti berdasarkan pola substitusi cincin aromatik (posisi R2–R6) dan jenis rantai samping (–COOH, –CH_2COOH, atau –CH=CHCOOH). Pendekatan ini dipilih agar setiap deskriptor memiliki makna kimia yang jelas dan perhitungannya dapat diperiksa ulang. Sebanyak 15 deskriptor dihitung, terdiri atas deskriptor konstitusional, deskriptor gugus fungsi berbasis mekanisme, dan deskriptor fisikokimia, seperti dirangkum pada Tabel 1.

Tabel 1. Deskriptor molekul yang digunakan

Kelompok	Deskriptor	Definisi
Konstitusional	nOH, nOMe, MW	Jumlah OH fenolik, jumlah OCH_3, massa molekul (g/mol)
Ikatan hidrogen	HBD, HBA	Jumlah donor dan akseptor ikatan hidrogen
Pola dihidroksi	OH_orto_OH, OH_para_OH, OH_meta_OH	Jumlah pasangan OH pada posisi orto, para, dan meta
Stabilisasi radikal	OH_orto_OMe	Jumlah OH yang bersebelahan dengan OCH_3
Stabilisasi radikal	EDG_op_OH	Jumlah gugus donor elektron (OH/OCH_3) pada posisi orto atau para terhadap OH fenolik
Posisi	OH_para_rantai	1 bila OH berada para terhadap rantai samping
Rantai samping	Konjugasi, Spacer_CH2	1 bila rantai –CH=CHCOOH atau –CH_2COOH
Fisikokimia	TPSA	Luas permukaan polar topologis (Å²) dari kontribusi fragmen
Elektronik	Sigma	Jumlah konstanta Hammett σ substituen meta/para terhadap rantai samping

Catatan: kontribusi TPSA mengikuti Ertl et al. (2000); konstanta Hammett mengikuti Hansch et al. (1991).

Sebagai contoh, asam ferulat memiliki OH pada posisi 4, OCH_3 pada posisi 3, dan rantai samping –CH=CHCOOH. Program menghasilkan nOH = 1, nOMe = 1, OH_orto_OMe = 1, EDG_op_OH = 1, OH_para_rantai = 1, Konjugasi = 1, HBD = 2, HBA = 4, MW = 194,18 g/mol ($C_{10}H_{10}O_4$), dan TPSA = 37,30 + 20,23 + 9,23 = 66,76 Å². Pada asam galat (OH pada posisi 3, 4, dan 5), terdapat dua pasangan OH orto dan satu pasangan OH meta. Gugus OH pada posisi 4 memiliki dua tetangga orto, sedangkan OH pada posisi 3 dan 5 masing-masing memiliki satu, sehingga EDG_op_OH = 4. Inti perhitungan deskriptor posisi ditunjukkan pada potongan kode berikut. Jarak dua posisi pada cincin dihitung sebagai jarak terpendek melingkar, sehingga nilai 1 berarti orto, 2 berarti meta, dan 3 berarti para.

```
ring = lambda a, b: min(abs(a - b), 6 - abs(a - b))
orto_oh = sum(1 for a, b in combinations(oh, 2) if ring(a, b) == 1)
meta_oh = sum(1 for a, b in combinations(oh, 2) if ring(a, b) == 2)
para_oh = sum(1 for a, b in combinations(oh, 2) if ring(a, b) == 3)
# donor elektron (OH/OCH3) orto/para terhadap OH fenolik
edg_op = sum(1 for a in oh for b in oh + ome
             if a != b and ring(a, b) in (1, 3))
tpsa = 37.30 + sum(TPSA[g] for g in substituen if g in TPSA)
```

Kebenaran rumus molekul dan massa molekul hasil program diperiksa dengan membandingkannya terhadap rumus yang tercantum pada data asli untuk seluruh 21 senyawa. Karena perhitungan tidak menggunakan pustaka kemoinformatika seperti RDKit, deskriptor yang dihasilkan terbatas pada sifat yang dapat diturunkan langsung dari pola substitusi. Konsekuensi keterbatasan ini dibahas pada bagian hasil.

This Article is licensed under a CC BY-SA 4.0 International License

Pembangunan dan Validasi Model

Data latih terdiri atas 15 senyawa dan data uji eksternal terdiri atas 6 senyawa. Lima algoritma dibangun dengan pustaka scikit-learn (Pedregosa et al., 2011). Regresi linear berganda (MLR) hanya menggunakan dua deskriptor, EDG_op_OH dan nOH, untuk menjaga rasio jumlah senyawa terhadap jumlah variabel. Empat model lain menggunakan seluruh 15 deskriptor yang telah distandarkan, yaitu PLS dengan dua komponen laten, k-NN dengan k = 3 dan pembobotan jarak, SVR dengan kernel RBF (C = 10, ε = 0,1), dan Random Forest dengan 500 pohon dan minimal dua sampel per daun.

Kinerja model dinilai dengan koefisien determinasi pada data latih (R^2), koefisien validasi silang leave-one-out (Q^2_{LOO}), galat akar kuadrat rata-rata (RMSE), dan koefisien prediksi eksternal pada set uji (R^2_{pred}). Rumus yang digunakan adalah:

$$Q^2 = 1 - \Sigma(y_i - \hat{y}_i)^2 / \Sigma(y_i - \bar{y}_{latih})^2 \quad (1)$$

dengan y_i nilai eksperimen, $\hat{y}_i$ nilai prediksi LOO atau prediksi set uji, dan $\bar{y}_{latih}$ rata-rata pTEAC data latih (Golbraikh & Tropsha, 2002). Selain regresi, ketepatan klasifikasi dihitung dengan membagi senyawa menjadi aktif (pTEAC < 1,5) dan kurang aktif (pTEAC ≥ 1,5). Kekokohan model diuji dengan 100 kali pengacakan respons (Rücker et al., 2007). Domain aplikabilitas model MLR ditentukan dengan nilai leverage, yaitu $h_i = x_i^T(X^TX)^{-1}x_i$, dengan batas peringatan $h^* = 3(p + 1)/n = 0{,}60$. Kontribusi deskriptor pada Random Forest diukur dengan permutation importance, yaitu penurunan R^2 ketika nilai satu deskriptor diacak.

Tabel 2. Pengaturan algoritma dan praproses

Model	Deskriptor masukan	Praproses	Parameter utama
MLR	EDG_op_OH, nOH	Standardisasi	Kuadrat terkecil biasa
PLS	15 deskriptor	Standardisasi	2 komponen laten
k-NN	15 deskriptor	Standardisasi	k = 3, bobot jarak
SVR	15 deskriptor	Standardisasi	Kernel RBF, C = 10, ε = 0,1, γ = scale
Random Forest	15 deskriptor	Tanpa standardisasi	500 pohon, min. 2 sampel per daun, random_state = 42

Pengaturan setiap algoritma dirangkum pada Tabel 2. Parameter model ditetapkan sebelum analisis dan tidak dioptimasi menggunakan data uji, karena data latih yang hanya 15 senyawa terlalu kecil untuk penyetelan parameter bersarang tanpa risiko kebocoran informasi. Standardisasi dilakukan di dalam pipeline sehingga rata-rata dan simpangan baku dihitung ulang pada setiap lipatan validasi silang. Kepentingan deskriptor pada Random Forest dihitung dengan permutation importance, yaitu penurunan R^2 ketika nilai satu deskriptor diacak, dengan 50 pengulangan (Breiman, 2001).

Penerapan pada Senyawa Baru

Model terbaik diterapkan pada enam asam fenolat yang tidak terdapat dalam data, yaitu asam 2,3-dihidroksibenzoat, asam 3,5-dihidroksibenzoat, asam isoferulat, asam 3,4-dimetoksisinamat, asam 3,4-dihidroksifenilasetat (DOPAC), dan asam 3-hidroksibenzoat. Senyawa-senyawa ini dipilih karena memiliki pola substitusi yang mirip dengan data latih, tetapi juga mencakup satu pola yang belum terwakili (meta-dihidroksi tanpa gugus donor orto/para) untuk menguji fungsi domain aplikabilitas. Hasil prediksi bersifat hipotesis dan tetap memerlukan konfirmasi eksperimen.

This Article is licensed under a CC BY-SA 4.0 International License

D. RESULT AND DISCUSSION (HASIL DAN PEMBAHASAN)

Gambaran Data dan Hubungan Struktur–Aktivitas

Tabel 3 menyajikan 21 senyawa beserta deskriptor utama dan nilai pTEAC eksperimen. Nilai pTEAC berkisar antara 0,168 (asam gentisat) hingga 3,700 (asam veratrat), dan sebarannya cenderung terbagi dua: sembilan senyawa aktif dengan pTEAC < 1,5 dan dua belas senyawa yang hampir tidak aktif dengan pTEAC sekitar 3,3–3,7. Pola dua kelompok ini sudah menunjukkan bahwa aktivitas tidak berubah secara bertahap, melainkan sangat ditentukan oleh ada atau tidaknya motif struktur tertentu.

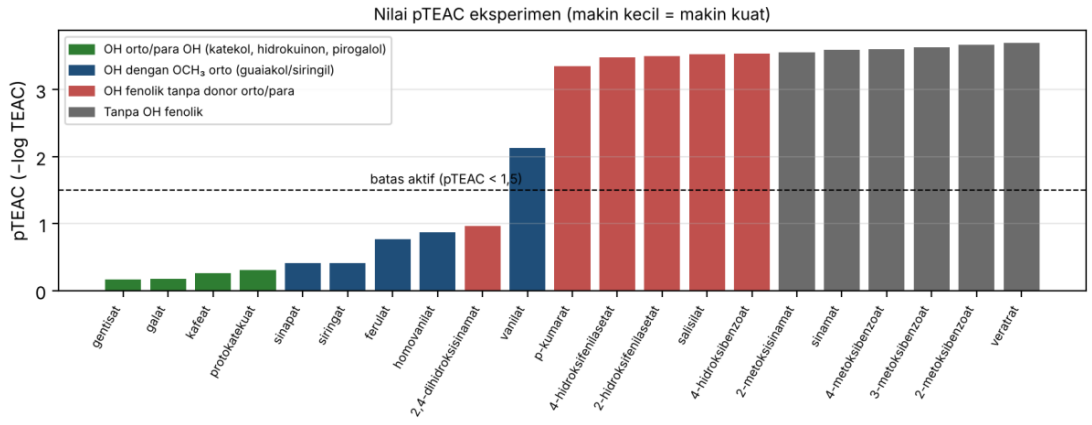
Tabel 3. Data senyawa, deskriptor utama, dan pTEAC eksperimen

No	Senyawa	Rumus	MW	nOH	nOMe	EDG_op_OH	TPSA	Konjugasi	pTEAC	Set
1	Asam vanilat	C8H8O4	168,15	1	1	1	66,76	Tidak	2,127	Uji
2	Asam protokatekuat	C7H6O4	154,12	2	0	2	77,76	Tidak	0,309	Latih
3	Asam 4-hidroksibenzoat	C7H6O3	138,12	1	0	0	57,53	Tidak	3,544	Latih
4	Asam veratrat	C9H10O4	182,17	0	2	0	55,76	Tidak	3,700	Latih
6	Asam 2-metoksibenzoat	C8H8O3	152,15	0	1	0	46,53	Tidak	3,667	Latih
7	Asam siringat	C9H10O5	198,17	1	2	2	75,99	Tidak	0,409	Latih
8	Asam 4-metoksibenzoat	C8H8O3	152,15	0	1	0	46,53	Tidak	3,608	Latih
9	Asam gentisat	C7H6O4	154,12	2	0	2	77,76	Tidak	0,168	Uji
10	Asam salisilat	C7H6O3	138,12	1	0	0	57,53	Tidak	3,535	Latih
11	Asam galat	C7H6O5	170,12	3	0	4	97,99	Tidak	0,179	Latih
12	Asam 3-metoksibenzoat	C8H8O3	152,15	0	1	0	46,53	Tidak	3,633	Latih
13	Asam 4-hidroksifenilasetat	C8H8O3	152,15	1	0	0	57,53	Tidak	3,482	Uji
14	Asam 2-hidroksifenilasetat	C8H8O3	152,15	1	0	0	57,53	Tidak	3,501	Latih
15	Asam homovanilat	C9H10O4	182,17	1	1	1	66,76	Tidak	0,876	Latih
16	Asam 2,4-dihidroksisinamat	C9H8O4	180,16	2	0	0	77,76	Ya	0,964	Uji
17	Asam p-kumarat	C9H8O3	164,16	1	0	0	57,53	Ya	3,349	Latih
18	Asam kafeat	C9H8O4	180,16	2	0	2	77,76	Ya	0,266	Uji
20	Asam sinamat	C9H8O2	148,16	0	0	0	37,30	Ya	3,594	Latih
22	Asam 2-metoksisinamat	C10H10O3	178,19	0	1	0	46,53	Ya	3,561	Uji
23	Asam ferulat	C10H10O4	194,19	1	1	1	66,76	Ya	0,773	Latih
24	Asam sinapat	C11H12O5	224,21	1	2	2	75,99	Ya	0,409	Latih

Sumber pTEAC: Jing et al. (2012); deskriptor: perhitungan penulis.

Pengelompokan menurut motif struktur pada Gambar 1 memperjelas pola tersebut. Empat senyawa dengan dua gugus OH yang saling orto atau para, yaitu asam gentisat, galat, kafeat, dan protokatekuat, merupakan kelompok paling aktif dengan pTEAC rata-rata 0,230 (rentang 0,168–0,309). Lima senyawa dengan OH yang bersebelahan dengan gugus metoksi (tipe guaiakol dan siringil) memiliki rata-rata 0,919 (0,409–2,127). Senyawa dengan satu OH tanpa gugus donor orto/para rata-rata bernilai 3,062, sedangkan senyawa tanpa OH fenolik rata-rata bernilai 3,627. Urutan ini sejalan dengan kajian struktur–aktivitas klasik yang menekankan peran katekol, pirogalol, dan metoksi orto (Rice-Evans et al., 1996; Cuvelier et al., 1992).

This Article is licensed under a CC BY-SA 4.0 International License



Gambar 1. Nilai pTEAC eksperimen 21 asam fenolat, diurutkan dari paling aktif, dengan warna menurut motif struktur

Ada dua pengecualian yang menarik. Pertama, asam vanilat (pTEAC 2,127) jauh kurang aktif dibandingkan senyawa bertipe guaiakol lain, seperti asam homovanilat (0,876) dan asam ferulat (0,773), padahal ketiganya memiliki pola OH dan OCH_3 yang sama. Perbedaannya terletak pada rantai samping. Pada asam vanilat, gugus karboksil yang menarik elektron terikat langsung pada cincin pada posisi para terhadap OH, sehingga melemahkan kemampuan donasi hidrogen. Pada asam homovanilat, efek ini diputus oleh gugus CH_2, sedangkan pada asam ferulat rantai propenoat justru memperluas delokalisasi radikal. Kedua, asam 2,4-dihidroksisinamat cukup aktif (0,964) meskipun kedua gugus OH-nya saling meta. Kedua OH tersebut berada pada posisi orto dan para terhadap rantai samping terkonjugasi, sehingga efek resonansinya kemungkinan berbeda dari pola meta-dihidroksi pada turunan asam benzoat.

Koefisien korelasi Pearson antara deskriptor dan pTEAC menunjukkan bahwa deskriptor terkait gugus hidroksil paling berpengaruh. TPSA ($r = -0{,}887$), EDG_op_OH ($r = -0{,}845$), HBA ($r = -0{,}832$), dan nOH ($r = -0{,}762$) berkorelasi negatif kuat, artinya makin besar nilainya, makin kuat aktivitas antioksidan. Sebaliknya, konstanta Hammett total terhadap rantai samping ($r = 0{,}139$), jumlah metoksi ($r = -0{,}043$), dan keberadaan konjugasi ($r = -0{,}154$) berkorelasi lemah bila dilihat sendiri-sendiri. Hasil ini menegaskan bahwa yang menentukan bukan jumlah substituen, melainkan posisi gugus donor elektron relatif terhadap OH fenolik.

Kinerja Model Prediksi

Tabel 4 merangkum kinerja kelima model. Random Forest memberikan kinerja validasi silang terbaik dengan $Q^2_{LOO} = 0{,}892$ dan RMSE LOO = 0,499, serta mengklasifikasikan seluruh senyawa latih dengan benar dalam validasi LOO. Pada set uji eksternal, Random Forest juga terbaik di antara model berbasis deskriptor 2D dengan $R^2_{pred} = 0{,}767$ dan RMSE = 0,730. Model k-NN dan SVR memiliki R^2 latih hampir sempurna (1,000 dan 0,996), tetapi nilai Q^2 dan R^2_{pred}-nya jauh lebih rendah. Selisih besar ini merupakan tanda overfitting yang lazim terjadi pada data kecil.

Tabel 4. Kinerja model pada data latih (n = 15) dan set uji eksternal (n = 6)

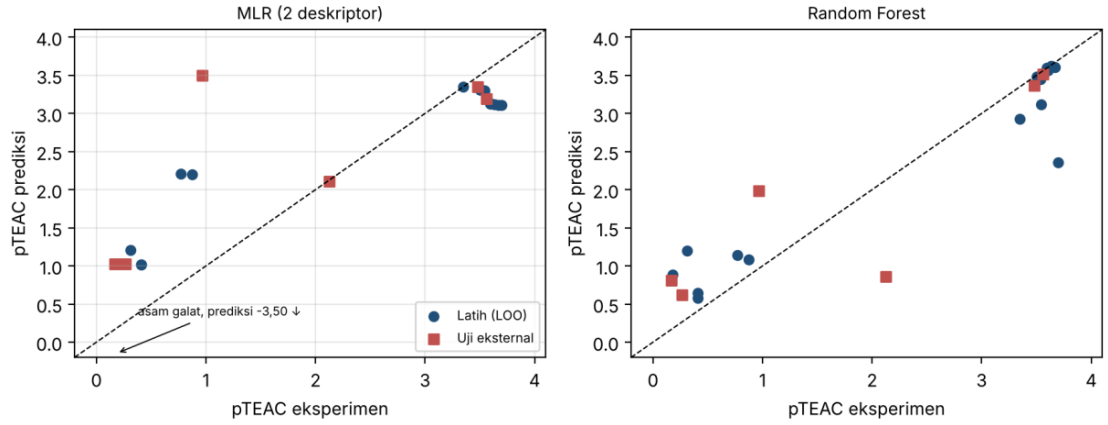
Model	R^2 latih	Q^2_{LOO}	RMSE LOO	R^2_{pred} uji	RMSE uji	Akurasi kelas LOO (%)	Akurasi kelas uji (%)
MLR (2 deskriptor)	0,782	0,409	1,167	0,425	1,148	86,7	83,3
PLS (2 komponen)	0,884	0,572	0,993	0,590	0,969	86,7	66,7
k-NN (k = 3)	1,000	0,768	0,731	0,430	1,142	86,7	66,7
SVR (kernel RBF)	0,996	0,614	0,943	0,678	0,859	73,3	50,0
Random Forest	0,985	0,892	0,499	0,767	0,730	100,0	66,7

This Article is licensed under a CC BY-SA 4.0 International License

Model	R^2 latih	Q^2_{LOO}	RMSE LOO	R^2_{pred} uji	RMSE uji	Akurasi kelas LOO (%)	Akurasi kelas uji (%)
CoMFA (Jing et al., 2012)	–	–	–	0,977	0,227	–	–
CoMSIA (Jing et al., 2012)	–	–	–	0,981	0,208	–	–

Catatan: nilai CoMFA dan CoMSIA dihitung dari prediksi yang dilaporkan Jing et al. (2012) pada enam senyawa uji yang sama.

Model MLR dua deskriptor menghasilkan persamaan pTEAC = 3,196 − 1,236·EDG_op_OH + 0,152·nOH. Koefisien EDG_op_OH yang negatif dan besar menunjukkan bahwa setiap gugus donor elektron pada posisi orto atau para terhadap OH fenolik menurunkan pTEAC sekitar 1,24 satuan, atau meningkatkan TEAC sekitar 17 kali. Setelah EDG_op_OH dimasukkan, koefisien nOH menjadi kecil dan positif. Artinya, jumlah OH saja tidak menambah aktivitas bila tidak disertai pola posisi yang tepat. Interpretasi ini sejalan dengan mekanisme stabilisasi radikal fenoksil oleh gugus donor elektron (Leopoldini et al., 2011). Namun, Q^2_{LOO} model MLR hanya 0,409 karena hubungan linear tidak mampu menangkap efek jenuh. Contohnya, asam galat dengan empat interaksi donor diprediksi jauh melampaui rentang data ketika dikeluarkan dari pelatihan.



Gambar 2. Nilai pTEAC eksperimen terhadap prediksi model MLR dan Random Forest untuk data latih (validasi LOO) dan set uji eksternal; prediksi LOO MLR untuk asam galat (−3,50) berada di luar sumbu

Gambar 2 memperlihatkan perbedaan kedua model secara visual. Pada Random Forest, sebagian besar titik berada dekat garis identitas, dengan galat LOO terbesar pada asam veratrat (residu +1,34) serta asam protokatekuat dan asam galat (−0,89 dan −0,70). Kedua senyawa terakhir sangat aktif, tetapi diprediksi sedikit lebih lemah karena Random Forest cenderung merata-ratakan nilai pada ujung rentang data. Pada MLR, prediksi untuk senyawa tidak aktif hampir seragam, sekitar 3,1–3,3, karena model hanya membedakan senyawa berdasarkan dua deskriptor.

Analisis Set Uji dan Perbandingan dengan Model 3D

Tabel 5 menyajikan prediksi untuk setiap senyawa uji. Ketiga model 2D memprediksi senyawa tidak aktif, yaitu asam 4-hidroksifenilasetat dan asam 2-metoksisinamat, dengan cukup baik. Kesulitan utama muncul pada asam 2,4-dihidroksisinamat. Model MLR memprediksinya sebagai tidak aktif (3,500), sedangkan nilai eksperimennya 0,964. Senyawa ini memiliki leverage 0,902 terhadap model MLR, jauh di atas batas $h^* = 0,60$, sehingga berada di luar domain aplikabilitas. Dengan kata lain, analisis domain berhasil memberi peringatan bahwa prediksi untuk senyawa ini tidak dapat dipercaya. Penyebabnya, pola dua OH yang saling meta tanpa gugus donor orto/para tidak terwakili dalam data latih, kecuali dalam asam galat yang juga memiliki pasangan orto.

This Article is licensed under a CC BY-SA 4.0 International License

Tabel 5. Prediksi pTEAC senyawa uji eksternal oleh model 2D dan model 3D

No	Senyawa	Eksperimen	MLR	Random Forest	SVR	CoMFA	CoMSIA	Leverage MLR
1	Asam vanilat	2,127	2,112	0,862	1,404	2,252	2,497	0,069
9	Asam gentisat	0,168	1,028	0,807	1,715	0,135	0,149	0,202
13	Asam 4-hidroksifenilasetat	3,482	3,348	3,366	2,644	3,297	3,602	0,208
16	Asam 2,4-dihidroksisinamat	0,964	3,500	1,988	1,728	0,622	1,090	0,902*
18	Asam kafeat	0,266	1,028	0,620	0,706	0,244	0,290	0,202
22	Asam 2-metoksisinamat	3,561	3,196	3,511	3,400	3,184	3,258	0,152

*Keterangan: * di luar domain aplikabilitas (h > h* = 0,60). Nilai CoMFA/CoMSIA dari Jing et al. (2012).*

Bila asam 2,4-dihidroksisinamat dikeluarkan dari evaluasi, R^2_{pred} model MLR naik dari 0,425 menjadi 0,876 (RMSE 0,543), dan Random Forest naik dari 0,767 menjadi 0,819 (RMSE 0,656). Hasil ini menunjukkan bahwa sebagian besar galat set uji berasal dari satu senyawa di luar domain. Kasus lain yang perlu dicatat adalah asam vanilat. Model MLR memprediksinya dengan tepat (2,112 dibandingkan 2,127), sedangkan Random Forest memprediksinya terlalu aktif (0,862). Random Forest tampaknya mengelompokkan asam vanilat bersama senyawa guaiakol lain yang aktif, karena deskriptor rantai samping yang tersedia belum cukup menangkap efek penarik elektron gugus karboksil yang terikat langsung pada cincin.

Model CoMFA dan CoMSIA dari Jing et al. (2012) menghasilkan R^2_{pred} 0,977 dan 0,981 pada set uji yang sama, lebih tinggi daripada seluruh model 2D dalam penelitian ini. Perbandingan ini perlu disampaikan secara terbuka. Model medan molekul tiga dimensi menangkap distribusi sterik dan elektrostatik secara lebih rinci, termasuk efek posisi substituen yang tidak terwakili oleh 15 deskriptor sederhana. Sebaliknya, model 2D tidak memerlukan penyejajaran struktur dan pemilihan konformasi, lebih cepat dihitung, dan persamaannya langsung dapat ditafsirkan. Model 2D dengan demikian lebih cocok untuk penapisan awal dan pembelajaran, sedangkan model 3D lebih tepat untuk optimasi senyawa yang strukturnya sudah sangat mirip.

Uji Pengacakan Respons dan Kekokohan Model

Hasil uji pengacakan respons disajikan pada Tabel 6. Untuk Random Forest, Q^2_{LOO} rata-rata dari 100 model dengan respons acak adalah −0,359 dengan nilai maksimum 0,503, dan tidak ada satu pun yang menyamai Q^2 model asli (0,892). Hasil ini menunjukkan bahwa kinerja Random Forest bukan kebetulan. Untuk MLR, nilai maksimum Q^2 acak (0,513) justru melampaui Q^2 model asli (0,409), dan dua dari 100 pengacakan menghasilkan Q^2 yang sama atau lebih tinggi. Dengan demikian, daya prediksi MLR secara statistik lemah, meskipun persamaannya tetap berguna untuk menjelaskan arah pengaruh deskriptor.

Tabel 6. Hasil uji pengacakan respons (100 kali)

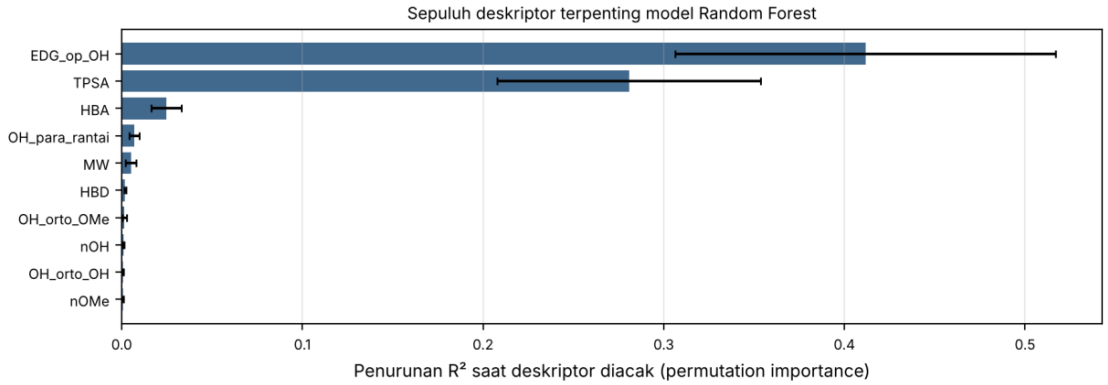
Model	Q^2_{LOO} asli	Rata-rata Q^2 acak	Q^2 acak maksimum	Jumlah Q^2 acak ≥ asli
MLR (2 deskriptor)	0,409	−0,379	0,513	2 dari 100
Random Forest	0,892	−0,359	0,503	0 dari 100

Nilai maksimum Q^2 acak yang cukup tinggi pada kedua model (sekitar 0,5) mencerminkan risiko korelasi kebetulan pada data dengan hanya 15 senyawa latih. Rücker et al. (2007) menekankan bahwa makin sedikit data dan makin banyak deskriptor, makin besar peluang terjadinya korelasi kebetulan. Karena itu, hasil penelitian ini sebaiknya dipandang sebagai bukti konsep, bukan sebagai model akhir untuk penggunaan luas.

This Article is licensed under a CC BY-SA 4.0 International License

Deskriptor Terpenting dan Interpretasi Mekanistik

Gambar 3 menunjukkan sepuluh deskriptor terpenting menurut permutation importance pada model Random Forest. Dua deskriptor sangat dominan, yaitu EDG_op_OH (penurunan R^2 0,412 ± 0,105) dan TPSA (0,281 ± 0,073). Deskriptor lain hanya berkontribusi kecil, seperti HBA (0,025), OH_para_rantai (0,007), dan MW (0,005). Peringkat yang sama juga diperoleh dari ukuran kepentingan berbasis pengotor (impurity), dengan nilai 0,429 untuk EDG_op_OH dan 0,337 untuk TPSA.



Gambar 3. Sepuluh deskriptor terpenting model Random Forest berdasarkan permutation importance (rata-rata ± simpangan baku dari 50 pengulangan)

Dominasi EDG_op_OH memiliki dasar mekanistik yang kuat. Deskriptor ini menghitung gugus donor elektron yang mampu menstabilkan radikal fenoksil melalui resonansi dan ikatan hidrogen intramolekul, yaitu dua faktor yang menurunkan entalpi disosiasi ikatan O–H (Leopoldini et al., 2011). Kajian teoretis Chen et al. (2015) pada asam fenolat juga menemukan bahwa asam kafeat memiliki entalpi disosiasi O–H yang jauh lebih rendah (75,1 kkal/mol) dibandingkan asam protokatekuat (83,2 kkal/mol). Temuan ini menunjukkan bahwa rantai samping terkonjugasi ikut menurunkan entalpi tersebut. TPSA, di sisi lain, merangkum jumlah total atom oksigen polar sehingga berkorelasi dengan jumlah OH dan OCH_3. Karena TPSA dan EDG_op_OH saling berkorelasi, kepentingan masing-masing tidak dapat dipisahkan sepenuhnya.

Deskriptor elektronik Hammett σ terhadap rantai samping hampir tidak berperan. Hal ini masuk akal karena σ dalam penelitian ini mengukur pengaruh substituen terhadap gugus karboksil, bukan terhadap gugus OH fenolik yang menjadi pusat reaksi. Penelitian lanjutan dapat menggunakan deskriptor elektronik yang dihitung relatif terhadap setiap OH, atau deskriptor kuantum seperti entalpi disosiasi ikatan O–H dan potensial ionisasi, yang telah terbukti berkorelasi dengan aktivitas DPPH (Chen et al., 2015).

Pengaruh Kelompok Deskriptor

Untuk menilai sumbangan setiap kelompok deskriptor, model Random Forest dengan pengaturan yang sama dibangun ulang menggunakan beberapa himpunan deskriptor. Hasilnya disajikan pada Tabel 7. Analisis ini bersifat eksploratif dan dilakukan setelah model utama selesai dievaluasi, sehingga tidak digunakan untuk memilih model akhir.

Tabel 7. Kinerja Random Forest dengan himpunan deskriptor yang berbeda

Himpunan deskriptor	Jumlah	Q^2_{LOO}	RMSE LOO	R^2_{pred} uji	RMSE uji
Konstitusional (nOH, nOMe, MW, HBD, HBA, TPSA)	6	0,844	0,599	0,833	0,618

This Article is licensed under a CC BY-SA 4.0 International License

Himpunan deskriptor	Jumlah	Q^2_{LOO}	RMSE LOO	R^2_{pred} uji	RMSE uji
Posisi/mekanistik (pola OH, EDG_op_OH, rantai samping)	8	0,919	0,431	0,465	1,107
Dua deskriptor teratas (EDG_op_OH, TPSA)	2	0,983	0,197	0,745	0,764
Tanpa EDG_op_OH	14	0,797	0,684	0,801	0,675
Seluruh deskriptor (model utama)	15	0,892	0,499	0,767	0,730

Catatan: pengaturan Random Forest sama dengan model utama (500 pohon, min. 2 sampel per daun).

Hasil pada Tabel 7 memperlihatkan pola yang perlu ditafsirkan dengan hati-hati. Himpunan dua deskriptor teratas memberikan Q^2_{LOO} sangat tinggi (0,983), tetapi kedua deskriptor tersebut dipilih berdasarkan kepentingan yang dihitung pada data latih yang sama. Nilai Q^2 ini karenanya bias ke atas akibat seleksi, dan pada set uji kinerjanya (R^2_{pred} = 0,745) tidak lebih baik daripada model utama. Himpunan deskriptor posisi/mekanistik juga menunjukkan selisih besar antara Q^2_{LOO} (0,919) dan R^2_{pred} (0,465). Selisih ini terutama berasal dari asam 2,4-dihidroksisinamat, yang pola posisinya tidak terwakili dalam data latih.

Sebaliknya, himpunan deskriptor konstitusional sederhana justru memberikan R^2_{pred} tertinggi (0,833) dengan Q^2_{LOO} yang tetap baik (0,844). Deskriptor jumlah seperti TPSA dan HBA tidak membedakan posisi substituen, sehingga senyawa di luar pola latih tidak diperlakukan secara ekstrem. Ketika EDG_op_OH dihapus, Q^2_{LOO} turun dari 0,892 menjadi 0,797, tetapi R^2_{pred} sedikit naik menjadi 0,801. Temuan ini menunjukkan adanya tarik-menarik antara daya jelas mekanistik dan kekokohan ekstrapolasi. Deskriptor posisi sangat informatif di dalam domain data, tetapi dapat menyesatkan untuk pola substitusi baru. Dengan hanya enam senyawa uji, perbedaan R^2_{pred} sebesar 0,05–0,07 juga belum dapat dianggap bermakna secara statistik.

Prediksi Aktivitas Senyawa Baru

Tabel 8 menyajikan prediksi untuk enam asam fenolat di luar data. Senyawa dengan motif katekol, yaitu asam 2,3-dihidroksibenzoat dan DOPAC, diprediksi aktif oleh kedua model, dengan pTEAC 0,79 dan 0,60 menurut Random Forest. Asam isoferulat diprediksi aktif oleh Random Forest (0,94) tetapi berada di dekat batas menurut MLR (2,11). Perbedaan ini mencerminkan ketidakpastian yang sama seperti pada asam vanilat. Asam 3,4-dimetoksisinamat dan asam 3-hidroksibenzoat diprediksi tidak aktif oleh kedua model, sesuai dengan tidak adanya OH fenolik pada senyawa pertama dan tidak adanya gugus donor orto/para pada senyawa kedua.

Tabel 8. Prediksi pTEAC asam fenolat di luar data

Senyawa	Rumus	MLR	Random Forest	Leverage	Dalam domain?	Prediksi kelas (RF)
Asam 2,3-dihidroksibenzoat	C7H6O4	1,03	0,79	0,202	Ya	Aktif
Asam 3,5-dihidroksibenzoat	C7H6O4	3,50	2,18	0,902	Tidak	Tidak dapat dipastikan
Asam isoferulat	C10H10O4	2,11	0,94	0,069	Ya	Aktif
Asam 3,4-dimetoksisinamat	C11H12O4	3,20	3,20	0,152	Ya	Kurang aktif
Asam 3,4-dihidroksifenilasetat (DOPAC)	C8H8O4	1,03	0,60	0,202	Ya	Aktif
Asam 3-hidroksibenzoat	C7H6O3	3,35	3,51	0,208	Ya	Kurang aktif

Asam 3,5-dihidroksibenzoat memiliki leverage 0,902, sama dengan asam 2,4-dihidroksisinamat, karena keduanya hanya memiliki pasangan OH meta. Pengalaman pada set uji menunjukkan bahwa prediksi untuk pola ini tidak dapat diandalkan, sehingga senyawa tersebut sebaiknya diuji secara eksperimen sebelum diambil kesimpulan. Contoh ini

This Article is licensed under a CC BY-SA 4.0 International License

memperlihatkan fungsi praktis domain aplikabilitas: model tidak hanya memberikan angka, tetapi juga menandai kapan angka tersebut tidak layak dipercaya (OECD, 2007).

Secara kualitatif, prediksi untuk senyawa katekol konsisten dengan data eksperimen dari sumber lain. Özdemir et al. (2026), misalnya, melaporkan bahwa asam galat dan asam kafeat memiliki IC50 DPPH yang rendah (4,45 dan 12,93 μmol/L), sedangkan senyawa tanpa gugus katekol seperti asam 4-hidroksibenzoat tidak mencapai 50% penghambatan pada rentang konsentrasi yang diuji. Karena metode dan satuannya berbeda, perbandingan ini hanya dapat dilakukan pada tingkat urutan aktivitas, bukan pada nilai absolut.

Implikasi dan Keterbatasan

Penelitian ini menunjukkan bahwa deskriptor sederhana yang dirancang berdasarkan mekanisme dapat menjelaskan sebagian besar variasi aktivitas antioksidan asam fenolat, dan Random Forest memberikan prediksi terbaik di antara model 2D yang diuji. Bagi peneliti bahan alam, pendekatan ini dapat digunakan untuk mengurutkan prioritas senyawa dari ekstrak tumbuhan sebelum diisolasi dan diuji, terutama untuk senyawa fenolik yang strukturnya telah teridentifikasi melalui analisis kromatografi dan spektrometri massa.

Keterbatasan utama penelitian ini adalah jumlah data yang kecil (21 senyawa) dan cakupan struktur yang terbatas pada asam fenolat sederhana. Model belum dapat diterapkan pada flavonoid, tanin, atau senyawa dengan gugus fungsi berbeda. Deskriptor dihitung dengan program yang disusun khusus berdasarkan pola substitusi, sehingga belum mencakup deskriptor topologi dan kuantum yang lebih lengkap seperti yang tersedia pada RDKit atau PaDEL. Selain itu, aktivitas hanya diukur dengan satu metode (DPPH), padahal urutan aktivitas dapat berbeda pada uji ABTS, FRAP, atau ORAC. Penelitian lanjutan disarankan menggunakan data yang lebih besar dan beragam dengan protokol yang seragam, menambahkan deskriptor kuantum, serta menguji prediksi secara eksperimen.

E. CONCLUSION (KESIMPULAN)

Penelitian ini telah membangun model prediksi aktivitas penangkapan radikal DPPH 21 asam fenolat berdasarkan 15 deskriptor molekul yang dirancang secara mekanistik. Analisis data menunjukkan bahwa senyawa dengan dua gugus OH pada posisi orto atau para merupakan kelompok paling aktif (pTEAC rata-rata 0,230), diikuti senyawa dengan OH yang bersebelahan dengan gugus metoksi (0,919), sedangkan senyawa tanpa pola tersebut hampir tidak aktif.

Di antara lima algoritma, Random Forest memberikan kinerja terbaik dengan $Q^2_{LOO} = 0{,}892$ dan $R^2_{pred} = 0{,}767$ pada set uji eksternal, serta lolos uji pengacakan respons. Deskriptor terpenting adalah jumlah gugus donor elektron pada posisi orto/para terhadap OH fenolik (EDG_op_OH) dan luas permukaan polar topologis (TPSA). Model MLR dua deskriptor mudah ditafsirkan, tetapi daya prediksinya lemah. Model 2D dalam penelitian ini belum menyamai akurasi model CoMFA/CoMSIA tiga dimensi (R^2_{pred} sekitar 0,98), terutama karena satu senyawa uji berada di luar domain aplikabilitas.

Secara keseluruhan, pendekatan machine learning berbasis deskriptor molekul yang sederhana dan transparan dapat digunakan untuk penapisan awal dan penjelasan mekanistik aktivitas antioksidan, asalkan dilengkapi validasi eksternal, uji pengacakan respons, dan analisis domain aplikabilitas. Penelitian lanjutan perlu memperluas data ke kelompok senyawa bioaktif lain, menambahkan deskriptor kuantum, dan memverifikasi prediksi melalui pengujian laboratorium.

This Article is licensed under a CC BY-SA 4.0 International License

F. BIBLIOGRAPHY (DAFTAR PUSTAKA)

Brand-Williams, W., Cuvelier, M. E., & Berset, C. (1995). Use of a free radical method to evaluate antioxidant activity. LWT – Food Science and Technology, 28(1), 25–30. https://doi.org/10.1016/S0023-6438(95)80008-5

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, Y., Xiao, H., Zheng, J., & Liang, G. (2015). Structure-thermodynamics-antioxidant activity relationships of selected natural phenolic acids and derivatives: An experimental and theoretical evaluation. PLoS ONE, 10(3), e0121276. https://doi.org/10.1371/journal.pone.0121276

Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., … Tropsha, A. (2014). QSAR modeling: Where have you been? Where are you going to? Journal of Medicinal Chemistry, 57(12), 4977–5010. https://doi.org/10.1021/jm4004285

Cuvelier, M.-E., Richard, H., & Berset, C. (1992). Comparison of the antioxidative activity of some acid-phenols: Structure-activity relationship. Bioscience, Biotechnology, and Biochemistry, 56(2), 324–325. https://doi.org/10.1271/bbb.56.324

Ertl, P., Rohde, B., & Selzer, P. (2000). Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. Journal of Medicinal Chemistry, 43(20), 3714–3717. https://doi.org/10.1021/jm000942e

Golbraikh, A., & Tropsha, A. (2002). Beware of q2! Journal of Molecular Graphics and Modelling, 20(4), 269–276. https://doi.org/10.1016/S1093-3263(01)00123-1

Halliwell, B., & Gutteridge, J. M. C. (2015). Free radicals in biology and medicine (5th ed.). Oxford University Press.

Hansch, C., Leo, A., & Taft, R. W. (1991). A survey of Hammett substituent constants and resonance and field parameters. Chemical Reviews, 91(2), 165–195. https://doi.org/10.1021/cr00002a004

Jing, P., Zhao, S.-J., Jian, W.-J., Qian, B.-J., Dong, Y., & Pang, J. (2012). Quantitative studies on structure-DPPH• scavenging activity relationships of food phenolic acids. Molecules, 17(11), 12910–12924. https://doi.org/10.3390/molecules171112910

Leopoldini, M., Russo, N., & Toscano, M. (2011). The molecular basis of working mechanism of natural polyphenolic antioxidants. Food Chemistry, 125(2), 288–306. https://doi.org/10.1016/j.foodchem.2010.08.012

OECD. (2007). Guidance document on the validation of (quantitative) structure-activity relationship [(Q)SAR] models (OECD Series on Testing and Assessment No. 69). OECD Publishing. https://doi.org/10.1787/9789264085442-en

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Rice-Evans, C. A., Miller, N. J., & Paganga, G. (1996). Structure-antioxidant activity relationships of flavonoids and phenolic acids. Free Radical Biology and Medicine, 20(7), 933–956. https://doi.org/10.1016/0891-5849(95)02227-9

Rücker, C., Rücker, G., & Meringer, M. (2007). y-Randomization and its variants in QSPR/QSAR. Journal of Chemical Information and Modeling, 47(6), 2345–2357. https://doi.org/10.1021/ci700157b

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Todeschini, R., & Consonni, V. (2009). Molecular descriptors for chemoinformatics (2nd ed.). Wiley-VCH. https://doi.org/10.1002/9783527628766

Tropsha, A. (2010). Best practices for QSAR model development, validation, and exploitation. Molecular Informatics, 29(6–7), 476–488. https://doi.org/10.1002/minf.201000061

This Article is licensed under a CC BY-SA 4.0 International License

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58(2), 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1

Özdemir, Z. Ö., Soy, M., Ataseven, S., Özer, A., & Elmastaş, M. (2026). Comparative antioxidant profiling of phenolic acids and flavonoids: Assay-resolved structure–activity relationships under harmonized in vitro conditions. Molecules, 31(9), 1478. https://doi.org/10.3390/molecules31091478

This Article is licensed under a CC BY-SA 4.0 International License