Jurnal Ekonomi, Teknologi dan Bisnis (JETBIS)
Volume 2, Number 10 October 2026
p-ISSN 2964-903X; e-ISSN 2962-9330

Penerapan *Machine Learning* untuk Klasifikasi Asal Anggur Berdasarkan Komposisi Kimia Guna Mendukung Autentikasi dan Pencegahan Pemalsuan dalam Industri Wine

Zeny Widianingsih[1], Aldi Algino[2]
Sekolah Tinggi Manajemen Informatika dan Komputer IKMI, Indonesia[1,2]
*Email: zenywidianingsih05@gmail.com[1], alginoaldi88@gmail.com[2]

ABSTRAK

Pemalsuan anggur (wine fraud) merupakan masalah bisnis serius yang menimbulkan kerugian ekonomi, merusak kepercayaan konsumen, dan mendistorsi harga pasar, terutama pada segmen wine premium yang nilainya bergantung pada keaslian asal dan kultivar. Penelitian ini bertujuan merancang dan membandingkan lima model machine learning untuk mengklasifikasikan kultivar anggur berdasarkan 13 parameter komposisi kimia serta mengidentifikasi parameter yang paling berpengaruh sebagai dasar autentikasi produk. Penelitian menggunakan pendekatan eksperimental kuantitatif dengan data sekunder Wine sebanyak 178 sampel dari tiga kultivar di Italia. Data dibagi secara stratified menjadi 142 data latih dan 36 data uji dengan rasio 80:20. Model yang dibandingkan meliputi Deep Neural Network (Multi-Layer Perceptron), Logistic Regression, Random Forest, Support Vector Machine, dan K-Nearest Neighbors. Hasil pengujian pada pembagian data tersebut menunjukkan Random Forest mencapai akurasi 100%, sedangkan empat model lainnya mencapai 97,22%. Feature importance Random Forest mengidentifikasi intensitas warna, flavonoid, dan prolin sebagai parameter kimia paling dominan. Temuan ini menunjukkan potensi komposisi kimia standar dan machine learning sebagai pendukung autentikasi yang objektif dan efisien bagi produsen serta lembaga sertifikasi untuk menjaga reputasi merek dan integritas pasar. Namun, karena data uji hanya 36 sampel dan belum menggunakan cross-validation, hasil tersebut perlu ditafsirkan sebagai bukti awal yang memerlukan validasi lebih luas sebelum penerapan operasional.

Kata Kunci: pembelajaran mesin; pembelajaran mendalam; autentikasi produk; industri anggur; pencegahan pemalsuan

ABSTRACT

Wine fraud is a serious business issue that causes economic losses, erodes consumer trust, and distorts market prices, particularly in the premium wine segment where value hinges on the authenticity of origin and cultivar. This study aims to design and compare five machine learning models to classify wine cultivars based on 13 chemical composition parameters and to identify the most influential parameters for product authentication. A quantitative experimental approach was employed using a secondary dataset comprising 178 wine samples from three Italian cultivars. The data were stratified into a training set of 142 samples and a test set of 36 samples, maintaining an 80:20 ratio. The models compared included Deep Neural Network (Multi-Layer Perceptron), Logistic Regression, Random Forest, Support Vector Machine, and K-Nearest Neighbors. Test results showed that the Random Forest model achieved 100% accuracy, while the other four models achieved 97.22%. Feature importance analysis within the Random Forest model identified color intensity, flavonoids, and proline as the most dominant chemical parameters. These findings demonstrate the potential of using standard chemical composition data and machine learning to support objective and efficient authentication for producers and certification bodies, thereby safeguarding brand reputation and market integrity. However, given the small test set (36 samples) and the absence of cross-validation, these results should be interpreted as preliminary evidence requiring broader validation prior to operational implementation.

Keywords: *machine learning; deep learning; product authentication; wine industry; fraud prevention*

PENDAHULUAN

Industri anggur (wine) global merupakan salah satu sektor bisnis dengan nilai ekonomi tinggi yang sangat rentan terhadap praktik pemalsuan (counterfeiting) dan adulterasi produk. Pemalsuan wine mencakup pemalsuan label asal (provenance), vintage, maupun kultivar anggur, dan diperkirakan menyebabkan kerugian tahunan hingga miliaran dolar bagi industri secara global. Kasus-kasus pemalsuan wine premium telah berulang kali mengguncang pasar, salah satu yang paling terkenal adalah skandal pemalsuan Sassicaia 2015, sebuah wine kelas premium asal Italia (Super Tuscan), di mana operasi pemalsuan yang terungkap pada 2020 berhasil mengedarkan botol-botol palsu dengan kualitas tinggi seharga sekitar sepertiga dari harga wine asli, sehingga merusak kepercayaan pasar terhadap keaslian produk (Kuang et al., 2024). Selain kerugian finansial langsung, pemalsuan semacam ini mendistorsi mekanisme harga pasar, merusak reputasi produsen yang sah, dan pada akhirnya dapat menurunkan kepercayaan konsumen terhadap keseluruhan industri.

Pada tingkat rantai pasok, pemalsuan pangan bukan hanya pelanggaran label, melainkan perilaku oportunistik yang dipengaruhi oleh insentif ekonomi, asimetri informasi, kompleksitas distribusi, serta keterbatasan pengawasan. Dampaknya dapat menjalar dari kerugian produsen dan distributor hingga erosi kepercayaan konsumen dan terganggunya integritas pasar (Brooks et al., 2021; Giannakas & Yiannaka, 2023). Literatur kebijakan juga menegaskan bahwa deteksi analitis perlu ditempatkan bersama tata kelola pencegahan karena karakter risiko berbeda menurut segmen dan posisi pelaku dalam rantai pasok (Lord et al., 2022; Van Ruth & Nillesen, 2021). Dalam industri wine premium, karakter produk sebagai credence good membuat konsumen sulit memverifikasi asal dan kualitas sebelum pembelian, sehingga mekanisme autentikasi objektif memiliki nilai ekonomi yang langsung.

Verifikasi keaslian asal dan kultivar anggur secara konvensional umumnya mengandalkan uji sensori oleh panel ahli (sommelier) atau dokumen sertifikasi rantai pasok, yang keduanya memiliki keterbatasan: penilaian sensori bersifat subjektif dan mahal untuk dilakukan secara masif, sementara dokumen sertifikasi rentan dipalsukan atau dimanipulasi sepanjang rantai distribusi. Kondisi ini mendorong kebutuhan akan metode verifikasi berbasis data objektif yang dapat diukur secara instrumental, seperti komposisi kimia anggur, sebagai dasar teknologi autentikasi yang lebih sulit dipalsukan dibandingkan label maupun kemasan.

Kajian autentikasi pangan menunjukkan bahwa verifikasi dapat dilakukan melalui teknik kromatografi, spektroskopi, pencitraan, dan pengukuran komposisi kimia yang dipadukan dengan kemometrik atau machine learning. Pada wine, spektroskopi dan NMR mampu menghasilkan fingerprint yang informatif, tetapi implementasinya bergantung pada ketersediaan instrumen, basis data pembanding, prosedur kalibrasi, dan validasi antarlaboratorium (Ranaweera, Capone, et al., 2021; Solovyev et al., 2021). Secara lebih luas, metode autentikasi harus selektif, sensitif, akurat, dapat direproduksi, dan cukup robust untuk menghadapi variasi alami produk (Cichna-Markl & Mafra, 2023; Zhang et al., 2024). Persyaratan tersebut menegaskan pentingnya menilai apakah parameter kimia standar yang lebih mudah tersedia dapat memberi kemampuan klasifikasi yang memadai tanpa meniadakan kebutuhan validasi lanjutan.

Perkembangan teknologi machine learning dan deep learning membuka peluang besar bagi industri wine untuk membangun sistem autentikasi berbasis data instrumental. Berbagai penelitian terbaru menunjukkan keberhasilan penerapan machine learning pada data spektroskopi maupun komposisi kimia anggur untuk keperluan identifikasi merek, asal, maupun deteksi pemalsuan. Pan et al., (2022) mengembangkan model klasifikasi berbasis spektroskopi visible dan near-infrared yang dikombinasikan dengan Bayes classifier untuk identifikasi multi-merek wine guna mencegah adulterasi dan pemalsuan. Ta et al., (2023) mengusulkan metode penelusuran komponen wine (wine

component tracing) berbasis fusi spektrum near-infrared dan machine learning, yang mampu meningkatkan akurasi klasifikasi asal produksi wine dari 98,3% menjadi 99,75% menggunakan algoritma PCA-SVM yang disempurnakan pada 300 sampel wine dari enam wilayah produksi.

Penelitian yang lebih baru oleh Kritzinger et al., (2026) bahkan berhasil mengembangkan teknik autentikasi wine non-invasif melalui botol tertutup menggunakan spektroskopi fluoresensi near-infrared, yang mampu mengklasifikasikan dua puluh jenis wine merah secara akurat tanpa membuka botol, menunjukkan tren riset yang terus bergerak ke arah solusi autentikasi yang semakin praktis dan dapat diterapkan langsung di titik distribusi maupun titik jual.

Dari sisi metodologis, kemometrik memungkinkan matriks data kimia diterjemahkan menjadi pola klasifikasi, diskriminasi, dan prediksi yang relevan bagi autentikasi dan traceability (González-Domínguez et al., 2022). Campmajó et al., (2022), misalnya, menunjukkan bahwa fingerprint kimia yang dipadukan dengan model kemometrik dapat membedakan asal wine merah setelah validasi eksternal. Pada konteks lain, fusi beberapa model klasifikasi juga dikembangkan untuk meningkatkan reliabilitas keputusan autentikasi pangan dan minuman (Kalivas & Lemos, 2023). Studi Ranaweera, Gilmore, et al., (2021a), 2021(b) memperlihatkan bahwa informasi spektrofluorometrik dan multi-elemen dapat mendukung autentikasi geografis maupun varietal.

Rangkaian temuan tersebut memperkuat relevansi machine learning, tetapi sekaligus menunjukkan bahwa jenis data, desain validasi, dan interpretabilitas model menentukan kekuatan bukti yang dihasilkan. Meskipun berbagai penelitian tersebut menunjukkan hasil yang menjanjikan, sebagian besar berfokus pada teknik akuisisi data yang relatif kompleks dan mahal seperti spektroskopi near-infrared khusus, sementara kajian yang mengeksplorasi potensi data komposisi kimia standar (yang lazim tersedia dari uji laboratorium konvensional) sebagai basis klasifikasi kultivar menggunakan berbagai algoritma machine learning secara komparatif, sekaligus mengaitkannya secara eksplisit dengan implikasi bisnis berupa pencegahan kerugian ekonomi akibat pemalsuan, masih relatif terbatas.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk merancang dan membandingkan kinerja lima algoritma machine learning, termasuk pendekatan deep learning berbasis Artificial Neural Network, dalam mengklasifikasikan kultivar anggur berdasarkan parameter komposisi kimia standar, serta mengidentifikasi parameter kimia yang paling berpengaruh sebagai dasar rekomendasi praktis bagi pengembangan sistem autentikasi produk di industri wine. Kebaruan penelitian ini terletak pada perbandingan komprehensif lima algoritma pembelajaran mesin pada data komposisi kimia standar (bukan data spektroskopi khusus), disertai analisis feature importance yang dapat memberikan panduan praktis parameter kimia mana yang paling krusial untuk diprioritaskan dalam pengembangan sistem verifikasi produk berbiaya efisien.

Celah yang hendak diisi bukan semata-mata perbedaan objek atau lokasi, melainkan keterbatasan bukti komparatif mengenai sejauh mana data komposisi kimia standar dapat menopang klasifikasi kultivar ketika diproses dengan model linear, berbasis jarak, kernel, ensemble, dan jaringan saraf dalam satu skema evaluasi. Literatur terdahulu lebih banyak menonjolkan instrumen khusus atau satu keluarga model, sementara hubungan antara akurasi teknis, parameter kimia dominan, dan keputusan autentikasi berbiaya efisien belum selalu dibahas secara terpadu. Karena ukuran sampel kecil dapat menghasilkan estimasi performa yang optimistis, hasil dari satu pembagian data perlu dipahami sebagai bukti awal yang memerlukan validasi lebih luas (Vabalas et al., 2019). Penegasan ini membatasi klaim penelitian tanpa mengurangi nilai eksploratif perbandingan model yang dilakukan.

Berdasarkan kesenjangan tersebut, penelitian ini merumuskan dua pertanyaan utama: sejauh mana lima algoritma machine learning mampu mengklasifikasikan tiga kultivar anggur berdasarkan 13 parameter

komposisi kimia pada dataset yang digunakan, dan parameter kimia apa yang paling dominan dalam keputusan klasifikasi Random Forest? Kedua pertanyaan ini mengarahkan tujuan penelitian untuk membandingkan performa model serta mengidentifikasi penciri kimia yang relevan bagi autentikasi produk. Rumusan tersebut juga membatasi interpretasi pada dataset dan rancangan pengujian yang digunakan, bukan pada seluruh kultivar atau wilayah produksi wine.

Secara teoretis, penelitian ini memberi bukti komparatif mengenai penggunaan beberapa keluarga algoritma pada data kimia standar serta memperjelas posisi feature importance sebagai dasar interpretasi awal. Secara praktis, hasilnya dapat membantu produsen, laboratorium pengujian, dan lembaga sertifikasi menilai parameter yang layak diprioritaskan ketika menyusun sistem verifikasi yang lebih terjangkau. Implikasi manajerialnya adalah tersedianya dasar untuk merancang tahap penyaringan risiko sebelum pemeriksaan instrumental yang lebih mahal, memperkuat kendali mutu, dan mendukung keputusan penelusuran produk yang dicurigai. Namun, penerapan operasional tetap memerlukan dataset representatif, validasi silang atau eksternal, pengujian signifikansi antarmodel, serta prosedur audit yang mengintegrasikan hasil model dengan bukti rantai pasok.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimental kuantitatif. Dataset yang digunakan adalah dataset Wine yang tersedia secara terbuka melalui UCI Machine Learning Repository dan terintegrasi pada pustaka scikit-learn, pertama kali dipublikasikan oleh Michele, et.al (1998) peneliti di Institute of Pharmaceutical and Food Analysis and Technologies, Genoa, Italia. Dataset ini terdiri atas 178 sampel anggur yang berasal dari tiga kultivar berbeda (class_0, class_1, class_2) yang ditanam di wilayah yang sama di Italia, dengan distribusi 59 sampel pada kultivar pertama, 71 sampel pada kultivar kedua, dan 48 sampel pada kultivar ketiga.

Setiap sampel diukur pada 13 parameter komposisi kimia hasil analisis laboratorium, meliputi kadar alkohol, asam malat, kadar abu (ash), alkalinitas abu, magnesium, total fenol, flavonoid, fenol non-flavonoid, proantosianin, intensitas warna, hue, rasio OD280/OD315 dari wine yang diencerkan, serta kadar prolin. Data penelitian bersifat sekunder; teknik pengumpulan data dilakukan melalui dokumentasi dan pemuatan dataset publik menggunakan fungsi load_wine pada scikit-learn, bukan melalui pengambilan sampel primer. Instrumen penelitian mencakup data dictionary 13 fitur kimia, skrip Python untuk pra-pemrosesan dan pemodelan, serta matriks evaluasi yang dihasilkan dari data uji. Pustaka scikit-learn dipilih karena menyediakan implementasi yang terdokumentasi dan dapat direproduksi untuk seluruh model pembanding (Pedregosa et al., 2018).

Tahap pertama adalah pembagian data menjadi data latih dan data uji dengan rasio 80:20 menggunakan teknik stratified sampling untuk menjaga proporsi ketiga kelas kultivar pada kedua subset, menghasilkan 142 sampel data latih dan 36 sampel data uji. Tahap kedua adalah pra-pemrosesan data berupa standardisasi fitur (z-score normalization) menggunakan StandardScaler, mengingat rentang nilai antarparameter kimia bervariasi sangat besar (misalnya kadar prolin berkisar ratusan hingga ribuan, sementara hue berkisar kurang dari 2), sehingga standardisasi penting untuk mencegah dominasi fitur berskala besar terhadap proses pembelajaran model, khususnya pada model berbasis jarak seperti K-Nearest Neighbors dan SVM.

Tahap ketiga adalah pelatihan dan perbandingan lima model klasifikasi. Model utama yang menjadi fokus penelitian ini adalah Deep Neural Network berupa Multi-Layer Perceptron (MLP) dengan dua lapisan tersembunyi berisi 32 dan 16 neuron, menggunakan fungsi aktivasi ReLU dan dilatih menggunakan optimizer Adam dengan skema early stopping berdasarkan galat validasi. Sebagai pembanding, penelitian ini juga melatih Logistic Regression, Random Forest (300 pohon

keputusan), Support Vector Machine dengan kernel Radial Basis Function (RBF), dan K-Nearest Neighbors (k=5) pada data dan skema pembagian yang identik. Tahap keempat adalah evaluasi kinerja model pada data uji menggunakan metrik akurasi, presisi makro, recall makro, F1-score makro (rata-rata metrik di ketiga kelas kultivar dengan bobot setara), dan confusion matrix.

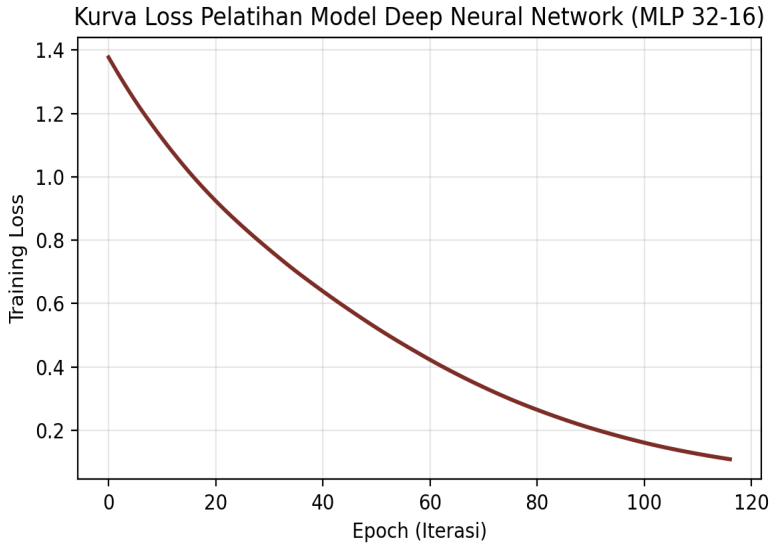
Tahap kelima adalah analisis feature importance menggunakan model Random Forest untuk mengidentifikasi parameter kimia mana yang paling berkontribusi terhadap keputusan klasifikasi, sebagai dasar rekomendasi praktis parameter kimia prioritas bagi pengembangan sistem autentikasi produk yang efisien secara biaya. Seluruh eksperimen dilaksanakan menggunakan bahasa pemrograman Python dengan pustaka scikit-learn versi 1.8, dengan random state tetap (42) pada seluruh tahapan untuk menjamin reproduksibilitas hasil.

Konfigurasi MLP 32-16 dipilih sebagai arsitektur ringkas agar kapasitas model tetap proporsional terhadap 13 fitur dan 142 data latih; 300 pohon pada Random Forest digunakan untuk menstabilkan agregasi ensemble, sedangkan k=5 pada K-Nearest Neighbors dipilih untuk menyeimbangkan sensitivitas lokal dan variasi prediksi. Prinsip agregasi pohon pada Random Forest mengikuti kerangka (Breiman, 2001). Seluruh konfigurasi ditetapkan sebelum evaluasi dan tidak dioptimalkan menggunakan data uji. Hasil yang dilaporkan merupakan evaluasi pada satu pembagian stratified 80:20; penelitian ini belum menghasilkan estimasi rerata dan standar deviasi dari k-fold cross-validation, sehingga angka kinerja tidak ditafsirkan sebagai estimasi generalisasi final (Vabalas et al., 2019).

HASIL DAN PEMBAHASAN

Dataset Wine yang digunakan terdiri atas 178 sampel dengan distribusi 59 sampel kultivar class_0 (33,1%), 71 sampel kultivar class_1 (39,9%), dan 48 sampel kultivar class_2 (27,0%). Setelah pembagian data secara stratified dengan rasio 80:20, diperoleh 142 sampel data latih dan 36 sampel data uji dengan proporsi kelas yang tetap konsisten dengan dataset asli.

Model Deep Neural Network (MLP 32-16) dilatih hingga mencapai kondisi konvergen, sebagaimana ditunjukkan oleh kurva penurunan training loss pada Gambar 1, yang menunjukkan penurunan loss yang stabil hingga melandai tanpa indikasi ketidakstabilan proses optimisasi.



Gambar 1. Kurva training loss model Deep Neural Network (MLP 32-16) pada dataset Wine.

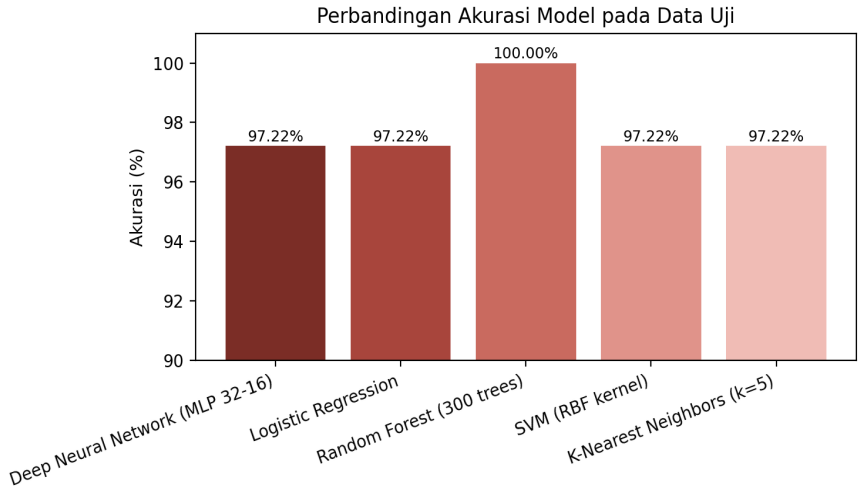
Tabel 1 menyajikan perbandingan kinerja kelima model klasifikasi pada data uji. Pada

pembagian data tunggal 80:20 dengan 36 sampel uji, model Random Forest mengklasifikasikan seluruh sampel secara benar, sehingga akurasi, presisi, recall, dan F1-score pada split tersebut masing-masing mencapai 100%; angka ini tidak dengan sendirinya membuktikan kinerja sempurna pada populasi yang lebih luas. Model Deep Neural Network menempati posisi kedua dengan akurasi 97,22%, presisi makro 97,44%, recall makro 97,62%, dan F1-score makro 97,43%. Model Logistic Regression, SVM, dan K-Nearest Neighbors masing-masing mencapai akurasi 97,22%, 97,22%, dan 97,22%.

Tabel 1. Perbandingan kinerja lima model klasifikasi pada 36 data uji dataset Wine, dievaluasi menggunakan skema pembagian data latih-uji 80:20 yang identik untuk seluruh model, diurutkan berdasarkan akurasi tertinggi.

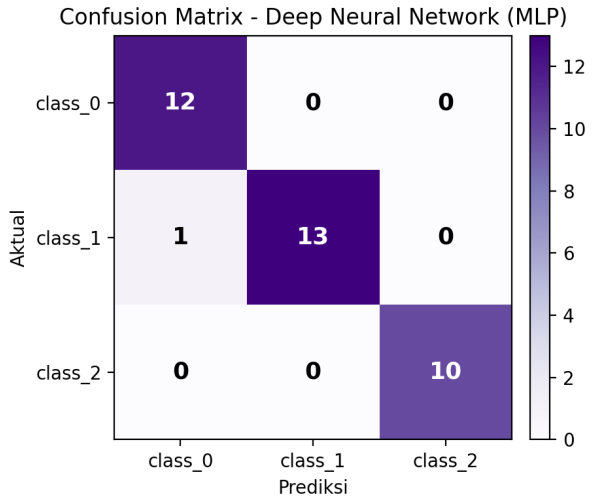
Model	Akurasi	Presisi (Makro)	Recall (Makro)	F1-Score (Makro)
Random Forest (300 pohon)	100,00%	100,00%	100,00%	100,00%
Deep Neural Network (MLP)	97,22%	97,44%	97,62%	97,43%
Logistic Regression	97,22%	97,78%	96,67%	97,10%
SVM (kernel RBF)	97,22%	97,78%	96,67%	97,10%
K-Nearest Neighbors (k=5)	97,22%	96,97%	97,62%	97,18%

Sumber: Data diolah

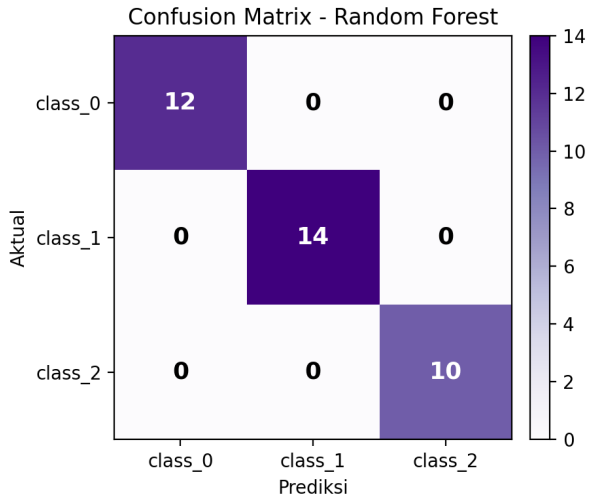


Gambar 2. Perbandingan Akurasi Kelima Model pada Data Uji

Gambar 3 dan Gambar 4 menyajikan confusion matrix model Deep Neural Network dan Random Forest secara berturut-turut. Model Random Forest berhasil mengklasifikasikan seluruh 36 sampel data uji secara benar tanpa kesalahan sama sekali, sementara model Deep Neural Network melakukan satu kesalahan klasifikasi, yaitu satu sampel kultivar class_1 yang salah diklasifikasikan sebagai class_0.

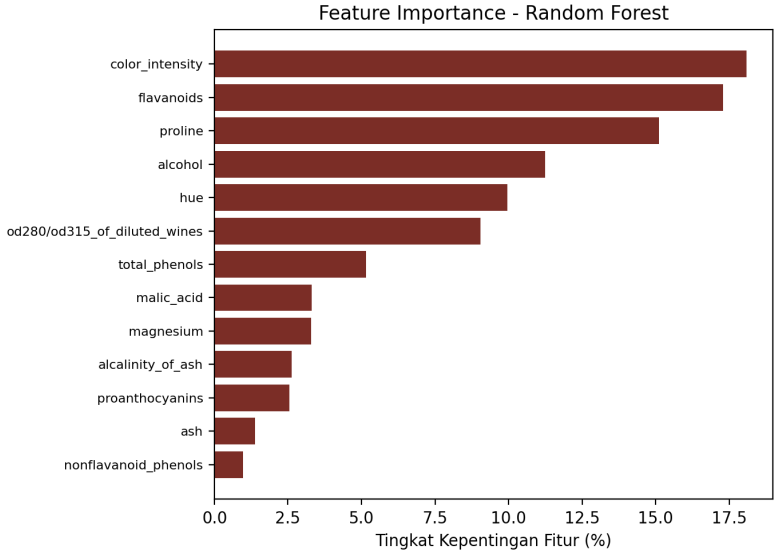


Gambar 3. Confusion *Matrix Model Deep Neural Network* (MLP) pada Data Uji



Gambar 4. Confusion *Matrix Model Random Forest* pada Data Uji

Analisis feature importance menggunakan model Random Forest, sebagaimana disajikan pada Gambar 5, mengidentifikasi lima parameter kimia paling berpengaruh secara berurutan, yaitu color_intensity, flavanoids, proline, alcohol, hue. Intensitas warna (color intensity) dan kadar flavonoid menyumbang porsi kepentingan tertinggi, masing-masing sekitar 18% dan 17% dari total kontribusi keputusan klasifikasi model, diikuti oleh kadar prolin sekitar 15%. Temuan ini sejalan dengan prinsip enologi bahwa intensitas warna dan profil senyawa fenolik seperti flavonoid sangat dipengaruhi oleh varietas anggur, kondisi tanah, dan proses fermentasi yang spesifik pada tiap kultivar, sehingga secara ilmiah masuk akal jika parameter-parameter tersebut menjadi penciri paling kuat asal kultivar. Karena ukuran kepentingan yang digunakan berasal dari penurunan impurity internal Random Forest, urutan tersebut bersifat model-specific dan belum diuji dengan permutation importance atau SHAP. Literatur menunjukkan bahwa ukuran kepentingan berbasis pohon dapat memiliki bias, sehingga temuan parameter dominan perlu diperlakukan sebagai indikasi awal dan dikonfirmasi melalui metode interpretasi agnostik pada penelitian lanjutan (Altmann et al., 2010).



Gambar 5. Feature Importance Parameter Kimia berdasarkan model Random Forest

Dibandingkan dengan penelitian terdahulu, hasil pada penelitian ini konsisten dengan tren yang dilaporkan Ta et al., (2023) yang mencapai akurasi hingga 99,75% dalam penelusuran asal produksi wine menggunakan kombinasi spektroskopi near-infrared dan machine learning, maupun Pan et al., (2022) yang berhasil mengidentifikasi multi-merek wine secara akurat menggunakan spektroskopi dikombinasikan dengan Bayes classifier. Yang menarik, penelitian ini menunjukkan bahwa akurasi pengujian yang tinggi (97-100%) pada split tunggal dapat dicapai bahkan hanya menggunakan 13 parameter komposisi kimia standar hasil uji laboratorium konvensional, tanpa memerlukan instrumen spektroskopi khusus seperti pada penelitian-penelitian tersebut maupun teknik non-invasif melalui botol tertutup sebagaimana dikembangkan oleh (Kritzinger et al., 2026). Hal ini mengindikasikan bahwa bagi produsen atau lembaga sertifikasi dengan keterbatasan anggaran untuk investasi instrumen spektroskopi canggih, pendekatan machine learning berbasis data uji laboratorium standar yang sudah lazim tersedia tetap dapat memberikan kinerja klasifikasi yang sangat kompetitif sebagai fondasi sistem autentikasi produk.

Dari sisi implikasi bisnis, tingkat akurasi klasifikasi yang sangat tinggi pada penelitian ini menegaskan bahwa komposisi kimia anggur menyimpan sidik jari (fingerprint) kimiawi yang secara efektif dapat membedakan kultivar, sehingga berpotensi diadopsi sebagai lapisan verifikasi tambahan dalam rantai pasok wine premium, melengkapi mekanisme autentikasi konvensional seperti sertifikat asal (appellation) dan segel kemasan yang selama ini terbukti rentan dipalsukan, sebagaimana ditunjukkan pada kasus Sassicaia 2015 (Kuang et al., 2024). Bagi produsen anggur premium, investasi pada sistem verifikasi berbasis data komposisi kimia dan machine learning berpotensi memberikan pengembalian yang signifikan mengingat besarnya nilai pasar wine premium yang berisiko terdampak pemalsuan; sementara bagi konsumen dan lembaga sertifikasi independen, ketersediaan alat verifikasi objektif berbasis data dapat meningkatkan kepercayaan pasar secara keseluruhan. Selain untuk pencegahan pemalsuan, identifikasi parameter kimia paling berpengaruh (seperti intensitas warna, flavonoid, dan prolin) juga dapat dimanfaatkan produsen sebagai indikator kendali mutu (quality control) selama proses produksi, membantu memastikan konsistensi karakteristik produk yang menjadi identitas merek dan dasar penetapan harga.

Penelitian ini memiliki sejumlah keterbatasan yang perlu diperhatikan. Pertama, dataset yang digunakan berjumlah relatif kecil (178 sampel) dan hanya mencakup tiga kultivar dari satu wilayah di Italia, sehingga generalisasi model terhadap variasi kultivar dan wilayah produksi wine yang jauh lebih luas secara global masih perlu diuji lebih lanjut. Kedua, akurasi sempurna (100%) yang dicapai model Random Forest pada data uji berukuran kecil (36 sampel) perlu diinterpretasikan secara hati-hati, karena berpotensi mengindikasikan pemisahan kelas yang memang sangat baik pada dataset ini, namun validasi lebih lanjut menggunakan skema cross-validation dan dataset independen berukuran lebih besar tetap diperlukan sebelum kesimpulan generalisasi ditarik secara luas.

Ketiga, penelitian ini berfokus pada klasifikasi kultivar berdasarkan komposisi kimia, belum mencakup deteksi adulterasi (pencampuran) wine yang secara teknis merupakan tantangan berbeda dan umumnya memerlukan data komposisi kimia yang lebih granular atau data spektroskopi tambahan. Keempat, selisih antara Random Forest dan empat model lain hanya setara dengan satu sampel uji, sementara uji signifikansi prediksi berpasangan seperti McNemar belum dilakukan; karena itu, hasil ini belum mendukung klaim bahwa Random Forest secara statistik lebih unggul. Kelima, feature importance belum dikonfirmasi melalui permutation importance atau SHAP, sehingga stabilitas peringkat fitur terhadap teknik interpretasi lain belum dapat dipastikan.

Ketersediaan Data Dan Pertimbangan Etik

Dataset yang digunakan dalam penelitian ini merupakan dataset publik yang tersedia secara terbuka melalui UCI Machine Learning Repository dan terintegrasi langsung pada pustaka scikit-learn versi 1.8 yang digunakan pada penelitian ini, tanpa memerlukan pengunduhan tambahan dari sumber eksternal. Dataset tidak memuat informasi identitas produsen atau merek dagang tertentu, sehingga tidak menimbulkan isu privasi maupun kerahasiaan bisnis. Kode program dan konfigurasi eksperimen yang digunakan untuk menghasilkan seluruh angka dan gambar pada penelitian ini disusun agar dapat direproduksi ulang (reproducible), dengan random state yang tetap pada seluruh tahapan pembagian data dan pelatihan model.

KESIMPULAN

Penelitian ini berhasil merancang dan membandingkan lima model machine learning, termasuk pendekatan deep learning berbasis Artificial Neural Network, untuk klasifikasi kultivar anggur berdasarkan 13 parameter komposisi kimia. Model Random Forest mencapai akurasi sempurna 100% pada data uji, diikuti oleh Deep Neural Network, Logistic Regression, Support Vector Machine, dan K-Nearest Neighbors yang masing-masing mencapai akurasi 97,22%. Analisis feature importance mengidentifikasi intensitas warna, kadar flavonoid, dan kadar prolin sebagai parameter kimia paling berpengaruh dalam membedakan kultivar anggur.

Hasil ini menunjukkan bahwa data komposisi kimia standar hasil uji laboratorium konvensional, tanpa memerlukan instrumen spektroskopi khusus, sudah mampu memberikan kinerja klasifikasi yang sangat kompetitif dibandingkan pendekatan berbasis spektroskopi pada literatur sejenis. Dari sisi implikasi bisnis, temuan ini menegaskan potensi adopsi sistem verifikasi berbasis machine learning sebagai lapisan tambahan autentikasi produk dalam rantai pasok industri wine premium, guna melindungi produsen dari kerugian akibat pemalsuan, menjaga kepercayaan konsumen, dan mempertahankan integritas harga pasar pada penelitian selanjutnya.

DAFTAR PUSTAKA

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

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

Brooks, C., Parr, L., Smith, J. M., Buchanan, D., Snioch, D., & Hebishy, E. (2021). A review of food fraud and food authenticity across the food supply chain, with an examination of the impact of the COVID-19 pandemic and Brexit on food industry. *Food Control*, *130*, 108171. https://doi.org/10.1016/j.foodcont.2021.108171

Campmajó, G., Saurina, J., & Núñez, O. (2022). FIA–HRMS fingerprinting subjected to chemometrics as a valuable tool to address food classification and authentication: Application to red wine, paprika, and vegetable oil samples. *Food Chemistry*, *373*, 131491. https://doi.org/10.1016/j.foodchem.2021.131491

Cichna-Markl, M., & Mafra, I. (2023). Techniques for Food Authentication: Trends and Emerging Approaches. *Foods*, *12*(6), 1134. https://doi.org/10.3390/foods12061134

Giannakas, K., & Yiannaka, A. (2023). Food Fraud: Causes, Consequences, and Deterrence Strategies. *Annual Review of Resource Economics*, *15*(1), 85–104. https://doi.org/10.1146/annurev-resource-101422-013027

González-Domínguez, R., Sayago, A., & Fernández-Recamales, Á. (2022). An Overview on the Application of Chemometrics Tools in Food Authenticity and Traceability. *Foods*, *11*(23), 3940. https://doi.org/10.3390/foods11233940

Kalivas, J. H., & Lemos, T. (2023). Automatic food and beverage authentication and adulteration detection by classification hybrid fusion. *Journal of Chemometrics*, *37*(3). https://doi.org/10.1002/cem.3371

Kritzinger, A., Mouthaan, R., Bruce, G. D., Wilkes, E., & Dholakia, K. (2026). *Through the bottle authentication of red wine using near-IR fluorescence spectroscopy*. https://arxiv.org/pdf/2601.08298

Kuang, Y., Tolhurst, T. N., & Alston, J. M. (2024). Unraveling the economic impact of wine counterfeiting: An analysis of the Sassicaia 2015 scandal and its consequences. *Journal of Wine Economics*, *19*(4), 365–374. https://doi.org/10.1017/jwe.2024.26

Lord, N., Flores Elizondo, C., Davies, J., & Spencer, J. (2022). Fault lines of food fraud: key issues in research and policy. *Crime, Law and Social Change*, *78*(5), 577–598. https://doi.org/10.1007/s10611-021-09983-w

Michele Forina, Riccardo Leardi, Armanino C, S. L. (1998). *Parvus : An Extendable Package of Programs for Data Exploration,Classification and Correlation (Diskettes/Ma*. https://www.researchgate.net/publication/263587482_PARVUS_An_Extendable_Package_of_Programs_for_Data_Exploration

Pan, T., Li, J., Fu, C., Chang, N., & Chen, J. (2022). Visible and Near-Infrared Spectroscopy Combined With Bayes Classifier Based on Wavelength Model Optimization Applied to Wine Multibrand Identification. *Frontiers in Nutrition*, *9*. https://doi.org/10.3389/fnut.2022.796463

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos,

A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2018). *Scikit-learn: Machine Learning in Python*. http://arxiv.org/abs/1201.0490

Ranaweera, R. K. R., Capone, D. L., Bastian, S. E. P., Cozzolino, D., & Jeffery, D. W. (2021). A Review of Wine Authentication Using Spectroscopic Approaches in Combination with Chemometrics. *Molecules*, *26*(14), 4334. https://doi.org/10.3390/molecules26144334

Ranaweera, R. K. R., Gilmore, A. M., Capone, D. L., Bastian, S. E. P., & Jeffery, D. W. (2021a). Authentication of the geographical origin of Australian Cabernet Sauvignon wines using spectrofluorometric and multi-element analyses with multivariate statistical modelling. *Food Chemistry*, *335*, 127592. https://doi.org/10.1016/j.foodchem.2020.127592

Ranaweera, R. K. R., Gilmore, A. M., Capone, D. L., Bastian, S. E. P., & Jeffery, D. W. (2021b). Spectrofluorometric analysis combined with machine learning for geographical and varietal authentication, and prediction of phenolic compound concentrations in red wine. *Food Chemistry*, *361*, 130149. https://doi.org/10.1016/j.foodchem.2021.130149

Solovyev, P. A., Fauhl-Hassek, C., Riedl, J., Esslinger, S., Bontempo, L., & Camin, F. (2021). NMR spectroscopy in wine authentication: An official control perspective. *Comprehensive Reviews in Food Science and Food Safety*, *20*(2), 2040–2062. https://doi.org/10.1111/1541-4337.12700

Ta, N., Wei, H., Hu, Z., Cao, X., & Xiao, M. (2023). Wine component tracing method based on near infrared spectrum fusion machine learning. *Frontiers in Sustainable Food Systems*, *7*. https://doi.org/10.3389/fsufs.2023.1197508

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLOS ONE*, *14*(11), e0224365. https://doi.org/10.1371/journal.pone.0224365

Van Ruth, S. M., & Nillesen, O. (2021). Which Company Characteristics Make a Food Business at Risk for Food Fraud? *Foods*, *10*(4), 842. https://doi.org/10.3390/foods10040842

Zhang, Z., Li, Y., Zhao, S., Qie, M., Bai, L., Gao, Z., Liang, K., & Zhao, Y. (2024). Rapid analysis technologies with chemometrics for food authenticity field: A review. *Current Research in Food Science*, *8*, 100676. https://doi.org/10.1016/j.crfs.2024.100676