# PERBANDINGAN ALGORITMA LIGHTGBM DAN ANN UNTUK MENENTUKAN KUALITAS ANGGUR MERAH

**Ahmad Rafi R, Ardiansyah Wahyu W, Esa Pillardien A.R,**
**Sofyano Fadilah R, Agung Mustika Rizki**
Prodi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jawa Timur
Jl. Rungkut Madya, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294
*22081010052@student.upnjatim.ac.id*

**ABSTRAK**

Penentuan kualitas anggur merah merupakan proses penting dalam industri minuman fermentasi untuk memastikan produk berkualitas tinggi. Penelitian ini bertujuan untuk mengevaluasi penerapan algoritma LightGBM *(Light Gradient Boosting Machine)* dan *Artificial Neural Networks* (ANN) dalam memprediksi kualitas anggur merah berdasarkan karakteristik kimia seperti tingkat keasaman, kadar alkohol, dan kandungan sulfur. Dataset yang digunakan mencakup 1599 sampel anggur merah dengan 11 fitur kimia sebagai variabel input dan label "*quality*" sebagai variabel target. Hasil penelitian menunjukkan bahwa algoritma LightGBM memberikan akurasi yang lebih baik dibandingkan ANN dalam menangani dataset ini, dengan akurasi mencapai 80% dan waktu komputasi yang lebih efisien. Sementara itu, ANN menunjukkan potensi untuk mengidentifikasi pola yang lebih kompleks tetapi memerlukan waktu pelatihan yang lebih lama. Penelitian ini menggarisbawahi pentingnya pemilihan algoritma yang tepat untuk klasifikasi kualitas anggur merah, sekaligus memberikan kontribusi pada pengembangan sistem prediksi berbasis *machine learning* di industri *wine.*

***Kata kunci :*** *Red wine, LightGBM, ANN, Machine learning.*

## 1. PENDAHULUAN

Anggur adalah buah populer yang dapat dikonsumsi langsung atau diolah menjadi berbagai produk, termasuk wine. Wine merupakan minuman beralkohol hasil fermentasi sari buah anggur, dengan durasi fermentasi bervariasi tergantung jenis dan metode pembuatannya. Jenis-jenis wine meliputi *Rose Wine, Sweet Wine, Sparkling Wine, Red Wine, White Wine, dan Fortified Wine* [1].

*Red wine* atau anggur merah adalah wine yang dibuat dari anggur merah atau hitam. Warna merahnya dihasilkan dari pencelupan kulit dan biji anggur ke dalam sari buah yang sudah diperas, kemudian difermentasi dalam periode tertentu [1].

Selain sebagai minuman, *red wine*sering dikaitkan dengan manfaat kesehatan jika dikonsumsi secara moderat. Kandungan antioksidan seperti resveratrol pada red wine dilaporkan membantu meningkatkan kesehatan jantung dan mengurangi risiko penyakit kronis. Namun, kualitas red wine dipengaruhi oleh komposisi kimia, teknik fermentasi, dan metode penyimpanan.

Penentuan kualitas wine menjadi fokus utama industri wine dalam beberapa dekade terakhir untuk memastikan produk memenuhi standar tinggi dan preferensi konsumen. Teknologi modern, termasuk machine learning, digunakan untuk mengklasifikasikan kualitas wine berdasarkan data kimia yang terukur.

Algoritma machine learning seperti Light Gradient Boosting Machine (LightGBM) menawarkan keunggulan dalam prediksi kualitas wine. LightGBM mampu menangani dataset besar dan fitur kompleks dengan akurasi tinggi serta efisiensi waktu komputasi yang baik [2].

Selain itu, Artificial Neural Network (ANN) juga sering digunakan dalam analisis data wine. Dengan kemampuannya menangani pola non-linear dan hubungan kompleks antar variabel, ANN menjadi alat efektif dalam prediksi kualitas wine. Algoritma ini bekerja melalui jaringan lapisan input, lapisan tersembunyi, dan lapisan output yang saling terhubung dan disesuaikan selama proses pelatihan.

Dalam prediksi kualitas anggur, ANN sering digunakan untuk menganalisis hubungan antara sifat fisikokimia anggur dan kualitasnya. ANN mampu mengungkap pola tersembunyi yang sulit ditemukan oleh metode tradisional. Namun, ANN memerlukan waktu pelatihan lebih lama dan sumber daya komputasi lebih besar dibandingkan algoritma seperti Support Vector Machine (SVM) atau Decision Tree (DT) [3].

Penelitian sebelumnya menunjukkan bahwa algoritma machine learning, termasuk ANN, efektif menganalisis properti fisikokimia anggur merah dan memprediksi kualitasnya dengan akurat. ANN mengevaluasi fitur penting seperti alkohol, keasaman volatil, dan sulfat, sehingga memberikan wawasan tentang faktor utama yang memengaruhi kualitas anggur [3].

Dataset yang digunakan adalah "Red Wine Quality," yang terdiri dari 1.599 observasi dan 12 atribut, termasuk variabel target kualitas anggur. Penelitian ini bertujuan untuk menentukan algoritma paling efisien dan akurat sekaligus berkontribusi pada pengembangan sistem klasifikasi berbasis machine learning yang dapat diterapkan di industri wine.

Penelitian ini diharapkan mampu meningkatkan efisiensi dan objektivitas dalam penilaian kualitas anggur merah, serta memberikan wawasan baru tentang penerapan machine learning di industri minuman fermentasi.

## 2. TINJAUAN PUSTAKA

### 2.1. Red Wine (Anggur Merah)

Anggur merupakan buah populer yang sering diolah menjadi produk fermentasi seperti wine. Wine adalah minuman beralkohol hasil fermentasi jus anggur, dengan jenis seperti Red Wine, White Wine, Rose Wine, Sweet Wine, Sparkling Wine, dan Fortified Wine. Red wine dibuat dari fermentasi anggur merah atau hitam, dengan warna merah diperoleh dari kulit dan biji anggur yang direndam selama fermentasi. Wine menjadi bagian budaya konsumsi dan berperan penting di negara beriklim dingin sebagai penghangat tubuh, serta dihidangkan dalam berbagai acara perayaan [1], [4].

Kualitas red wine dipengaruhi oleh komposisi kimia seperti alkohol, keasaman, gula residu, dan senyawa bioaktif. Penilaian kualitas secara manual membutuhkan keahlian khusus dan sulit untuk skala besar. Pendekatan berbasis teknologi seperti algoritma machine learning kini banyak diterapkan untuk mempermudah proses penilaian [5].

Algoritma seperti Random Forest, Decision Tree, Light Gradient Boosting Machine (LGBM), dan Artificial Neural Network (ANN) digunakan dalam klasifikasi kualitas red wine. Random Forest terbukti memiliki performa terbaik dalam beberapa kasus, sementara LGBM unggul dalam efisiensi waktu pelatihan dan akurasi pada dataset besar. Di sisi lain, ANN mampu mempelajari pola kompleks secara mendalam, menjadikannya pilihan yang kuat untuk data dengan hubungan non-linear [1], [5].

Sebagai produk konsumsi, red wine juga menjadi objek penelitian, termasuk di bidang kedokteran gigi. Red wine diketahui dapat mengubah warna bahan restorasi gigi seperti resin komposit karena kandungan tanin, antosianin, dan sifat keasamannya yang mempercepat degradasi resin [1], [4].

Dengan algoritma data mining, kualitas red wine dapat dievaluasi lebih cepat berdasarkan atribut penting seperti alkohol, keasaman, dan senyawa bioaktif. Teknologi ini membantu produsen meningkatkan standar produk di pasar, mendukung efisiensi produksi, dan memastikan wine berkualitas tinggi [5].

### 2.2. LightGBM

LightGBM (Light Gradient Boosting Machine) adalah algoritma berbasis Gradient Boosting Decision Tree (GBDT) yang dikembangkan oleh Microsoft Research Asia [6]. Algoritma ini unggul dalam kecepatan pelatihan, efisiensi memori, dan mendukung data besar melalui pembelajaran paralel serta GPU [7].

Pendekatan unik LightGBM terletak pada metode leaf-wise tree growth, yang memilih cabang dengan pengurangan error terbesar. Metode ini meningkatkan akurasi dan efisiensi dibandingkan pendekatan level-wise, meskipun berisiko overfitting jika tidak diatur dengan baik [6].

Penelitian Rizky et al. [6] menunjukkan LightGBM unggul dalam menangani dataset dengan ketidakseimbangan kelas. Algoritma ini mencatat spesifisitas tinggi dengan rata-rata 80,41%, menjadikannya andal untuk aplikasi modern seperti analisis data medis dan perilaku pengguna.

LightGBM fleksibel untuk berbagai tugas seperti klasifikasi, regresi, dan perankingan, serta mudah diimplementasikan melalui platform seperti Python. Dengan keunggulan teknis dan kemudahan penggunaan, algoritma ini menjadi solusi andal di berbagai industri [7].

### 2.3. Artifical Neural Network (ANN)

Artificial Neural Network (ANN) adalah algoritma yang terinspirasi dari struktur otak manusia. ANN memodelkan hubungan non-linear antara input dan output melalui lapisan seperti input layer, hidden layer, dan output layer, dengan fungsi aktivasi seperti sigmoid, ReLU, dan softmax. Metode ini sering digunakan dalam klasifikasi dan prediksi karena kemampuannya mempelajari pola kompleks tanpa pengkodean manual [8].

Dalam analisis kualitas anggur merah, ANN menggunakan dataset dari UCI Machine Learning Repository yang mencakup 11 atribut fisikokimia, seperti alkohol dan pH. ANN menunjukkan performa kompetitif dibandingkan Ridge Regression dan Gradient Boosting Regressor, meskipun Gradient Boosting Regressor memiliki MSE (Mean Squared Error) dan MAPE (Mean Absolute Percentage Error) yang lebih rendah [9]. Teknik seleksi fitur, yang mengidentifikasi atribut penting seperti alkohol dan pH, juga membantu meningkatkan akurasi prediksi [9], [10].

Teknologi hyperspectral imaging berbasis UAV menunjukkan potensi besar dalam klasifikasi varietas anggur. Dalam studi López et al., Convolutional Neural Networks (CNN) mencapai akurasi hingga 99% dalam mengklasifikasi 17 varietas anggur merah dan putih, dengan memanfaatkan fitur spektral dan spasial melalui lapisan perhatian spasial dan blok Inception [10].

Meski ANN memiliki keterbatasan seperti kebutuhan data besar dan waktu pelatihan signifikan, teknik seperti normalisasi data dan tuning hyperparameter dapat meningkatkan performanya. Pendekatan ini tetap menjanjikan untuk analisis kualitas anggur karena fleksibilitasnya dalam menangani hubungan antar-parameter yang kompleks [9], [10].

### 2.4. Confusion Matrix

Confusion matrix merupakan representasi matriks dari hasil klasifikasi model yang digunakan untuk mengevaluasi performa suatu algoritma klasifikasi. Matriks ini memetakan hasil prediksi terhadap nilai aktual dari data uji dengan format sebagai berikut:

Tabel 1. Data Confusion Matrix

| | **Prediksi Positif** | **Prediksi Negatif** |
|---|---|---|
| **Aktual Positif** | True Positif (TP) | FalseNegative (FN) |
| **Aktual Negatif** | True Negative (TN) | False Positive(FP) |

Confusion matrix merupakan alat evaluasi yang digunakan untuk memetakan hasil prediksi terhadap nilai aktual pada data uji dalam model klasifikasi. Matriks ini terdiri dari empat komponen utama, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). True Positive (TP) adalah jumlah prediksi benar untuk data positif, sedangkan True Negative (TN) adalah jumlah prediksi benar untuk data negatif. False Positive (FP) terjadi ketika data negatif salah diprediksi sebagai positif (*Type I error*), dan False Negative (FN) terjadi ketika data positif salah diprediksi sebagai negatif (*Type II error*).

Persamaan evaluasi kerja dapat dihitung sebagai berikut:

- Akurasi: Mengukur seberapa sering model

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN}$$

- Presisi: Mengukur seberapa tepat prediksi positif yang dihasilkan oleh model:

$$\text{Presisi} = \frac{TP}{TP + FP}$$

- Recall: Mengukur seberapa baik model mendeteksi data positif:

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

- F1-Score: Kombinasi rata-rata harmonik antara presisi dan recall:

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

- Metrik-metrik ini digunakan untuk mengevaluasi performa algoritma seperti LightGBM dan ANN, terutama untuk memahami kemampuan model dalam memprediksi kualitas anggur merah pada dataset yang digunakan.

## 3. METODE PENELITIAN

### 3.1. Pengumpulan Data

Tahap awal dalam penelitian ini adalah pengumpulan data yang akan digunakan untuk membangun model prediksi kualitas anggur merah. Dataset yang digunakan merupakan kumpulan data wine quality yang berisi 1.599 sampel anggur merah. Setiap sampel memiliki sebelas parameter fisikokimia seperti keasaman tetap, keasaman volatil, kandungan asam sitrat, gula residu, klorida, sulfur dioksida bebas dan total, kepadatan, pH, sulfat, dan alkohol. Selain itu, terdapat satu nilai target yang merepresentasikan kualitas anggur dengan skala diskrit, biasanya berkisar dari 0 hingga 10.

Data ini merupakan data sekunder yang telah tersedia dalam format terstruktur. Sumber data ini relevan dan cukup representatif untuk menganalisis hubungan antara fitur-fitur fisikokimia dan kualitas anggur. Data ini penting karena memberikan informasi berbasis sains yang memungkinkan model machine learning mempelajari pola-pola yang ada di dalamnya. Dengan demikian, hasil prediksi yang dihasilkan nantinya dapat digunakan untuk mendukung keputusan dalam proses produksi anggur.

Selanjutnya, data ini akan diperiksa untuk memahami distribusinya. Tahap ini melibatkan analisis deskriptif awal, seperti menghitung statistik dasar (mean, median, standar deviasi) dan visualisasi distribusi setiap fitur. Analisis awal ini membantu dalam mengenali karakteristik data, seperti apakah ada ketidakseimbangan dalam nilai target atau apakah fitur tertentu memiliki rentang nilai yang sangat bervariasi. Hasil dari analisis ini akan menjadi dasar untuk proses pra-pemrosesan data.

### 3.2. Pra-Proses Data

Setelah data terkumpul, langkah berikutnya adalah pra-pemrosesan untuk memastikan bahwa data siap digunakan dalam pelatihan model. Proses ini melibatkan beberapa langkah penting, dimulai dengan pemeriksaan nilai kosong atau missing values. Jika ada nilai yang hilang, langkah selanjutnya adalah memutuskan bagaimana cara menanganinya, apakah dengan menghapus data yang tidak lengkap atau mengisinya menggunakan metode imputasi seperti rata-rata atau median.

Setelah itu, dilakukan normalisasi pada semua fitur numerik. Hal ini penting karena dataset ini memiliki fitur dengan skala yang sangat berbeda, seperti "*residual sugar*" yang bernilai dalam satuan gram, sementara "*chlorides*" dalam nilai desimal kecil. Normalisasi dilakukan dengan cara mengubah data ke dalam rentang tertentu, misalnya 0 hingga 1, untuk meningkatkan performa model ANN yang sangat sensitif terhadap perbedaan skala pada input.

Selain itu, dataset dibagi menjadi dua subset: data pelatihan dan data pengujian. Pembagian ini dilakukan dengan rasio 70:30 untuk memastikan bahwa model memiliki data yang cukup untuk belajar dan tetap memiliki data yang independen untuk pengujian. Pemisahan data ini dilakukan secara acak, tetapi memastikan bahwa distribusi nilai target pada kedua subset tetap proporsional. Langkah ini penting untuk mendapatkan hasil evaluasi yang representatif dan realistis.

### 3.3. Arsitektur Model LightGBM

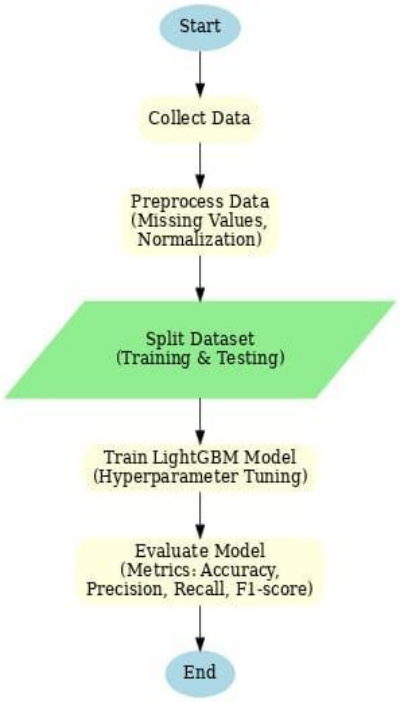
Model LightGBM (Light Gradient Boosting Machine) adalah algoritma pembelajaran mesin berbasis pohon keputusan yang dirancang untuk efisiensi dan kecepatan, terutama pada dataset berukuran besar. LightGBM menggunakan teknik gradient boosting untuk secara iteratif membangun pohon keputusan, di mana setiap iterasi memperbaiki kesalahan prediksi dari pohon sebelumnya, sehingga meningkatkan akurasi secara progresif.

Keunggulan utama LightGBM meliputi kemampuannya menangani fitur dengan banyak kategori, membagi data berdasarkan histogram, dan mendukung paralelisasi, yang membuat proses pelatihannya lebih cepat dibanding metode lain seperti XGBoost. Pada dataset *wine quality*, model LightGBM dirancang dengan parameter seperti jumlah pohon (*estimators*), kedalaman maksimum

(*max_depth*), dan *learning rate* yang dioptimalkan melalui tuning hyperparameter.

LightGBM juga memiliki fitur auto-tuning untuk menangani ketidakseimbangan data, yang penting untuk distribusi target yang tidak merata. Dengan efisiensinya, LightGBM menjadi alternatif yang ringan dan efektif untuk memprediksi kualitas anggur merah, terutama pada dataset dengan dimensi yang relatif kecil.

Parameter utama model, seperti jumlah pohon *(estimators)*, kedalaman maksimum *(max_depth)*, dan learning rate, akan ditentukan melalui eksperimen dan tuning hyperparameter. LightGBM juga memiliki fitur auto-tuning untuk menangani ketidakseimbangan data, yang penting dalam kasus nilai target dengan distribusi tidak merata. Dengan demikian, penggunaan LightGBM memberikan alternatif yang lebih ringan dan efisien dibandingkan ANN, terutama untuk pengolahan dataset dengan dimensi yang lebih rendah seperti ini.




Gambar 1. Flowchart LGBM

Proses penggunaan algoritma LightGBM dimulai dengan inisialisasi metode LightGBM. Dataset yang digunakan adalah "Red Wine Quality," yang terdiri dari 11 fitur kimia seperti kadar alkohol, keasaman, dan kandungan sulfur, serta satu label target berupa kualitas anggur. Data tersebut terlebih dahulu diproses dengan menangani nilai kosong menggunakan metode imputasi dan menormalisasi fitur agar memiliki skala yang sebanding. Selanjutnya, dataset dibagi menjadi dua bagian: 70% untuk pelatihan dan 30% untuk pengujian. Model LightGBM kemudian dilatih menggunakan parameter seperti *learning_rate*, *max_depth*, dan *n_estimators* dengan metode boosting yang meningkatkan akurasi prediksi secara iteratif. Setelah pelatihan selesai, performa model diukur menggunakan metrik seperti akurasi, precision, recall, dan F1-score untuk mengevaluasi kemampuan prediksi kualitas anggur. Pada tahap akhir, model siap digunakan untuk keperluan prediksi atau analisis lebih lanjut.

### 3.4. Arsitektur Model ANN

Model yang digunakan dalam penelitian ini adalah *Artificial Neural Network* (ANN), yang merupakan salah satu metode pembelajaran mendalam yang populer. Arsitektur ANN dirancang dengan lapisan input yang memiliki 11 neuron, masing-masing mewakili satu fitur fisikokimia dari dataset. Lapisan input ini menjadi titik awal di mana data mentah dimasukkan ke dalam jaringan untuk diproses.

Model ini memiliki beberapa lapisan tersembunyi (*hidden layers*) untuk menangkap hubungan kompleks antara fitur-fitur dalam data. Dalam penelitian ini, digunakan dua hingga tiga lapisan tersembunyi dengan jumlah neuron yang bervariasi, seperti 64 pada lapisan pertama dan 32 pada lapisan kedua. Fungsi aktivasi ReLU diterapkan pada setiap lapisan tersembunyi untuk menambahkan non-linearitas, memungkinkan model mempelajari pola yang lebih kompleks.

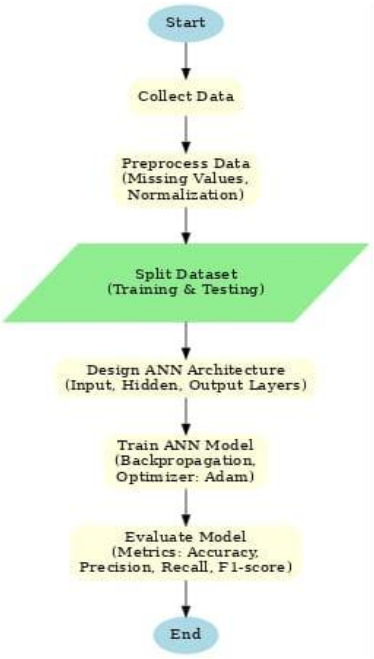
Lapisan output pada model ini memiliki satu neuron jika masalah dianggap sebagai regresi, atau beberapa neuron jika dianggap sebagai klasifikasi (misalnya, jika kualitas diperlakukan sebagai kategori). Fungsi aktivasi yang digunakan pada lapisan output adalah *sigmoid* atau *softmax*, tergantung pada pendekatan yang diambil. Desain arsitektur ini dipilih untuk memberikan keseimbangan antara kapasitas model dan kemampuannya untuk generalisasi.

### 3.5. Optimasi Model

Setelah arsitektur model dirancang, langkah berikutnya adalah mengoptimalkan model untuk meningkatkan performa. Proses optimasi dimulai dengan memilih fungsi loss yang sesuai. Jika masalah dianggap sebagai regresi, fungsi loss yang digunakan adalah *Mean Squared Error* (MSE). Sebaliknya, jika model dirancang untuk klasifikasi, digunakan *Categorical Cross-Entropy* sebagai fungsi *loss*.

Proses penerapan algoritma ANN diawali dengan pengumpulan dataset "Red Wine Quality," yang memiliki fitur dan label target serupa dengan LightGBM. Dataset ini melalui proses imputasi untuk menangani nilai kosong, diikuti dengan normalisasi fitur ke dalam rentang tertentu (misalnya 0–1) untuk meningkatkan stabilitas pelatihan. Dataset kemudian dibagi menjadi data pelatihan (70%) dan data pengujian (30%). Model ANN dirancang dengan 11 neuron di lapisan input (sesuai jumlah fitur), menggunakan fungsi aktivasi ReLU pada lapisan tersembunyi, dan fungsi softmax pada lapisan output. Pelatihan model dilakukan dengan algoritma *backpropagation*, menggunakan optimizer Adam untuk mempercepat konvergensi, dan penyesuaian hyperparameter seperti learning rate melalui eksperimen. Kinerja model diuji menggunakan akurasi, precision, recall, dan F1-score untuk

mengevaluasi hasil prediksi. Setelah proses selesai, model siap digunakan untuk analisis lebih lanjut.




Gambar 2. Flowchart ANN

### 3.6. Evaluasi Model

Setelah model dilatih, evaluasi dilakukan untuk menilai performa dan kemampuannya dalam memprediksi kualitas anggur. Evaluasi dilakukan dengan menggunakan data pengujian yang telah disisihkan sebelumnya, sehingga model diuji pada data yang belum pernah dilihat sebelumnya. Hal ini memberikan gambaran yang realistis tentang kemampuan model untuk melakukan generalisasi.

Beberapa metrik evaluasi digunakan tergantung pada jenis masalah. Jika model digunakan untuk regresi, metrik seperti *Mean Absolute Error* (MAE) atau *Root Mean Squared Error* (RMSE) digunakan untuk mengukur sejauh mana prediksi model berbeda dari nilai sebenarnya. Jika model digunakan untuk klasifikasi, akurasi dan matriks kebingungan

Hasil evaluasi divisualisasikan dalam bentuk grafik, seperti *learning curve*, untuk memahami performa pelatihan dan pengujian selama iterasi. Jika model menunjukkan tanda-tanda *overfitting* atau *underfitting*, seperti perbedaan besar antara akurasi pelatihan dan pengujian, dilakukan penyesuaian ulang pada arsitektur atau parameter model. Langkah ini memastikan bahwa model yang dihasilkan tidak hanya akurat tetapi juga dapat digunakan dalam konteks nyata.

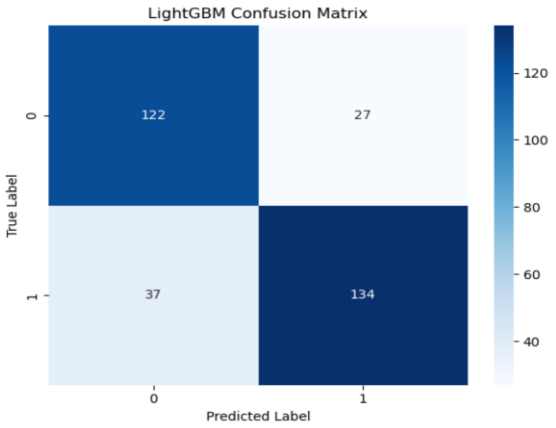
## 4. HASIL DAN PEMBAHASAN

### 4.1. Hasil Eksperimen Algoritma LightGBM

Model LightGBM menunjukkan performa yang sangat baik dalam klasifikasi kualitas anggur merah. Evaluasi pada data uji menghasilkan akurasi sebesar 82.00% dengan F1-Score 0.80, yang menunjukkan keseimbangan yang baik antara presisi dan recall.

Proses metode dalam penentuan kualitas menggunakan LightGBM melibatkan beberapa tahapan penting. Pertama, data awal yang mencakup 11 fitur fisikokimia dan satu label target diproses melalui tahap pra-pemrosesan. Pada tahap ini, data diperiksa untuk nilai kosong yang kemudian diatasi menggunakan metode imputasi rata-rata. Selanjutnya, semua fitur numerik dinormalisasi ke dalam rentang 0 hingga 1 untuk memastikan keseragaman skala dan mengoptimalkan kinerja model.

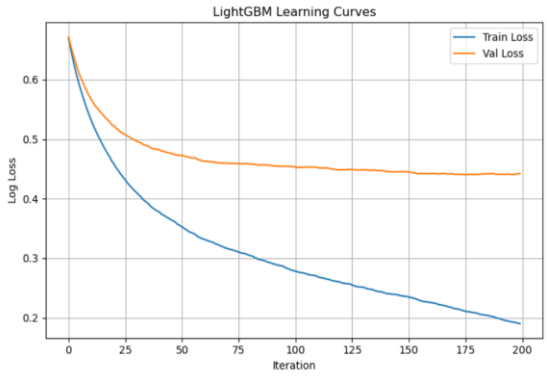
Setelah data diproses, dataset dibagi menjadi dua bagian: 70% untuk pelatihan dan 30% untuk pengujian. Model LightGBM kemudian dilatih dengan parameter seperti jumlah pohon (estimators), kedalaman maksimum (max_depth), dan learning rate. Parameter-parameter ini ditentukan melalui eksperimen tuning hyperparameter untuk memastikan performa optimal. Proses pelatihan menggunakan teknik boosting, di mana setiap iterasi bertujuan untuk memperbaiki kesalahan prediksi pada iterasi sebelumnya.




Gambar 3. Confusion Matrix Model dengan LightGBM

Berdasarkan *confusion matrix* pada Gambar 1, dapat dianalisis bahwa dari 159 sampel anggur berkualitas rendah, 122 sampel (76.73%) berhasil diklasifikasikan dengan benar sebagai kualitas rendah *(True Negative)*, dan hanya 37 sampel (23.27%) yang salah diklasifikasi sebagai kualitas tinggi (*False Negative*). Sementara itu, dari 161 sampel anggur berkualitas tinggi, 134 sampel (83.23%) berhasil diklasifikasikan dengan benar sebagai kualitas tinggi *(True Positive)*, dan 27 sampel (16.77%) salah diklasifikasi sebagai kualitas rendah *(False Positive)*.

Hasil ini menunjukkan tingkat kesalahan yang cukup rendah untuk kedua kelas, terutama untuk klasifikasi anggur berkualitas tinggi yang memiliki error rate sebesar 16.77%. Tingkat *error* ini dapat lebih ditekan dengan mengoptimalkan *hyperparameter* model atau meningkatkan representasi fitur dataset.



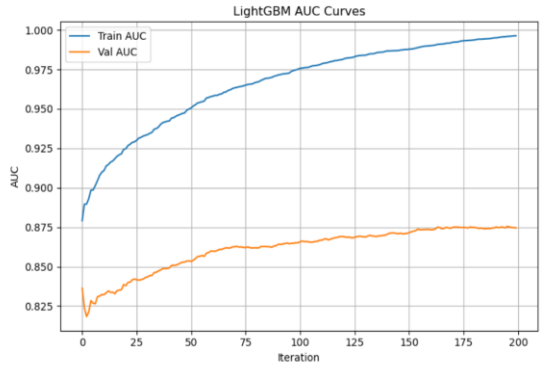

Gambar 4. Kurva Loss Model Algoritma LightGBM

Model LightGBM menunjukkan performa pelatihan yang sangat stabil dalam klasifikasi kualitas anggur merah. Berdasarkan *learning curve* pada Gambar 2, dapat diamati bahwa nilai *log loss* pada data pelatihan (*train loss*) terus menurun secara signifikan hingga iterasi ke-200, sedangkan nilai *log loss* pada data validasi (*val loss*) stabil setelah iterasi ke-50.

Dari 200 iterasi, terlihat bahwa pada tahap awal (0-50 iterasi), *train loss* menurun tajam dari nilai awal di atas 0.6 hingga mendekati 0.4. Setelah itu, laju penurunan *train loss* menjadi lebih lambat hingga mencapai nilai akhir mendekati 0.2 pada iterasi ke-200. Hal ini menunjukkan bahwa model mampu belajar secara efektif dari data pelatihan tanpa mengalami stagnasi.

Sementara itu, *val loss* menunjukkan pola penurunan yang serupa pada tahap awal (0-50 iterasi), dari nilai awal di atas 0.6 hingga stabil pada nilai sekitar 0.4. Stabilitas *val loss* setelah iterasi ke-50 menunjukkan bahwa model mencapai keseimbangan antara pelatihan dan validasi, tanpa indikasi *overfitting* meskipun *train loss* terus menurun.

Hasil ini menunjukkan kemampuan LightGBM untuk menggeneralisasi dengan baik terhadap data validasi, sekaligus memanfaatkan informasi pada data pelatihan secara optimal. Performa ini mencerminkan potensi model yang kuat untuk tugas klasifikasi kualitas anggur merah dengan risiko *overfitting* yang rendah.



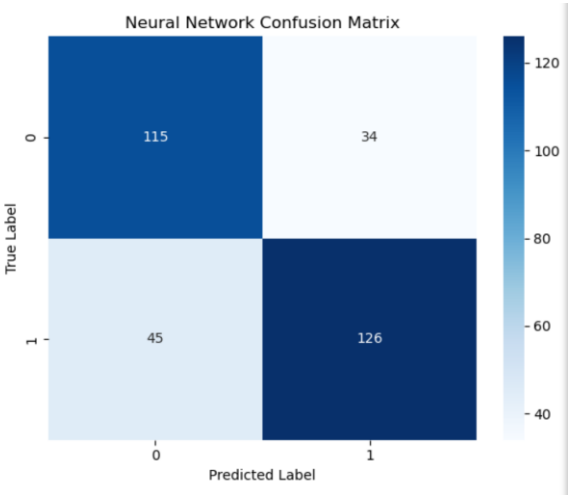

Gambar 5. Kurva Akurasi Model LightGBM

Model LightGBM menunjukkan performa yang memuaskan dalam klasifikasi kualitas anggur merah, dengan akurasi sebesar 82.00% dan F1-Score 0.8000 pada data uji. Berdasarkan Gambar 3, kurva AUC untuk data pelatihan meningkat secara konsisten hingga mendekati 1.0, sementara kurva validasi stabil di sekitar 0.87, mencerminkan kemampuan generalisasi yang baik. Dari analisis confusion matrix, model berhasil mengklasifikasikan 76.73% sampel anggur berkualitas rendah dengan benar sebagai kualitas rendah (*True Negative*) dan 83.23% sampel berkualitas tinggi sebagai kualitas tinggi (*True Positive*).

Meski begitu, error rate pada kelas berkualitas tinggi (16.77%) dan rendah (23.27%) menunjukkan adanya ruang untuk perbaikan, seperti optimalisasi *hyperparameter* atau peningkatan representasi data, agar model dapat mencapai performa yang lebih optimal.

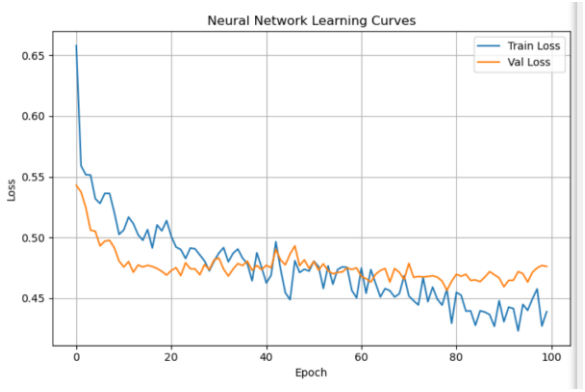
### 4.2. Hasil Eksperimen Algoritma ANN

*Confusion matrix* pada model *neural network* menunjukkan distribusi prediksi model terhadap data uji. Sebanyak 115 sampel kelas 0 berhasil diprediksi dengan benar sebagai kelas 0 (*True Negative*), sementara 126 sampel kelas 1 juga diprediksi dengan benar sebagai kelas 1 (*True Positive*). Namun, terdapat 34 sampel kelas 0 yang salah diprediksi sebagai kelas 1 (*False Positive*), serta 45 sampel kelas 1 yang salah diklasifikasikan sebagai kelas 0 (*False Negative).*



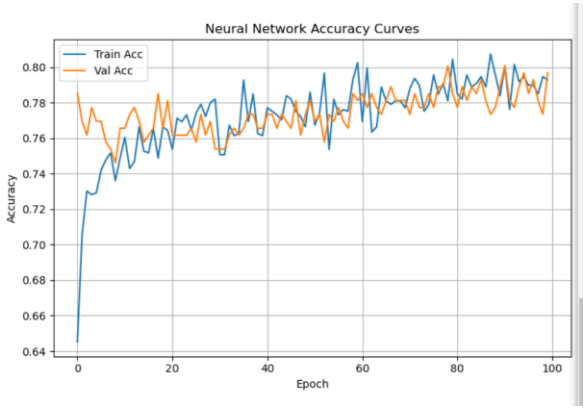

Gambar 6. Confusion Matrix Model dengan ANN

Distribusi hasil ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang cukup baik, tetapi masih terdapat ruang untuk perbaikan, terutama dalam mengurangi kesalahan prediksi pada kedua kelas. Kesalahan *False Negative* yang relatif tinggi (45 kasus) menunjukkan bahwa model terkadang gagal mengenali sampel kelas 1, yang bisa berpengaruh pada performa jika aplikasi model ini membutuhkan deteksi yang presisi untuk kelas tersebut. Pelatihan lebih lanjut dengan pengaturan *hyperparameter* atau penyeimbangan dataset dapat menjadi langkah untuk meningkatkan akurasi model.




Gambar 7. Kurva Loss Model dengan ANN

Kurva *loss* menunjukkan perubahan nilai *loss* selama pelatihan model. Terlihat bahwa training *loss* menurun secara konsisten dari nilai awal sekitar 0.65 hingga di bawah 0.45 pada epoch terakhir. Hal ini mengindikasikan bahwa model berhasil belajar untuk meminimalkan kesalahan pada data pelatihan. Sementara itu, validation loss juga menunjukkan tren penurunan, meskipun fluktuatif, dengan nilai yang mendekati 0.45 pada akhir pelatihan. Pola ini mengindikasikan bahwa model secara umum mampu generalisasi dengan baik, meskipun adanya fluktuasi pada validation loss menunjukkan adanya sedikit inkonsistensi dalam proses validasi.




Gambar 8. Kurva Akurasi Model dengan ANN

Kurva akurasi menunjukkan peningkatan yang signifikan pada training accuracy dari nilai awal sekitar 0.64 hingga mendekati 0.80 pada epoch ke-100. *Validation accuracy* juga mengikuti tren yang serupa dengan nilai akhir yang mendekati akurasi pelatihan. Hal ini menunjukkan bahwa model tidak hanya mampu belajar dari data pelatihan tetapi juga mempertahankan performa yang baik pada data validasi. Beberapa fluktuasi kecil pada *validation accuracy* menunjukkan variasi dalam kinerja model terhadap batch validasi yang berbeda, tetapi secara keseluruhan model menunjukkan stabilitas yang baik.

### 4.3. Analisis Perbandingan Kedua Algoritma

Tabel di bawah ini menyajikan perbandingan hasil evaluasi model menggunakan algoritma *Artificial Neural Network* (ANN) dan LightGBM pada data testing:

Tabel 2. Analisis Perbandingan

| Metrik | ANN | LGBM |
|---|---|---|
| Akurasi | 75.31% | 80.00% |
| *Precision* (Kelas 0) | 72% | 77% |
| *Precision* (Kelas 1) | 79% | 83% |
| *F1-Score* (Rata-rata) | 74% | 80% |
| *Macro Average* | 75% | 80% |
| *Weighted Average* | 75% | 80% |

Dari tabel tersebut, terlihat bahwa algoritma LightGBM memiliki performa lebih baik dibandingkan dengan ANN pada semua metrik evaluasi. LightGBM mencapai akurasi sebesar 80.00%, lebih tinggi dibandingkan ANN yang hanya 75.31%. Selain itu, metrik rata-rata seperti precision, recall, dan *F1-score* menunjukkan keunggulan serupa untuk LightGBM.

Hal ini menunjukkan bahwa LightGBM lebih efektif dalam menangani data testing untuk kasus ini dibandingkan ANN. Keunggulan LightGBM dapat disebabkan oleh kemampuannya dalam menangani data dengan fitur yang kompleks serta penggunaan boosting yang meningkatkan performa secara iteratif.

## 5. KESIMPULAN DAN SARAN

Penelitian ini membandingkan algoritma LightGBM dan Artificial Neural Networks (ANN) dalam memprediksi kualitas anggur merah berdasarkan parameter fisikokimia dari dataset "Red Wine Quality". Hasil pengujian menunjukkan bahwa algoritma LightGBM memberikan performa terbaik dengan akurasi sebesar 82% dan F1-Score 0,80, serta mampu mencapai keseimbangan yang baik antara presisi dan recall pada kedua kelas kualitas anggur. LightGBM juga memiliki keunggulan dalam efisiensi waktu pelatihan, dengan proses pelatihan yang stabil dan risiko overfitting yang rendah. Di sisi lain, algoritma ANN menunjukkan akurasi sebesar 75,31% dan F1-Score 0,74, dengan kemampuan mempelajari pola data yang lebih kompleks, meskipun memerlukan waktu pelatihan yang lebih lama dan mengalami fluktuasi pada validation loss. Analisis perbandingan menunjukkan bahwa LightGBM lebih efektif dalam menangani dataset dengan fitur kompleks melalui pendekatan boosting, sementara ANN memiliki potensi yang dapat ditingkatkan dengan optimasi lebih lanjut. Berdasarkan evaluasi keseluruhan, LightGBM lebih disarankan untuk implementasi dalam klasifikasi kualitas anggur merah, khususnya dalam konteks efisiensi dan akurasi.

Untuk penelitian selanjutnya, disarankan untuk melakukan optimasi lebih lanjut pada ANN, seperti tuning hyperparameter yang lebih mendalam, penggunaan teknik augmentasi data, atau penerapan regularisasi seperti dropout untuk mengurangi overfitting. Selain itu, pengayaan dataset dengan menambahkan lebih banyak sampel dan fitur kimia lainnya, seperti kandungan polifenol atau mineral,

dapat meningkatkan akurasi prediksi. Pendekatan model hybrid, yang menggabungkan keunggulan LightGBM dan ANN melalui metode ensemble atau stacked models, juga patut dieksplorasi. Di sisi praktis, penerapan model LightGBM pada sistem kontrol kualitas di pabrik wine dapat memastikan prediksi kualitas dilakukan secara real-time dan efisien, dengan potensi penerapan pada varietas anggur lainnya untuk meningkatkan fleksibilitas algoritma.

**DAFTAR PUSTAKA**

[1] R. Supriyadi, W. Gata, N. Maulidah, A. Fauzi, I. Komputer, and S. Nusa Mandiri Jalan Margonda Raya No, "Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah," vol. 13, no. 2, pp. 67–75, 2020, [Online]. Available: http://journal.stekom.ac.id/index.php/E-Bisnis■page67

[2] K. Jain, K. Kaushik, S. K. Gupta, S. Mahajan, and S. Kadry, "Machine learning-based predictive modelling for the enhancement of wine quality," *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-44111-9.

[3] S. Zaza, M. Atemkeng, and S. Hamlomo, "Wine feature importance and quality prediction: A comparative study of machine learning algorithms with unbalanced data," Oct. 2023, [Online]. Available: http://arxiv.org/abs/2310.01584

[4] N. Wayan, P. Y. Praditya, K. Kunci, and J. S. Komputer, "Prediksi Kualitas Red Wine dan White Wine Menggunakan Data Mining," *JOURNAL SHIFT VOL*, vol. 3, 2023.

[5] F. Priscilia Dinata Tobing, B. Ongki Iskandar, and R. Stefani, "Pengaruh perendaman dengan red wine terhadap perubahan warna restorasi resin komposit one shade," *Jurnal Kedokteran Gigi Terpadu*, vol. 6, no. 1, pp. 138–142, Aug. 2024, doi: 10.25105/jkgt.v6i1.20968.

[6] P. Septiana Rizky, R. Haiban Hirzi, U. Hidayaturrohman, U. Hamzanwadi Selong Jl TGKH Muhammad Zainuddin Abdul Madjid Pancor, and L. Timur, "Perbandingan Metode LightGBM dan XGBoost dalam Menangani Data dengan Kelas Tidak Seimbang," 2022. [Online]. Available: www.unipasby.ac.id

[7] K. Nisa, "Klasifikasi Penyakit Gangguan Mental dengan Algoritma LightGBM," *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, vol. 9, no. 2, pp. 1086–1094, 2024, [Online]. Available: https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

[8] G. Astray, J. X. Castillo, J. A. Ferreiro-Lage, J. F. Gálvez, and J. C. Mejuto, "Artificial neural networks: a promising tool to evaluate the authenticity of wine Redes neuronales: Una herramienta prometedora para evaluar la autenticidad del vino," *CYTA - Journal of Food*, vol. 8, no. 1, pp. 79–86, 2010, doi: 10.1080/19476330903335277.

[9] K. R. Dahal, J. N. Dahal, H. Banjade, and S. Gaire, "Prediction of Wine Quality Using Machine Learning Algorithms," *Open J Stat*, vol. 11, no. 02, pp. 278–289, 2021, doi: 10.4236/ojs.2021.112015.

[10] A. López, C. J. Ogayar, F. R. Feito, and J. J. Sousa, "Classification of grapevine varieties using UAV hyperspectral imaging," Jan. 2024, doi: 10.3390/rs16122103