

IMPLEMENTASI KUALITAS MINUMAN WINE MENGGUNAKAN METODE KLASIFIKASI

Johan Ryan Hutajulu

¹⁾ Teknik Informatika Universitas Tarumanagara
Jl. Letjen S. Parman No. 1, Jakarta Barat 11440
email : johan.535210097@stu.untar.ac.id

ABSTRACT

This study aims to evaluate wine quality using three different classification algorithms: K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), and Polynomial regression. The data used in this study comes from the Wine Quality repository, which includes various chemical attributes of wine, such as alcohol content, acidity, and sugar. Each algorithm is evaluated based on several performance metrics, including precision, recall, f1-score, and accuracy.

Keywords

classification, Wine, KNN, ANN, Polynomial, Wine

1. Pendahuluan

Anggur biasanya dikonsumsi langsung atau bisa juga diolah dari buah anggur menjadi produk seperti makanan dan minuman fermentasi yang disebut dengan wine.[1] Saat ini permintaan terhadap minuman wine semakin meningkat dan trending karena populeritasnya di kalangan Masyarakat Indonesia, mulai dari wisatawan domestik hingga wisatawan mancanegara.[2] hal ini disebabkan oleh perubahan gaya hidup dan tingginya minat meminum wine mendorong berdirinya Perusahaan produksi wine menggunakan varieties anggur lokal maupun anggur impor.[3] Proses fermentasi wine dimulai dengan melibatkan banyak proses biokimia kompleks yang terlibat. Proses ini melibatkan peran banyak enzim dari beberapa mikroorganisme seperti ragi yang berperan besar dalam pembentukan alcohol melalui proses fermentasi primer. Anggur dibuat dengan memfermentasikan atau mengubah gula yang dari anggur menjadi alcohol waktu fermentasi adalah waktu yang dibutuhkan mikroorganisme untuk menguraikan menjadi produk yang berkualitas[4][5]

Penilaian kualitas wine merupakan elemen utama dalam menjamin kualitas seperti pengukuran kepadatan, kandungan alcohol atau pH untuk melakukan pengujian biasanya dilakukan otentikasi dengan evaluasi tes fisikokimia dan sensorik.[6] Metode tradisional dalam menilai kualitas wine melibatkan evaluasi sensorik oleh

ahlinya tentu ahlinya perlu mencicipi sampel Wine. Jika bukan ahlinya maka sulit untuk menentukan kualitas wine yang difermentasi.[7]

Dengan berkembangnya era digital dan teknologi saat ini telah membantu dalam meningkatkan pengujian otomatis dengan era teknologi saat ini identifikasi dapat dilakukan dengan waktu singkat dan detail. Dengan menerapkan machine learning untuk analisis dataset yang dimiliki[8]. Machine Learning (ML) adalah metode untuk analisis data yang otomatis membuat model analitik, yang memungkinkan computer dapat belajar dari data, mengenali pola, dengan machine learning komputasi menangani situasi baru melalui pembelajaran mandiri, pengalaman, analisis dan observasi[9]

2. Sistem Persamaan Linier

2.1 Data

Pada penelitian ini pengujian dilakukan menggunakan metode klasifikasi dengan membandingkan tiga algoritma yaitu, KNN, ANN, Polynomial, untuk pengujian algoritma tersebut dilakukan dengan system. Dengan membandingkan data alcohol dan pH dalam minuman wine data yang diperoleh melalui *Machine Learning repository Wine Quality* diakses dengan link <https://archive.ics.uci.edu/dataset/186/wine+quality>.

2.2 Algoritma KNN

Algoritma KNN (K-Nearest Neighbor) merupakan algoritma yang mengklasifikasikan objek berdasarkan jarak terdekatnya dengan data latih. Algoritma ini telah diperkenalkan dalam studi observasional dan mengekstraksi sebagian besar kelas dari data perkiraan untuk menentukan kelas yang di prediksi.[10] KNN dapat mengklasifikasikan objek berdasarkan kemiripan dengan data latih terdekat yang menghitung jarak titik yang ingin diklasifikasikan dengan titik dalam dataset. [11] Proses KNN diawali dengan pengumpulan dan pelabelan data, dilanjutkan dengan pembersihan data untuk menjamin kualitas data. KNN kemudian mencari K titik data terdekat pada set pelatihan berdasarkan pengukuran jarak. [12]

2.2.1 Rumus matematika pada Algoritma K-nearest neighbor

$$K = \sqrt{(x_2 - y_1)^2 + (x_1 - y_2)^2}$$

Untuk metode matematis penggunaan metode K-Nearest Neighbor seperti yang telah disebutkan diatas, contoh berikut dapat dijelaskan.

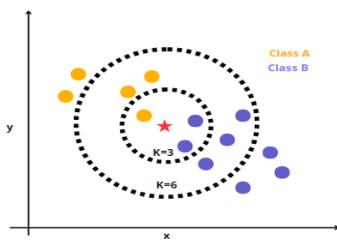
Tabel 1.keterangan Simbol KNN

Kategori	Keterangan
K	Jarak terdekat
x	Data sampel
y	Data uji

Model ini menemukan jarak terpendek antara data yang dievaluasi dan terdekatnya dengan data latih

2.2.2 Diagram algoritma KNN

Diagram dari algoritma KNN dapat Digambar dari gambar berikut [13]



Gambar 1.Diagram kerja KNN

Gambar tersebut menunjuka bagaimana KNN bekerja dalam melakukan klasifikasi:

- Poin merah adalah data baru yang ingin diklasifikasikan
- Class A dan Class B adalah data yang sudah diketahui kelasnya
- Lingkaran menunjukan radius yang berbeda k=3 dan k=6, dimana KNN akan menghitung tetangga terdekat dari data baru berdasarkan jarak

2.3 Algoritma ANN

ANN (Artificial Neural Network) adalah model komputasi yang terinspirasi oleh cara kerja otak manusia dan digunakan untuk memproses informasi dan mengenali pola. ANN terdiri dari lapisan masukan, lapisan tersembunyi dan lapisan keluaran, dimana setiap neuron pada satu lapisan dihubungkan dengan neuron pada lapisan lainnya melalui bobot yang dapat disesuaikan. ANN banyak digunakan dalam berbagai aplikasi termasuk klasifikasi, regresi, dan pengenalan pola, seperti mendeteksi varieties anggur dan Tingkat

kematangan dari spektrum daun[14]ANN bekerja dengan meniru cara otak manusia memproses informasi yang terdiri dari lapisan neuron yang saling berhubungan, di mana setiap neuron menerima masukan, memprosesnya melalui fungsi aktivasi, dan mengirimkan keluaran ke neuron berikutnya. Proses pembelajaran dilakukan melalui algoritma seperti backpropagation yang mengoptimalkan bobot koneksi berdasarkan error keluaran yang dihasilkan[15]

2.3.1 Rumus Matematika ANN

Karena privasi dan ketahanan model pelatihan ANN yang dikembangkan, data uji biasanya diestimasi. Regularisasi data pelatihan dianggap dapat meningkatkan kinerja jaringan yang dilatih. Pada penelitian ini menggunakan objek input dan output sebagai nilai pengujian data. Dengan Neuron menghitung jumlah input dan meneruskan hasilnya ke fungsi secara matematis dijelaskan seperti berikut [16]

$$x = y \sum_i^n (wja - z)$$

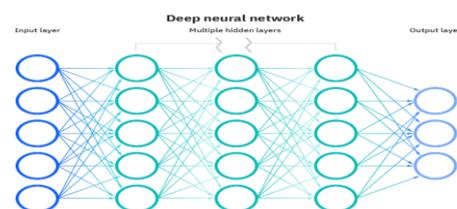
Berikut keterangan dari rumus diatas:

Tabel 2. Keterangan Simbol ANN

Kategori	Keterangan
X	Output neuron
Y	Fungsi transformasi
w	Bobot
a	Input neuron
z	Nilai batas neuron

2.3.2 Diagram Algoritma ANN

Berikut diagram dari Algoritma dari ANN sebagai berikut: [17]



Gambar 1. Struktur ANN sumber <https://dqlab.id/4-metode-deep-learning-yang-digunakan-dalam-data-science>

Adapun penjelasan dari gambar diatas adalah bahwa jenis ANN memiliki beberapa lapisan tersembunyi. Berikut penjelasan bagian-bagian dari diagram tersebut:

- Input layer adalah lapisan pertama yang menerima data dari luar. Setiap neuron pada lapisan input mewakili satu fitur dari data input.

- Multiple input layers adalah lapisan diantara input dan lapisan output
- Output layer adalah lapisan terakhir dalam jaringan, yang menghasilkan prediksi atau hasil akhir
- Connections adalah garis-garis yang menghubungkan neuron-neuron anara lapisan menunjukan aliran data atau informasi

2.4 Algoritma PolyNomial

Algoritma Polynomual adalah metode yang membuat semua prediksi polynomial berdasarkan data yang diketahui dan menganalisis prediksi tersebut [18] Polynomial terdiri dari variable dan koefisien, yang dihubungkan oleh operasi penjumlahan, pengurangan dan perkalian serta memiliki derajat tertentu. Yang umum nya digunakan untuk membangun model prediksi berdasarkan data yang diketahui dengan memilih derajat polynomial yang optimal [19]

2.4.1 Rumus Matematika Polynomial

Rumus dari dari algoritma Polynomial sebagai berikut

$$1.2 \quad y = X\beta + \epsilon$$

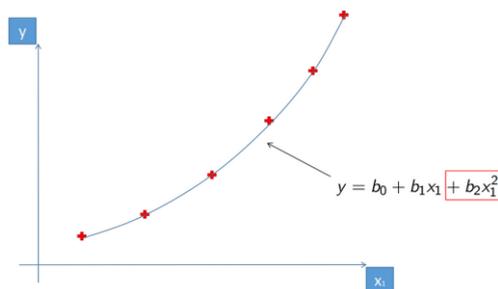
Adapun dari rumus matematika diatas dapat dijelaskan sebagai berikut:

Tabel 3. Keterangan Simbol Polynomial

Kategori	Keterangan
y	Vector respons
x	Matriks predictor
β	Vector koefisien regresi
ϵ	Vector kesalahan

2.4.2 Diagram Algoritma Polynomial

Dalam gambar diagram Algoritma Polynomial sebagai berikut: [20]



Gambar 2. Diagram Algoritma Polynomial

Gambar diatas diperoleh melalui web page diakses melalui [link https://medium.com/@farhanopen/polynomial-regression-tinjauan-dari-perspektif-machine-learning](https://medium.com/@farhanopen/polynomial-regression-tinjauan-dari-perspektif-machine-learning)

db5a470b9a92. Adapun penjelasan dari gambar adalah menampilkan sebuah grafik yang mengilustrasikan model polynomia, dimana hubungan antara variable independent (x_1) dan variable (y) bukan linear melainkan berbentuk kuadratik.

3. Hasil Percobaan

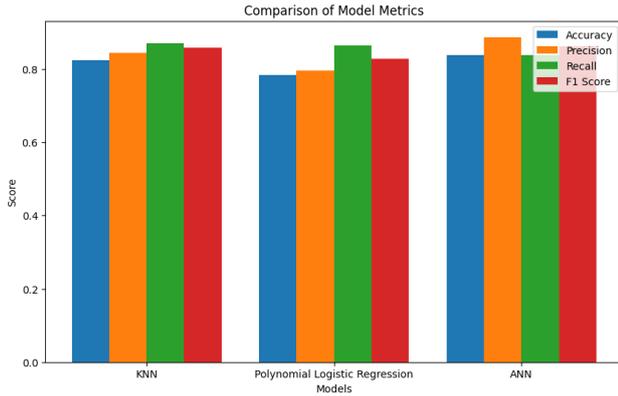
Hasil dari klasifikasi menggunakan ketiga metode algoritma tersebut menghasilkan laporan klasifikasi yang memberikan informasi detail mengenai performa model

Tabel 4. Hasil Klasifikasi

	K-Nearest Neighbor		ANN		Polynomial	
	0	1	0	1	0	1
Precision	0.79	0.84	0.77	0.89	0.76	0.80
Recall	0.75	0.77	0.84	0.84	0.66	0.86
F1-score	0.77	0.86	0.80	0.86	0.71	0.83
Accuracy	0.82		0.84		0.78	

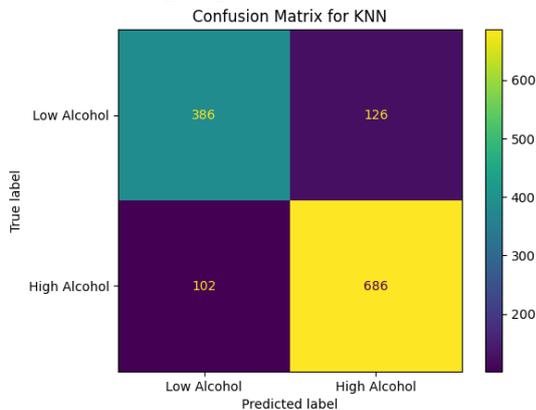
Hasil laporan klasifikasi dibagi menjadi 4 bagian yaitu:

- Precision: mengukur proporsi prediksi positif yang benar untuk kelas 0 pada model KNN adalah 0.79 dan untuk kelas 1 adalah 0.84 model ANN, precision untuk kelas 0 adalah 0.77 dan untuk kelas 1 adalah 0.89 dan untuk kelas 0.89 pada model Polynomial, precision untuk kelas 0 adalah 0.76 dan untuk kelas 1 adalah 0.80
- Recall: Mengukur Proporsi sebenarnya dari kelas positif yang berhasil diprediksi dengan benar pada model KNN, kelas 0 adalah 0.75 dan untuk kelas 1 adalah 0.77 pada model ANN, recall untuk kelas 0 adalah 0.84 dan untuk kelas 1 adalah 0.84 pada model polynomial untuk kelas 0 adalah 0.66 dan untuk kelas 1 adalah 0.86
- F1-score: f1-score merupakan rata-rata harmonis dari precision dan recall, memberikan kinerja keseluruhan model dalam memprediksi kelas positif pada model KNN, kelas 0 adalah 0.77 dan kelas 1 adalah 0.86 pada model ANN kelas 0 adalah 0.80 dan untuk kelas 1 adalah 0.86 model Polynomial kelas 0 adalah 0.71 dan untuk kelas 1 adalah 0.83.
- Accuracy: mengukur proporsi total prediksi dengan benar di semua dataset dalam model KNN memiliki akurasi sebesar 0.82, model ANN memiliki akurasi sebesar 84, polynomial memiliki akurasi keseluruhan sebesar 0.78

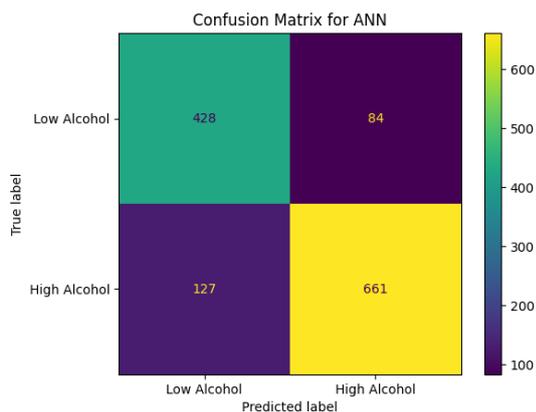


Gambar 3. Grafik perbandingan

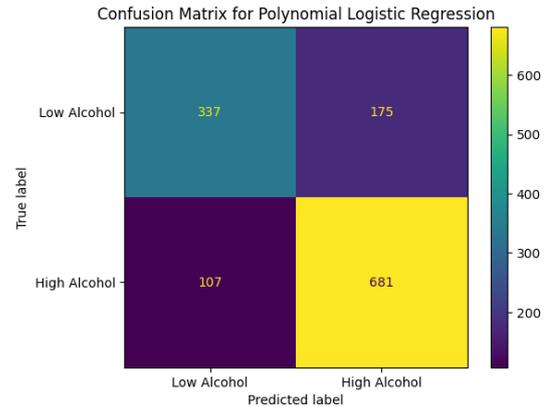
Grafik diatas menunjukkan bahwa metode ANN merupakan nilai yang optimal dibandingkan ke dua algoritma lainnya jika kinerja diukur berdasarkan keempat metrik seperti precision, recall, f1-score dan accuracy. Berikut dibawah ini merupakan confusion matrix dari ke ketiga algoritma tersebut:



Gambar 4. Confusion Matriks KNN



Gambar 5. Confussion matriks ANN



Gambar 6. Confussion Matiks Polinomial

Peroleh nilai dari ketiga metode tersebut seperti precision, recal, f1-score accuracy bahkan matriks diperoleh melalui system

4. Kesimpulan

Penelitian ini membahas tentang penggunaan metode klasifikasi untuk menilai kualitas wine dengan membandingkan tiga algoritma, yaitu K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), dan Polynomial. Berdasarkan data alkohol dan pH, hasil penelitian menunjukkan bahwa algoritma ANN memberikan performa terbaik dibandingkan dengan KNN dan Polynomial, dengan nilai precision, recall, F1-score, dan accuracy yang lebih tinggi. Algoritma KNN menunjukkan performa yang cukup baik, namun masih di bawah ANN dalam hal akurasi dan metrik lainnya. Sementara itu, algoritma Polynomial memiliki performa yang paling rendah di antara ketiga algoritma yang diuji. Secara keseluruhan, penelitian ini menyimpulkan bahwa metode ANN lebih efektif untuk klasifikasi kualitas wine dibandingkan dengan KNN dan Polynomial. Implementasi machine learning dalam penilaian kualitas wine dapat meningkatkan efisiensi dan akurasi dibandingkan dengan metode tradisional.

REFERENSI

- [1] G. Slowiński, "Dry Beans Classification Using Machine Learning," 2021. [Online]. Available: <http://ceur-ws.org>
- [2] M. Salauddin Khan *et al.*, "Comparison of multiclass classification techniques using dry bean dataset," *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 6–20, Jun. 2023, doi: 10.1016/j.ijcce.2023.01.002.
- [3] H. Vaidya, K. Swamy Karekal, K. V Prasad, and K. Kumar Swamy, "Multiclass Classification of Dry Beans using Artificial Neural Network," 2023. [Online]. Available: <https://www.researchgate.net/publication/373523354>

- [4] A. Mehta, P. Sengupta, D. Garg, H. Singh, and Y. S. Diamand, "Benchmarking the Effectiveness of Classification Algorithms and SVM Kernels for Dry Beans," Jul. 2023, [Online]. Available: <http://arxiv.org/abs/2307.07863>
- [5] E. Dönmez, S. Kılıçarslan, and A. Diker, "Classification of hazelnut varieties based on bigtransfer deep learning model," *European Food Research and Technology*, vol. 250, no. 5, pp. 1433–1442, May 2024, doi: 10.1007/s00217-024-04468-1.
- [6] M. Koklu and I. A. Ozkan, "Multiclass classification of dry beans using computer vision and machine learning techniques," *Comput Electron Agric*, vol. 174, Jul. 2020, doi: 10.1016/j.compag.2020.105507.
- [7] S. Krishnan, S. K. Aruna, K. Kanagarathinam, and E. Venugopal, "Identification of Dry Bean Varieties Based on Multiple Attributes Using CatBoost Machine Learning Algorithm," *Sci Program*, vol. 2023, 2023, doi: 10.1155/2023/2556066.
- [8] I. Wardhana, Musi Ariawijaya, Vandri Ahmad Isnaini, and Rahmi Putri Wirman, "Gradient Boosting Machine, Random Forest dan Light GBM untuk Klasifikasi Kacang Kering," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 1, 2022, doi: 10.29207/resti.v6i1.3682.
- [9] H. Hananti and K. Sari, "Perbandingan Metode Support Vector Machine (SVM) dan Artificial Neural Network (ANN) pada Klasifikasi Gizi Balita," *Seminar Nasional Official Statistics*, vol. 2021, no. 1, 2021, doi: 10.34123/semnasoffstat.v2021i1.1014.
- [10] E. Matel, F. Vahdatikhaki, S. Hosseinyalamdary, T. Evers, and H. Voordijk, "An artificial neural network approach for cost estimation of engineering services," *International Journal of Construction Management*, vol. 22, no. 7, 2022, doi: 10.1080/15623599.2019.1692400.
- [11] K. Sinha, Z. Uddin, H. I. Kawsar, S. Islam, M. J. Deen, and M. M. R. Howlader, "Analyzing chronic disease biomarkers using electrochemical sensors and artificial neural networks," 2023. doi: 10.1016/j.trac.2022.116861.
- [12] A. Almomany, W. R. Ayyad, and A. Jarrah, "Optimized implementation of an improved KNN classification algorithm using Intel FPGA platform: Covid-19 case study," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, 2022, doi: 10.1016/j.jksuci.2022.04.006.
- [13] M. Suyal and P. Goyal, "A Review on Analysis of K-Nearest Neighbor Classification Machine Learning Algorithms based on Supervised Learning," 2022. doi: 10.14445/22315381/IJETT-V70I7P205.
- [14] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 6, no. 2, 2021, doi: 10.31294/ijcit.v6i2.10438.
- [15] T. A. Q. Putri, A. Triayudi, and R. T. Aldisa, "Implementasi Algoritma Decision Tree dan Naïve Bayes Untuk Klasifikasi Sentimen Terhadap Kepuasan Pelanggan Starbucks," *Journal of Information System Research (JOSH)*, vol. 4, no. 2, pp. 641–649, Jan. 2023, doi: 10.47065/josh.v4i2.2949.
- [16] M. B. Abdul Majid, Y. M. Cani, and U. Enri, "Penerapan Algoritma K-Means dan Decision Tree Dalam Analisis Prestasi Siswa Sekolah Menengah Kejuruan," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 2, p. 355, Dec. 2022, doi: 10.30865/json.v4i2.5299.
- [17] F. Y. Pamuji and V. P. Ramadhan, "Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotherapy," *Jurnal Teknologi dan Manajemen Informatika*, vol. 7, no. 1, 2021, doi: 10.26905/jtmi.v7i1.5982.
- [18] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, "PENGUKURAN KINERJA SISTEM KUALITAS UDARA DENGAN TEKNOLOGI WSN MENGGUNAKAN CONFUSION MATRIX," *Jurnal Informatika Upgris*, vol. 6, no. 2, 2021, doi: 10.26877/jiu.v6i2.6552.
- [19] M. Hasnain, M. F. Pasha, I. Ghani, M. Imran, M. Y. Alzahrani, and R. Budiarto, "Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2994222.
- [20] G. Zeng, "On the confusion matrix in credit scoring and its analytical properties," *Commun Stat Theory Methods*, vol. 49, no. 9, 2020, doi: 10.1080/03610926.2019.1568485.