

Prediksi Kualitas Minuman Anggur Putih Berdasarkan pH dan Alkohol Yang Terkandung Didalamnya Menggunakan Metode K-Nearest Neighbor

Ahmad Muhajir Ridwan¹

Universitas Serang Raya¹
Email : 900ahmadmuhajir@gmail.com

ABSTRAKSI

Penelitian ini bertujuan untuk mengembangkan model prediksi kualitas minuman anggur putih dengan menggunakan dua parameter utama, yaitu pH dan kadar alkohol. Kualitas anggur putih sering kali dipengaruhi oleh keseimbangan antara keasaman (pH) dan kandungan alkohol, yang dapat mempengaruhi rasa, aroma, dan kesegaran produk akhir. Dalam penelitian ini, data sampel anggur putih yang mencakup variasi pH dan kadar alkohol dianalisis untuk menentukan korelasi antara kedua parameter tersebut dengan persepsi kualitas yang ditentukan oleh panelis. Metode K-Nearest Neighbor digunakan untuk membangun model prediksi yang dapat memperkirakan kualitas anggur berdasarkan nilai pH dan alkohol. Hasil penelitian menunjukkan bahwa pH dan kadar alkohol memiliki hubungan signifikan terhadap kualitas anggur putih, dengan pH yang lebih rendah dan kadar alkohol tertentu berkontribusi pada peningkatan kualitas anggur. Model yang dikembangkan dapat digunakan oleh produsen anggur untuk mengontrol dan memperbaiki kualitas produk mereka, serta memberikan dasar bagi penelitian lebih lanjut mengenai faktor-faktor lain yang mempengaruhi kualitas anggur.

Kata Kunci: KNN, minuman, anggur putih, alkohol, ph, kualitas

ABSTRACT – dalam bahasa inggris

This study aims to develop a white wine quality prediction model using two main parameters, namely pH and alcohol content. The quality of white wine is often influenced by the balance between acidity (pH) and alcohol content, which can affect the flavor, aroma and freshness of the final product. In this study, white wine sample data that included variations in pH and alcohol content were analyzed to determine the correlation between these two parameters and the perceived quality determined by panelists. The K-Nearest Neighbor method was used to build a prediction model that could estimate wine quality based on pH and alcohol values. The results showed that pH and alcohol content have a significant relationship to white wine quality, with lower pH and certain alcohol levels contributing to improved wine quality. The developed model can be used by wine producers to control and improve the quality of their products, as well as provide a basis for further research into other factors affecting wine quality.

Keywords: KNN, beverage, white wine, alcohol, ph, quality

Penulis Korespondensi

Ahmad Muhajir Ridwan

Tanggal Submit : 01/02/2025
Tanggal Diterima : 21/12/2025
Tanggal Terbit : 23/12/2025

This is an open access article under the [CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0) International License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Publisher's Note: JPPM stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

I. PENDAHULUAN

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 [1].

Anggur putih merupakan salah satu jenis minuman beralkohol yang banyak dikonsumsi di berbagai belahan dunia, dengan kualitas yang sangat dipengaruhi oleh proses fermentasi dan komposisi kimianya. Kualitas anggur putih sering kali menjadi perhatian utama produsen, karena dapat mempengaruhi citra merek dan kepuasan konsumen. Salah satu faktor utama yang mempengaruhi kualitas anggur adalah keseimbangan

antara pH dan kadar alkohol yang terkandung dalam minuman tersebut. Kedua parameter ini dapat mempengaruhi rasa, aroma, serta karakteristik organoleptik lainnya dari anggur putih.

pH anggur putih berperan penting dalam menentukan tingkat keasaman, yang dapat memberikan kesegaran dan keseimbangan rasa. Di sisi lain, kadar alkohol yang terkandung dalam anggur juga berperan dalam membentuk tubuh dan kehangatan rasa anggur. Keduanya saling berinteraksi untuk menciptakan profil rasa yang diinginkan, yang pada akhirnya berpengaruh pada penilaian kualitas oleh para konsumen dan ahli anggur. Oleh karena itu, penting untuk memahami bagaimana pH dan alkohol berkontribusi terhadap kualitas keseluruhan anggur putih.

II. PENELITIAN YANG TERKAIT

Beberapa penelitian sebelumnya telah mengeksplorasi penggunaan metode data mining untuk memprediksi kualitas anggur merah atau white wine. Salah satunya adalah penelitian yang dilakukan oleh [2] yang mengembangkan model prediksi kualitas anggur menggunakan data mining. Fokus penelitian ini adalah untuk menentukan algoritma mana yang paling efektif dalam mengklasifikasikan kualitas anggur merah berdasarkan komposisi kimianya.

Penelitian lain seperti Implementasi Metode Imputasi Mean dan Single Center Imputation Chained Equation (SICE) Terhadap Hasil Prediksi Linear Regression pada Data Numerik oleh [1], Prediction of Wine Quality Using Machine Learning Algorithms oleh [3], Selection of important features and predicting wine quality using machine learning techniques oleh [4], dan Wine Quality Analysis Using machine learning algorithms oleh [5] menunjukkan bahwa kualitas minuman anggur atau wine berpengaruh oleh beberapa faktor diantaranya pH dan alkohol yang terkandung didalamnya.

III. METODE PENELITIAN

Tahap pertama yaitu Business Understanding pada tahap ini memahami kebutuhan atau tujuan dari dataset yang akan diteliti. Berikut tahapan dalam bussines understanding :

1. Menentukan tujuan bisnis, yaitu memprediksi kualitas white wine berdasarkan komposisi atau ingredients yang terkandung di dalam white wine.
2. Menilai situasi, beberapa produk wine memiliki kualitas yang berbeda dari kualitas 6, 7, dan 8 yang pengaruhi oleh kandungan bahan di dalamnya, jadi perlu mengetahui kandungan apa yang paling berpengaruh terhadap kualitas dari white wine.
3. Menentukan tujuan data mining, tujuan dari penelitian ini adalah meningkatkan pengetahuan tentang kualitas wine untuk memenuhi kebutuhan

dan keinginan konsumen dan memberi petunjuk tentang kemungkinan, dan kesediaan konsumen untuk membeli anggur dengan campuran bahan-bahan tertentu serta memberikan keunggulan bagi produsen dibandingkan pesaing lainnya.

Tahap kedua yaitu Data Understanding, dataset yang digunakan yaitu berasal dari UC Irvine Machine Learning Repositories tentang wine quality dari Vinho Verde Portuguese wine. Pada dataset memiliki input dan output yang menetapkan nilai kualitas mulai dari 3 sampai 8. Ada 4.899 data sampel observasi dengan 12 atribut yaitu fixed acidity, citric acid, volatile acidity, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol dan quality. Dalam penerapan kali ini peneliti menggunakan perhitungan manual sekaligus aplikasi Orange Mining untuk memvalidasi hasil akhir dari perhitungan manual.

Tahap ketiga yaitu Data Preparation, yang dimana data disiapkan untuk dilakukan proses pelatihan. Tahapan pengolahan data mencakup pemilihan kelas, atribut-atribut data serta transformasi data yang kemudian dilakukan proses pembersihan dan regulasi data, data yang diambil hanya 120 data yang berisi 40 data teratas yang memiliki kualitas 6 sampai 8.

Tahap keempat yaitu Modelling, algoritma yang digunakan pada penelitian kali ini adalah K-Nearest Neighbor. Data yang akan digunakan pada penelitian ini juga akan dibagi menjadi dua bagian, yaitu 114 data training dan 6 data sebagai data testing. Diharapkan penggunaan KNN dapat menghasilkan hasil terbaik dari model yang lain sehingga memberikan hasil yang optimal.

Tahap kelima yaitu Evaluasi, dari hasil penelitian evaluasi yang digunakan pada penelitian kali ini adalah Confusion Matrix yang menggambarkan hasil dari prediksi dan hasil sebenarnya pada akhir penelitian.

IV. HASIL DAN PEMBAHASAN

Dataset yang digunakan pada penelitian kali ini adalah sebagai berikut :

Tabel 1. Dataset Setelah Preparation

No	pH	Alcohol	Quality
1	3	8,8	6
2	3	8,8	6
3	2,95	11,4	7
4	2,98	9	7
5	2,98	9	7
...
116	3,47	10,2	6

117	3,47	10,5	8
118	3,53	12,5	8
119	3,54	12,4	8
120	3,55	11,2	7

Untuk pembagian data testing, peneliti menentukan 6 angka random dari range 1 sampai 120 dan angka yang muncul adalah 72, 107, 15, 93, 52, dan 103 yang berarti data pada nomor tersebut akan dijadikan sebagai data testing dan data sisanya akan dijadikan data training sebagaimana mestinya.

Selanjutnya, masuk ke metode K-Nearest Neighbor. Data pertama yang akan dihitung adalah data ke-72 dengan nilai sebagai berikut :

Tabel 2. Data White Wine Nomor 72

No	pH	Alcohol	Quality
72	3,3	9,5	6

Setelah data tersebut ditemukan, selanjutnya adalah melakukan perhitungan dengan rumus sebagai berikut :

$$Euclidean = \sqrt{(A_1 - A_t)^2 + (B_1 - B_t)^2}$$

A_1 = Variabel Pertama pada Data Training

A_t = Variabel Pertama pada Data Testing

B_1 = Variabel Kedua pada Data Training

B_t = Variabel Kedua pada Data Testing

Dari rumus tersebut, didapatkan 5 data terdekat antara lain sebagai berikut :

Tabel 3. Hasil Perhitungan Data Nomor 72

pH	Alcohol	Quality	Euclidean	Urutan
3,3	9,5	6	0	1
3,24	9,5	6	0,06	2
3,22	9,5	6	0,08	3
3,32	9,6	8	0,10198039	4
3,24	9,6	6	0,116619038	5

Dari tabel diatas, maka kualitas yang terdapat pada data nomor 72 adalah 6 yang berarti sesuai dengan kualitas yang sudah ada. Setelah perhitungan untuk data

nomor 72 ditemukan, selanjutnya adalah menghitung data nomor 107, data tersebut adalah sebagai berikut :

Tabel 4. Data White Wine Nomor 107

No	pH	Alcohol	Quality
107	3,41	11,9	8

Setelah dilakukan perhitungan manual menggunakan rumus yang ditampilkan diatas, 5 data terdekat dengan data nomor 107 adalah sebagai berikut:

Tabel 5. Hasil Perhitungan Data Nomor 107

pH	Alcohol	Quality	Euclidean	Urutan
3,38	12	7	0,104403065	1
3,31	12	7	0,141421356	2
3,25	11,9	8	0,16	3
3,25	11,9	8	0,16	4
3,25	11,9	8	0,16	5

Dari tabel diatas, maka kualitas yang terdapat pada data nomor 107 adalah 8 yang berarti sesuai dengan kualitas yang sudah ada. Setelah perhitungan untuk data nomor 107 ditemukan, selanjutnya adalah menghitung data nomor 15, data tersebut adalah sebagai berikut :

Tabel 6. Data White Wine Nomor 15

No	pH	Alcohol	Quality
15	3,11	9,4	6

Setelah dilakukan perhitungan manual menggunakan rumus yang ditampilkan diatas, 5 data terdekat dengan data nomor 15 adalah sebagai berikut :

Tabel 7. Hasil Perhitungan Data Nomor 15

pH	Alcohol	Quality	Euclidean	Urutan
3,1	9,4	6	0,01	1
3,22	9,5	6	0,148660687	2
3,24	9,5	6	0,164012195	3
3,09	9,6	6	0,200997512	4
3,18	9,6	6	0,211896201	5

Dari tabel diatas, maka kualitas yang terdapat pada data nomor 15 adalah 6 yang berarti sesuai dengan kualitas yang sudah ada. Setelah perhitungan untuk data nomor 15 ditemukan, selanjutnya adalah menghitung data nomor 93, data tersebut adalah sebagai berikut :

Tabel 8. Data White Wine Nomor 93

No	pH	Alcohol	Quality
93	3,34	11,1	8

Setelah dilakukan perhitungan manual menggunakan rumus yang ditampilkan diatas, 5 data terdekat dengan data nomor 93 adalah sebagai berikut :

Tabel 9. Hasil Perhitungan Data Nomor 93

pH	Alcohol	Quality	Euclidean	Urutan
3,34	11,1	7	0	1
3,3	11,1	7	0,04	2
3,3	11,1	7	0,04	2
3,25	11,1	8	0,09	4
3,34	11	7	0,1	5

Dari tabel diatas, maka kualitas yang terdapat pada data nomor 93 adalah 7 yang berarti tidak sesuai dengan kualitas yang sudah ada. Setelah perhitungan untuk data nomor 93 ditemukan, selanjutnya adalah menghitung data nomor 52, data tersebut adalah sebagai berikut :

Tabel 10. Data White Wine Nomor 52

No	pH	Alcohol	Quality
52	3,25	12,4	8

Setelah dilakukan perhitungan manual menggunakan rumus yang ditampilkan diatas, 5 data terdekat dengan data nomor 52 adalah sebagai berikut :

Tabel 11. Hasil Perhitungan Data Nomor 52

pH	Alcohol	Quality	Euclidean	Urutan
3,26	12,4	7	0,01	1
3,26	12,4	8	0,01	2
3,29	12,4	7	0,04	3
3,29	12,4	7	0,04	4
3,35	12,4	8	0,1	5

Dari tabel diatas, maka kualitas yang terdapat pada data nomor 52 adalah 7 yang berarti tidak sesuai dengan

kualitas yang sudah ada. Setelah perhitungan untuk data nomor 52 ditemukan, selanjutnya adalah menghitung data nomor 103, data tersebut adalah sebagai berikut :

Tabel 12. Data White Wine Nomor 103

No	pH	Alcohol	Quality
103	3,36	12,8	8

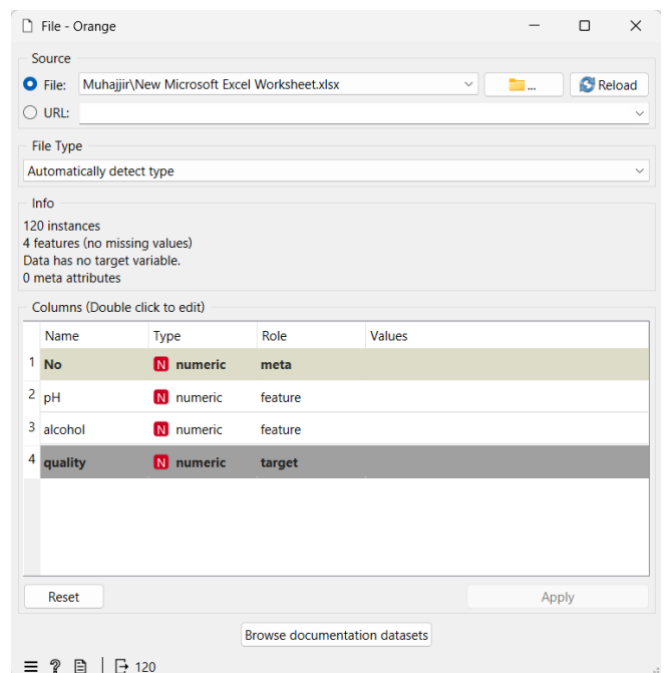
Setelah dilakukan perhitungan manual menggunakan rumus yang ditampilkan diatas, 5 data terdekat dengan data nomor 103 adalah sebagai berikut :

Tabel 13. Hasil Perhitungan Data Nomor 103

pH	Alcohol	Quality	Euclidean	Urutan
3,33	12,8	8	0,03	1
3,33	12,8	8	0,03	2
3,43	12,9	8	0,122066	3
3,26	12,7	7	0,141421	4
3,26	12,7	8	0,141421	5

Dari tabel diatas, maka kualitas yang terdapat pada data nomor 103 adalah 8 yang berarti sesuai dengan kualitas yang sudah ada. Selanjutnya adalah penerapan aplikasi Orange Mining menggunakan fitur kNN yang ada didalamnya.

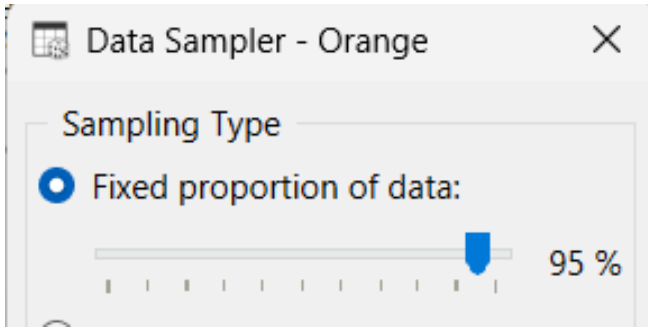
Langkah pertama adalah melakukan menambahkan fitur "File" pada aplikasi Orange Mining dan pilih fitur tersebut kemudian arahkan ke file dataset yang ingin digunakan. Setelah file tersebut dipilih, kemudian mengubah quality menjadi target agar output yang dihasilkan adalah target.



Gambar 1. Pemilihan File, Fitur, dan Target

Setelah file, fitur, dan target dipilih kemudian close menu File – Orange dan tarik fitur tersebut dan

tambahkan “Data Sampler” yang berguna untuk memisahkan antara 114 data training dan 6 data testing.



Gambar 2. Custom Setting Pada Data Sampler

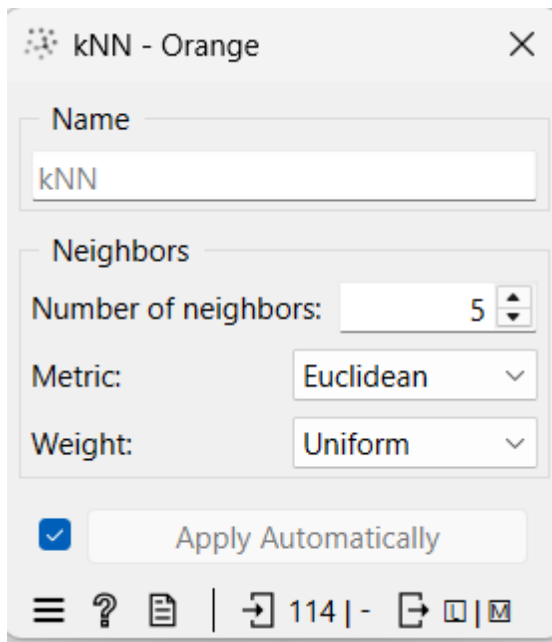
Setelah data terpisah, data yang terbentuk adalah sebagai berikut :

Data Sample: New Microsoft Excel Worksheet: 114 instances, 4 variables					
Features: 2 numeric (no missing values)					
Target: numeric					
Metas: numeric					
	quality	No	pH	alcohol	
1	6	45	3.24	9.5	
2	7	48	3.25	11.4	
3	7	5	2.98	9.0	
4	8	56	3.25	11.1	

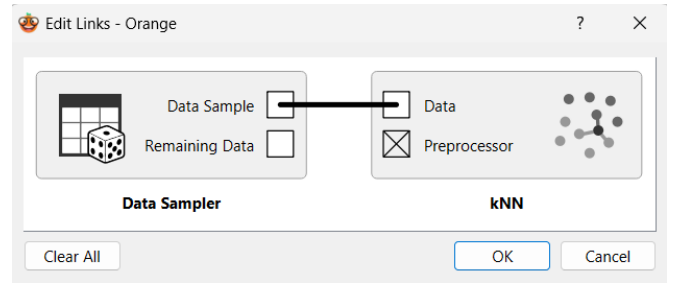
Remaining Data: New Microsoft Excel Worksheet: 6 instances, 4 variables					
Features: 2 numeric (no missing values)					
Target: numeric					
Metas: numeric					
	quality	No	pH	alcohol	
1	6	72	3.30	9.5	
2	8	107	3.41	11.9	
3	6	15	3.11	9.4	
4	8	93	3.34	11.1	

Gambar 3. Pemisahan Data Training dan Data Testing

Setelah itu, tarik fitur data sampler kemudian pilih fitur “kNN” dan masukan data sample sebagai data ke kNN.

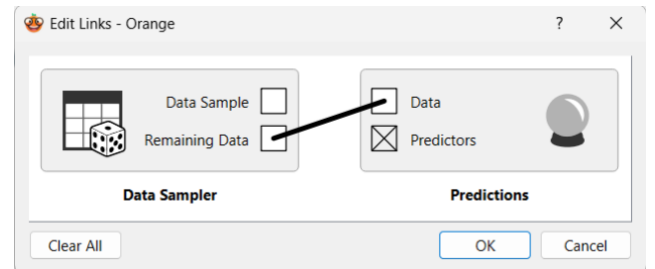


Gambar 4. Setting Pada Fitur kNN



Gambar 5. Pemilihan Data Sample Sebagai Data kNN

Setelah itu, tarik dari kNN dan tambahkan fitur “Predictions” dan tarik data sampler kemudian hubungkan ke fitur “Predictions” juga.



Gambar 6. Pemilihan Data Testing Sebagai Data Predictions

Maka, hasil yang akan muncul jika menekan fitur Predictions adalah sebagai berikut :

kNN	error	quality	No	pH	alcohol	
1	6	0	6	72	3.30	9.5
2	8	-0	8	107	3.41	11.9
3	6	0	6	15	3.11	9.4
4	7	-1	8	93	3.34	11.1
5	7	-1	8	52	3.25	12.4
6	8	-0	8	103	3.36	12.8

Model	MSE	RMSE	MAE	MAPE	R2
kNN	0.227	0.476	0.400	0.053	0.745

Gambar 7. Hasil Fitur Predictions

Evaluasi yang dilakukan dengan menggunakan aplikasi Orange Mining dengan Confusion Matrix adalah sebagai berikut :

		Predicted			
		6	7	8	Σ
Actual	6	25	10	5	40
	7	6	24	10	40
	8	4	12	24	40
Σ		35	46	39	120

Gambar 8. Hasil Confusion Matrix

V. KESIMPULAN

Berdasarkan hasil pengujian dengan 6 data tes, dapat disimpulkan bahwa model prediksi yang diterapkan menunjukkan kinerja yang cukup baik, meskipun ada beberapa ketidaksesuaian dalam hasilnya. Dari 6 data tes yang digunakan, 4 data di antaranya sesuai dengan prediksi yang dihasilkan oleh model, sementara 2 data lainnya tidak sesuai dengan prediksi. Hal ini menunjukkan bahwa model memiliki tingkat akurasi yang cukup tinggi, namun masih perlu adanya penyempurnaan untuk meningkatkan ketepatannya, terutama dalam kasus data yang tidak sesuai.

Ketidaksesuaian yang terjadi pada 2 data tes dapat disebabkan oleh beberapa faktor, seperti kualitas data yang kurang lengkap atau adanya variabilitas dalam atribut yang tidak dapat diprediksi dengan sempurna oleh model. Untuk meningkatkan akurasi model, diperlukan peningkatan dalam preprocessing data dan pemilihan atribut yang lebih tepat. Meskipun demikian, model ini tetap memberikan kontribusi yang berarti dalam memprediksi kualitas produk, dengan potensi untuk diimplementasikan lebih lanjut dalam proses klasifikasi kualitas anggur.

DAFTAR PUSTAKA

- [1] M. R. Baihaqi, T. N. Padilah, and M. Jajuli, "Implementasi metode imputasi mean dan single center imputation chained equation (SICE) terhadap hasil prediksi linear regression pada data numerik," *Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 7, no. 4, pp. 661-671, 2023.
- [2] R. Supriyadi, et al., "Penerapan algoritma Random Forest untuk menentukan kualitas anggur merah," *E-Bisnis: Jurnal Ilmiah Ekonomi dan Bisnis*, vol. 13, no. 2, pp. 67-75, 2020.
- [3] K. R. Dahal, et al., "Prediction of wine quality using machine learning algorithms," *Open Journal of Statistics*, vol. 11, no. 2, pp. 278-289, 2021.
- [4] Y. Gupta, "Selection of important features and predicting wine quality using machine learning techniques," *Procedia Computer Science*, vol. 125, pp. 305-312, 2018.
- [5] Mahima, et al., "Wine quality analysis using machine learning algorithms," in *Micro-Electronics and Telecommunication Engineering: Proceedings of 3rd ICMETE 2019*, Springer Singapore, 2020.