

## Klasifikasi Penyakit Diabetes dengan Algoritma *Decision Tree* dan Naïve Bayes

Maulana Muhammad Jogo Samodro<sup>1</sup>, Muhammad Kunta Biddinika<sup>2\*)</sup>, Abdul Fadlil<sup>3</sup>

<sup>1)2\*)</sup> Magister Informatika Universitas Ahmad Dahlan, Yogyakarta

<sup>3)</sup> Teknik Elektro Universitas Ahmad Dahlan, Yogyakarta  
Jl. Ringroad Selatan, Tamanan, Bantul

Email: <sup>1)</sup>2108048022@webmail.uad.ac.id, <sup>2)</sup>muhammad.kunta@mti.uad.ac.id <sup>2)</sup>fadlil@mti.uad.ac.id

### ABSTRAK

*Diabetes merupakan salah satu penyakit yang dapat menyebabkan kematian prematur. Pada tahun 2019 ada 463 juta orang terkena penyakit tersebut. Di tahun 2021 penderita diabetes bertambah 74 juta menjadi 537 juta. Maka perlu adanya pengklasifikasian guna mendeteksi dini penyakit diabetes. Machine learning adalah suatu sistem dengan latar belakang ilmu statistika dan matematika. ML mampu untuk melakukan pengklasifikasian kelas sehingga dapat diketahui prediksi terkena penyakit diabetes. Penelitian menggunakan 2 algoritma dalam machine learning yaitu naïve bayes dan decision tree. Data penelitian sebanyak 520 data, dengan 300 data training, 220 data testing. Hasil penelitian menunjukkan dengan algoritma naïve bayes menghasilkan tingkat akurasi sebesar 90,45%. Algoritma decision tree mempunyai tingkat akurasi sebesar 96,36%.*

**Kata Kunci:** Diabetes, Naïve Bayes, Decision Tree

### ABSTRACT

*Diabetes is a disease that can cause premature death. In 2019 there were 463 million people affected by the disease. In 2021 people with diabetes will increase by 74 million to 537 million. So there is a need for classifiers to detect diabetes early. Machine learning is a system with a background in statistics and mathematics. ML is able to classify classes so that predictions of diabetes can be known. The study uses 2 algorithms in machine learning, namely naïve Bayes and decision trees. The research data is 520 data, with 300 training data, 220 testing data. The results of the research show that the naïve Bayes algorithm produces an accuracy rate of 90.45%. The decision tree algorithm has an accuracy rate of 96.36%.*

**Keywords:** Diabetes, Naïve Bayes, Decision Tree

## 1 PENDAHULUAN

Berdasarkan Organisasi International Diabetes (IDF) pada tahun 2019 sedikitnya 463 juta orang dengan usia sekitar 20-79 tahun di dunia menderita diabetes atau setara dengan angka prevalensi 9,3% dari penduduk seluruh dunia pada usia yang sama. Di tahun 2021 terjadi kenaikan 16% penderita diabetes menjadi 537 juta [1]. Penyakit diabetes selain menyebabkan kematian prematur, juga menyebabkan beberapa penyakit lainnya seperti kebutaan, penyakit jantung, dan gagal ginjal. IDF memproyeksikan jumlah penderita penyakit diabetes di sekitar umur 20-79 tahun, beberapa negara tersebut adalah Cina, India dan Amerika Serikat dengan jumlah kasus tertinggi di dunia. Indonesia masuk ke dalam kategori 10 besar negara dengan penderita terbanyak. Penderita diabetes di Indonesia mencapai 10,7 juta yang membuat Indonesia berada nomor 7 di dunia dan menjadi negara asia tenggara dengan penderita diabetes terbanyak[2].

Diabetes adalah penyakit kronis berupa gangguan metabolisme tubuh yang ditandai dengan kadar gula darah melebihi batas normal manusia. Terdapat 2 tipe

penyakit diabetes melitus, yaitu diabetes melitus tipe 1 dan diabetes melitus tipe 2. Diabetes melitus tipe 1 disebabkan karena kenaikan gula darah yang diakibatkan dari kerusakan sel beta pancreas. Kendala tersebut mengakibatkan tidak adanya produksi insulin pada tubuh. Diabetes melitus tipe 2 disebabkan karena penurunan sekresi insulin yang rendah oleh kelenjar pankreas sehingga menyebabkan kenaikan gula darah.

Riset Kesehatan Dasar (Riskesdas) tahun 2018 melakukan pengumpulan data penderita diabetes melitus dengan umur diatas 15 tahun. Kriteria ini mengacu pada consensus Perkumpulan Endokrinologi Indonesia (PERKENI) yang mengadopsi kriteria dari American Diabetes Association (ADA). Dikatakan mengalami diabetes melitus jika kadar glukosa darah saat puasa  $\geq 126$  mg/dl, atau glukosa darah 2 jam pasca bebanan  $\geq 200$  mg/dl, atau glukosa darah sewaktu  $\geq 200$ mg/dl dengan gejala sering lapar, sering haus, sering buang air kecil dalam jumlah banyak, dan berat badan turun [3].

Penyakit diabetes perlu dideteksi sejak dini. Penyakit ini dapat menyebabkan gangguan pada penyakit lainnya. Selain menyebabkan gangguan menyakit ini juga dapat menimbulkan jenis penyakit baru pada penderitanya. Maka diperlukan suatu sistem klasifikasi guna untuk mendeteksi dini penyakit diabetes.

Pembelajaran mesin dikembangkan berdasarkan disiplin ilmu seperti statistika, matematika dan data mining sehingga mesin dapat belajar dengan menganalisa data tanpa perlu di program ulang atau diperintah [4]. Machine learning berasal dari sensor yang diubah menjadi keputusan sehingga mampu memprediksi selanjutnya[5]. Pembelajaran ML digunakan untuk mengajari mesin agar data terolah seefisien mungkin. Terkadang data tidak dapat diinterpretasikan dengan baik, pembelajaran mesin mampu untuk meningkatkan intepretasi data dengan baik. Banyak penelitian-penelitian untuk membuat mesin mampu belajar sendiri. Para matematikawan dan programmer menerapkan beberapa pendekatan. Pendekatan tersebut diantaranya Supervised Learning, Unsupervised Learning, Reinforcement Learning [6].

Dengan sistem learning mampu untuk membuat pekerjaan manusia menjadi lebih efisien. Peran machine learning sudah merambah di berbagai bidang dalam kegiatan manusia. Seperti pengenalan suara [7], pengenalan gambar [8]. Selain itu ML juga membantu peran manusia [9] dalam bidang biofarma [10], kedokteran [11] seperti mendeteksi dini penyakit kanker [12] dan diabetes [13], meteorologi [14] [15], ketahanan pangan [16][17].

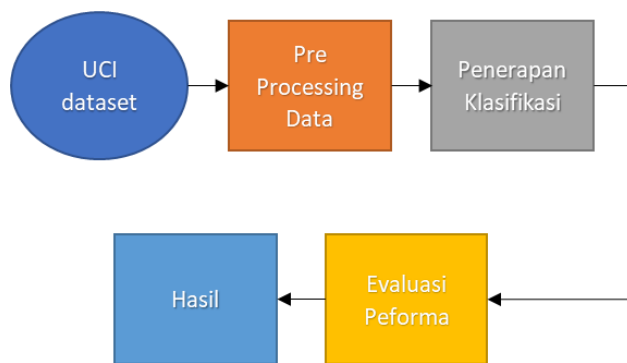
Algoritma yang dipakai pada machine learning ini adalah Naïve Bayes dan Dececion Tree. Naïve bayes adalah Teknik klasifikasi yang berlandaskan dari teorema bayes. Naïve bayes mempunyai atribut independen yang memberikan probabilitas pada kelas. Klasifikasi ini termasuk kedalam supervised learning, atau pembelajaran diawasi. Naïve bayes membutuhkan data training untuk menentukan data testing. Klasifikasi ini sering digunakan untuk diagnose dengan bantuan komputer [18][19][20].

Sistem klasifikasi ini akan digunakan untuk mengklasifikan kelas diabetes. Sistem diharapkan dapat membantu kinerja manusia untuk memilah antara penderita dan tidak penderita. Terdapat beberapa atribut untuk menentukan bahwa terkena penyakit diabetes atau tidak. Sehingga sistem ini mampu untuk memprediksi kemungkinan diabetes.

## 2 METODOLOGI

Awan ditunjukkan pada Gambar 1 Machine Learning (ML) merupakan mesin yang dikembangkan untuk bisa belajar dengan sendirinya

tanpa arahan penggunanya. ML merupakan jenis metodologi dari kecerdasan buatan (artificial intelligent) yang membantu memperoleh data dengan program tertentu. Tujuan utama dari machine learning adalah untuk mengenali pola dalam sebuah data, sistem ini menginformasikan cara menangani masalah yang tidak terlihat [21]. Pada penyusunan penelitian terlebih dahulu dilakukan penyusunan kerangka penelitian. Kerangka penelitian ditunjukkan pada gambar 1 dalam bentuk diagram.



Gambar 1. Blok Diagram

Blok Diagram dari penelitian ini dapat dilihat pada gambar 1. Langkah awal penelitian ini yaitu membuat data set sebagai data pelatihan. Selanjutnya dilakukan pre-processing data guna melakukan pengacakan data pada data penelitan. Setelah dilakukan pre-processing data diklasifikasi menggunakan algoritma machine learning yaitu naïve bayes dan decision tree. Dilakukan evaluasi peforma dengan pengukuran akurasi, Precision, Recall dan F-Measure.

### Naïve Bayes

Algoritma bayes digunakan untuk memprediksi peluang di masa yang akan datang berdasarkan pengalaman di masa sebelumnya. Naïve bayes terdiri dari atribut dan kelas. Atribut dari naïve bayes dapat terdiri dari 1 atau lebih atribut guna untuk menentukan kelas. Kelas digunakan untuk mengklasifikan suatu data. Kelas akan menentukan data pada atribut masuk kedalam kategori yang sudah ditentukan [22].

$$P(\mathbf{x}|y) = \prod_{i=1}^a P(x_i|x_1, x_2, \dots, x_{i-1}, y) \quad (1)$$

$P(x_i|x_1, x_2, \dots, x_{i-1}, y)$  tidak mampu untuk di estimasi secara akurat pada dataset dengan atribut yang banyak. Setiap atribut naïve bayes bersifat independen, tidak bergantung dengan atribut lainnya. Maka, algoritma naïve bayes menyimpulkan persamaan (1) menjadi:

$$P(\mathbf{x} | y) = \prod_{i=1}^a P(x_i|y) \quad (2)$$

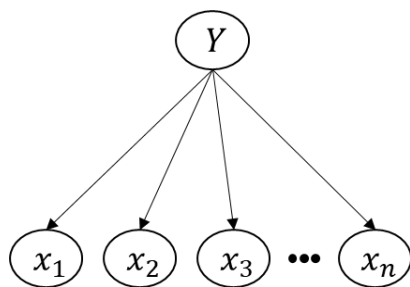
Berdasarkan rumus (2), untuk menghitung probabilitas algoritma naïve bayes menggunakan rumus (3):

$$P_{NB}(y, \mathbf{x}) = P(y) \prod_{i=1}^a P(x_i|y) \quad (3)$$

Maka untuk melakukan klasifikasi bayes pada  $\mathbf{x}$ :

$$\operatorname{argmax} \left( \hat{P}_y(y) \prod_{i=1}^a \hat{P}(x_i|y) \right) \quad (4)$$

$\hat{P}(y)$  dan  $\hat{P}(x_i|y)$  adalah turunan dari masing-masing atribut. Turunan tersebut merupakan representasi dari sampel pelatihan. Estimasi  $P(x|y)$   $P(y)$  diperoleh berdasarkan frekuensi dalam data training.

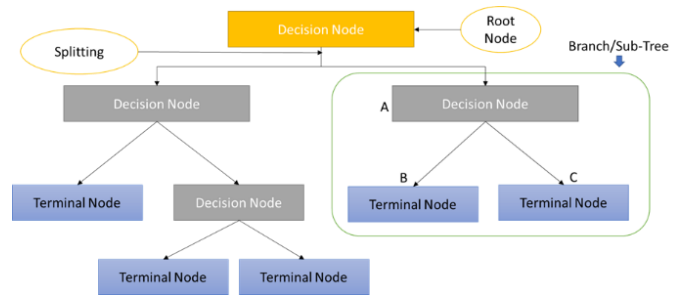


Gambar 2. Naïve Bayes

Gambar 2 merupakan sistematika kinerja dari algoritma naïve bayes.  $Y$  merupakan kelas yang akan diklasifikasikan, dan  $x_1, x_2, x_3, \dots, x_n$  merupakan atribut kelas.

### Decision Tree

Decision tree (DT) adalah salah satu algoritma machine learning yang menggunakan sistem supervised learning. Sistem yang menggunakan data training untuk melakukan uji pada data testing [23][24]. Metode ini digunakan untuk memprediksi nilai kelas yang akan menjadi tujuan dari nilai variabelnya. Perhitungan decision tree berdasarkan nilai tertinggi dari entropi. Input dari data ini yaitu data training yang akan diklasifikasikan untuk menentukan kelas [25].



Gambar 3. Algoritma Decision Tree

Gambar 3 merupakan proses dari Algoritma DT. Algoritma ini terdiri dari beberapa bagian, yaitu Root Node, Branch/Sub-Tree, Leaf Node/Terminal Node. Root node adalah node yang tidak memiliki inputan akan tetapi dia memiliki kondisi output. Kondisi dari root node akan terpisah menjadi brachnode. Brachnode terdiri dari inputan dan output. Terminal node adalah keluaran hasil keputusan dari algoritma decision tree [26].

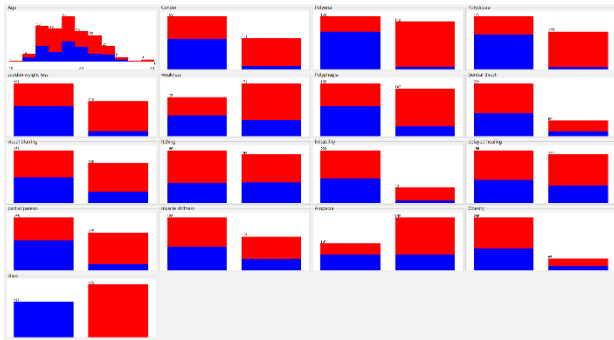
### 3 HASIL DAN PEMBAHASAN

Penelitian menggunakan aplikasi weka versi 3.8.6 menggunakan komputer AMD Ryzen 5 dengan RAM 8 GB. Data set penelitian ini menggunakan UCI dataset yang didapatkan dari pasien diabetes Rumah Sakit Sylhet, Bangladesh. Terdapat 17 atribut data set dengan total jumlah 520 data. Data tersebut dibagi menjadi 2, yaitu data training dan data testing. Data training terdiri dari 300 data, untuk data testing terdiri dari 220 data. Tabel 2 adalah deskripsi dari dataset.

Tabel 1. Deskripsi Dataset

Atribut	Nilai
Age	20-65
Gender	Male, Female
Polyuria	Yes, No
Polydipsia	Yes, No
sudden weight loss	Yes, No
weakness	Yes, No
Polyphagia	Yes, No
Genital thrush	Yes, No
visual blurring	Yes, No
Itching	Yes, No
Irritability	Yes, No
delayed healing	Yes, No
partial paresis	Yes, No
muscle stiffness	Yes, No
Alopecia	Yes, No
Obesity	Yes, No
class	Positive, Negative

Output dari klasifikasi penyakit diabetes ini yaitu positif atau negatif. Data tersebut di import dari bentuk file dengan bentuk .csv.



Gambar 4. Grafik Klasifikasi Penderita Diabetes

Gambar 4 merupakan klasifikasi penderita penyakit diabetes pada data training. Warna merah pada grafik ditunjukkan untuk penderita diabetes. Warna biru pada grafik ditunjukkan bagi penderita non diabetes. Jumlah penderita penyakit diabetes pada data training yaitu 179 orang. Sedangkan yang tidak menderita diabetes sebanyak 121 orang. Jumlah jenis kelamin laki-laki pada data training yaitu 189, untuk jumlah perempuan sebanyak 111. Terlihat pada tabel 1 umur responden berkisar dari 20-65 tahun.

Pengujian akurasi klasifikasi penyakit diabetes menggunakan confusion matrix. Confusion matrix terdiri dari 4 tabel kombinasi. Keempat kombinasi tersebut adalah True Positif (TP), False True Negatif (TN), False Positif (FP), dan False Negatif (FN).

Tabel 2. Confusion Matrix

	a	b	Naïve Bayes	a1	b1	Decision Tree
a	72	7		79	0	
b	14	127	8	133		

Tabel 2 merupakan confusion matrix dari perhitungan data testing yang sebanyak 220. Kolom a baris a adalah kelas negative yang keluarannya bernilai negative yaitu sebanyak 72. Kolom a baris b adalah kelas Negatif yang keluarannya bernilai positif yaitu sebanyak 14. Kolom b baris a adalah kelas positive yang keluarannya bernilai negative yaitu sebanyak 7. Kolom b baris b adalah kelas positive yang keluarannya bernilai positive. Kolom a dan b adalah untuk nilai keluaran pada algoritma naïve bayes. Berdasarkan data tabel 2 dapat ditentukan nilai akurasinya. Penentuan nilai akurasi didapatkan dari:

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (2)$$

$$Precision = \frac{TP}{FP + TP} \quad (3)$$

$$Recall = \frac{TP}{FN + TP} \quad (4)$$

$$F - Measure = 2 * \frac{precision * recall}{precision + recall} \quad (5)$$

- TP = True Positif
- TN = True Negatif
- FP= False Positif
- FN = False Negatif

Hasil dari nilai akurasi tersebut dapat dilihat pada tabel 3.

Tabel 3. Hasil Akurasi Perbandingan Algoritma Bayes dan Decision Tree

	TP Rate	FP Rate	Class	Accuracy	Algoritma
Rata-Rata	0.911	0.099	Negative	90,45%	Naïve Bayes
	0.901	0.089	Positive		
Rata-rata	0.905	0.092		96,36%	Decision Tree
	1.000	0.057	Negative		
Rata-rata	0.943	0.000	Positive		
	0.964	0.020			
	Precision	Recall	F-Measure	Class	Algoritma
Rata-rata	0.837	0.911	0.873	Negative	Naïve Bayes
	0.948	0.901	0.924	Positive	
Rata-rata	0.908	0.905	0.905		Decision Tree
	0.908	1.000	0.952	Negative	
Rata-rata	1.000	0.943	0.971	Positive	
	0.967	0.964	0.964		

Pada tabel 3 terlihat jika pada algoritma naïve bayes memiliki tingkat akurasi 90,45%. TP rate pada kelas positif yaitu 0,911. Pada kelas negative memiliki TP rate 0,901. Rata-rata TP rate kelas positive maupun kelas negative pada algoritma adalah 0,905. Sedangkan pada decision tree memiliki tingkat akurasi 96,45%. TP decision tree bernilai 1 pada kelas negative. Artinya pada kelas negative decision tree mampu mengklasifikasikan dengan baik. Pada kelas positive memiliki TP 0,943. Rata-rata TP rate dari decision tree adalah 0,964. FP rate Decision tree pada kelas negative sebesar 0,057. FP rate pada kelas positive sebesar 0,000. Kedua hasil tersebut kemudian dirata-rata mendapatkan nilai sebesar 0,020.

Tingkat precision algoritma naïve bayes pada kelas negative sebesar 0,837. Sedangkan pada kelas positif adalah 0,948 dengan rata-rata keduanya sebesar 0,908. Pada algoritma decision tree mempunyai tingkat precision sebesar 0,908 pada kelas negative. Kelas positive decision tree mempunyai tingkat precision sebesar 1,000. Rata-rata tingkat precision decision tree adalah 0,967.

Recall atau tingkat keberhasilan dalam menjalankan sistem pada algoritma naïve bayes kelas negative sebesar 0,911. Kelas positive sebesar 0,901. Rata-rata tingkat keberhasilan algoritma naïve bayes adalah 0,905. Tingkat keberhasilan algoritma decision tree kelas negative adalah 1,000. Kelas positive decision tree mempunyai tingkat keberhasilan sebesar 0,943. Rata-rata dari algoritma decision tree mempunyai tingkat keberhasilan 0,964.

F-measure yaitu nilai dari perhitungan precision dan recall. Nilai f-measure pada kelas negative algoritma naïve bayes adalah 0,873. Kelas positif algoritma naïve bayes adalah 0,924. Rata-rata dari f-measure naïve bayes adalah 0,905. Sedangkan f-measure kelas negative decision tree adalah 0,952. Kelas positive decision tree f-measurenya adalah 0,971. Rata-rata dari f-measure decision tree adalah 0,964.

Berdasarkan perhitungan kedua algoritma tersebut dengan perhitungan accuracy, precision, recall dan f-measurenya. Algoritma decision tree mempunyai hasil yang lebih baik dibandingkan dengan algoritma naïve bayes. Decision tree mempunyai akurasi sebesar 96,36% sedangkan naïve bayes sebesar 90,45%.

#### 4 KESIMPULAN

Data Diabetes yang menggunakan 520 data. Data tersebut dibagi menjadi 2, yaitu 300 data training dan 220 data. Pada hasil percobaan terlihat algoritma Decision Tree menghasilkan tingkat akurasi 96,36%. Hasil tersebut lebih baik dari menggunakan algoritma naïve bayes. Naïve bayes mempunyai akurasi sebesar 90,45%. Sehingga Algoritma decision tree mempunyai akurasi yang lebih besar dibandingkan dengan algoritma naïve bayes dalam menentukan prediksi diabetes.

#### DAFTAR PUSTAKA

- [1] M. Diana, E. Boyko, and I. Diabetes, *International Diabetes Federation*, vol. 10, no. 2. 2021.
- [2] R. W. et. Al, *IDF Diabetes Atlas 9th*, vol. 9. 2019.
- [3] Kementerian Kesehatan RI., “Infodatin tetap produktif, cegah, dan atasi Diabetes Melitus 2020,” *Pusat Data dan Informasi Kementerian Kesehatan RI*. pp. 1–10, 2020.
- [4] Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin, “Learning From Data.” p. 213, 2012.
- [5] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, and L. Vogt-maranto, “Machine learning and the physical sciences \*,” *Rev. Mod. Phys.*, vol. 91, no. 4, p. 45002, 2019, doi: 10.1103/RevModPhys.91.045002.
- [6] M. Batta, “Machine Learning Algorithms - A Review,” *Int. J. Sci. Res. (IJ)*, vol. 9, no. 1, pp. 381–386, 2020, doi: 10.21275/ART20203995.
- [7] A. Tahseen Ali, H. S. Abdullah, and M. N. Fadhil, “WITHDRAWN: Voice recognition system using machine learning techniques,” *Mater. Today Proc.*, Apr. 2021, doi: 10.1016/j.matpr.2021.04.075.
- [8] Sunardi, A. Yudhana, and S. Saifullah, “Identity analysis of egg based on digital and thermal imaging: Image processing and counting object concept,” *Int. J. Electr. Comput. Eng.*, vol. 7, no. 1, pp. 200–208, 2017, doi: 10.11591/ijece.v7i1.pp200-208.
- [9] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine learning in agriculture: A review,” *Sensors*, vol. 18, no. 8, pp. 1–29, 2018, doi: 10.3390/s18082674.
- [10] S. D. Mackowiak dkk., “Extensive identification and analysis of conserved small ORFs in animals,” *Genome Biol.*, vol. 16, no. 1, pp. 1–21, 2015, doi: 10.1186/s13059-015-0742-x.
- [11] B. Zhang dkk., “Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma,” *Cancer Lett.*, vol. 403, pp. 21–27, 2017, doi: 10.1016/j.canlet.2017.06.004.
- [12] A. U. Mazlan, N. A. b. Sahabudin, M. A. Remli, N. S. N. Ismail, M. S. Mohamad, and N. B. A. Warif, “Supervised and Unsupervised Machine Learning for Cancer Classification: Recent Development,” in *2021 IEEE International Conference on Automatic Control & Intelligent Systems (I2CACIS)*, 2021, pp. 392–395, doi: 10.1109/I2CACIS52118.2021.9495888.
- [13] T. Chauhan, S. Rawat, S. Malik, and P. Singh, “Supervised and Unsupervised Machine Learning based Review on Diabetes Care,” in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2021, vol. 1, pp. 581–585, doi: 10.1109/ICACCS51430.2021.9442021.
- [14] J. Rhee and J. Im, “Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data,” *Agric. For.*

- Meteorol.*, vol. 237–238, pp. 105–122, 2017, doi: 10.1016/j.agrformet.2017.02.011.
- [15] A. Aybar-Ruiz *dkk.*, “A novel Grouping Genetic Algorithm-Extreme Learning Machine approach for global solar radiation prediction from numerical weather models inputs,” *Sol. Energy*, vol. 132, pp. 129–142, 2016, doi: 10.1016/j.solener.2016.03.015.
- [16] R. Fragni, A. Trifirò, A. Nucci, A. Seno, A. Allodi, and M. Di Rocco, “Italian tomato-based products authentication by multi-element approach: A mineral elements database to distinguish the domestic provenance,” *Food Control*, vol. 93, no. April, pp. 211–218, 2018, doi: 10.1016/j.foodcont.2018.06.002.
- [17] C. Maione and R. M. Barbosa, “Recent applications of multivariate data analysis methods in the authentication of rice and the most analyzed parameters: A review,” *Crit. Rev. Food Sci. Nutr.*, vol. 59, no. 12, pp. 1868–1879, 2019, doi: 10.1080/10408398.2018.1431763.
- [18] F. Gorunescu, *Data Mining : Concepts, Models and Techinques*. Blue Publishing House, 2006.
- [19] D. Dumitru, “Prediction of recurrent events in breast cancer using the Naive Bayesian classification,” *Ann. Univ. Craiova - Math. Comput. Sci. Ser.*, vol. 36, pp. 92–96, 2009.
- [20] F. Hosseinzadeh, A. H. Kayvanjoo, and M. Ebrahimi, “Prediction of lung tumor types based on protein attributes by machine learning algorithms,” *Springerplus*, vol. 2, no. 1, pp. 1–14, 2013, doi: 10.1186/2193-1801-2-238.
- [21] P. Mehta *dkk.*, “A high-bias, low-variance introduction to Machine Learning for physicists,” *Phys. Rep.*, vol. 810, pp. 1–124, 2019, doi: 10.1016/j.physrep.2019.03.001.
- [22] S. Chen, G. I. Webb, L. Liu, and X. Ma, “A novel selective naïve Bayes algorithm,” *Knowledge-Based Syst.*, vol. 192, p. 105361, Mar. 2020, doi: 10.1016/j.knosys.2019.105361.
- [23] R. Agrawal, S. Ghosh, T. Imielinski, B. Iyer, and A. Swami, “An Interval Classifier for Database,” *Science (80-. )*, 1992.
- [24] J. R. Quinlan and R. L. Rivest, “Inferring Decision Description Trees Using the Minimum Length Principle,” *Inf. Comput.*, vol. 80, no. 1989, pp. 227–248, 1989.
- [25] G. Sarailidis, T. Wagener, and F. Pianosi, “Integrating Scientific Knowledge into Machine Learning using Interactive Decision Trees,” *Comput. Geosci.*, vol. 170, no. March 2022, p. 105248, 2021, doi: 10.1016/j.cageo.2022.105248.
- [26] N. Sneha and T. Gangil, “Analysis of diabetes mellitus for early prediction using optimal features selection,” *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0175-6.