



Prediksi Morfologi Jamur Menggunakan Algoritma C5.0

Lita Karlitasari¹⁾, I Wayan Sriyasa²⁾, Irfan Wahyudi³⁾, Heri Bambang Santosi⁴⁾

^{1,2,3,4}*Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Pakuan*
^{1,2,3,4}*Jl. Pakuan No.1, Ciheuleut, Bogor, 16143*

Email: ¹*lita.karlitasari@unpak.ac.id*, ²*iwayan.sriyasa@unpak.ac.id*, ³*irfan.wahyudi@unpak.ac.id*,
⁴*heri.bambang santosi@unpak.ac.id*

Abstract

Mushrooms are divided into two, namely edible mushrooms and poisonous mushrooms. to distinguish between types of mushrooms that can be consumed from those that are poisonous is a complicated matter because of the nearly identical forms of the various species of mushrooms. One way to identify mushrooms is by way of morphology later which will be able to know what class a fungus belongs to. The algorithm is used because it is faster, because it is multithreaded, but otherwise produces exactly the same classifiers. shows that the C5.0 algorithm has a very good accuracy value of 99-100% with a fairly good amount of power, using both split validation and cross-validation. And also almost all attributes have a significant relationship or influence, because they have an expected value of less than 5%. Using keywords related to the research topic is recommended.

Keyword: *Prediction, Mushroom Morphology, C5.0. Algorithm*

Abstrak

Jamur terbagi dua yaitu jamur yang dapat dikonsumsi serta jamur yang beracun. untuk membedakan antara jenis jamur yang dapat dikonsumsi dengan yang beracun merupakan hal yang rumit karena bentuk yang hampir sama dari beragam spesies jamur tersebut. Salah satu cara untuk mengidentifikasi jamur yaitu dengan cara morfologi nantinya yang akan dapat diketahui termasuk kelas apakah suatu jamur tersebut. Algoritma digunakan karena lebih cepat, karena multithreaded, tetapi sebaliknya menghasilkan pengklasifikasi yang persis sama. menunjukkan bahwa algoritma C5.0 memiliki nilai akurasi yang sangat baik yaitu sebesar 99-100% dengan jumlah daya yang cukup baik, baik menggunakan split validation maupun cross-validation. Dan juga hampir semua atribut memiliki hubungan atau pengaruh yang signifikan, karena memiliki nilai expected kurang dari 5%. dalam abstrak. Menggunakan kata kunci yang terkait dengan topik penelitian direkomendasikan.

Kata Kunci: *Prediksi, Morfologi jamur, Algoritma c5.0*

1. PENDAHULUAN

Jamur merupakan tumbuhan yang termasuk kedalam kingdom *Fungi*. Pada umumnya, jamur terbagi dua yaitu jamur yang dapat dikonsumsi serta jamur yang beracun. Jamur yang dapat dikonsumsi menjadi bahan baku pembuatan makanan di Indonesia. Namun, untuk membedakan antara jenis jamur yang dapat dikonsumsi dengan yang beracun merupakan hal yang rumit karena bentuk yang hampir sama dari beragam spesies jamur tersebut [1]. Salah satu cara untuk mengidentifikasi jamur yaitu dengan cara morfologi. Ciri morfologi yang dimaksud adalah bentuk payung, warna, tekstur payung, dan ciri lain yang terlihat. Ekstraksi ciri morfologi ini digunakan untuk membantu identifikasi jamur, sehingga nantinya akan dapat diketahui termasuk kelas apakah suatu jamur tersebut [2].

Data mining merupakan salah satu alternatif sebuah proses yang dilakukan dalam menyelesaikan sebuah permasalahan. Penerapan proses data mining akan membantu dalam memproses keseluruhan data yang sudah tersimpan untuk menghasilkan sebuah informasi yang baru dan bermanfaat [3], [4]. Data mining merupakan sebuah proses ekstraksi dari sekelompok data yang akan menghasilkan sebuah informasi yang sangat berharga bagi sebuah organisasi baik instansi pemerintahan, instansi swasta, maupun perusahaan. Dalam proses pengumpulan data serta ekstraksi informasi dalam data mining dapat dilakukan dengan menggunakan statistik, matematika, atau *artificial intelligence* (AI). Untuk mengaplikasikan *data mining* tentunya memiliki beberapa tahapan proses di dalamnya. Tahapan-tahapan tersebut bersifat saling aktif yang dapat di artikan bahwa pengguna atau user mendapatkan peran secara langsung atau dengan perantara pengetahuan dasar yang di dapatkan.

Klasifikasi dalam data mining merupakan sebuah proses dalam menemukan kesamaan karakteristik dalam sebuah kelompok atau kelas. Metode klasifikasi ini dilakukan dengan tujuan untuk memprediksi kelas dalam sebuah objek yang



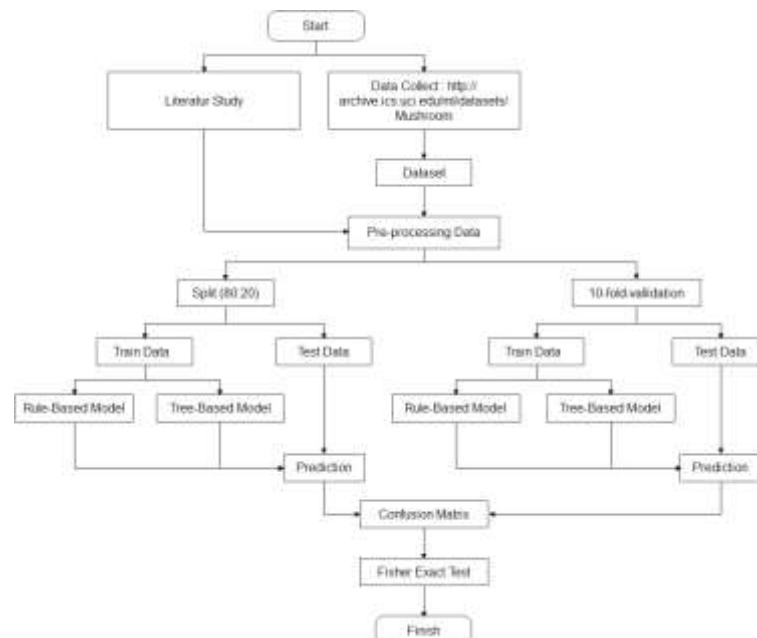
belum diketahui label dari objek tersebut. Klasifikasi dapat digunakan untuk melakukan prediksi terhadap morfologi jamur, salah satu bentuk klasifikasi adalah metode *decision tree* yang memiliki kemampuan memproses data yang diskrit dan kontinyu adalah C5.0 *algorithm*. Keunggulan dari C5.0 *algorithm* adalah lebih cepat, karena *multithreaded*, tetapi sebaliknya menghasilkan pengklasifikasi yang persis sama. Selain itu juga dapat menghasilkan kumpulan aturan, dan peningkatannya cukup besar [5].

Decision tree merupakan sebuah algoritma dalam *machine learning* dengan menggunakan aturan untuk membuat atau menentukan sebuah keputusan[1], [6]. *Decision tree* memiliki struktur seperti pohon yang sangat memungkinkan hasil, biaya, utility, serta kemungkinan resiko. Konsep dari *decision tree* yaitu dengan cara menampilkan sebuah algoritma dengan pernyataan bersyarat dari sebuah *node* atau cabang untuk mewakili dari langkah-langkah dalam menghasilkan sebuah keputusan. Algoritma C5.0 merupakan salah satu metode klasifikasi dari *decision tree* sehingga menerapkan struktur pohon atau struktur berhierarki dalam mengambil keputusan[7], [8]. Algoritma C5.0 merupakan pengembangan dari metode *decision tree* yang sebelumnya, yaitu algoritma ID3 dan C4.5. Sehingga kinerja algoritma ini lebih unggul dari algoritma sebelumnya dengan adanya fase *boosting* di proses terakhir dalam perhitungan algoritma C5.0.

Dari beberapa penelitian yang dilakukan, semuanya menunjukkan tingkat akurasi 100% pada algoritma, seperti ID3, AdaBoost, dan KNN. Tetapi data sampel yang digunakan masih sedikit dan waktu yang digunakan masih cukup lama [1], [9], [2]. Sementara itu algoritma C5.0 telah cukup banyak digunakan pada obyek penelitian lain, dengan tingkat akurasi antara 80-99% [7], [10], [8], [11]. Dengan demikian maka dibuatlah penelitian yang memprediksi morfologi jamur menggunakan algoritma C5.0. dimana dengan dataset yang ada akan coba dibuktikan berapa tingkat akurasi dan pengaruh setiap atribut terhadap class yang dituju.

2. METODE PENELITIAN

Metode penelitian yang digunakan dalam tugas penelitian ini adalah menggunakan tahapan melalui tahapan[12] pada gambar 1:



Gambar 1. Metode Penelitian

2.1 Classification

Model klasifikasi bertujuan untuk mengelompokkan data ke dalam kelas-kelas yang telah memiliki label dengan membangun suatu model yang berdasarkan kepada suatu *data mining* serta memprediksi kelas dari suatu data baru. *Data training* adalah suatu set data yang telah diketahui klasifikasinya. Untuk melakukan validasi model perlu menghitung segala bentuk tingkat kesalahan prediksi pada model. Terdapat dua teknik validasi yang bisa digunakan yaitu : *split validation* dan *cross validation*. *Split validation* dilakukan dengan membagi dataset secara acak menjadi dua data terpisah, yaitu data latuh dan data uji. Sementara *cross validation* melakukan validasi berulang-ulang dimana dataset dibagi menjadi



banyak subset data latih dan validasi. Setiap iterasi memvalidasi satu subset dengan subset yang tersisa sebagai data latih [5]

2.2. Decision Tree

Decision Tree adalah salah satu metode klasifikasi berupa pohon keputusan secara *top-down*. Caranya dengan mengevaluasi semua atribut menggunakan suatu ukuran statistik, berupa *information gain*, untuk mengukur efektifitas suatu atribut dalam mengklasifikasikan himpunan sampel data. Algoritma *decision tree* secara otomatis akan menentukan variabel yang terpenting berdasarkan kemampuan mereka menyortir data menjadi kelas yang benar. Kelebihan *decision tree* adalah pohon yang dihasilkan dapat disajikan ke dalam bentuk yang dipahami manusia, mampu meranking yang paling baik sebagai *predictor*. Setelah model pohon terbentuk, proses komputasi cepat untuk mengklasifikasikan data baru menggunakan model. Serta dapat menangani data numerik dan data kategorikal dengan baik [5]

2.3. Algoritma C5.0

Algoritma C5.0 merupakan pengembangan dari algoritma C4.5 yang merupakan algoritma berbasis *decision tree*. Algoritma C5.0 cocok untuk kumpulan data besar. Algoritma 5.0 lebih baik daripada C4.5 dari sisi kecepatan, memori dan efisiensi. Jika pada algoritma C4.5 berhenti sampai perhitungan *gain*, maka pada algoritma C5.0 akan melanjutkannya dengan menghitung *gain ratio* dengan menggunakan *gain* dan *entropy* yang telah ada [5]

2.4. Confusion Matrix

Confusion Matrix merepresentasikan prediksi dan kondisi sebenarnya (aktual) dari data yang dihasilkan oleh algoritma ML. Berdasarkan Confusion Matrix, kita bisa menentukan Accuracy, Precision, Recall dan Specificity [5].

Accuracy merupakan rasio prediksi Benar (positif dan negatif) dengan keseluruhan data.

$$Accuracy = (TP + TN) / (TP + FP + FN + TN) \quad (1)$$

Precision merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif.

$$Precision = (TP) / (TP + FP) \quad (2)$$

Recall merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif.

$$Recall = (TP) / (TP + FN) \quad (3)$$

Specificity merupakan kebenaran memprediksi negatif dibandingkan dengan keseluruhan data negatif.

$$Specificity = (TN) / (TN + FP) \quad (4)$$

F-Measure merupakan perbandingan rata-rata presisi dan recall yang dibobotkan

$$F1\ Measure = 2 * (Recall * Precision) / (Recall + Precision) \quad (5)$$

2.5. Chi-Square dan Fisher Exact Test

Chi-square (Kai Kuadrat) berguna untuk menguji hubungan atau pengaruh dua buah variabel nominal dan mengukur kuatnya hubungan antara variabel satu dengan variabel nominal lainnya. Syarat uji *chi-square* adalah tidak ada *cell* dengan frekuensi kenyataan (*observed*) sebesar 0 (nol). Apabila bentuk tabel kontigensi 2x2, maka tidak boleh ada 1 *cell* yang memiliki frekuensi harapan (*expected*) kurang dari 5. Apabila bentuk tabel lebih dari 2x2, misali 2x3, maka jumlah *cell* dengan *expected* yang kurang dari 5 tidak boleh dari 20%, dan diganti dengan *fisher exact test*. Rumus chi square :

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad (6)$$

Dimana χ^2 = nilai chi square

O = observed

E = expected

Dengan derajat kebebasan $df = (row-1)(column-1)$

Kemudian dibuat hipotesis terlebih dahulu dengan merumuskan H_0 untuk independen dan H_a untuk tidak independen [5]

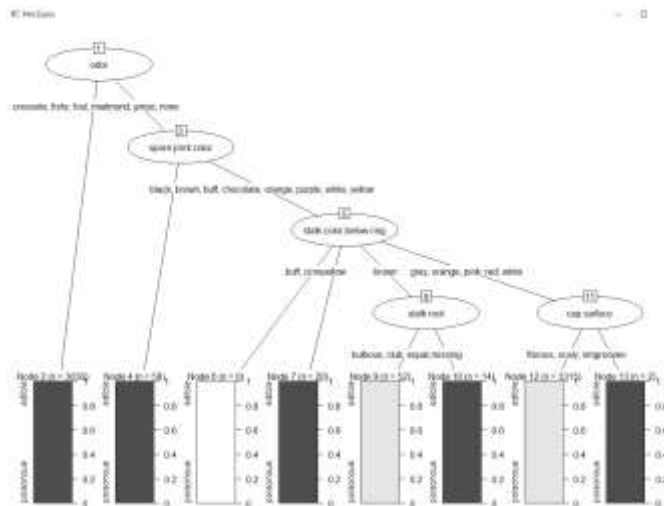
3. HASIL DAN PEMBAHASAN

Data jamur yang akan dianalisis merupakan seluruhnya adalah atribut kategorikal, terlebih dahulu dilakukan *pre-processing data*, dengan tujuan membersihkan data agar tidak ada data yang hilang atau tidak konsisten. Kalau pun ada



data yang hilang maupun tiak konsisten, maka dilakukan penanganan terlebih dahulu. Dari dataset tersebut terdapat 8124 atribut dengan jumlah atribut 22 termasuk kelas yang akan dituju. Ada pun atribut-atribut tersebut adalah *cap-shape*, *cap-surface*, *cap-color*, *bruisses*, *odor*, *gill-attachement*, *gill-spacing*, *gill-size*, *gill-color*, *stalk-shape*, *stalk-root*, *stalk-surface-above-ring*, *stalk-surface-bellow-ring*, *stalk-color-above-ring*, *stalk-color-bellow-ring*, *veil-color*, *ring-number*, *ring-type*, *spore-print color*, *population*, dan *habitat*. Apabila data sudah bersih, tahap selanjutnya dilakukan proses klasifikasi menggunakan Algoritma C5.0. Terdapat 2 teknik validasi yang digunakan yaitu *split validation*, yang kali ini menggunakan rasio 80% untuk data latih dan 20% untuk data uji, dan *10-fold-validation*. Penetapan kelas telah dilakukan terlebih dahulu dengan atribut *classes*, yang berisi apakah jamur beracun (*poisonous*) atau tidak beracun (*edible*).

Untuk teknik *split validation* 80:20, pada tree-based model dengan menggunakan 6491 data dan 21 predictor, dihasilkan *trees-size* sebesar 8, seperti pada gambar 2.



Gambar 2. Split validation 80:20 tree-based model

Adapun atribut terbentuk adalah odor (100%), spore-print-color (53,322%), stalk-color-bellow-ring (52,43%), cap-surface (51,10%), dan stalk-root (1,02%). Waktu yang dihasilkan 0 detik. Untuk prediksi data uji, 844 jamur benar masuk ke dalam jamur yang edible, dan hanya 1 saja yang masuk ke dalam jamur poisonous, sementara 788 jamur semuanya benar termasuk ke dalam jamur *poisonous*. Untuk rule-based model dengan menggunakan 6491 data dan 21 predictor, dihasilkan *trees-size* sebesar 6 rule, seperti gambar 3.

```

rule 1: (1315/3, lift 1.0)
  cap.surface in {fibrous, scaly, smooth}
  odor in {almond, anise, none}
  stalk.color.below.ring in {gray, orange, pink, red, white}
  spore.print.color in {black, brown, buff, chocolate, orange, purple,
    white, yellow}
  => class edible [0.999]

rule 2: (3818/11, lift 1.0)
  odor in {almond, anise, none}
  stalk.root in {bulbous, club, equal, rooted}
  spore.print.color in {black, brown, purple, white}
  => class edible [0.990]

rule 3: (3030, lift 2.1)
  odor in {creosote, fishy, foul, musty, pungent, spicy}
  => class poisonous [1.000]

rule 4: (58, lift 2.0)
  spore.print.color = green
  => class poisonous [0.981]

rule 5: (20, lift 2.0)
  stalk.color.below.ring = yellow
  => class poisonous [0.955]

rule 6: (14, lift 1.9)
  stalk.root = missing
  stalk.color.below.ring = brown
  => class poisonous [0.938]

Default class: edible

```

Gambar 3. Split validation 80:20 rules-based model

Adapun atribut terbentuk adalah odor (98,68%), spore-print-color (52,89%), stalk-color-bellow-ring (51,59%), cap-surface (51,07%), dan stalk-root (43, 63%). Waktu yang dihasilkan 0 detik. Untuk prediksi data uji, 844 jamur benar masuk ke dalam jamur yang edible, dan 5 masuk ke dalam jamur poisonous, sementara 784 jamur semuanya benar termasuk ke dalam jamur *poisonous*.



TABEL 1. Confusion Matrix

	Accuracy	Error Rate	Precision	Recall/ Sensitivity	Spesifisitas	F- Measure
C5.0 Tree-based model (Split 80:20)	100%	52%	100%	100%	100%	100%
C5.0 Rule-based model (Split 80:20)	100%	52%	99%	100%	99%	100%
C5.0 Tree-based model (10-fold validation)	100%	44%	100%	100%	100%	100%
C5.0 Rule-based model (10-fold validation)	99%	44%	98%	100%	99%	99%

Dan untuk melihat keeratan antara atribut dengan kelas, maka telah dilakukan uji *chi-square* dan *fisher exact test*

TABEL 2. Ringkasan Uji Chi-Square dan Uji Fisher Exact Test

	Nilai statistik chi-square	df	p-value (asyp)	p-value (exact)	Hasil	Analisa
cap-shape	489,92	5	< 0,001	-	2 cells (16,7%) expected < 5	Terdapat hubungan
cap-surface	3155,043	3	< 0,001	< 0,001	2 cells (25%) expected < 5	Terdapat hubungan
cap-color	387,598	9	< 0,001	-	0 cells (0%) expected < 5	Terdapat hubungan
bruisses	2043,452	1	0,000	0,000	0 cells (0%) expected < 5	Terdapat hubungan
odor	7659,727	7	0,000	-	0 cells (0%) expected < 5	Terdapat hubungan
gill-attachment	135,611	1	< 0,001	< 0,001	0 cells (0%) expected < 5	Terdapat hubungan
gill-spacing	987,037	1	< 0,001	< 0,001	0 cells (0%) expected < 5	Terdapat hubungan
gill-size	2369,172	1	0,000	0,000	0 cells (0%) expected < 5	Terdapat hubungan
gill-color	2765,714	11	0	-	0 cells (0%) expected < 5	Terdapat hubungan
stalk-shape	84,554	1	< 0,001	< 0,001	0 cells (0%) expected < 5	Terdapat hubungan
stalk-root	1344,441	4	< 0,001	-	0 cells (0%) expected < 5	Terdapat hubungan
stalk-surface-above-ring	2808,286	3	0,000	0,000	0 cells (0%) expected < 5	Terdapat hubungan
stalk-surface-bellow-ring	2684,274	3	0,000	0,000	0 cells (0%) expected < 5	Terdapat hubungan
stalk-color-above-ring	0,247	2	0,884	0,884	0 cells (0%) expected < 5	Terdapat hubungan
stalk-color-bellow-ring	2152,391	8	0,000	0,000	0 cells (0%) expected < 5	Terdapat hubungan
veil-color	191,224	3	< 0,001	< 0,001	2 cells (25%) expected < 5	Terdapat hubungan
ring-number	374,737	2	< 0,001	< 0,001	0 cells (0%) expected < 5	Terdapat hubungan
ring-type	2956,619	4	0,000	-	0 cells (0%) expected < 5	Terdapat hubungan
spore-printt-color	4602,033	8	0,000	-	0 cells (0%) expected < 5	Terdapat hubungan
population	1929,741	5	0,000	-	0 cells (0%)	Terdapat



habitat	1573,777	6	0,000	-	expected < 5 0 cells (0%) expected < 5	hubungan Terdapat hubungan
---------	----------	---	-------	---	--	----------------------------------

Berdasarkan hasil Analisa uji chi-square dan uji fisher di atas, menunjukkan Sebagian besar atribut menunjukkan adanya hubungan significant atau pengaruh terhadap atribut *classes*, kecuali atribut *stalk-color-above-ring*, karena nilai *expected* lebih besar dari 5. Yang bisa menggunakan nilai chi-square hanya atribut *bruises*, *gill-attachment*, *gill-spacing*, *gill-size*, dan *stalk shape* karena memiliki tabel kontigensi 2x2, dan tidak terdapat cell dengan dengan nilai *expected* kurang dari 5. Sedangkan yang lain memiliki table kontigensi lebih dari 2x2 sehingga menggunakan uji *fisher test*, tetapi yang memiliki cell dengan nilai *expected* kurang dari 5 tidak lebih dari 20%, hanya atribut *cap-shape*.

4. KESIMPULAN

Berdasarkan hasil dan pembahasan di atas, menunjukkan bahwa algoritma C5.0 memiliki nilai akurasi yang sangat baik yaitu sebesar 99-100% dengan jumlah data yang cukup baik, baik menggunakan *split validation* maupun *cross-validation*. Dan juga hampir semua atribut memiliki hubungan atau pengaruh yang signifikan, karena memiliki nilai *expected* kurang dari 5%. Dengan demikian algoritma C5.0 bisa digunakan untuk data yang sangat banyak. Ke depannya, bisa digunakan *boosting*, yang merupakan pendekatan iteratif untuk menghasilkan pengklasifikasi yang kuat, yang mampu mencapai kesalahan training seminimal mungkin dari sekelompok pengklasifikasi yang lemah.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini, diantaranya :

1. Dekan Fakultas Matematika dan Pengetahuan Alam (FMIPA) – Universitas Pakuan,
2. Wakil Dekan Bidang Akademik dan Kemahasiswaan – FMIPA – Universitas Pakuan,
3. Asisten Dekan Bidang Penelitian, Pengabdian pada Masyarakat, dan Publikasi – FMIPA - Universitas Pakuan,,
4. Ketua Program Studi Ilmu Komputer – FMIPA – Universitas Pakuan,,
5. Divisi Penelitian dan Pengabdian pada Masyarakat, Prodi Ilmu Komputer – FMIPA - Universitas Pakuan,
6. Dosen-dosen Prodi Ilmu Komputer,

Yang telah banyak membantu hingga terselesaikan penelitian ini.

DAFTAR PUSTAKA

- [1] N. Syafitri and J. E. Sari, "Sistem Klasifikasi Jamur Dengan Algoritma Iterative Dichotomiser 3," *It J. Res. Dev.*, vol. 1, no. 1, pp. 27–37, 2017, doi: 10.25299/itjrd.2016.vol1(1).672.
- [2] A. Zubair and A. R. Muslikh, "Identifikasi jamur menggunakan metode k-nearest neighbor dengan ekstraksi ciri morfologi," *Semin. Nas. Sist. Inf.*, no. September, pp. 965–972, 2017.
- [3] A. R. Isnain, A. I. Sakti, D. Alita, and N. S. Marga, "SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM," *J. Data Min. dan Sist. Inf.*, vol. 2, no. 1, pp. 31–37, 2021.
- [4] F. Gorunescu, *Data mining: Concepts, models and techniques*, vol. 12. 2011.
- [5] M. J. Zaki and W. J. Meira, *Data Mining and Machine Learning Fundamental Concepts and Algorithms*, vol. 53, no. 9. 2020.
- [6] A. Nurkholis and T. Susanto, "Algoritme Spatial Decision Tree Untuk Evaluasi Kesesuaian Lahan Padi Sawah Irigasi," *J. RESTI (Rekayasa Sist. Dan Teknol. Informasi)*, vol. 4, no. 5, pp. 978–987, 2020.
- [7] R. Revathy, S. Balamurali, and R. Lawrance, "Classifying Agricultural Crop PestData Using Hadoop MapReduceBased C5.0 Algorithm," *J. Cyber Secur. Mobil.*, vol. 8, pp. 393–408, 2019.
- [8] N. H. Harani and F. S. Damayanti, "Implementasi Algoritma C5.0 Untuk Menentukan Pelanggan Potensial Di Kantor Pos Cimahi," *J. SITECH Sist. Inf. dan Teknol.*, vol. 4, no. 1, pp. 69–76, 2021, doi: 10.24176/sitech.v4i1.6281.
- [9] I. P. Ihsan and F. Yusuf, "Analisis Jamur Beracun Berdasarkan Ciri Menggunakan Algoritma AdaBoost," *Pros.*



- Konf. Ilmu Komput. Nas. 2013*, no. ISSN 2338-2899, pp. 1–6, 2013.
- [10] Y. Susanti *et al.*, “Classification of teak wood production in Central Java using the C5.0 algorithm,” *AIP Conf. Proc.*, vol. 2202, no. December, pp. 0–6, 2019, doi: 10.1063/1.5141707.
- [11] M. S. Sungkar and M. T. Qurohman, “Penerapan Algoritma C5.0 Untuk Prediksi Kelulusan Pembelajaran Mahasiswa Pada Matakuliah Arsitektur Sistem Komputer,” *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 1166, 2021, doi: 10.30865/mib.v5i3.3116.
- [12] S. Ahdan and S. Setiawansyah, “Android-Based Geolocation Technology on a Blood Donation System (BDS) Using the Dijkstra Algorithm,” *IJAIT (International J. Appl. Inf. Technol.)*, pp. 1–15, 2021.