

DIAGNOSA HAMA DAN PENYAKIT PADA TANAMAN PADI MENGGUNAKAN METODE *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOR*

Restanti Marlina Bianome¹, Derwin R. Sina² dan Yelly Y. Nabuasa³

^{1,2,3} Program Studi Ilmu Komputer, Fakultas Sains dan Teknik, Universitas Nusa Cendana

¹Email: tanntvbianome@gmail.com

²Email: derwinilkom@gmail.com

³Email: yelly.yosiana.n@gmail.com

ABSTRAK

Penelitian ini membangun sistem *Case Based Reasoning* (CBR) untuk mendiagnosis hama dan penyakit pada tanaman padi menggunakan algoritma *Naïve Bayes* dan *K-nearest neighbor*. CBR merupakan salah satu metode pemecahan masalah dengan pengambilan keputusan kasus baru berdasarkan solusi dari kasus-kasus sebelumnya dengan cara menghitung tingkat kemiripan (*similarity*). Topik permasalahan yang dikaji terdiri dari 13 jenis hama dan 10 jenis penyakit tanaman padi. Tingkat kemiripan dapat ditentukan dengan cara *indexing* dan *nonindexing*. *Indexing* merupakan proses pengelompokkan kasus yang ada berdasarkan kelas yang telah ditentukan, sedangkan *nonindexing* merupakan proses tanpa pengelompokkan kasus. Berdasarkan pengujian menggunakan *cross validation* didapatkan nilai rata-rata akurasi 92,88% terhadap 153 data uji pada pengujian menggunakan *indexing* dan nilai rata-rata akurasi 89,63% terhadap 153 data uji pada pengujian menggunakan *nonindexing*. Kata kunci : *case based reasoning*, padi, hama dan penyakit, *k-nearest neighbor*, *naïve Bayes*, *indexing*, *nonindexing*

ABSTRACT

This study builds systems *Case Based Reasoning* (CBR) to diagnose pests and diseases in rice plants using *Naïve Bayes* algorithm and *K-nearest neighbor*. CBR is one method of solving the problem with new cases of decision making based on the solution of previous cases by calculating the degree of similarity (*similarity*). The case consists of 13 species and 10 types of disease pests of rice plants. The degree of similarity can be determined by *indexing* and *nonindexing*. *Indexing* is the process of grouping the cases by classes that have been determined, while *nonindexing* a process without grouping cases. Based on *cross validation* testing using average values obtained accuracy of 92.88% to 153 test data on testing using the *indexing* and the average value of 89.63% accuracy of the test data in the test 153 using *nonindexing*.

Keywords: *case based reasoning*, rice, pests and diseases, *k-nearest neighbor*, *Naïve Bayes*, *indexing*, *nonindexing*

1. PENDAHULUAN

Padi merupakan bahan baku pangan pokok bagi rakyat Indonesia dan setiap faktor yang mempengaruhi tingkat produksinya sangat penting diperhatikan, salah satunya faktor hama dan penyakit atau organisme mengganggu tanaman (OPT). Terhadap OPT, perlindungan tanaman pangan berperan dalam mengelola OPT agar tidak menimbulkan kerusakan dan kerugian.

Terdapat beberapa masalah yang sering terjadi di lapangan yaitu minimnya pengetahuan yang dimiliki oleh petani. Sesuai dengan data dari Unit Pelaksana Teknis Daerah Proteksi Tanaman Pangan dan Hortikultura Dinas Pertanian dan Perkebunan Provinsi NTT tahun 2018, tenaga pengamat hama yang tersedia hanya berjumlah 4 orang. Hal ini berdampak pada penyebaran tenaga pengamat yang tersebar di Kabupaten Kupang tidak merata. Peranan pengamat hama sangat dibutuhkan karena serangan hama penyakit yang terjadi harus langsung ditangani oleh pakar (pengamat hama). Hal ini mengakibatkan pengamat hama yang ada tidak dapat menjangkau semua petani di seluruh wilayah Kabupaten Kupang dengan maksimal dalam rentang waktu singkat.

Adapun dipilihnya algoritma *naïve Bayes* dan *k-nearest neighbor* sebagai algoritma klasifikasi yang dibandingkan karena tingkat akurasi dari kedua algoritma ini relatif tinggi.

2. MATERI DAN METODE

Data Penelitian

Data yang diperoleh sebanyak 153 data yang terdiri dari 10 jenis hama sesuai tabel 1, ada 13 jenis penyakit sesuai tabel 2 dan 70 gejala sesuai tabel 3. Data yang diambil pada tahun 2016-2018. Dari 153 data yang digunakan 70% data merupakan data latih dan 30% data merupakan data uji.

Tabel 1. Jenis hama padi

Hama Padi	Kode
Penggerek batang padi	H1
Hama putih	H2
Hama putih palsu	H3
Wereng coklat	H4
Walang sangit	H5
Kepinding tanah	H6
Lalat pucuk padi	H7
Ganjur	H8
Keong mas/siput	H9
Anjing tanah	H10

Tabel 2. Jenis penyakit padi

Penyakit Padi	Kode
Blas	P1
Hawar pelepah	P2
Bercak daun coklat	P3
Bercak daun bergaris	P4
Kerdil hampa	P5
Kerdil rumput	P6
Bakanae	P7
Bakteri daun jingga	P8
Tungro	P9
Lapuk daun	P10
Bakteri daun bergaris	P11
Kresek/hawar daun	P12
Nematoda	P13

Tabel 3. Gejala

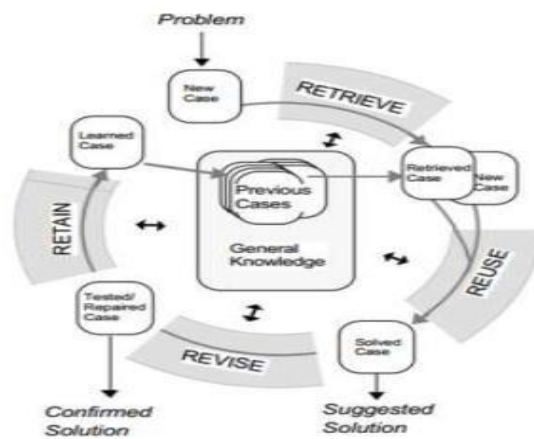
Gejala	Kode
Sebelum padi berbunga disebut sundep	g1
Padi berbunga/malai disebut beluk	g2
Telur diletakkan pada batang tanaman padi	g3
Pucuk batang padi kering-mudah di cabut dan berwarna kuning	g4
Bulir padi hampa	g5
Daun berwarna putih kering	g6
Tinggal tulang daun	g7
Pucuk daun terlihat berbekas potong	g8
Terdapat ulat pada gulungan daun sampai jadi pupa	g9
terdapat pada lipatan daun	g10
menyerang daun bendera, sehingga hanya tinggal tulang daun	g11
terdapat spot-spot pada tanaman padi	g12
menghisap cairan sel tanaman sehingga tanaman kering dan menimbulkan efek seperti terbakar	g13
siklus hidup sekitar 25 hari	g14
menghisap bulir padi yang sedang masak susu	g15
malai yang dihisap menjadi hampa	g16

Gejala	Kode
hilangnya cairan menjadi hampa dan berwarna cokelat kehitaman	g17
menghisap cairan pelepah dan batang yang menyebabkan warna cokelat disekitar bagian yang dihisap	g18
tanaman tumbuh terhambat, berwarna kekuning-kuningan dan akhirnya mati membusuk	g19
telur diletakkan satu persatu pada permukaan atas atau bawah daun	g20
larva tinggal dipucuk daun yang belum terbuka	g21
larva serangga ganjur memakan tanaman padi pada titik tumbuh yang menyebabkan daun tumbuh berbentuk gulungan seperti daun bawang	g22
tunas padi yang tumbuh berbentuk pentil	g23
anakan yang terserang ganjur tidak dapat membentuk malai	g24
menyerang memakan daun, batang dan akar tanaman padi yang masih muda	g25
tanaman padi menjadi terpotong dan putus	g26
tanaman menjadi layu dan akhirnya mati	g27
lebih menyukai tanaman berakar serabut	g28
aktif pada malam hari	g29
tanaman yang terserang biasanya yang paling pinggir atau gundukan-gundukan tanaman ditengah sawah	g30
Pada daun timbul bercak oval/ellips	g31
Ujung-ujungnya meruncing mirip belah ketupat	g32
Dapat menyerang buku tanaman, malai dan gabah	g33
Stadia kritis tanaman terjadi mulai umur 1 bulan(padi gogo, anakan maksimum, bunting dan awal berbunga	g34
Bercak berbentuk lonjong, berwarna kelabu berukuran mulai 10mm-3cm	g35
Sklerotia yang berbentuk bulat berwarna putih sampai cokelat	g36
Stadia kritis tanaman adalah mulai persemaian sampai menjelang panen	g37
Bercak berwarna cokelat tua, berbentuk oval sampai bulat, berukuran sebesar biji wijen dengan titik tengah berwarna abu-abu atau putih	g38
Dapat pula muncul pada malai, pelepah dan gabah	g39
Stadia kritis tanaman terjadi mulai persemaian (dari gabah teinfeksi berat) sampai menjelang panen	g40
Daun timbul bercak sempit dan berwarna cokelat	g41
Bercak makin ketepi, warna daun makin pucat	g42
Pada varietas rentan, bercak lebih besar dengan pucat bercak yang lebih kecil dan berwarna terang	g43
Pada vase vegetatif tanaman yang terinfeksi menjadi kerdil	g44
Pada vase generatif daun bendera terpuntir, robek-robek atau berombak	g45
Anakan bercabang, pembungaan terhambat, pertumbuhan malai tidak sempurna dan hampa	g46
Virus ditularkan oleh wereng cokelat	g47
Tanaman kerdil, jumlah anakan banyak, tidak produktif dan sempit	g48
Berwarna kekuning-kuningan dengan bercak-bercak cokelat	g49
Kehilangan hasil makin tinggi bila tanaman terinfeksi pada umur kurang dari 60 hst	g50
Bibit yang terinfeksi terkadang 2 kali lebih tinggi dari pada tananam normal	g51
Daun tipis hijau kekuningan	g52
Tanaman memanjang mati, gejalanya berupa tanaman kerdil dan membusuk pangkalnya atau bibit mati	g53
Daun bendera yang hijau pucat berbeda jelas dibandingkan warna tanaman umumnya	g54
Daun ditandai oleh timbulnya bercak berbentuk bulat atau bulat telur berwarna merah kekuningan atau merah cokelat kekuningan	g55
Daun bergaris merah dan mengering	g56
Masa pembungaan menjadi tidak serempak sehingga proses pematangan bulir menjadi tidak serempak	g57
Pertumbuhan tanaman terhambat, kerdil dan jumlah anakan berkurang	g58
Daun menguning sampai jingga dari mulai pucuk kearah pangkal	g59

Gejala	Kode
Ada titik-titik sklerotium(bintik hitam)	g60
Bercak bercincin	g61
Bercak berbentuk jorong	g62
Terdapat bercak pada daun	g63
Munculnya bintik-bintik pada daun yang sempit dan gelap/gelap kuning hijau dan buram	g64
Pucuk daun menguning kemudian menjalar melalui pingir daun hingga ke pangkal	g65
Pada serangan berat, daun padi akan tampak mengering	g66
Munculnya puru akar, luka pada akar, ujung akar rusak dan akar membusuk	g67
Pertumbuhan tanaman terhambat karena tidak dapat menyerap unsur hara	g68
Busuk akar/umbi	g69
Percabangan akar yang berlebihan dan ujung akar yang mati	g70

Case Base Reasoning

Case-Based Reasoning adalah suatu pendekatan untuk menyelesaikan suatu permasalahan (*problem solving*) berdasarkan solusi dari permasalahan sebelumnya. Penjelasan singkat siklus CBR terlihat pada gambar 1.



Gambar 1. Siklus CBR

Case-Based Reasoning ada empat siklus yang meliputi

1. *Retrieve*
Mendapatkan/memperoleh kembali kasus yang paling menyerupai/relevan (*similar*) dengan kasus yang baru.
2. *Reuse*
Memodelkan/menggunakan kembali pengetahuan dan informasi kasus lama berdasarkan bobot kemiripan yang paling relevan ke dalam kasus yang baru.
3. *Revise*
Meninjau kembali solusi yang diusulkan kemudian mengetesnya pada kasus nyata (simulasi).
4. *Retain*
Mengintegrasikan/menyimpan kasus baru yang telah berhasil mendapatkan solusi agar dapat digunakan oleh kasus-kasus selanjutnya yang mirip dengan kasus tersebut.

Metode

Pada penelitian ini tahap *indexing* menggunakan metode *naïve Bayes* dan *k-nearest neighbor*. Algoritma *naïve Bayes* merupakan sebuah metode klasifikasi menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma *Naïve Bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema *Bayes*. Algoritma menggunakan teorema *Bayes* dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Metode ini hanya membutuhkan jumlah data pelatihan yang kecil untuk menentukan estimasi

parameter yang diperlukan dalam proses pengklasifikasian. Secara umum metode *naïve Bayes* ditunjukkan dalam persamaan (1).

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \dots\dots\dots (1)$$

dimana: $P(E|H) * P(H) = P(H \cap E)$ sehingga $P(H|E) = \frac{P(H \cap E)}{P(E)}$

Keterangan:

H = hipotesis H

E = *evidence* E

P(H|E) = Probabilitas hipotesis H terjadi jika *evidence* E terjadi

P(E|H) = Probabilitas munculnya *evidence* E, jika hipotesis H terjadi

P(H) = Probabilitas hipotesis H tanpa memandang *evidence* apapun

P(E) = Probabilitas *evidence* E tanpa memandang hipotesis apapun

K-Nearest Neighbor

Algoritma *K-nearest neighbor* merupakan pendekatan untuk mencari kasus dengan kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada kecocokan bobot sejumlah fitur yang ada Metode ini mencari jarak terhadap tujuan dari data yang telah disimpan sebelumnya. Setelah didapatkan jaraknya kemudian dicari jarak terdekat. Jarak terdekat tersebut yang digunakan untuk mencari identitas tujuan.

Rumus *K-nearest neighbor* yang digunakan untuk menghitung tingkat kemiripan (*similarity*) antara kasus baru yang dimasukan dengan kasus yang terdapat dalam basis kasus yaitu dengan menggunakan konsep *similarity* yang diperkenalkan oleh Amos Tversky 1977 yaitu dengan menggunakan persamaan 2.

$$SM_{pq} = \frac{\alpha(\text{common})}{\alpha(\text{common}) + \beta(\text{different})} \dots\dots\dots (2)$$

di mana :

SMPq = *similarity measure* dengan p adalah kasus baru dan q adalah kasus yang tersimpan di *casebase*.

Common = jumlah atribut yang sama.

Different = jumlah atribut yang beda.

Untuk suatu atribut yang mana pada kasus lama dan kasus baru memiliki nilai yang sama akan diberi nilai 1 sedangkan bila atribut kasus yang baru dan kasus yang lama tidak sama maka diberikan nilai 0.

3. HASIL DAN PEMBAHASAN

Penelitian ini telah menghasilkan sebuah sistem pengujian hama dan penyakit pada tanaman padi yang sesuai dengan tujuan penelitian. Berdasarkan data yang diperoleh di UPTD Proteksi Tanaman Pangan dan Holtikultura Dinas Pertanian dan Perkebunan Provinsi NTT sebanyak 153 data kasus pada tahun 2016-2018 terdapat 23 jenis hama dan penyakit tanaman padi. Penelitian ini dapat dibuktikan dengan hasil pengujian sistem yang telah dilakukan berdasarkan data yang diperoleh. Hasil pengujian ini dilakukan dengan 2 cara yaitu cara *indexing* dan *nonindexing*, dimana setiap pengujian akan dilihat tingkat akurasi.

Pada pengujian sistem terdapat 2 hasil pengujian yaitu hasil pengujian menggunakan *indexing* dan *nonindexing*. Setiap proses pengujian akan divalidasi dengan membagi data menjadi 2 kelompok yaitu data latih dan tes, data akan divalidasi menggunakan *K-fold cross validation* dimana data akan dibagi sebanyak K partisi secara acak untuk mempercepat proses komputasi. Pada proses validasi sistem pengujian hama/penyakit padi dipakai K sebesar 10 partisi yang setiap partisi akan mendapatkan nilai akurasi. Nilai akurasi didapat dari jumlah data benar dalam partisi dibagi total data partisi. Hasil pengujian *cross validation (Indexing)* dapat dilihat pada tabel 4 sedangkan hasil pengujian *cross validation (Nonindexing)* dapat dilihat pada tabel 5.

Berdasarkan tabel 4, diperoleh nilai rata-rata akurasi 92,88%, nilai didapat dari rata-rata pengujian menggunakan *indexing*. Jumlah data uji dari setiap fold adalah 15 dan 16 data. Nilai akurasi pada setiap partisi berkisar 80% hingga 100%. Sedangkan pada tabel 5, diperoleh nilai rata-rata akurasi 89,63%. Jumlah data uji dari setiap fold adalah 15 dan 16 data. Nilai akurasi pada setiap partisi berkisar 70% hingga 100%.

Sesuai tabel 4 dan 5 terdapat perbedaan dimana hasil akurasi lebih baik adalah menggunakan *indexing*. Proses pengujian menggunakan *indexing* lebih singkat karena hanya mencari hasil yang sesuai dengan hasil *index*. Berbeda dengan pengujian *indexing*, hasil yang diperoleh dari pengujian *nonindexing* cenderung mendapatkan hasil yang berbeda, dimana terdapat banyak data yang tidak sesuai dengan data aktual. Hal ini menunjukkan bahwa pengujian menggunakan *indexing* lebih akurat dibandingkan pengujian *nonindexing*.

Tabel 4. *Cross validation (indexing)*

Fold	Jumlah Data Uji	Jumlah Data Latih	Jumlah Data Benar	Akurasi
1	16	137	15	93.75 %
2	16	137	15	93.75 %
3	16	137	13	81.25 %
4	15	138	14	93.33 %
5	15	138	14	93.33 %
6	15	138	15	100.00 %
7	15	138	13	86.67 %
8	15	138	15	100.00 %
9	15	138	13	86.67 %
10	15	138	15	100.00 %

Tabel 5. *Cross validation (nonindexing)*

Fold	Jumlah Data Uji	Jumlah Data Latih	Jumlah Data Benar	Akurasi
1	16	137	13	81.25 %
2	16	137	14	87.50 %
3	16	137	14	87.50 %
4	15	138	11	73.33 %
5	15	138	12	80.00 %
6	15	138	15	100.00 %
7	15	138	14	93.33 %
8	15	138	15	100.00 %
9	15	138	14	93.33 %
10	15	138	15	100.00 %

4. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil pengujian pada sistem, pengujian dengan *indexing* menggunakan metode *Naïve Bayes* dan *k-nearest neighbor* menghasilkan nilai rata-rata akurasi 92,88% terhadap 153 data uji. Pada pengujian dengan *nonindexing* menggunakan metode *k-nearest neighbor* menghasilkan nilai rata-rata akurasi 89,63% terhadap 153 data uji. Hasil pengujian dari *indexing* dan *nonindexing* dengan menggunakan *10-fold cross validation* terhadap 153 data uji dapat disimpulkan bahwa nilai akurasi pada pengujian *indexing* lebih tinggi dari pengujian *nonindexing* dikarenakan kasus yang diuji terlebih dahulu dihitung peluang kelasnya sehingga perhitungan kemiripan hanya melihat pada kelas yang memiliki peluang lebih besar.

Saran

Adapun saran yang dapat disampaikan untuk pengembangan sistem ini yaitu:

1. Penelitian selanjutnya diharapkan dapat memperhatikan atribut-atribut menjadi lebih spesifik dan sesuai dengan hama penyakit masing-masing.
2. Dalam mengembangkan sistem ini, metode *naïve Bayes* bukan satu-satunya metode yang digunakan untuk menghitung *indexing* hasil diagnosa hama dan penyakit tanaman padi salah satu yang bisa di jadikan pilihan adalah algoritma C4.5.
3. Peneliti selanjutnya bisa memanfaatkan sumber data yang banyak tentang kasus serangan hama dan penyakit tanaman padi ini dengan menggunakan metode-metode yang lain (misalkan metode *K-means*, *cosine similarity*,) untuk mengasihkan penelitian-penelitian lain yang bermanfaat.

DAFTAR PUSTAKA

- [1] A. Aamodt and E.Plaza, Case-Based Reasoning: Foundational Issues, Methodological Variations, System Approaches, *AICom - Artificial Intelligence Communications*. IOS Press, vol.7,no.1,pp. 39-59, 1994
- [2] A.Tversky, Features of Similarity, *Journal of American Psychological Association*, 1977

- [3] Iskandar and Setiadi, Analisis B-Tree Indexed File, Bandung: Institut Teknologi Bandung, 2015
- [4] Kusrini and E.T.Luthfi, Algoritma Data Mining, Yogyakarta: Andi, 2009
- [5] R.L. Mantaras, D.Mcsherry,D.Bridge, B.Smyth, S.Craw, B.Falting, M.L.Maher,M.T. Cox, K.Forbus, M.Keane, A.Aamodt, and I.Watson, Retrieval, Reuse, Revision Retention In Case Based Reasoning, The Knowledge Engineering Review, United Kingdom : Cambridge University Press, 2006
- [6] T. R. Patil and S. S. Sherekar , Performance Analysis of Naïve Bayes and J48 Classification Algorithm for Data Classification, *International Journal of Computer Science and Applications*, vol.6, no.2, pp. 256 -261, 2013