

Implementasi Metode Naïve Bayes dan Information Gain Untuk Klasifikasi Penyakit dan Hama Tanaman Jagung

Eza Rahmanita¹, Yudha Dwi Putra Negara², Yeni Kustiyahningsih^{3*}, Verdi Sasmeka⁴, Bain Khusnul Khotimah⁵

^{1,2,3,4,5} Jurusan Teknik Informatika, Fakultas Teknik, Universitas Trunojoyo, Madura, Jawa Timur

Email: ¹ eza.rahmanita@trunojoyo.ac.id, ² yudha.putra@trunojoyo.ac.id, ^{3*} ykustiyahningsih@trunojoyo.ac.id, ⁴ verdysas@gmail.com, ⁵ bain@trunojoyo.ac.id

(Naskah masuk: 17 Aug 2023, direvisi: 21 Sep 2023, diterima: 27 Sep 2023)

Abstrak

Jagung (*Zea mays ssp. mays*) adalah tanaman pangan ketiga terbesar setelah gandum dan beras, dan di Indonesia menempati posisi kedua setelah padi. Jagung dapat ditanam di daerah dengan suhu tinggi dan rendah serta curah hujan dan irigasi yang cukup. Namun jagung sangat rentan terhadap penyakit selama siklus hidupnya, yang dapat menurunkan kualitas dan kuantitasnya. Di Sumenep, Jagung dapat dikatakan sebagai bahan pangan pokok untuk sebagian masyarakat pedesaan atau pelosok. Penyebab terjadinya serangan pada tanaman jagung adalah ketidaktahuan petani dalam pencegahan dan penanganannya sehingga menyebabkan produksi jagung mengalami penurunan. Dinas pertanian kabupaten Sumenep juga belum mempunyai sistem untuk klasifikasi hama dan penyakit jagung. Tujuan penelitian ini adalah klasifikasi penyakit dan hama tanaman jagung menggunakan metode naïve bayes dengan *information gain*. Naïve Bayes digunakan untuk mengolah nilai-nilai probabilitas setiap gejala, dan nilai persentase dari setiap hama dan penyakit. *Information Gain* untuk menyeleksi bobot gejala yang paling berpengaruh dalam menentukan hama dan penyakit jagung. Hasil uji coba, akurasi *naïve bayes* dengan *information gain* dapat meningkatkan akurasi rata-rata sebesar 3,17 % dibanding klasifikasi tanpa seleksi fitur. Akurasi terbaik diperoleh dengan metode *information gain* dan naïve bayes sebanyak 15 fitur dari 47 fitur dengan akurasi sebesar 98,47 %. Penelitian ini merekomendasikan 15 fitur, dengan 3 fitur terbesar adalah tidak berbuah, daun berklorosis sebagian atau seluruh daun dan adanya bekas gigitan pada batang.

Kata Kunci : *Naïve Bayes, Information Gain, Klasifikasi, Jagung, Hama dan Penyakit*

Implementation of Naïve Bayes Method and Information Gain for Classification of Diseases and Pests of Corn Plants

Abstract

Corn (*Zea mays ssp. mays*) is the third largest food crop after wheat and rice, and in Indonesia it ranks second after rice. Corn can be grown in areas with high and low temperatures and sufficient rainfall and irrigation. However, maize is very susceptible to disease during its life cycle, which can reduce its quality and quantity. In Sumenep, corn can be said to be a staple food for some rural or remote communities. The cause of the attack on corn plants is the ignorance of farmers in prevention and handling, causing a decrease in corn production. The Sumenep district agriculture office also does not yet have a system for classifying corn pests and diseases. The purpose of this research is to classify diseases and pests of corn plants using the naïve Bayes method with information gain. Naïve Bayes is used to process the probability values of each symptom, and the percentage values of each pest and disease. Information Gain to select the most influential symptom weight in determining maize pests and diseases. The test results show that the accuracy of Naïve Bayes with information gain can increase accuracy by an average of 3.17% compared to classification without feature selection. The best accuracy was obtained using the information gain and naïve Bayes methods with 15 features out of 47 features with an accuracy of 98.47%. This research recommends 15 features, with the 3 biggest features being no fruit, partially or completely chlorotic leaves and bite marks on the stem.

Keywords: *Naïve Bayes, Information Gain, Classification, Corn, Pests and Diseases.*

I. PENDAHULUAN

Madura terdiri dari 4 kabupaten, yaitu Kabupaten Bangkalan, Sampang, Pamekasan dan Sumenep. Dari beberapa kabupaten, Sumenep merupakan kawasan jagung terluas di Madura seluas 14.000 hektar [1]. Pada maret-april 2020 luas panen tanaman jagung mencapai 99.250 Ha mempunyai hasil rata-rata 3,03 ton/ha dengan total perkiraan produksi kurang lebih 300.727 ton. Jagung tidak hanya menjadi makanan pokok masyarakat, tetapi juga digunakan sebagai pakan ternak dan bahan baku industri. Konsumsinya memiliki tingkat pertumbuhan tahunan rata-rata 5,16%, dan bahan baku industri dan peternakan memiliki tingkat pertumbuhan tahunan rata-rata 10,87%. Potensi serangan penyakit dan hama pada tanaman jagung dapat menyerang kapan saja. Penyakit pada tanaman jagung terjadi akibat beberapa komponen seperti patogen (parasit), inang (induknya), dan lingkungan. Melihat Banyaknya serangan penyakit pada tanaman jagung dapat merugikan petani hingga menyebabkan gagal panen jika tidak segera diatasi, maka perlu perlakuan khusus agar dapat mencegah atau mengurangi penyebaran penyakit. Penyebab terjangkitnya penyakit pada tanaman budidaya jagung adalah kurangnya pengetahuan tentang pencegahan dan pengobatan penyakit yang sedang dialami. Penyakit tanaman jagung dapat diketahui oleh seorang ahli seperti penyuluh pertanian [2]. Sedangkan, pada UU No 19 Tahun 2013 disebutkan bahwa pada setiap desa harus memiliki minimal paling sedikit satu penyuluh.

Teknologi saat ini telah berkembang pesat dan telah banyak digunakan oleh berbagai bidang seperti bidang perbankan, pemerintahan, industri, kesehatan, pendidikan, pertanian, dan masih banyak lagi. Metode yang dapat digunakan sebagai metode klasifikasi adalah *Naïve Bayes*. *Naïve Bayes* merupakan metode untuk memprediksi tingkat kemungkinan dari suatu kejadian. *Naïve Bayes* merupakan metode probabilistik klasifikasi sederhana berdasarkan teorema *bayes* menggunakan data latih dari beberapa data secara efisien untuk mengatasi ketidakpastian data [3][4]. Kebanyakan pada situasi nyata yang kompleks, metode *Naïve Bayes* bekerja lebih baik dari yang diharapkan. Keuntungan *Naïve Bayes* adalah tidak membutuhkan banyak data latih dalam pemrosesan klasifikasi [5][6]. Proses klasifikasi *Naïve Bayes* akan membutuhkan waktu yang lebih jika fitur yang digunakan cukup banyak, maka dari itu dibutuhkan sebuah metode seleksi fitur. Metode yang akan digunakan dalam penelitian ini adalah seleksi fitur dengan menggunakan *information gain*. *Information gain* tidak kalah jauh dari metode seleksi fitur yang umum seperti *Gain Ratio* (GR), *Correlation-Based* (CB), dan *Particle Swarm Optimization* (PSO) dari segi kecepatan dan ketepatan dengan menggunakan beberapa metode klasifikasi [3][7][8][9]. Kelebihan *information gain* adalah penerapannya sederhana dibandingkan metode seleksi fitur yang lain, sehingga banyak digunakan dalam pembuatan aplikasi [10]. *Information Gain* merupakan metode yang dapat memberikan bobot sebagai perankingan pada fitur. *Information gain* dapat menentukan fitur-fitur yang memiliki informasi terbanyak, Sehingga ditemukan fitur-fitur yang paling dominan [11][12]. Fitur

yang akan terpilih dilihat dari nilai *information gain* terbesar. Seleksi fitur *gain ratio* dan algoritma *K-Nearest Neighbor* juga digunakan pada klasifikasi penyakit diabetes mellitus mendapat kenaikan akurasi dengan *threshold* 0,152 dengan mempertahankan 4 fitur dari 8 [13][14]. Penelitian Klasifikasi Menggunakan Metode *Naïve Bayes Classifier* sudah banyak dilakukan diantara perbandingan antara metode naïve bayes dengan KNN untuk klasifikasi artikel Bahasa Indonesia[15], hasilnya naïve bayes lebih baik karena akurasinya lebih dari KNN. Penelitian selanjutnya mengenai klasifikasi penyakit anak, menggunakan *Naïve Bayes*, dengan fitur 16 tanpa menggunakan metode seleksi fitur. Metode *Naïve Bayes* diketahui sedikit lebih unggul dibandingkan metode yang lain (KNN) [3].

Naïve Bayes juga digunakan untuk klasifikasi penyakit darah tinggi, hasilnya menunjukkan akurasi sebesar 89,65% [16]. Untuk Gejala yang digunakan sebanyak 19 gejala yang didapat dari seorang pakar pertanian di Universitas Brawijaya Malang. Akurasi yang dihasilkan dari aplikasi yang dibuat mencapai 96% dengan 50 data *testing*[6]. Untuk seleksi fitur menggunakan *information gain*. *Information gain* merupakan metode yang dapat memberikan bobot sebagai perankingan pada fitur. *Information gain* dapat menentukan fitur-fitur yang memiliki informasi terbanyak, Sehingga ditemukan fitur-fitur yang paling dominan. Seleksi fitur menggunakan metode *information gain* dapat meningkatkan nilai akurasi klasifikasi.

Berdasarkan penelitian sebelumnya belum terdapat metode klasifikasi penyakit dan hama tanaman jagung menggunakan seleksi fitur, indikator yang digunakan juga berbeda, *software* yang digunakan juga belum bisa secara dinamis melakukan klasifikasi jagung, sehingga diperlukan aplikasi ini untuk membantu petani jagung dan dinas dalam mengidentifikasi hama dan penyakit jagung. Pada penelitian ini akan membangun sistem klasifikasi penyakit tanaman jagung menggunakan metode *naïve bayes* dan seleksi fitur *information gain* untuk klasifikasi hama dan penyakit tanaman jagung.

II. METODE PENELITIAN

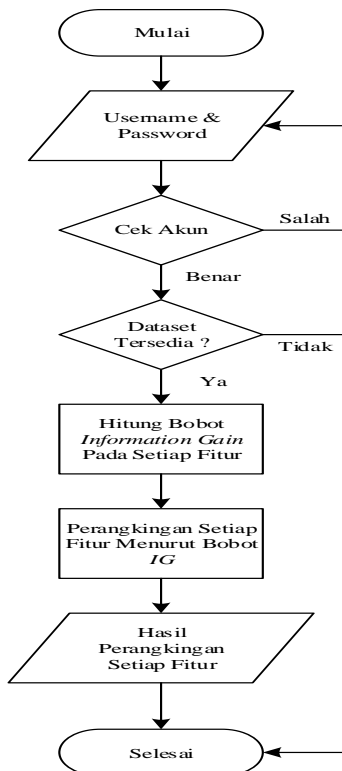
Metode penelitian merupakan tahapan penelitian secara keseluruhan untuk menjawab permasalahan yang terdapat pada bab pertama. Penelitian ini dimulai dengan kajian awal model klasifikasi hama dan penyakit tanaman jagung pada kabupaten Sumenep Madura. Tahapan penelitian ini adalah:

1. Studi Literatur, Wawancara dan Kuisisioner

Pada tahap ini dilakukan studi *literature* dan jurnal mempelajari teori-teori mengenai sistem pendukung keputusan, data mining, klasifikasi, *Artificial Intelligence* (AI), *data preprocessing*, metode *naïve bayes*, metode *z-Score*, metode *information gain*. Wawancara dan kuisisioner dilakukan kepada dinas pertanian tanaman pangan, hortikultura dan perkebunan kabupaten Sumenep. Hasil data dari dinas dan petani garam adalah terdiri dari 35 gejala dalam menentukan hama dan penyakit tanaman jagung.

2. Pengumpulan Data. Pada tahap pengumpulan data-data terdiri dari:
 - Data yang digunakan sebagai objek penelitian ini adalah data penyakit dan hama serta gejala yang terjadi pada tanaman jagung tahun 2022 sejumlah 300 data yang terdiri dari 2 serangan yaitu penyakit dan hama serta 47 gejala yang didapat dari 11 desa pada Kecamatan Rubaru antara lain: Banasare, Basoka, Bunbarat, Duko, Klebengan, Karangnangka, Mandala, Pankondang, Matanair, Rubaru, dan Tambaksari.
 - Penelitian ini menggunakan data yang didapat dari Dinas Pertanian Tanaman Pangan, Holtikultura dan Perkebunan di Kabupaten Sumenep.
3. Tahap Pemodelan Klasifikasi Hama dan Penyakit Tanaman Jagung.

Tahap ini merupakan tahap identifikasi masalah dengan menentukan jumlah variabel yang akan digunakan dalam penelitian. Selanjutnya adalah tahap *preprocessing*. Sebelum data di klasifikasikan, data tersebut akan dilakukan *preprocessing* untuk menyeragamkan data. Setelah data ternormalisasi maka akan dilakukan klasifikasi dengan metode *naïve bayes*. Metode *naïve bayes* untuk menentukan klasifikasi penyakit dan hama tanaman jagung dengan 47 fitur dan *information gain* untuk menentukan bobot fitur dalam klasifikasi. Metode ini dapat memberikan bobot pada setiap fitur sehingga ditemukan fitur yang paling dominan dan mengurangi fitur yang tidak diperlukan. *Flowchart* metode pembobotan fitur dengan metode *information gain* dapat dilihat pada Gambar 1.



Gambar 1. *Flowchart* Sistem Pembobotan Fitur Metode *Information Gain*

Penjelasan *flowchart* pada Gambar 1 adalah *Login khusus admin* ke dalam sistem, sistem akan otomatis memeriksa akun yang masuk kedalam sistem terdaftar atau tidak. Pengecekan terlebih dahulu *dataset* tersedia atau tidak. Jika *dataset* tidak tersedia maka proses pembobotan fitur tidak dapat dilakukan. Sebaliknya, jika *dataset* tersedia maka proses selanjutnya dapat dilakukan. Proses selanjutnya adalah menghitung pembobotan fitur dengan metode *information gain*, dan perangkingan fitur. Langkah selanjutnya adalah menentukan *flowchart* sistem *navie bayes* untuk klasifikasi serangan tanaman jagung, terdapat pada Gambar 2.



Gambar 2. *Flowchart* Sistem Implementasi Metode *Naïve Bayes*

Penjelasan *Flowchart* pada Gambar 2. adalah sebelum melakukan diagnosis, *user* diharuskan mengisi data diri terlebih dahulu, dan login sebagai petani, bukan petani, atau penyuluh. *User* akan ditampilkan daftar peringkat dari yang tertinggi hingga terendah setiap fitur. *User* menentukan jumlah fitur yang akan digunakan dengan memilih antara 5, 10, 15, 20, 25, 30, 35, 40, atau semua fitur. Selanjutnya sistem akan otomatis membuat dataset baru yang didapat dari dataset asli yang berada di dalam database dengan fitur yang terpilih. *Dataset* baru ini nantinya akan digukan sebagai data *training* untuk model perhitungan metode *Naïve Bayes*. Proses selanjutnya adalah proses klasifikasi dengan metode *Naïve Bayes* dengan menggunakan data *Training* dan data *testing*.

Information Gain

Information gain adalah metode seleksi fitur sederhana untuk memberikan peringkat atau ranking pada atribut / fitur yang banyak diterapkan pada berbagai aplikasi analisis. Adapun langkah dalam pembobotan fitur dengan Information gain [7], yaitu:

- a. Langkah pertama adalah dengan menentukan entropy dengan menggunakan persamaan (1), entropy untuk mengukur ketidakpastian kelas dengan menggunakan ketidakpastian kejadian atau atribut tertentu [7].

$$Entropy(S) = \sum_i^c -P_i \log_2 P_i \tag{1}$$

Dimana:

- c = Total nilai yang ada pada kelas klasifikasi
- P_i = Total sample untuk kelas i

- b. Langkah kedua adalah menghitung information gain dengan menggunakan persamaan (2)

$$Gain(S,A) = Entropy(S) - \sum_{Val(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{2}$$

Dimana:

- A = Atribut / Fitur
- v = Nilai untuk atribut A
- Value(A) = Kumpulan nilai atribut A
- |S_v| = Total sample untuk nilai v
- |S| = Total keseluruhan data sample
- Entropy(S_v) = Entropy untuk sample nilai v

Setelah pembobotan pada setiap fitur, maka fitur akan diranking menurut nilai information gain yang telah dihitung sebelumnya.

- c. Klasifikasi, merupakan model diaman fungsinya adalah mengelompokkan atau mengkategorikan pada suatu data, klasifikasi digunakan untuk menemukan konsep kelas yang tidak diketahui dalam data, yang dianalisis berdasarkan data latih kelas yang diketahui sebelumnya.
- d. Naïve Bayes, klasifikasi bayes didasarkan pada torema bayes atau bisa dikenal sebagai Naïve Bayes merupakan klasifikasi statistik yang dapat memprediksi suatu kelas pada suatu data [7]. Mengenal setiap kondisi/peristiwa dengan asumsi yang kuat merupakan ciri utama dari Naïve Bayes. Adapun langkah-langkah perhitungan Naïve Bayes sebagai berikut [21]:
Hitung nilai class prior probability pada masing-masing kelas dengan menghitung average pada masing-masing kelas dengan persamaan 3.

$$P(A) = \frac{\text{total kemunculan kejadian A}}{\text{total keseluruhan data}} \tag{3}$$

Hitung nilai likelihood pada masing-masing kelas dengan persamaan 4

$$P(B|A) = \frac{\text{total kemunculan kejadian B bila A terjadi}}{\text{total kemunculan kejadian A}} \tag{4}$$

Hitung nilai predictor prior probability dengan persamaan 5.

$$P(B) = \frac{\text{total kemunculan kejadian B}}{\text{total keseluruhan data}} \tag{5}$$

Hitung nilai posterior probability:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \tag{6}$$

Dimana:

- P = Probabilitas
 - A = Hipotesis data kelas spesifik
 - B = Data belum diketahui kelasnya
 - P(A) = Probabilitas A (class prior probability)
 - P(B|A) = Probabilitas B berdasarkan hipotesis A
 - P(B) = Probabilitas B
 - P(A|B) = Probabilitas A berdasarkan kondisi B
- Langkah pertama hingga langkah ketiga merupakan proses pembelajaran (learning) dimana proses tersebut bertujuan untuk mempelajari pola dari sebuah data. Langkah keempat merupakan proses klasifikasi (classify) yang nantinya hasil dari proses tersebut akan akan menemukan label / class dengan nilai tertinggi.

- e. Laplacian Correction

Naïve Bayes merupakan metode klasifikasi yang bertumpu pada probabilitas. Laplacian Correction merupakan teknik untuk menyiasati apabila nilai probabilitas pada perhitungan Naïve Bayes bernilai 0 yang sebabkan tidak adanya data pada untuk kategori tertentu dalam suatu label/class [17].

III. HASIL DAN PEMBAHASAN

- A. Tahapan Hasil Penelitian

Tahapan dari hasil dan pembahasan dari penelitian ini adalah:

1. Menentukan jumlah dataset
Data yang didapat dari Dinas Pertanian Tanaman Pangan, Holtikultura dan Perkebunan Kabupaten Sumenep sebanyak 300 data dengan record kelas hama sebanyak 158 dan penyakit 142. Dataset tersebut akan dibagi menjadi 2 bagian dataset yaitu data training dan data testing.
2. Pembobotan menggunakan information gain
Dataset yang akan digunakan untuk perhitungan pembobotan information gain adalah dipakai sebanyak 300 data yang terdiri dair 47 fitur. Hasil pembobotan dengan information gain dari yang tertinggi hingga terendah seperti yang ditunjukkan pada Tabel 1.

Tabel 1. Hasil Pembobotan Information Gain

Kode	Nama Gejala	Hasil IG
G41	Tidak berbuah	0,161
G20	Daun berklorosis sebagian atau seluruh daun	0,145
G01	Adanya bekas gigitan pada batang	0,100
G28	Layu pada pucuk daun	0,100
G32	Permukaan daun berwarna coklat	0,088

Kode	Nama Gejala	Hasil IG
G16	Daun tanaman muda rusak	0,08
G03	Adanya bekas gigitan pada daun	0,075
G14	Batang yang mudah patah	0,075
G24	Daun tampak bercak bergaris kuning	0,071
G25	Daun tampak bergaris kuning panjang	0,068
G35	Tanaman cepat mati atau mengering	0,064
G22	Adanya ulat ditongkol jagung	0,06
G37	Tanaman menjadi layu	0,059
G45	Tulang daun rusak	0,058
G34	Daun berlubang bahkan tingaal tulang-tulang saja	0,052
G30	Sebagian biji jagung yang bengkak tersembul keluar	0,052
G06	Terdapat serbuk berwarna kuning kecoklatan	0,051
G10	Bagian atas layu dan mengering	0,048
G15	Daun bagian bawah dan atas tidak terdapat bubuk spora	0,048
G47	Batang busuk	0,048
G40	Tanaman mudah rebah	0,048
G44	Tongkolnya tidak normal	0,034
G04	Biji akan rusak dan membusuk, bahkan tongkolnya pun dapat berjatuhan	0,034
G13	Permukaan biji ditutupi miselium berwarna abu-abu hingga hitam	0,033
G46	Tongkol pembungkus rusak	0,032
G18	Tongkol Rusak	0,026
G09	Tanaman menjadi kerdil	0,025
G08	Bercak melebar pada daun	0,022
G39	Pelepah berwarna merah keabu-abuan	0,022
G19	Terdapat titik merah kecoklatan seperti karat	0,02
G21	Adanya bekas gigitan pada buah/biji	0,019
G31	Adanya lubang kecil pada daun	0,015
G12	Garis putus-putus pendek pada tulang daun	0,011
G17	Daun menjadi transparan	0,007
G42	Pangkal batang atau tongkol berwarna merah muda, coklat kemerahan atau coklat	0,002
G43	Bagian batang kulit luarnya tipis	0,001
G36	Adanya kotoran-kotoran di tongkol jagung	0

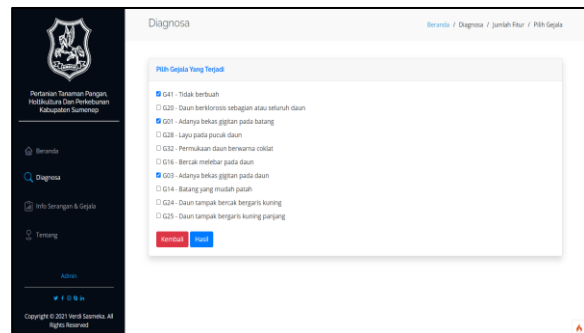
3. Menghitung klasifikasi *Naïve Bayes*

a. *Dataset* yang digunakan sebelumnya sebanyak 300 data yang terdiri dair 47 fitur (gejala) dengan 2 kelas yaitu hama dan penyakit yang ditunjukkan pada gambar 2. Perhitungan klasifikasi setelah proses pemobotan fitur *Information Gain* jumlah fitur yang akan digunakan adalah 10. *Dataset* baru akan terbentuk yang terdiri dari 10 fitur tertinggi setelah dibobotkan berserta labelnya. *Dataset* ini akan digunakan sebagai data *training* untuk perhitungan metode *Naïve Bayes*.

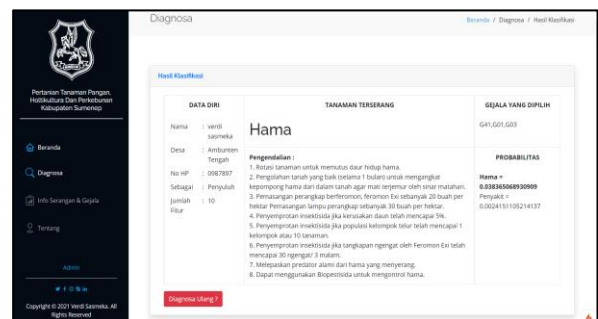
- b. Pembersihan data (*Cleaning data*) adalah tahap dimana menyeleksi data yang tidak mempunyai gejala tetapi mempunyai kelas setelah dilakukan 10 pemilihan fitur dengan bobot tertinggi. Data tersebut akan dihapus karena dapat berdampak pada proses pemodelan / pembelajaran *naïve bayes*. Setelah proses *cleaning data* dilakukan maka *dataset* yang awalnya berjumlah 300 menjadi 199 data.
- c. Data *testing* yang digunakan dalam perhitungan algoritma *naïve bayes* adalah data yang tidak ada pada data *training* dan tidak diketahui labelnya. Jika dilihat gejala yang terjadi pada data *testing* yang di tunjukkan pada Tabel 1 dengan gejala yang dipilih G41 merupakan ciri penyakit, sedangkan G01 dan G03 merupakan ciri dari hama maka serangan yang dominan adalah hama.

B. Hasil Implementasi Sistem

Hasil implementasi sistem akan dilakukan uji kepakaran dengan membandingkan hasil diagnosa dari sistem dengan perhitungan secara manual. Data *testing* yang digunakan adalah gejala yang terjadi belum pernah ada pada data *training*. Data *testing* tersebut sama dengan data *testing* yang digunakan pada analisa perhitungan sistem yang ditunjukkan pada Gambar 3. Jumlah fitur yang digunakan adalah 10 tertinggi setelah dibobotkan dengan *information gain*. Gambar 4 menunjukkan hasil dari klasifikasi hama dan penyakit tanaman jagung.



Gambar 3. Pilih Gejala



Gambar 4. Hasil Klasifikasi

Dari gejala yang dipilih yaitu G41, G01, dan G03 seperti yang ditunjukkan pada Gambar 1, maka hasil klasifikasi dari gejala tersebut adalah hama dengan nilai probabilitas **0,03836** yang dapat ditunjukkan pada Gambar 2. dapat disimpulkan hasil klasifikasi dari sistem yang dibangun sama dengan analisa perhitungan sistem.

Tabel 2. Perbandingan Hasil Akurasi

Jumlah Fitur	Naïve Bayes tanpa IG	Naïve Bayes dengan IG	Peningkatan Akurasi
5	96,16%	98,20%	2,04%
10	90,82%	98,46%	7,64%
15	88,49%	98,47%	9,98%
20	96,22%	98,24%	2,02%
25	96,52%	98,27%	1,75%
30	95,89%	98,28%	2,39%
35	96,07%	97,98%	1,91%
40	94,95%	97,98%	3,03%
45	96,00%	96,97%	0,97%
47	96,33%	96,33%	0,00%
Akurasi			3,17 %

Pada Tabel 2 merupakan hasil akurasi setiap percobaan dengan jumlah fitur yang berbeda, perhitungan akurasi *naïve bayes* tanpa IG, di tentukan berdasarkan urutan gejala atau fitur Untuk perhitungan akurasi *naïve bayes* dengan IG, akurasi berdasarkan seleksi bobot gejala dengan IG, kemudian dihitung akurasinya. Penggunaan jumlah fitur terbaik sebelum menggunakan pembobotan *Information Gain* adalah 25 dengan akurasi 96,52%, sedangkan penggunaan jumlah fitur terbaik setelah menggunakan pembobotan *Information Gain* adalah fitur 15 dengan akurasi 98,47%. Rata-rata peningkatan akurasi tanpa menggunakan *information gain* dan tanpa *information gain* adalah 3,17 %

IV. KESIMPULAN

Berdasarkan implementasi dan pengujian yang telah dilakukan dalam penelitian ini yaitu proses setelah pembobotan *information gain* yaitu *cleaning data* yang bertujuan untuk menghapus data yang tidak mempunyai gejala tetapi mempunyai *label/class* pada *dataset*. Proses *cleaning* tersebut dapat mempengaruhi waktu proses klasifikasi disebabkan perbedaan jumlah *dataset* yang dapat diproses. Dapat disimpulkan bahwa waktu proses klasifikasi pada data *training* tergantung jumlah data yang dapat diproses. Pembobotan *information gain* dan klasifikasi *naïve bayes* dapat diimplementasikan dengan baik untuk mengklasifikasi serangan pada tanaman jagung. Setiap percobaan pengujian akurasi mengalami kenaikan setelah menggunakan pembobotan fitur *information gain*. Pengujian akurasi dengan *5-fold* sebelum menggunakan pembobotan fitur *information gain* didapat akurasi tertinggi 97,62% dengan penggunaan jumlah fitur 25, sedangkan pengujian setelah menggunakan pembobotan *information gain* didapat akurasi tertinggi 99,46% dengan penggunaan 15 fitur. Dapat disimpulkan bahwa penggunaan pembobotan fitur *information gain* dapat meningkatkan tingkat akurasi serta penggunaan jumlah fitur yang lebih sedikit.

REFERENSI

- [1] Haryanto, “Jadikan Sumenep Corong Industri Pertanian, Pusat Bantu Rp 35 Miliar,” *Radar Madura*, 2019. .
- [2] R. P. Ramadhan and N. L. Marpaung, “Identifikasi jenis penyakit daun tanaman jagung menggunakan jaringan saraf tiruan berbasis backpropagation,” *Jom FTEKNIK*, vol. 6, no. 1, pp. 1–5, 2019.
- [3] N. Sagala and H. Tampubolon, “Komparasi Kinerja Algoritma Data Mining pada Dataset Konsumsi Alkohol Siswa,” *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 4, no. 2, p. 98, 2018, doi: 10.23917/khif.v4i2.7061.
- [4] D. W. Nugraha, A. Y. E. Dodu, and N. Chandra, “Klasifikasi Penyakit Stroke Menggunakan Metode Naive Bayes Classifier (Studi Kasus Pada Rumah Sakit Umum Daerah Undata Palu),” *semanTIK*, vol. 3, no. 2, pp. 13–22, 2017.
- [5] R. Wibowo and H. Indriyawati, “Top-k Feature Selection Untuk Deteksi Penyakit Hepatitis Menggunakan Algoritme Naïve Bayes,” *J. Buana Inform.*, vol. 11, no. 1, p. 1, 2020, doi: 10.24002/jbi.v11i1.2456.
- [6] A. Syarifudin, N. Hidayat, and L. Fanani, “Sistem Pakar Diagnosis Penyakit Pada Tanaman Jagung Menggunakan Metode Naive Bayes Berbasis Android,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 4, pp. 1492–1500, 2018.
- [7] A. A. Syafitri Hidayatul AA, Yuita Arum S, “Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 9, pp. 2546–2554, 2018.
- [8] S. Sen, L. Sahoo, K. Tiwary, V. Simic, and T. Senapati, “Wireless Sensor Network Lifetime Extension via K-Medoids and MCDM Techniques in Uncertain Environment,” *Appl. Sci.*, vol. 13, no. 5, 2023, doi: 10.3390/app13053196.
- [9] A. Harris and A. E. Mintaria, “Komparasi Information Gain , Gain Ratio , CFs-Bestfirst dan CFs-PSO Search Terhadap Performa Deteksi Anomali,” vol. 5, pp. 332–343, 2021, doi: 10.30865/mib.v5i1.2258.
- [10] M. Eghtesadifard, P. Afkhami, and A. Bazyar, “An integrated approach to the selection of municipal solid waste landfills through GIS, K-Means and multi-criteria decision analysis,” *Environ. Res.*, vol. 185, no. March, p. 109348, 2020, doi: 10.1016/j.envres.2020.109348.
- [11] M. R. Hasibuan and Marji, “Pemilihan Fitur dengan Information Gain untuk Klasifikasi Penyakit Gagal Ginjal menggunakan Metode Modified K-Nearest Neighbor (MKNN),” vol. 3, no. 11, pp. 10435–10443, 2019.
- [12] K. M. Khoirunnisak, “Sistem Pakar Diagnosa Penyakit Pada Tanaman Jagung Dengan Metode Dempster Shafer,” 2020.
- [13] E. Sugiharto, I. -, and I. D. Wijaya, “Sistem Rekomendasi Tempat Wisata di Malang Raya Dengan Metode Fuzzy Berbasis Web,” *J. Apl. Dan Inov. Ipteks “Soliditas,”* vol. 4, no. 1, p. 8, 2021, doi: 10.31328/js.v4i1.1731.
- [14] Y. Kustiyahningsih, “Feature Selection and K-nearest Neighbor for Diagnosis Cow Disease,” *Int. J. Sci. Eng. Inf.*

-
- Technol.*, vol. 5, no. 02, pp. 249–253, 2021, doi: 10.21107/ijseit.v5i02.10218.
- [15] M. Marlina, W. Saputra, B. Mulyadi, B. Hayati, and J. Jaroji, “Aplikasi sistem pakar diagnosis penyakit ispa berbasis speech recognition menggunakan metode naive bayes classifier,” *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 8, no. 1, pp. 58–70, 2017, doi: 10.31849/digitalzone.v8i1.629.
- [16] Indriyanti, D. Sugianti, and M. A. Al Karomi, “Peningkatan Akurasi Algoritma KNN dengan Seleksi Fitur Gain Ratio untuk Klasifikasi Penyakit Diabetes Mellitus,” *IC-Tech*, vol. 7, no. 2, pp. 1–6, 2017.
- [17] M. Rizki, M. Arhami, and Huzeni, “Perbaikan Algoritma Naive Bayes Classifier Menggunakan Teknik Laplacian Correction,” *J. Teknol.*, vol. 21, no. 1, p. 7, 2021.