Jurnal SimanteC
Vol. 12, No. 1 Desember 2023
P-ISSN : 2088-2130
E-ISSN : 2502-4884

PENERAPAN METODE *NAÏVE BAYES CLASSIFIER* UNTUK KLASIFIKASI JUDUL BERITA

APPLICATION OF THE NAÏVE BAYES CLASSIFIER METHOD FOR NEWS TITLE CLASSIFICATION

[1]Hendri Hartono, [2]**Alyauma Hajjah,** [3]**Yulvia Nora Marlim**

[123] Prodi Teknik Informatika, Fakultas Ilmu Komputer Institut Bisnis dan Teknologi Pelita Indonesia
Jl. Ahmad Yani No.78-88, Pekanbaru, Riau
E-mail: henhartono014@gmail.com, alyauma.hajjah@lecturer.pelitaindonesia.ac.id, yulvia.nora@lecturer.pelitaindonesia.ac.id
*e-mail: alyauma.hajjah@lecturer.pelitaindonesia.ac.id

Abstrak

Berita merupakan media informasi utama di dunia. Beragamnya berita yang disajikan oleh media digital saat ini, mencakup berbagai aspek seperti olahraga, hiburan, politik, kesehatan, keuangan, teknologi, dan lain-lain. Dengan beragamnya berita, maka diperlukan pengelompokan berita tersebut, demi memudahkan masyarakat dalam mendapatkan informasi yang diinginkan. *Naïve Bayes* merupakan metode klasifikasi yang sering digunakan dalam kasus pengelompokan. *Naïve Bayes* adalah metode klasifikasi dengan melakukan *preprocessing* pada judul berita,kemudian menghitung probabilitas setiap kelasnya. Kelas yang dipakai dalam metode ini adalah kategori berita. Kategori berita meliputi, Olahraga, Hiburan, Kesehatan, Politik, dan Teknologi.Dari 500 data latih yang dijadikan acuan untuk menghitung probabilitas, setelah data uji dimasukkan maka akan dihitung probabilitas setiap kata yang digunakan dan akan menghasilkan suatu kategori, dari 50 data yang diuji sebanyak 43 dokumen yang berhasil sesuai dengan kategori yang tepat yaitu sebesar 86% dan sebanyak 7 dokumen dengan kesalahan kategori sebesar 14%.

Kata kunci *:* Berita, Klasifikasi, Naïve Bayes, Probabilitas.

Abstract

News is the main medium of information in the world. The variety of news presented by digital media today covers various aspects such as sports, entertainment, politics, health, finance, technology, and others. With the variety of news, it is necessary to group the news, in order to facilitate the public in getting the desired information. Naïve Bayes is a classification method that is often used in clustering cases. Naïve Bayes is a classification method by preprocessing news headlines, then calculating the probability of each class. The classes used in this method are news categories. News categories include, Sports, Entertainment, Health, Politics, and Technology. From 500 training data that is used as a reference for calculating probabilities, after the test data is entered, the probability of each word used will be calculated and will produce a category, from 50 data tested as many as 43 documents that successfully fit the right category which is 86% and as many as 7 documents with a category error of 14%.

.

Keywords: *News, Classification, Naïve Bayes,Probabilities.*

1 PENDAHULUAN

Perkembangan zaman serta teknologi pada era ini membawa dampak yang sangat penting, hampir setiap manusia sangat membutuhkan teknologi sebagai alat yang dapat membantu pekerjaannya. Perkembangan teknologi juga merupakan hal umum yang dapat ditemukan dalam berbagai aspek kebutuhan manusia, yaitu sebagai alat komunikasi, visualisasi, pendidikan,bisnis, dan lain-lainnya. Dalam menjalankan suatu usaha/bisnis sangat diperlukan kemampuan dari teknologi itu sendiri agar dapat menyajikan sesuatu yang pasti tanpa terdapat data-data yang tidak valid [1].

Berita adalah cerita atau uraian tentang suatu kejadian atau peristiwa yang bersifat factual, penting, dan diminati sebagian besar pembaca, serta berkaitan dengan kepentingan mereka [2]. Pada perusahaan-perusahaan atau agen pengelola berita tentu saja sangat sering menyajikan berita-berita atau data yang berupa fakta untuk dijadikan sebagai sumber informasi. Namun jika dilihat pada zaman ini masih banyak pengelola berita yang masih menggunakan cara manual. Tentu saja ini berdampak signifikan terhadap pengguna/pembaca apabila terdapat kesalahan data/ kesalahan kategori pada judul-judul berita yang tersedia pada kanal berita. Oleh sebab itu diperlukan klasifikasi berita. Klasifikasi menerapkan metode dalam melakukan prediksi akan berhasil sangat baik karena adanya perhitungan secara matematis dan logis berdasarkan dengan informasi kriteria-kriteria dengan perhitungan yang sesuai dengan aturan pada tahapan pengambilan keputusan[3].

Klasifikasi merupakan salah satu bidang data mining. Proses pengklasifikasian adalah mempelajari dan menganalisa suatu dokumen teks baru yang belum memiliki kelas [4]. Proses pencarian model atau fungsi yang menjelaskan atau membedakan konsep atau data kelas sebagai klasifikasi. Tujuan dari klasifikasi adalah untuk mendapatkan kemampuan dalam memperkirakan kelas sebuah objek yang namanya tidak diketahui [5][6]. Metode *Naïve Bayes Classifier* merupakan salah satu metode klasifikasi yang memiliki akurasi sangat baik. Selain itu dalam proses perhitungannya metode *Naïve Bayes* ini menerapkan Theorema Bayes sehingga perhitungannya berjalan cepat, sederhana dan memiliki akurasi tinggi [7].

2 TINJAUAN PUSTAKA

Mengumpulkan data adalah langkah utama dalam proses *text mining* dan data mining. Data adalah informasi yang benar dan nyata, dan merupakan informasi atau bahan nyata yang dapat dijadikan dasar penelitian [1]. Data mining menggunakan data yang distrukturkan, sedangkan *text mining* menggunakan data yang tidak distrukturkan. Data yang dipakai dalam *text mining* tidak terstruktur yang menyebabkan adanya proses dalam pengolahan data. Data yang diperoleh tersebut bersifat ambigu (tidak jelas), tidak lengkap, atau data data yang diperoleh adalah data yang tidak diperlukan [8].

Pembersihan yang tepat dilakukan pada tahap *text preprocessing*. Mempersiapkan teks menjadi data yang akan mengalami pengolahan yang lebih lanjut merupakan tujuan awal pemrosesan [9]. Operasi berikut akan dilakukan pada tahap ini yaitu *text mining. Text mining* umunya disebut sebagai metode ekstraksi informasi dimana pengguna berintegrasi dengan sekelompok dokumen menggunakan alat analisis, salah satunya adalah kategorisasi [10]. Mendapatkan informasi yang relevan dari kumpulan dokumen adalah tujuannya[11]. Variasi data mining disebut juga *text mining* mencari pola yang menarik dalam sejumlah besar data teks [12].

Salah satu metode yang digunakan dalam prosedur klasifikasi adalah Metode *Naïve Bayes*. Klasifikasi menggunakan Metode *Naïve Bayes* menghasilkan akurasi yang sangat bagus[13]. Dalam penelitian ini, data uji adalah data berita yang akan digunakan sebagai perbandingan dalam mendapatkan hasil dari klasifikasi berdasarkan berita olahraga, hiburan, kesehatan, polhukam, dan teknologi. Teorema Bayes, yang menyatakan bahwa peluang masa depan dapat diprediksi berdasarkan pengalaman masa lalu, digunakan dalam

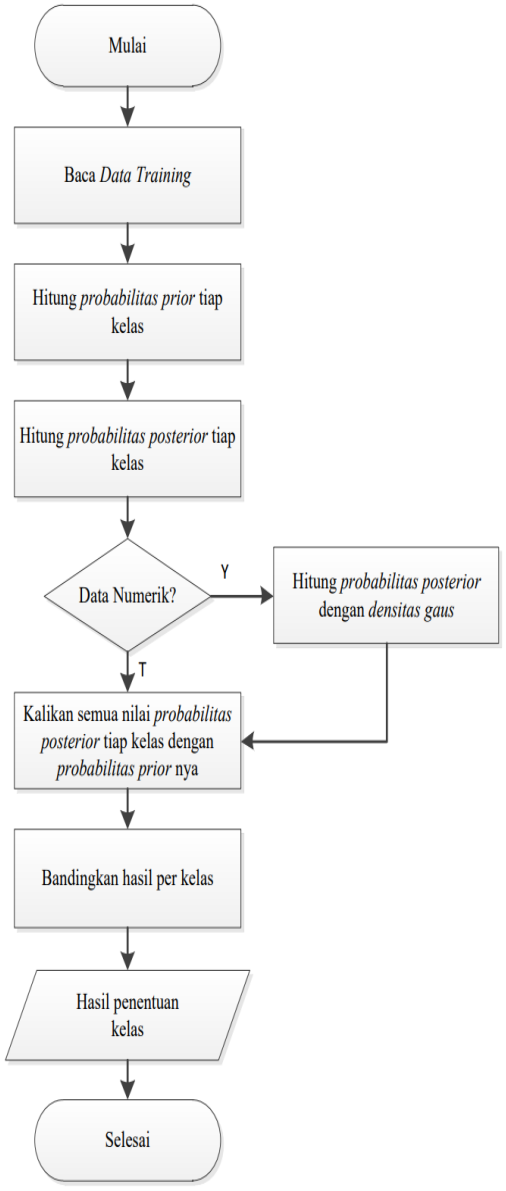
pendekatan klasifikasi Naïve Bayes, yang dikembangkan oleh ilmuwan Inggris Thomas Bayes. Teorema yang dikombinasikan dengan Naïve, membuat anggapan bahwa kriteria yang mengatur karakteristik tidak bergantung satu sama lain. Klasifikasi Naïve Bayes membuat asumsi bahwa kepemilikan kelas atau kurangnya atribut tertentu tidak ada hubungannya dengan karakteristik kelas lainnya [14].

Berikut ini adalah bentuk persamaan Teorema Naïve Bayes [15]:

$$P(H|X) = \frac{\mathrm{P(X|H).\,P(H)}}{\mathrm{P(X)}} \qquad \ldots (\mathbf{1})$$

3 METODE PENELITIAN

Pada metode penelitian ini dilakukan pengelompokan berita mengikuti diagram alur penelitian seperti yang terlihat pada gambar 1 berikut ini.



Gambar 1. Alur Metode Naïve Bayes Classifier

Pada Gambar 1 menjelaskan bahwa proses pengelompokan berita berdasarkan judul beritanya dengan menggunakan metode Naïve Bayes Classifier, adapun tahap-tahapannya adalah pengumpulan data yaitu mempersiapkan data-data berita selanjutnya melakukan data traning, kemudian dianalisa menggunakan metode Naïve Bayes Classifier untuk menentukan pengelompokan judul berdasarkan berita olahraga, hiburan, kesehatan, dan teknologi.

4 HASIL DAN PEMBAHASAN

Pada penelitian ini menerapkan metode Naïve Bayes Classifier dan Agile Model dalam klasifikasi kategori judul berita:

a. Tahapan Metode Naïve Bayes Classifier

Adapun tahapan metode Naïve Bayes Classifier adalah sebagai berikut.

1. Melakukan *preprocessing data*

Tabel 1. Data Training Judul Berita

No	Judul Berita	Kategori
1	Harry Kane bernafsu pecahkan rekor gol Wayne Rooney	Olahraga
2	Agensi Konser THE BOYZ umumkan pembatalan dua tur di AS,ini alasannya	Hiburan
3	Aksi Keren Eric Bailly: Overlap Awali Gol MU, Gocek Thiago Pula	Olahraga
4	Kata Paul Pogba soal Bakal Hadapi Mourinho di Liga Italia	Olahraga
5	Komposer Film “James Bond” meninggal dunia dalam usia 94 tahun	Hiburan
6	Usai MU Bantai Liverpool, Malacia Tak Sabar Main di Liga Inggris	Olahraga
7	Grup band pop rock Amerika Serikat, Maroon 5 akan konser di Korsel	Hiburan
8	Film “Stranger Things” sudah ditonton selama lebih dari 1 miliar jam	Hiburan
9	Pesona aktris asal Korea Selatan, Jang Nara di hari pernikahannya	Hiburan

10	Cavani: Gabung MU Bukanlah Kesalahan	Olahraga

Dari tabel berita diatas akan dilakukan tahap *preprocessing* yaitu *case folding, tokenizing, filtering,* dan *stemming* sehingga memunculkan data yang telah selesai di *preprocessing* dapat dilihat pada tabel dibawah ini.

Tabel 2. Proses *Case Folding* dan *Tokenizing*

No	Judul Berita	Kategori
1	harry kane bernafsu pecah rekor gol wayne rooney	Olahraga
2	agensi konser the boyz umum batal dua tur as alas an	Hiburan
3	aksi eric bailly awal gol mu gocek thiago	Olahraga
4	paul pogba hadapi mourinho liga italia	Olahraga
5	komposer film james bond ninggal dunia usia tahun	Hiburan
6	mu bantai liverpool malacia sabar main liga inggris	Olahraga
7	grup band pop rock amerika serikat maroon konser korsel	Hiburan
8	film stranger things tonton miliar jam	Hiburan
9	pesona aktris asal korea selatan jang nara hari nikahan	Hiburan
10	cavani gabung mu bukan salah	Olahraga

2. Menentukan atribut dengan menghitung kemunculan kata

Dari Tabel 2. data judul berita yang telah di *preprocessing* data berita akan di hitung kemunculan kata untuk menentukan sebuah atribut. Pada proses ini nilai minimum kemunculan dari sebuah kata adalah 3, maka kata yang kemunculannya < 3 akan dibuang. Hasil tersebut akan menjadi atribut. Perhitungan kemunculan jumlah kata dapat dilihat pada Tabel 3, Kata disingkat dengan huruf K dan total disingkat dengan huruf T.

Tabel 3. Perhitungan Jumlah Kata

K	T	K	T	K	T
Harry	1	The	1	bailly	1
Kane	1	Boyz	1	Awal	1
Nafsu	1	Umum	1	Mu	3
Pecah	1	Batal	1	Gocek	1
Rekor	1	Dua	1	thiago	1

K	T	K	T	K	T
Gol	3	Tur	1	Paul	1
wayne	1	As	1	pogba	1
rooney	1	Alas an	1	hadapi	1
agensi	1	aksi	1	Mourinho	1
konser	3	eric	1	Liga	3

K	T	K	T	K	T	K	T
italia	1	liverpool	1	serikat	1	aktris	1
komposer	1	malacia	1	Maroon	1	Asal	1
Film	3	sabar	1	korea	1	Jang	1
James	1	main	1	selatan	1	Nara	1
Bond	1	inggris	1	stranger	1	Hari	1
Ninggal	1	grup	1	things	1	nikahan	1
Dunia	1	band	1	tonton	1	cavani	1
Usia	1	pop	1	miliar	1	gabung	1
Tahun	1	rock	1	Jam	1	bukan	1
Bantai	1	amerika	1	pesona	1	salah	1

Setelah itu akan diperhitungkan kata unik dan kata-kata pada setiap kategori pada tabel 3 diatas .

Tabel 4. Perhitungan Jumlah Kata Unik Tiap Kategori

Jumlah Token/ Kata Unik	70
Olahraga	35
Hiburan	42

Setelah menghitung kemunculan kata maka proses berikutnya akan dilakukan penghapusan kata yang tidak mencapai nilai *minimum.* Kata yang akan digunakan sebagai atribut yakni, gol,liga,konser, dan film. Gol dan liga merupakan atribut untuk olahraga, sedangkan konser dan film merupakan atribut dari kategori hiburan.

3. Melakukan perhitungan dengan atribut yang telah didapatkan

Setelah atribut didapatkan pada masing masing kategori maka dapat dilakukan perhitungan dengan *Naïve bayes* menggunakan *data testing* "Liga Inggris tonton paling banyak sejarah sepakbola". Dari berita testing diatas akan dicari probabilitas setiap kata dari kategori sepakbola maupun hiburan.

Tabel 5. Perhitungan Kelas Kategori

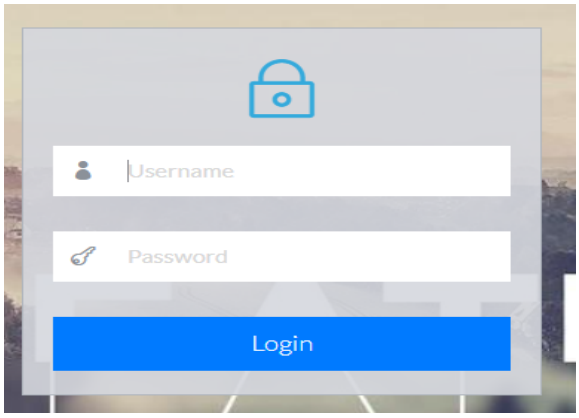
Probabilitas kemunculan kata pada kategori Olahraga		
P|Liga		0,038095
P|Inggris		0,019047
P| tonton		0,009523
P|paling		0,009523
P|banyak		0,009523
P|sejarah		0,009523
P|sepakbola		0,009523
TOTAL SCORE		5,68E-14

Probabilitas kemunculan kata pada kategori Hiburan		
P|Liga		0,008928
P|Inggris		0,008928
P| tonton		0,017857
P|paling		0,008928
P|banyak		0,008928
P|sejarah		0,008928
P|sepakbola		0,008928
TOTAL SCORE		9,04E-15

Pada tabel diatas dapat diketahui bahwa judul berita "Liga Inggris tonton paling banyak sejarah sepakbola" termasuk kedalam kategori olahraga.

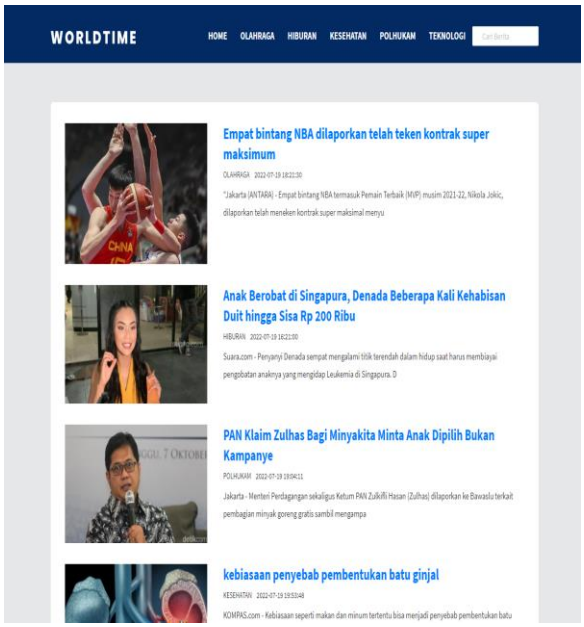
b. Implementasi Sistem

1. Halaman Login



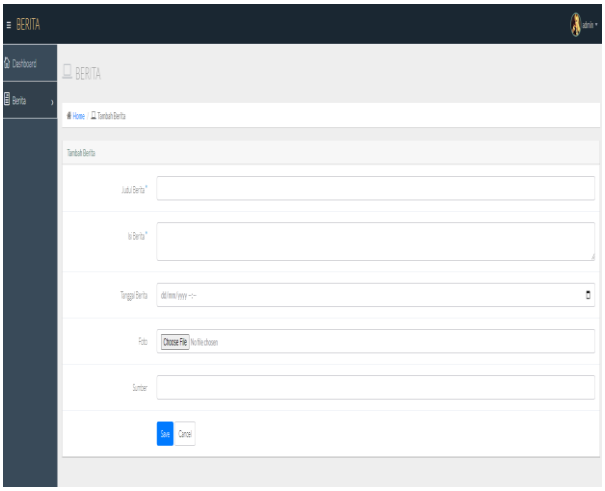
Gambar 2. Halaman Login

2. Halaman Utama



Gambar 3. Halaman Utama

3. Halaman Input Berita



Gambar 4. Halaman Input Berita

4. Halaman Pencarian Berita

Gambar 5. Halaman Pencarian Judul Berita

Pada pengujian dan analisis sistem yang dirancang. Terdapat data latih/*data training* yaitu sebanyak 500 judul berita yang akan digunakan untuk menjadi acuan untuk menghitung probabilitas judul dari data uji/*data testing* yang akan dicoba di input kedalam sistem, disini *data testing* yang di input sebanyak 50 judul berita.

Dari 50 data uji diatas yang di input ke database/sistem terdapat kesalahan kategori,hal ini sangat wajar karena setiap metode pasti memiliki kekurangannya, sehingga dapat dinilai akurasi dari sistem sebagai berikut:

Didapatkan nilai akurasi,dan nilai kesalahan pada penentuan judul berita.

Accuration Score =(10+8+9+7+9)/(10+10+10+10+10)x100% = 86%

Error rate : 100% - 86% = 14%.

5 KESIMPULAN

Dari hasil pengujian dan analisis dalam penelitian ini dapat disimpulkan bahwa teknik klasifikasi NBC dapat digunakan untuk mengkategorikan judul berita. Berdasarkan pengujian dengan metode *black box testing*, akurasi sistem dalam mengenali judul berita sebesar 86%, dengan tingkat kesalahan/ *error rate* sebesar 14%.

DAFTAR PUSTAKA

[1] W. Novia, “Kamus Besar Bahasa Indonesia,” 2006.

[2] Juwito, “Menulis Berita Dan Features,” *Menulis Ber. dan Featur.*, p. 148, 2008.

[3] Yadi, “Implementation Algorithm C4.5 Classification Of Prospective Scholarship Recipients,” *J. SimanteC*, vol. 11, no. 1, pp. 27–32, 2022.

[4] M. Sholih ’afif, M. Muzakir, M. I. Al, and G. Al Awalaien, “Text Mining Untuk Mengklasifikasi Judul Berita Online Studi Kasus Radar Banjarmasin Menggunakan Metode Naïve Bayes,” *Kumpul. J. Ilmu Komput.*, vol. 08, no. 2, pp. 199–208, 2021.

[5] M. Han, Jiawei, & Kamber, *Data mining: Data mining concepts and techniques*. 2006. doi:

10.1109/ICMIRA.2013.45.

[6] F. A. Mufarroha and D. A. Fatah, "KLASIFIKASI JENIS REMPAH PENGHASIL MINYAK ATSIRI MENGGUNAKAN METODE MACHINE LEARNING," *J. SimanteC*, vol. 11, no. 1, pp. 123–130, 2022.

[7] T. Arifin and D. Ariesta, "Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization," *J. Tekno Insentif*, vol. 13, no. 1, pp. 26–30, 2019, doi: 10.36787/jti.v13i1.97.

[8] A.-H. Tan, "Text Mining: The state of the art and the challenges," *Proc. PAKDD 1999 Work. Knowl. Disocovery from Adv. Databases*, vol. 8, pp. 65–70, 1999, doi: 10.1.1.38.7672.

[9] and P. S. A. Usama Fayyad, Gregory Piatetsky-Shapiro, "Mining association rules in graphs based on frequent cohesive itemsets," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9078, no. 3, pp. 637–648, 1996, doi: 10.1007/978-3-319-18032-8_50.

[10] D. Santi, J. Nangi, and N. Ransi, "Implementasi Naïve bayes Clasifier dalam Klasifikasi Jenis Berita," *Foristek*, vol. 10, no. 1, pp. 20–25, 2020, doi: 10.54757/fs.v10i1.52.

[11] K. Sihotang and R. Ghaniy, "Penerapan Metode Naïve Bayes Classifier Untuk Penentuan Topik Tugas Akhir," *Teknois J. Ilm. Teknol. Inf. dan Sains*, vol. 9, no. 1, pp. 63–72, 2019, doi: 10.36350/jbs.v9i1.7.

[12] R. Feldman and J. Sanger, "The Text Mining Handbook," 2007.

[13] M. M. Suhadi, M. A. Helmi, and W. Setiawan, "Simulasi Klasifikasi Hama Dan Penyakit Pada Jagung Dengan Naive Bayes," *J. Simantec*, vol. 10, no. 1, pp. 1–8, 2021, doi: 10.21107/simantec.v10i1.11686.

[14] C. Wijaya and A. Hajjah, "PENERAPAN ALGORITMA NAÏVE BAYES UNTUK REKOMENDASI GENSET," *J. Tek. Inform. Kaputama*, vol. 7, no. 1, pp. 53–60, 2023, doi: 10.31311/ji.v6i1.4685.

[15] M. Y. Ilham, R. Wulanningrum, I. N. Farida, M. Ayu, and D. Widyadara, "Citra Telapak Tangan Dengan Metode Naïve Bayes Application of Intelligent System Classification of Palm Image With Naïve," *J. SimanteC*, vol. 11, no. 2, pp. 139–146, 2023.