

Prototype Penggabungan Fuzzy Dan Naive Bayes dalam Analisis Citra Dokumen Untuk Penilaian Margin Laporan Praktikum UINSU

Hafiz Aryanda¹⁾, Muhammad Randy Fachrezi^{2✉)}, Arif Dennis Walidein³⁾
Dimas Aqila Aptanta⁴⁾ Alwi Syahputra⁵⁾

^{1),2),3),4),5)} Ilmu Komputer, Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara

¹⁾ aryandahafiz281@gmail.com, ²⁾ randyfachrezy6@gmail.com✉,

³⁾ ariefdennis11@gmail.com, ⁴⁾ dimasaqila71@gmail.com,

⁵⁾ alwiisyahputra0@gmail.com

ABSTRACT

Manual assessment of practicum report margins often faces challenges due to variations in document layout and the large number of reports that must be reviewed, resulting in an inefficient and subjective evaluation process. Therefore, this study aims to develop a prototype that serves as a guideline for automated practicum report assessment. The proposed research introduces a hybrid model integrating Fuzzy Logic and Naive Bayes. Evaluation results show that the model without fuzzy achieved an accuracy of 73.33% but exhibited bias toward the majority class, with low recall for the "Rejected" class. In contrast, the Fuzzy Naive Bayes model improved accuracy to 80% and produced more balanced classification performance, with significant increases in recall and F1 Score for the minority class. The integration of fuzzy logic effectively enhances the detection of margin non-compliance.

Keywords: Compliance of margins, image analysis, evaluation, fuzzy, naive bayes

ABSTRAK

Penilaian margin laporan praktikum secara manual seringkali menghadapi tantangan akibat variasi tata letak dokumen dikarenakan banyaknya jumlah laporan yang harus diperiksa secara manual, akibatnya pemeriksaan menjadi tidak objektif dan tidak efisien. Oleh karena itu, tujuan dari penelitian ini adalah untuk membuat prototype yang bisa menjadi pedoman pemeriksaan laporan praktikum otomatis. Penelitian ini mengusulkan model hybrid yang mengintegrasikan Logika Fuzzy dan Naive Bayes. Hasil evaluasi menunjukkan bahwa model tanpa fuzzy memiliki akurasi 73,33% dan cenderung bias terhadap kelas mayoritas, dengan recall rendah pada kelas "Tidak Diterima". Sementara itu, model Fuzzy Naive Bayes meningkatkan akurasi menjadi 80% dan menghasilkan performa klasifikasi yang lebih seimbang, dengan peningkatan signifikan pada recall dan F1 Score untuk kelas minoritas. Integrasi fuzzy terbukti membantu dalam menangkap ketidaksesuaian margin secara lebih tepat.

Kata kunci: Kepatuhan Margin, Analisis Citra, Evaluasi, Fuzzy, Naive Bayes

I. PENDAHULUAN

Karya ilmiah memerlukan tata letak yang rapi dan konsisten untuk memenuhi standar akademik. Salah satu aspek penting dalam tata letak adalah margin, yaitu batas area antara tepi halaman dan teks [1]. Margin yang sesuai tidak hanya meningkatkan estetika dokumen tetapi juga memudahkan pembacaan dan menunjukkan kepatuhan terhadap pedoman penulisan. Di lingkungan akademik, aturan margin menjadi salah

satu bagian penting dari format laporan praktikum yang wajib dipenuhi oleh mahasiswa.

Standar margin karya ilmiah diterapkan secara ketat di Universitas Islam Negeri Sumatera Utara (UINSU). UINSU memiliki beberapa program dalam menciptakan karya ilmiah, salah satunya adalah program laporan praktikum. Program ini melibatkan mahasiswa sebagai penulis dan laboran atau asisten laboran sebagai reviewer.

Selama ini, penilaian kepatuhan margin di UINSU dilakukan secara manual dengan memeriksa hardcopy laporan. Proses ini tidak efisien, rentan tidak objektif, dan memakan waktu, terutama ketika jumlah laporan yang harus dinilai cukup banyak. Selain itu, variasi tata letak dokumen misalnya perbedaan margin pada halaman isi, kesimpulan, dan lembar pengesahan menambah kompleksitas penilaian.

Untuk mengatasi masalah ini, penelitian ini mengusulkan prototipe otomatis berbasis analisis citra yang menggabungkan Logika *Fuzzy* dan *Naive Bayes*. Logika *Fuzzy* dipilih untuk menangani ketidakpastian dalam pemeriksaan margin, sedangkan *Naive Bayes* digunakan untuk mengklasifikasikan kepatuhan dokumen berdasarkan fitur margin yang telah diekstraksi.

Penelitian yang dilakukan oleh Bilal Bataineh [2], membahas berbagai metode untuk mengubah gambar dokumen berwarna atau abu-abu menjadi hitam-putih (binarisasi), terutama untuk mempermudah pengenalan teks. Penelitian ini membandingkan teknik lama seperti thresholding dengan metode modern berbasis AI seperti deep learning. Namun, penelitian ini memiliki beberapa kelemahan yaitu hanya merangkum penelitian orang lain tanpa melakukan percobaan baru sendiri. Oleh karena itu, tidak ada bukti langsung mana metode yang benar-benar terbaik [2].

Selain itu penelitian sebelumnya yang dilakukan oleh Dessy Tri Anggraeni [3], bertujuan membandingkan kinerja tiga teknik thresholding dalam meningkatkan kualitas visual citra hasil pindai dokumen. Penelitian ini menggunakan evaluasi kuantitatif MSE (*Mean Squared Error*) dan PSNR (*Peak Signal-to-Noise Ratio*) menilai kualitas hasil filterisasi. Namun, secara kritis, penelitian ini hanya berfokus pada aspek pra-pemrosesan citra tanpa memperluas aplikasinya pada analisis dokumen lebih lanjut, seperti segmentasi konten, struktur layout, atau klasifikasi elemen dokumen. Selain itu, penelitian ini tidak mengintegrasikan teknik pembelajaran mesin atau metode cerdas lainnya untuk analisis konten dokumen secara kontekstual [3].

Meskipun penelitian sebelumnya oleh Bilal Bataineh [2] dan Dessy Tri Anggraeni [3] telah membahas berbagai metode pra-pemrosesan citra dokumen seperti binarisasi dan thresholding, keduanya masih memiliki keterbatasan dalam hal penerapan praktis pada analisis struktur dokumen secara menyeluruh. Bataineh hanya membandingkan metode secara konseptual tanpa validasi eksperimental, sedangkan Anggraeni terbatas pada peningkatan kualitas visual tanpa melibatkan segmentasi, klasifikasi, atau teknik cerdas berbasis machine learning. Ketidadaan pendekatan dalam penilaian otomatis dokumen menjadi celah yang diisi oleh penelitian ini.

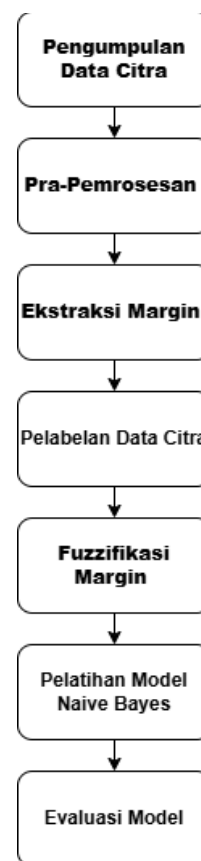
Penelitian ini menawarkan pendekatan baru dengan membangun sistem klasifikasi berbasis citra margin laporan praktikum menggunakan kombinasi logika *fuzzy* dan algoritma *Naive Bayes*. Keunggulan dari

penelitian ini terletak pada perbandingan integrasi *fuzzy* dengan *naive bayes* dan juga *naive bayes* tanpa *fuzzy* sehingga bisa menilai kontribusi dan efektivitas terhadap logika *fuzzy* yang benar – benar mampu meningkatkan akurasi.

II. METODE

Penelitian ini menggunakan pendekatan pengolahan citra digital, dimana citra dokumen laporan praktikum dimaksudkan sebagai citra digital [4]. Citra dokumen ini yang akan diubah menjadi data latih model agar model dapat digunakan untuk mengevaluasi kepatuhan atau ketaatan batas tepi (margin) sesuai aturan yang ada pada pedoman laporan praktikum UINSU.

Metode yang digunakan dalam penelitian ini adalah eksperimen kuantitatif, di mana sistem evaluasi margin dibangun, diuji, dan dianalisis performanya. Proses penelitian terdiri dari beberapa tahap, mulai dari pengumpulan citra, pra-pemrosesan, ekstraksi fitur margin, penerapan logika *fuzzy*, hingga evaluasi menggunakan *Naive Bayes*. Untuk alur penelitian ini akan digambarkan pada *flowchart* berikut ini:



Gambar 1. *Flowchart* Metode Penelitian

A. Pengumpulan Data Citra

Penelitian ini menggunakan data citra dari laporan praktikum mahasiswa Ilmu Komputer yang telah melalui tahap pemeriksaan oleh asisten laboratorium. Hasil pemeriksaan cukup beragam, ada yang menunjukkan laporan yang diterima (margin benar) dan ditolak (margin salah).

Laporan-laporan tersebut pada awalnya berformat dokumen PDF. Kemudian dokumen PDF ini dikonversi ke format gambar JPEG, sehingga setiap halaman dalam satu file PDF menjadi satu buah citra digital. Proses konversi ini menghasilkan 120 gambar digital, dimana setiap gambarnya adalah satu halaman laporan praktikum. Setelah dikonversi, 120 gambar ini dituangkan pada CSV sebagai dataset.

Dikarenakan dataset yang berekstensi CSV, maka dataset tidak dapat menyimpan gambar. karena CSV adalah format teks yang menyimpan data dalam bentuk nilai-nilai yang dipisahkan koma, bukan data biner. Oleh karena itu peneliti melakukan tindakan untuk menyimpan refrensi gambarnya saja dalam bentuk path atau alamat file gambar berada. Dengan cara ini, gambar tidak disimpan secara langsung di dalam file CSV, tetapi file CSV berfungsi sebagai daftar atau indeks yang menunjuk ke gambar-gambar yang disimpan terpisah.

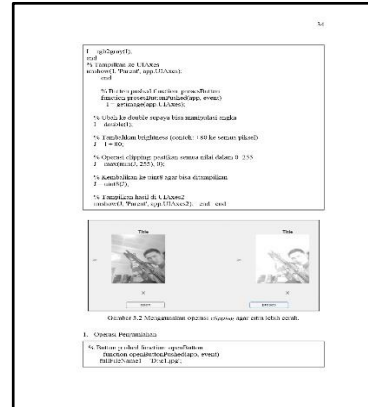
Setiap laporan praktikum terdiri dari satu modul yang mencakup beberapa bagian penting, seperti judul, sub-bab utama (tujuan, dasar teori, hasil praktikum), kesimpulan, dan lembar pengesahan (LP). Pada masing-masing bagian, memiliki aturan margin yang berbeda tergantung kondisinya. Sub-bab utama biasanya memiliki margin bawah yang konsisten karena teks mengisi halaman penuh, sementara margin bawah kesimpulan lebih fleksibel karena panjang teksnya bervariasi (kadang hanya sampai pertengahan halaman). LP juga memiliki pola margin yang berbeda, terutama pada margin kanan dan bawah, karena kontennya sering beragam. Untuk mempermudah analisis, Penulis mengklasifikasikan halaman menjadi tiga jenis: halaman isi (sub-bab utama), halaman kesimpulan, dan halaman LP. Pembagian ini membantu menetapkan aturan margin yang sesuai setelah ekstraksi fitur. Kita bisa lihat klasifikasi datanya dari tabel berikut:

Tabel 1. Klasifikasi data

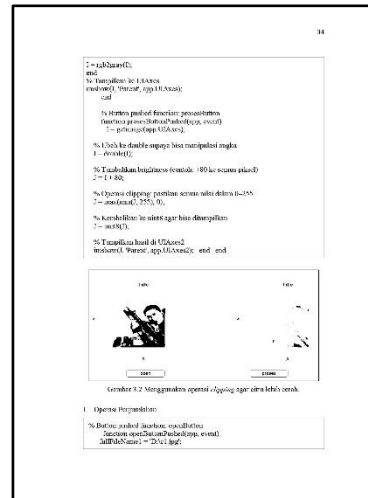
Jenis Halaman	Jumlah Gambar
Isi	98
Kesimpulan	10
Lembar Pengesahan (LP)	12
Total	120

B. Pra-Pemrosesan Data

Pada tahap ini masing-masing data citra akan olah secara pra-pemrosesan dengan tujuan untuk menghasilkan citra yang bersih, sehingga dapat dimaksimalkan dalam proses yang akan dilakukan setelahnya [5]. Pada penelitian ini fokus tahap pra-pemrosesan adalah membersihkan tampilan citra agar dapat dianalisis struktur visual citranya. Dengan beberapa tahapan utama yaitu konversi grayscale dan binerisasi. Bias dilihat citra sebelum dilakukan pra-pemrosesan pada Gambar 2 dan perbedaanya setelah dilakukan pra-pemrosesan pada Gambar 3.



Gambar 2. Tampilan Citra Sebelum Pra-Pemrosesan.



Gambar 3. Tampilan Citra Setelah Pra-Pemrosesan

C. Ekstraksi Margin

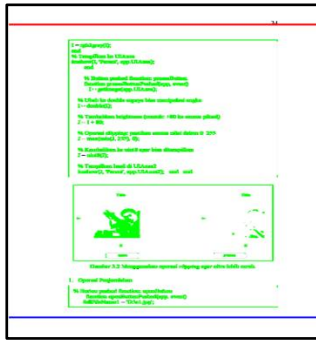
Pada tahap ini masing-masing data citra akan dianalisis margin dengan pendekatan citra biner. Untuk memastikan ketelitian pengukuran margin, pengukuran menggunakan sistem yang dapat menghitung resolusi aktual gambar dengan 300 DPI berdasarkan lebar piksel gambar lalu mengkonversikannya ke sentimeter. Setelah itu, sistem melakukan deteksi piksel teks dengan menyeleksi seluruh piksel berwarna hitam (dalam gambar biner), sekaligus menyaring *noise* yang biasanya muncul di pinggiran gambar

Sistem juga dibangun untuk menghindari deteksi area header sebesar 1.65 cm dan area footer sebesar 2.1 cm. Hal ini dimaksudkan supaya sistem tidak mendeteksi nomor halaman yang ada diatas dan nomor halaman yang ada dibawah sebagai margin. Jadi jika output deteksi margin atas yang keluar sebesar 1.35 cm dan jika output deteksi margin bawah keluar sebesar 0,9 cm maka:

$$1.65 + 1.35 = 3 \text{ cm} \rightarrow \text{Diterima}$$

$$2.1 + 0.9 = 3 \text{ cm} \rightarrow \text{Diterima}$$

Hasil tersebut diterima karena sesuai dengan pedoman laporan praktikum UINSU, yaitu margin atas, bawah, dan kanan masing-masing berukuran 3 cm, sedangkan margin kiri 4 cm.



Gambar 4. Tampilan citra setelah pra-pemrosesan

D. Pelabelan Data Citra

Setelah semua data citra telah dilakukan ekstraksi fitur margin, margin yang sesuai dengan aturan pedoman laporan UINSU dapat dilabeli sebagai “diterima” dan yang tidak sesuai dapat dilabeli dengan “tidak_diterima”. Hal ini dimaksudkan untuk model *naïve bayes* nantinya. Dapat diketahui *naïve bayes* adalah model yang dikategorikan sebagai *supervised learning*. Dari semua data citra yang ada dapat dilihat pada tabel 2 untuk jumlah yang diterima dan yang tidak diterima.

Tabel 2. Jumlah Data citra yang diterima dan tidak diterima dalam penetapan marginnya

Jenis Halaman	Jumlah Gambar	Diterima	Tidak Diterima
Isi	98	67	31
Kesimpulan	10	4	6
Lembar Pengesahan (LP)	12	6	6
Total	120	77	43

E. Fuzzifikasi Margin

Pada hakikatnya logika *fuzzy* digunakan untuk mengatasi ketidakpastian dengan memodelkan dan menggabungkan informasi yang tidak sepenuhnya jelas [6][7]. *Fuzzy Logic* atau logika *fuzzy* merupakan pendekatan yang dibuat untuk mengatasi ketidakpastian dan ketidaktepatan dalam penilaian, terutama ketika informasi yang diperoleh tidak bersifat tegas atau pasti [8].

Dalam mengatasi ketidakpastian dan ketidaktepatan dalam penilaian, *fuzzy* memperkenalkan beberapa metode. Salah satu metode itu adalah *fuzzy Mamdani*. Metode ini pertama kali diperkenalkan oleh Ibrahim Mamdani pada tahun 1975 dan sangat populer karena kemampuannya meniru cara manusia dalam mengambil keputusan berdasarkan data yang tidak pasti atau kabur [9]. *Fuzzy Mamdani* ini bekerja dimulai dengan Fuzzifikasi lalu penerapan aturan atau yang disebut *inference* dan yang terakhir Aggregasi dan Defuzzifikasi [10].

Dalam penelitian ini logika *fuzzy* dimaksudkan untuk mengatasi ketidakpastian dari nilai-nilai yang

terdapat toleransi untuk batas margin yang tidak bersifat kaku. Hal ini sangat penting mengingat dalam evaluasi laporan praktikum ini, tidak semua margin memiliki nilai yang tepat sama, namun masih dapat diterima jika berada dalam rentang tertentu.

Ditambah struktur laporan praktikum terdapat perbedaan jenis halaman, seperti yang sudah dibahas sebelumnya. Dan setiap jenis halaman ini memiliki aturan margin yang berbeda, yang didasarkan pada fungsi dan struktur tata letak halaman tersebut. Misalnya, pada halaman LP, margin kanan bisa lebih bervariasi karena adanya informasi nama atau judul yang panjangnya berbeda-beda. Sementara pada halaman kesimpulan, margin bawah cenderung sulit ditentukan karena bagian ini adalah subbab terakhir sehingga posisi akhir teks bisa lebih fleksibel.

Proses fuzzikasi ini memerlukan beberapa tahap. Sebelum menjalankan tahap pertama diperlukan konversi satuan margin yaitu sentimeter (cm) ke pixel (px). Hal ini dilakukan karena dalam konteks pengolahan citra digital, semua pengukuran posisi dan jarak pada gambar dihitung dalam satuan piksel. Alasan kenapa dari awal tidak mendeteksi langsung menjadi pixel karena dalam menghitung sentimeter akan menjadi lebih mudah untuk mengklasifikasikan label pada data citra didalam dataset.

Nilai-nilai pixel dapat dicari dengan mengacu pada standar 300 DPI, maka satu inci setara dengan 300 piksel sehingga 1 cm = 118,11px. Maka, rumusnya adalah nilai margin in cm x 118,11. Sebagai contoh, jika 1,35 cm nilai margin atas, 0,9 cm nilai margin bawah, 4 cm nilai margin kiri dan 3 cm nilai margin kanan yang diterima maka:

$$\begin{aligned}
 1,35 \times 118,11 &= 159.44\text{px} \text{ (margin atas yang diterima)} \\
 0,9 \times 118,11 &= 106.29\text{px} \text{ (margin bawah yang diterima)} \\
 4 \times 118,11 &= 472.44\text{px} \text{ (margin kiri yang diterima)} \\
 3 \times 118,11 &= 354.33 \text{ (margin kanan yang diterima)}
 \end{aligned}$$

Setelah margin dalam dataset dikonversi ke satuan piksel, nilai-nilai tersebut kemudian dimasukkan ke dalam fungsi keanggotaan *fuzzy*. Tahapan ini disebut fuzzifikasi. Fungsi keanggotaan (*Membership Function/MF*) ini didefinisikan untuk masing-masing variabel margin (atas, bawah, kiri, kanan) sesuai dengan jenis halaman laporan, lalu dapat ditentukan *range* dari tiap tiap batas fungsi keanggotaanya. *Membership function* digunakan dalam logika *fuzzy* untuk mengukur seberapa jauh suatu nilai termasuk dalam kategori tertentu. sedangkan *range* adalah nilai-nilai yang membentuk bentuk dan posisi dari kurva tersebut [11]. Pada keanggotaan *fuzzy* terdapat beberapa jenis yaitu *trimf* dan *trapmf*. Dimana *trimf* adalah fungsi yang memiliki 3 parameter (a, b, c). *trimf* cocok untuk nilai tengah pasti. Sedangkan *trapmf* adalah fungsi keanggotaan trapesium yang memiliki 4 parameter (a, b, c, d). *Trapmf* cocok untuk *range* yang luas [12]. Untuk lebih detail dapat dilihat pada tabel dibawah:

Tabel 3. Fungsi keanggotaan fuzzy untuk jenis halaman isi.

Variabel	MF	Range
Margin atas	Kurang	[79.79 - 79.79 - 141.2 - 146.1]
	ideal	[147.7 - 159.7 - 173.7]
	lebih	[173.5 - 270.1 - 277 - 291.2]
Margin bawah	Kurang	[1.18 - 30.56 - 69.68 - 103.9]
	ideal	[106.3 - 112.3 - 121.6]
	lebih	[123 - 202.1 - 502.1 - 660.2]
Margin kiri	Kurang	[374.4 - 413.3 - 453.5 - 459.4]
	ideal	[460.1 - 472 - 472.9]
	lebih	[473.7 - 500.5 - 525.9 - 558.3]
Margin kanan	Kurang	[277.5 - 329.5 - 331.8 - 337.7]
	ideal	[342.5 - 351.9 - 354.3]
	lebih	[355.5 - 360.5 - 425.5 - 472.4]

Tabel 4. Fungsi keanggotaan fuzzy untuk jenis halaman kesimpulan

Variabel	MF	Range
Margin atas	Kurang	[79.79 - 79.79 - 141.2 - 146.1]
	ideal	[147.7 - 159.7 - 173.7]
	lebih	[173.5 - 270.1 - 277 - 291.2]
Margin bawah	Kurang	[1.18 - 30.56 - 69.68 - 103.9]
	ideal	[106.3 - 112.3 - 121.6]
	lebih	[123.3 - 1288 - 1400 - 1900]
Margin kiri	Kurang	[374.4 - 413.3 - 453.5 - 459.4]
	ideal	[460.1 - 472 - 472.9]
	lebih	[473.7 - 500.5 - 525.9 - 558.3]
Margin kanan	Kurang	[277.5 - 329.5 - 331.8 - 337.7]
	ideal	[342.5 - 351.9 - 354.3]
	lebih	[355.5 - 360.5 - 425.5 - 472.4]

Tabel 5. Fungsi keanggotaan fuzzy untuk jenis halaman LP

Variabel	MF	Range
Margin atas	Kurang	[79.79 - 79.79 - 141.2 - 146.1]
	ideal	[147.7 - 159.7 - 173.7]
	lebih	[173.5 - 270.1 - 277 - 291.2]
Margin bawah	Kurang	[1.18 - 30.56 - 69.68 - 103.9]
	ideal	[485 - 500.7 - 546.8]
	lebih	[547.7 - 560.9 - 631.3 - 660]
Margin kiri	Kurang	[374.4 - 413.3 - 453.5 - 459.4]
	ideal	[460.1 - 472 - 472.9]
	lebih	[473.7 - 500.5 - 525.9 - 558.3]
Margin kanan	Kurang	[277.5 - 329.5 - 331.8 - 337.7]
	ideal	[342.5 - 351.9 - 354.3]
	lebih	[355 - 657.1 - 771.4 - 891.7]

Selanjutnya setelah proses fuzzifikasi dilakukan, kita akan melakukan tahap inferensi fuzzy. Dimana setiap aturan disusun dalam bentuk pernyataan logika

IF-THEN, yang menggabungkan kondisi dari masing-masing margin (atas, bawah, kiri, kanan) dan menghasilkan output berupa tingkat kepatuhan yang bersifat linguistik seperti diterima dan tidak diterima. Aturan-aturan ini disesuaikan untuk masing-masing jenis halaman dalam laporan. Misalnya pada aturan fuzzy berdasarkan jenis halaman kesimpulan.

If (margin_atas is ideal) and (margin_bawah is not kurang) and (margin_kiri is ideal) and (margin_kanan is ideal) then (evaluasi is diterima) (1)

If (margin_atas is not ideal) and (margin_bawah is ideal) and (margin_kiri is ideal) and (margin_kanan is ideal) then (evaluasi is tidak diterima) (2)

Setelah itu adalah tahap defuzzifikasi. Tahap defuzzifikasi bertujuan untuk mengubah hasil evaluasi fuzzy dari bentuk kategori menjadi skor kepatuhan margin dalam rentang antara 0 hingga 1. Nilai ini merepresentasikan tingkat kepatuhan margin terhadap pedoman penulisan laporan praktikum. Semakin tinggi skor fuzzy yang dihasilkan, semakin baik margin halaman tersebut dianggap sesuai dengan aturan.

Metode defuzzifikasi yang digunakan adalah Centroid (*Center of Gravity*). Metode ini mampu menghasilkan nilai yang stabil dan representatif dari distribusi fuzzy. Skor kepatuhan yang dihasilkan berada dalam rentang 0 hingga 1 dan akan digunakan sebagai fitur tambahan dalam proses klasifikasi *Naïve Bayes*. Proses defuzzifikasi dilakukan secara otomatis oleh sistem FIS (*Fuzzy Inference System*) yang digunakan.

F. Pelatihan Model *Naïve bayes*

Naïve Bayes adalah algoritma klasifikasi yang termasuk dalam kategori metode peluang. Keunggulannya terletak pada kemampuannya untuk pemodelan secara luas pada berbagai jenis dataset, baik kecil maupun besar, serta beragam jenis data. sehingga kita dapat berasumsi bahwa penelitian ini dengan dataset yang cukup terbilang kecil dapat diimplementasikan dengan algoritma ini. Dengan mengandalkan prinsip probabilitas dan statistik, algoritma ini menghitung kemungkinan suatu data termasuk dalam kelas tertentu berdasarkan fitur-fitur yang dimilikinya [13][14].

Model *naïve bayes* menghitung kemungkinan suatu data termasuk ke dalam sebuah kelas berdasarkan probabilitas bersyarat (*conditional probability*) dengan menerapkan Teorema Bayes, yang dirumuskan sebagai berikut: [15]

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

dengan $P(A|B)$ adalah probabilitas suatu kelas A diberikan fitur B, $P(B|A)$ adalah probabilitas fitur B muncul dalam kelas A, $P(A)$ adalah probabilitas awal (prior) kelas A, dan $P(B)$ adalah probabilitas gabungan dari fitur B di semua kelas.

Dalam penelitian ini, algoritma *Naive Bayes* digunakan untuk mengklasifikasikan halaman laporan menjadi dua kategori: Diterima dan Tidak Diterima. Fitur yang digunakan sebagai input terdiri dari nilai margin dalam satuan piksel (atas, bawah, kiri, kanan) serta output dari sistem *fuzzy* yang merepresentasikan tingkat kepatuhan margin terhadap format laporan. Dalam penelitian ini juga nantinya akan melakukan perbandingan antara evaluasi dengan output *fuzzy* dengan tanpa *output fuzzy*. Pemilihan algoritma *Naive Bayes* didasarkan pada suatu pertimbangan yaitu karena *naive bayes* sangat efektif digunakan pada dataset dengan jumlah fitur yang relatif sedikit dan bersifat numerik, seperti margin dan skor *fuzzy* yang digunakan dalam penelitian ini.

G. Evaluasi

Walaupun dalam penelitian ini menggunakan pelatihan kombinasi antara logika *fuzzy* dan *naive bayes*. Evaluasi memiliki tujuan yang sama seperti evaluasi *naive bayes* pada umumnya. Fokusnya adalah untuk menilai performa model yang dibangun, memahami kelebihan dan kekurangan dari model serta memastikan bahwa model dapat melakukan generalisasi dengan baik terhadap data yang baru.

Dalam penelitian ini terdapat beberapa jenis evaluasi yang digunakan. Diantaranya adalah akurasi yang dimana akurasi dimaksudkan untuk mengukur dari seluruh data yang dikumpulkan dan diklasifikasikan lalu dipersantasekan. Sehingga hasil ini dapat mewakili performa dari model. Lalu selanjutnya ada presisi dimana presisi dimaksudkan sebagai nilai rasio jumlah data yang benar diklasifikasikan sebagai suatu kelas tertentu terhadap semua data yang diklasifikasikan sebagai kelas tersebut. Jenis evaluasi selanjutnya ada *recall*. *Recall* atau sensitivitas dimaksudkan sebagai nilai Rasio jumlah data yang benar diklasifikasikan sebagai suatu kelas tertentu terhadap seluruh data yang sebenarnya termasuk dalam kelas tersebut. *Recall* ini menunjukkan seberapa baik model mampu menangkap semua data positif dari kelas tertentu. Jenis evaluasi yang terakhir adalah *F1-score*. *F1-score* merupakan harmonisasi dari presisi dan recall, yang membantu menyeimbangkan kedua aspek tersebut untuk mendapatkan gambaran kinerja yang lebih lengkap.

Untuk melakukan evaluasi dari model klasifikasi *fuzzy-naive bayes* ini digunakan pendekatan *k-fold cross validation* agar dapat menghindari *overfitting* dan memberikan estimasi performa model yang lebih andal. Metode *cross validation* memiliki keunggulan dalam membagi data ke dalam beberapa bagian (*fold*), di mana setiap bagian akan bergiliran menjadi data uji dan data latih. Hal ini memberikan gambaran performa model yang lebih stabil karena mengurangi pengaruh pembagian data yang tidak merata, serta menjaga agar setiap data digunakan dalam proses pelatihan dan pengujian. Untuk kasus ini digunakan *5-fold* dalam proses pemodelan.

Penggunaan *5-fold* dalam teknik *cross-validation* dipilih untuk akurasi evaluasi dan efisiensi komputasi.

Dalam metode ini, seluruh data dibagi secara acak menjadi lima bagian (*fold*) yang berukuran sama. Proses pelatihan dan pengujian dilakukan sebanyak lima kali, di mana pada setiap putaran satu *fold* digunakan sebagai data uji, sedangkan empat *fold* lainnya digunakan untuk melatih model.

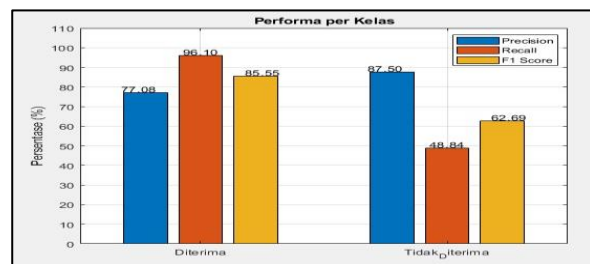
III. HASIL DAN PEMBAHASAN

Untuk melihat hasil evaluasi model dapat dilihat pada tabel dibawah.

Tabel 6. Hasil dari evaluasi fuzzy naive bayes

Jenis Evaluasi	Nilai Kelas yang Diterima	Nilai kelas yang Tidak Diterima
<i>Precision</i>	77%	88%
<i>Recall</i>	96%	49%
<i>F1 Score</i> :	86%	63%
Akurasi	80%	

Untuk visualisasi grafik bar terhadap hasil evaluasi *fuzzy naive bayes* dapat dilihat pada gambar dibawah ini:



Gambar 5. Visualisasi bar chart hasil evaluasi dengan fuzzy

Berdasarkan hasil test evaluasi dengan menggunakan kombinasi *fuzzy* dan *naive bayes* diatas, persantasi akurasi berada di 80% untuk masing-masing kelas dan untuk kelas yang "Diterima" memiliki persantasi jenis evaluasi precision dengan nilai 77%, *recall* dengan nilai 96%, dan *F1 Score* dengan nilai 86%. Sedangkan untuk kelas yang "Tidak Diterima" memiliki perbedaan persenstasi jenis evaluasi yang cukup signifikan dengan nilai *precision* sebesar 88%, *recall* 49% dan *F1 Score* 63%.

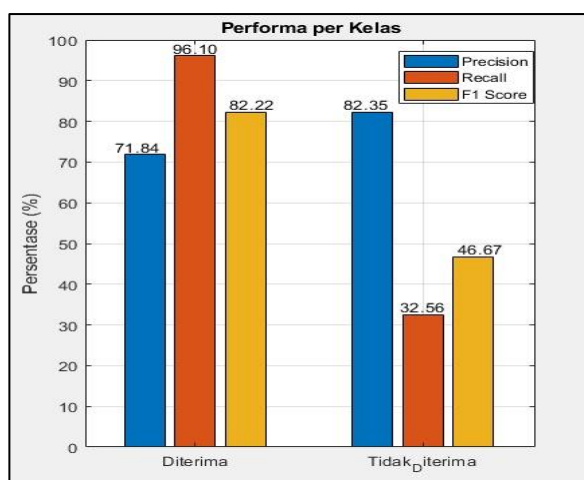
Model menghasilkan akurasi sebesar 80%, yang berarti mayoritas prediksi sistem sesuai dengan label aktual pada data. Nilai *recall* yang tinggi menunjukkan bahwa model sangat efektif dalam mengenali laporan yang memang sesuai dengan format margin, meskipun *precision*-nya sedikit lebih rendah, menandakan masih ada sebagian kecil laporan yang sebenarnya tidak sesuai tetapi diklasifikasikan sebagai "Diterima". Namun, performa model untuk mendeteksi laporan yang tidak sesuai format (kelas Tidak Diterima) masih belum optimal *recall*-nya belum mencapai angka ideal, nilai 48,84%. Hal ini berpotensi menimbulkan *false positive*, di mana sistem meloloskan laporan yang sebenarnya tidak sesuai ketentuan margin. *F1 score* pada kelas "Tidak Diterima" yang hanya mencapai 62,69% menunjukkan ketidakseimbangan performa antara presisi dan *recall*. Ini menandakan bahwa meskipun sebagian laporan ditolak dengan benar, banyak juga yang terlewatkan, sehingga perlu adanya peningkatan akurasi untuk kelas minoritas ini.

Setelah ini akan dilakukan evaluasi lagi dengan model klasifikasi *naïve bayes*, namun berbeda dengan tes evaluasi sebelumnya, jika tes sebelumnya menggunakan pemodelan kombinasi *fuzzy naïve bayes*. Maka, kali ini hanya menggunakan pemodelan dari performa klasifikasi *naïve bayes* saja. Untuk hasil evaluasi model dapat dilihat pada tabel dibawah

Tabel 7. Hasil dari evaluasi *naïve bayes*

Jenis Evaluasi	Nilai Kelas yang Diterima	Nilai kelas yang Tidak_Diterima
<i>Precision</i>	72%	82%
<i>Recall</i>	96%	33%
<i>F1 Score</i>	82%	47%
Akurasi	73%	

Untuk visualisasi grafik bar terhadap hasil evaluasi *naïve bayes* tanpa *fuzzy* dapat dilihat pada gambar dibawah ini:

Gambar 6. Visualisasi bar chart hasil evaluasi tanpa *fuzzy*

Berdasarkan hasil test evaluasi model dengan menggunakan evaluasi *naïve bayes* diatas, persantasi akurasi berada di 73% untuk masing-masing kelas dan untuk kelas yang "Diterima" memiliki persantasi jenis evaluasi *precision* dengan nilai 77%, *recall* dengan nilai 96%, dan *F1 Score* dengan nilai 82%. Sedangkan untuk kelas yang "Tidak Diterima" memiliki perbedaan persenstasi jenis evaluasi yang hampir sama dengan evaluasi model kombinasi *fuzzy* dan *naïve bayes* yaitu nilai diantara kedua kelas cukup signifikan dengan nilai kelas yang tidak diterima memiliki *precision* sebesar 82%, *recall* 33% dan *F1 Score* 47%.

Model *Naïve Bayes* tanpa fuzzifikasi memberikan akurasi sebesar 73,33%, yang secara umum menunjukkan bahwa model mampu mempelajari pola dari data margin yang digunakan. Nilai akurasi ini mencerminkan bahwa sekitar 3 dari 4 laporan berhasil diklasifikasikan dengan benar berdasarkan ukuran margin saja. Untuk kelas "Diterima", model menunjukkan performa yang sangat baik dengan *recall* sebesar 96%, yang berarti hampir seluruh laporan yang memang sesuai aturan margin berhasil dikenali. *F1 Score* sebesar 82% menunjukkan keseimbangan antara *precision* dan *recall*, yang mencerminkan bahwa model

cukup andal untuk mengenali laporan yang seharusnya diterima.

Hasil untuk kelas "Tidak Diterima" hampir sama dengan evaluasi model kombinasi *naïve bayes*. Beberapa *score* menunjukkan bahwa model ini dapat dikatakan masih belum optimal. *Recall* yang rendah yaitu 33% mengindikasikan bahwa banyak laporan yang seharusnya ditolak malah diklasifikasikan sebagai "Diterima" oleh model. Begitu juga dengan jenis evaluasi *F1 Score* yang menunjukkan nilai untuk kelas ini hanya mencapai 47%, yang menandakan ketidakseimbangan yang signifikan dalam kemampuan model untuk mendeteksi laporan yang tidak sesuai format. Meskipun *precision* cukup tinggi (82%), rendahnya *recall* membuat performa keseluruhan pada kelas "Tidak Diterima" masih kurang dapat diandalkan.

Dari hasil perbandingan, dapat dipahami bahwa penambahan fitur *fuzzy* memberikan kontribusi signifikan terhadap peningkatan akurasi keseluruhan dibandingkan model tanpa fitur *fuzzy*. Fitur *fuzzy* dapat merepresentasikan tingkat kepatuhan margin dalam bentuk skor terukur, membantu model menangkap nuansa ketidaksesuaian margin yang sebelumnya sulit dibedakan hanya dengan nilai piksel. Hal ini juga terbukti dari peningkatan *recall* pada kelas yang tidak diterima, yang menunjukkan bahwa model menjadi lebih sensitif dalam mengenali laporan bermasalah tanpa mengorbankan presisi secara drastis.

IV. KESIMPULAN

Hasil evaluasi menunjukkan bahwa akurasi model tanpa *fuzzy* mencapai 73,33%, dengan *precision* dan *recall* tinggi untuk kelas "Diterima" yaitu masing-masing 72% dan 96%, serta *F1 Score* sebesar 82%. Namun, performa terhadap kelas "Tidak Diterima" masih rendah, dengan *recall* hanya 33% dan *F1 Score* 47%, menandakan bahwa model ini cenderung bias terhadap kelas mayoritas (Diterima). Sementara itu, pada model *Fuzzy Naïve Bayes*, akurasi menjadi lebih tinggi yaitu 80%, dan terjadi peningkatan pada kemampuan mengenali kelas "Tidak Diterima", ditandai dengan kenaikan *recall* menjadi 49% dan *F1 Score* menjadi 63%. Alhasil model ini memberikan performa yang lebih seimbang antar kelas.

Dalam proses fuzzifikasi ketika mendata fungsi keanggotaan disarankan menggunakan nilai data yang tidak terlalu lebar rentangnya sehingga system dapat menentukan nilai ideal yang cukup valid. Alhasil output yang dapat diberikan setelah defuzzifikasi menjadi lebih efisien. Penelitian ini juga tidak menutup kemungkinan bahwa adanya peningkatan atau keterbaruan untuk penelitian selanjutnya. Kemungkinan untuk keterbaruan seperti pendeteksian spasi atau jarak antar kata, paragraf, judul ataupun subjudul.

REFERENSI

- [1] U. P. Pengaraian, "Penulisan skripsi," 2022, [Online]. Available: <https://pbio.uinsu.ac.id/>
- [2] B. Bataineh, "A Comprehensive Review on Document Image (DIBCO) Database," *A Compr. Rev. Doc. Image Bin.*, vol. 557,

- no. 1, pp. 1–30, 2025, doi: 10.1088/1757-899X/557/1/012006.
- [3] D. T. Anggraeni, “Perbaikan Citra Dokumen Hasil Pindai Menggunakan Metode Simple, Adaptive-Gaussian, dan Otsu Binarization Thresholding,” *Expert J. Manaj. Sist. Inf. dan Teknol.*, vol. 11, no. 2, p. 71, 2021, doi: 10.36448/expert.v11i2.2170.
- [4] M. K. Gibran, M. I. Rifki, A. H. Hasugian, A. T. A. A. Siahaan, A. Sahputra, and R. Ong, “Sentiment Analysis of Platform X Users on Starlink Using Naive Bayes,” vol. 15, no. June, pp. 113–119, 2024, [Online]. Available: <https://scholar.archive.org/work/te4arwrevjhhdg47urjqwoha2y/access/wayback/https://journalinstal.cattleyadf.org/index.php/Instal/article/download/8/8>
- [5] F. Azmi, M. K. Gibran, A. Ridwan, and A. Saleh, “Enhancing Water Potability Assessment Using Hybrid Fuzzy-Naïve Bayes,” *Indones. J. Comput. Sci.*, vol. 12, no. 3, pp. 1032–1043, 2023, doi: 10.33022/ijcs.v12i3.3232.
- [6] T. P. Negara, “PENERAPAN IMPLEMENTASI NAÏVE BAYES CLASSIFIER DAN METODE FUZZY AHP DALAM PENENTUAN BUKU LAYAK TERBIT MENGGUNAKAN ANDROID Teguh,” *Semin. Nas. Teknol. Inf. dan Komun. (SeNTIK STI&K)*, 2024.
- [7] D. C. M. Wijaya, B. Rahmat, and E. Y. Puspaningrum, “Sistem Pendukung Keputusan Berbasis Interval Type-2 Fuzzy Sugeno Pada Kendali pH Air,” *InComTech J. Telekomun. dan Komput.*, vol. 12, no. 3, p. 226, 2022, doi: 10.22441/incomtech.v12i3.15453.
- [8] T. Tundo and S. Saifullah, “Fuzzy Inference System Mamdani dalam Prediksi Produksi Kain Tenun Menggunakan Rule Berdasarkan Random Tree,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 3, pp. 443–452, 2022, doi: 10.25126/jtiik.2022924212.
- [9] R. Riskaputra, *Implementasi logika fuzzy mamdani dalam optimasi harga jasa distribusi produk (studi kasus: Kantor Konsultan Marketing RnD Enterprises)*. 2023. [Online]. Available: [https://repository.uinjkt.ac.id/dspace/handle/123456789/70749%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/70749/1/RAFFYANDA RISKAPUTRA-FST.pdf](https://repository.uinjkt.ac.id/dspace/handle/123456789/70749%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/70749/1/RAFFYANDA%20RISKAPUTRA-FST.pdf)
- [10] P. S. Saputra, “Perbandingan Algoritma Fuzzy C-Means Dan Algoritma Naive Bayes Dalam Menentukan Keluarga Penerima Manfaat (Kpm) Berdasarkan Status Sosial Ekonomi (Sse) Terendah,” *JST (Jurnal Sains dan Teknol.)*, vol. 10, no. 1, pp. 1–8, 2021, doi: 10.23887/jstundiksha.v10i1.23340.
- [11] A. S. Putra, S. Budiprayitno, and L. P. Rahayu, “Perancangan Sistem Kontrol pH dan Suhu Air Menggunakan Metode Fuzzy dan Terintegrasi dengan Internet of Things (IoT) pada Budidaya Ikan Hias,” *J. Tek. ITS*, vol. 10, no. 2, pp. 444–449, 2021, doi: 10.12962/j23373539.v10i2.74902.
- [12] A. Fadhl Rachman and Nuryuliani, “Analisis Tingkat Kepuasan Karyawan terhadap Aplikasi Service Desk menggunakan Metode Fuzzy Service Quality (Studi Kasus: Perusahaan Kliring di Jakarta) Analysis of Employee Satisfaction Levels with Service Desk Applications using the Fuzzy Service Quality,” *J. Bina Komputer*, vol. 4, no. 1, pp. 23–32, 2022.
- [13] A. Saleh, A. Ridwan, and M. K. Gibran, “Machine Learning and Fuzzy C-Means Clustering for the Identification of Tomato Diseases,” *Indones. J. Comput. Sci.*, vol. 12, no. 5, pp. 2401–2413, 2023, doi: 10.33022/ijcs.v12i5.3379.
- [14] J. Saputra, Y. Sa, V. Yoga Pudya Ardhana, and M. Afriansyah, “RESOLUSI: Rekayasa Teknik Informatika dan Informasi Klasifikasi Kematangan Buah Alpukat Mentega Menggunakan Metode K-Nearest Neighbor Berdasarkan Warna Kulit Buah,” *Media Online*, vol. 3, no. 5, pp. 347–354, 2023, [Online]. Available: <https://djournals.com/resolusi>
- [15] D. T. Anggraeni and C. Wibawa, “Perbaikan Citra Tanda Tangan Digital Menggunakan Metode Otsu Thresholding dan Sauvola,” *J. Ilm. Matrik*, vol. 25, no. 1, pp. 28–34, 2023, doi: 10.33557/jurnalnatrik.v25i1.2324.