

Unjuk Kerja Algoritma Support Vector Machine (SVM) dan Naïve Bayes Dalam Pengklasifikasian Berita Hoaks Pada Twitter Tentang Aksi Cepat Tanggap (ACT)

Amir Hasan^{1*}, Munirul Ula², Rini Meiyanti³

^{1,2,3} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Malikussaleh

*Corresponding Email: amir.170170068@mhs.unimal.ac.id

ABSTRAK

Twitter merupakan satu dari banyaknya media sosial yang populer di kalangan masyarakat. Terkadang informasi yang beredar di twitter merupakan berita palsu yang tidak dapat dibuktikan kebenarannya (hoaks). Penelitian ini menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM) untuk menentukan berita yang beredar di platform twitter mengenai Aksi Cepat Tanggap (ACT) termasuk ke dalam berita hoaks atau berita faktual. Proses klasifikasi dimulai dengan pengumpulan data dengan Teknik Scraping dan setelah itu dilakukan pelabelan untuk mengklasifikasi data latih. Data yang telah diberi label kemudian diproses melalui text pre-processing dan dilanjutkan dengan klasifikasi menggunakan metode Naïve Bayes dan Support Vector Machine (SVM). Jumlah data yang digunakan dalam penelitian ini sebanyak 1425 data dan dibagi ke dalam kategori fakta dan kategori hoaks. Pada proses klasifikasi algoritma Naïve Bayes mendapat nilai akurasi 66,76%, presisi 70,13%, dan recall 58,38%. Sedangkan hasil evaluasi klasifikasi Support Vector Machine (SVM) memiliki tingkat akurasi 65,22%, presisi 71,37%, dan recall 50,84%. Sehingga dapat disimpulkan performa algoritma Naïve Bayes memiliki performa yang lebih bagus dari algoritma Support Vector Machine.

Kata kunci: Aksi cepat tanggap, Klasifikasi, Twitter, Naïve Bayes, Support Vector Machine, Decision Support Systems

ABSTRACT

Twitter is one of the many social media platforms popular among the public. Sometimes the information circulating on Twitter is fake news that cannot be verified (hoaxes). This study uses the Naïve Bayes and Support Vector Machine (SVM) algorithms to determine whether news circulating on the Twitter platform regarding Aksi Cepat Tanggap (ACT) is classified as hoax or factual news. The classification process begins with data collection using scraping techniques, followed by labeling to classify the training data. The labeled data is then processed through text pre-processing and followed by classification using the Naïve Bayes and Support Vector Machine (SVM) methods. The amount of data used in this study is 1425 data points, divided into fact and hoax categories. In the classification process, the Naïve Bayes algorithm achieved an accuracy of 66.76%, a precision of 70.13%, and a recall of 58.38%. Meanwhile, the evaluation results of the Support Vector Machine (SVM) classification have an accuracy rate of 65.22%, a precision of 71.37%, and a recall of 50.84%. Thus, it can be concluded that the performance of the Naïve Bayes algorithm is better than that of the Support Vector Machine algorithm.

Keywords: Aksi Cepat Tanggap, Classification, Twitter, Naïve Bayes, Support Vector Machine

1. PENDAHULUAN

Aksi cepat tanggap adalah salah satu lembaga yang bekerja untuk menyebarkan program sosial dan kemanusiaan. Ini ditunjukkan oleh konsistensi dan keberadaan organisasi ini selama 14 tahun di bidang sosial dan kemanusiaan. Perkembangan teknologi semakin memudahkan manusia untuk mengakses berbagai hal, salah satunya melalui media sosial yang memungkinkan penggunaanya berbagai, bersosialisasi, atau bertukar

pesan dengan orang lain dengan mudah. Di Indonesia, pertumbuhan media sosial sangat pesat, terutama Twitter.

Hoaks adalah gabungan dari kebohongan dan kepalsuan. Di era penggunaan media sosial saat ini, penyebaran berita hoaks dan berita faktual terjadi dengan sangat mudah. Berita hoaks yang beredar di masyarakat dapat menyesatkan dan menimbulkan kesalahpahaman.

Penelitian ini merujuk pada penelitian terdahulu, salah satunya [1] yang meneliti Klasifikasi Artikel Hoaks Menggunakan Support Vector Machine Linear Dengan Pembobotan Frekuensi Term — Frekuensi Dokumen Inverse, yang memiliki tingkat akurasi sebesar 95.8333%.

Salah satu penelitian sebelumnya yang menjadi acuan penulis penelitian ini [2], yang berjudul Analisis Data Mining Klasifikasi Berita Hoax COVID 19 Menggunakan Algoritma Naive Bayes. Penelitian ini menghasilkan akurasi 86.3% dan 80–90% dari berita hoax memenuhi kriteria klasifikasi yang baik.

Temuan-temuan di atas juga didukung oleh penelitian [3] dengan judul Analisa dan Deteksi Konten Hoaks pada Media Berita Indonesia Menggunakan Machine Learning. Hasil penelitian tersebut menyimpulkan bahwa penggunaan algoritma Random Forest memberikan hasil klasifikasi yang terbaik dibandingkan klasifikasi yang dilakukan dengan Multilayer Perceptron, Naive Bayes, dan Support Vector Machine, dengan mencapai tingkat akurasi mencapai 75.37%.

Penelitian terdahulu [4] Penelitian ini menggunakan data dengan dua pengklasifikasi yang berbeda yaitu Naive Bayes dan SVM. Dua pengklasifikasi memprediksi label dalam dataset berdasar pada pola data dalam dataset. Hasilnya menunjukkan accuracy dari Naive Bayes adalah 91.48%, sedangkan SVM 85.47%. Hasil precision dari Naive Bayes adalah 89.28% dan hasil dari SVM adalah 90.95%. Sementara hasil recall dari Naive Bayes adalah 91.58% sedangkan SVM 76.18%. Jadi dapat disimpulkan bahwa metode Naive Bayes lebih baik dari SVM karena memberikan prediksi yang lebih akurat dan tepat. Perbedaan tersebut terjadi dikarenakan oleh kesesuaian karakteristik dari dataset dan parameter-parameter yang berbeda dari masing-masing algoritma. Dalam penelitian ini, Naive Bayes bekerja pada variabel independen, yaitu data teks tweet. Dari kesesuaian jenis data dan variabel ini, maka Naive Bayes dapat bekerja lebih baik dibanding Support Vector Machine.

Salah satu penelitian sebelumnya yang menjadi acuan penulis penelitian ini [5] Berdasarkan hasil pengujian menggunakan Naive Bayes Classifier dan Support Vector Machine maka ada beberapa hal yang dihasilkan, antara lain:

1) Hasil ketepatan klasifikasi menggunakan metode Naive Bayes Classifier diperoleh accuracy sebesar 81%, precision sebesar 80%, recall sebesar 99%, dan f1_score sebesar 89%.

2) Hasil ketepatan klasifikasi menggunakan metode Support Vector Machine diperoleh accuracy sebesar 87%, precision sebesar 88%, recall sebesar 96%, dan f1_score sebesar 92%.

3) Secara keseluruhan perbandingan performansi metode Naive Bayes Classifier dan Support Vector Machine menunjukkan hasil bahwa Support Vector Machine lebih baik dalam mengklasifikasi.

Salah satu penelitian sebelumnya yang menjadi acuan penulis penelitian ini [6] Berdasarkan hasil dari analisis sentimen pada penelitian ini, terlihat jasa transportasi online yang telah diprediksi dengan menggunakan metode Naive Bayes Classifier dan SVM menunjukan hasil yang signifikan. GrabId menggunakan metode SVM memberikan hasil yaitu class precision 66.57%, class recall 57.14%, Accuracy 84.08%. Sedangkan GojekIndonesia menggunakan metode SVM memberikan hasil yaitu class precision 54.78%, class recall 52.46%, Accuracy 69.50%. Disimpulkan metode SVM memiliki performance yang baik dibandingkan metode Naive Bayes Classifier.

2. METODOLOGI PENELITIAN

Metode yang digunakan dalam penelitian ini terdiri dari 5 tahapan, yaitu sebagai berikut:

A. Crawling Data

Pengambilan data pada platform twitter menggunakan Google Colab dan bahasa pemrograman python. Google Colab yang juga dikenal dengan sebutan Google Collaboraty merupakan platform berbasis awan yang dapat digunakan untuk menulis, menyimpan, dan membagikan kode python melalui website [7].

B. Study Literatur

Penelitian kepustakaan ialah proses meninjau literatur yang berkaitan dengan subjek yang sedang di kaji. Sebagai contoh, saat meneliti tentang perancangan sistem informasi dan analisis kelayakan sistem

informasi, maka peneliti akan mengumpulkan literatur berupa data, jurnal, kertas, makalah, bahan bacaan, dan artikel yang membahas mengenai topik tersebut.

C. Perancangan Sistem

Ada beberapa tahapan yang dilakukan dalam perancangan sistem yang ingin dibuat, yaitu desain skema, desain database dan desain antarmuka. Desain skema menghasilkan data flow diagram (DFD) dan diagram konteks, desain database menghasilkan entity relationship diagram (ERD).

D. Implementasi Sistem

Implementasi sistem merujuk pada penyelesaian masalah menggunakan sistem yang telah dirancang dengan bahasa pemrograman serta data base yang mendukung kinerja sistem tersebut. Penelitian ini menggunakan bahasa pemrograman PHP, sedangkan data base MYSQL.

E. Pengujian Sistem

Pengujian sistem digunakan untuk mengevaluasi sistem yang telah dijalankan. Melalui pengujian sistem dapat ditemukan kekurangan yang pada sistem dan melakukan perbaikan agar sistem yang dibuat benar-benar dapat memenuhi kebutuhan yang ada.

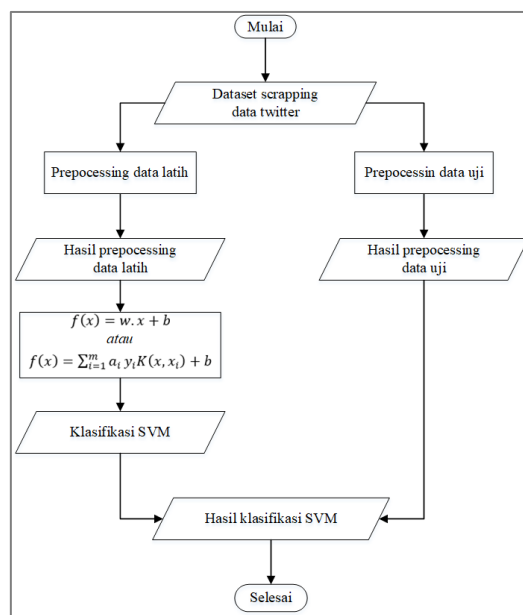
2.1 Support Vector Machine (SVM)

support vector machine ialah sebuah sistem pembelajaran dengan fungsi linier didalam ruang hipotesa dengan dimensi tinggi yang diturunkan dalam teori belajar statistik menggunakan algoritma pembelajaran optimasi [8]. Model Algoritma learning tinggi dalam hal analisis data untuk melakukan kalsifikasi dan regresi data [9]. Rumus Support Vector Machine :

$$f(x) = w \cdot x + b \text{ atau } f(x) = \sum_{i=1}^m a_i y_i K(x, x_i) + b \quad (1)$$

Penjelasan:

$f(x)$ = fungsi keputusan
 w = bobot dot product
 x = data ke- x
 b = bias



Gambar 1. Skema SVM

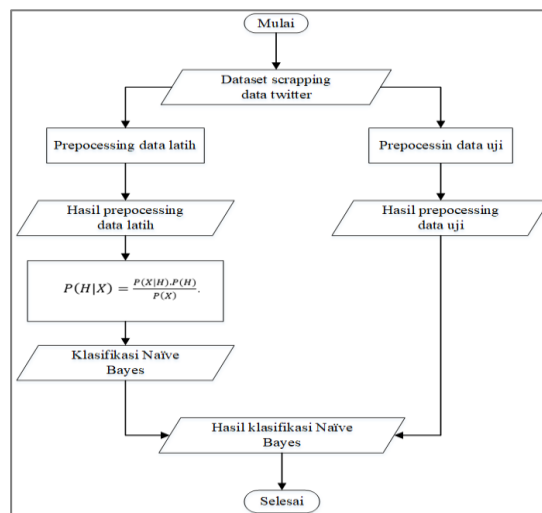
2.2 Naïve Bayes

Naïve Bayes merupakan satu dari banyaknya algoritma yang digunakan dalam teknik klasifikasi. Klasifikasi dengan metode Naive Bayes ditemukan oleh seorang ilmuwan inggris bernama Thomas Bayes. Metode ini juga dikenal dengan sebutan Teorema Bayes dan dikombinasikan dengan Naive yang mengansumsikan bahwa kondisi antar atribut saling independen. Menurut klasifikasi Naive Bayes, keberadaan atau tidaknya suatu karakteristik dalam membentuk suatu kelas tidak bergantung pada keberadaan karakter lainnya [7]. Rumus Naïve Bayes :

$$p(w_i|c_j) = \frac{1+n_j}{n + |kosakata|} \quad (2)$$

dimana:

$(w_i c_j)$	: probabilitas kata dari setiap kategori
n_j	: frekuensi kata dari setiap kategori
n	: jumlah seluruh kata pada dokumen
$ kosakata $	: jumlah total kata di semua data pelatihan



Gambar 2. Skema Naïve Bayes

III. HASIL DAN PEMBAHASAN

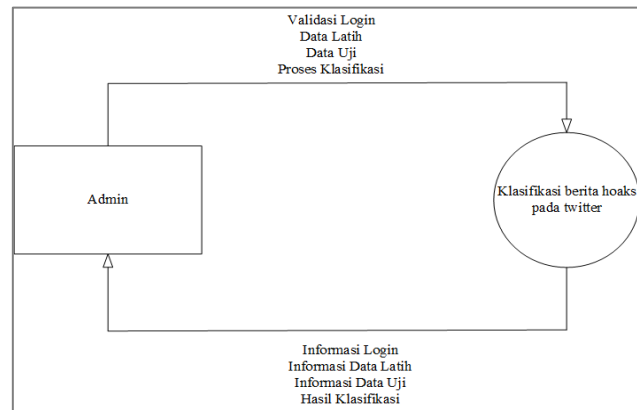
Pada penelitian ini penulis melakukan pengujian metode *Naïve Bayes* dan *Support Vector Machine* (SVM) untuk mengidentifikasi *tweet* berita yang disampaikan oleh pengguna media sosial khususnya *Twitter* tentang berita Aksi Cepat Tanggap (ACT).

3.1 Manajemen Basis Model

Pada tahap ini Data Flow Diagram (DFD) akan digunakan dalam perancangan aplikasi, dan perancangan ini akan diimplementasikan dalam bahasa pemrograman sebagai contoh alur kerja aplikasi.

A. Diagram Konteks

Diagram konteks menjelaskan secara keseluruhan *input*, *process*, dan *output* secara lengkap yang terdapat di dalam sistem. Berikut adalah gambar diagram konteks yang memperlihatkan proses sistem klasifikasi berita hoaks pada *Twitter*.

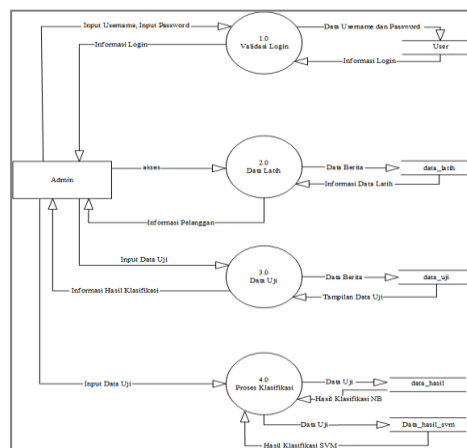


Gambar 3. Diagram Konteks

B. DFD Level 0

DFD level 0 merupakan suatu diagram yang menggambarkan interaksi sistem dengan *external entity*.

Berikut merupakan gambar dari DFD level 0.



Gambar 4. DFD Level 0

3.2 Text Processing

Berikut ini adalah pembahasan dari hasil klasifikasi tweet yang telah dikumpulkan dari platform Twitter dengan metode *Naïve Bayes* dan *Support Vector Machine* dan proses *Text Processing*.

A. Scraping

Tabel 1. Scraping Data

No	Berita	Label
1	Eks Presiden ACT Ibnu Khajar Divonis 3 Tahun Penjara	Fakta
2	Gelapkan Dana Rp117 Miliar, Pendiri ACT Hanya Divonis 3,5 Tahun Penjara	Fakta
3	ACT telah menjadi organisasi yang mendanai terorisme dengan mengatasnamakan aksi kemanusiaan.	Hoaks
4	Kucuran dana ke beberapa anggota yang pernah disinyalir bagian anggota Al-Qaeda di Turki	Hoaks

B. Case Folding

Tabel 2. Case Folding

No	Berita
1	eks presiden act ibnu khajar divonis 3 tahun penjara
2	gelapkan dana rp117 miliar pendiri act hanya divonis 3 5 tahun penjara
3	act telah menjadi organisasi yang mendanai terorisme dengan mengatasnamakan aksi kemanusiaan.
4	kucuran dana ke beberapa anggota yang pernah disinyalir bagian anggota al qaeda di turki

C. Stopword Removal

Tabel 3. Sropword Removal

No	Berita
1	eks presiden act ibnu khajar divonis 3 tahun penjara
2	gelapkan dana rp117 miliar pendiri act divonis 3 5 tahun penjara
3	act telah menjadi organisasi mendanai terorisme mengatasnamakan aksi kemanusiaan
4	kucuran dana beberapa anggota pernah disinyalir bagian anggota al qaeda turki

D. Stemming

Tabel 4. Stemming

No	Berita
1	eks presiden act ibnu khajar vonis 3 tahun penjara
2	gelap dana rp117 miliar pendiri act vonis 3 5 tahun penjara
3	act telah menjadi organisasi mendanai teroris atasnama aksi manusia
4	kucuran dana berapa anggota pernah sinyalir bagian anggota al qaeda turki

3.3 Klasifikasi

Setelah dilakukan pembobotan dan perhitungan pada tahap sebelumnya, maka dokumen yang telah diolah akan dilanjutkan ke tahap klasifikasi. Penelitian ini menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* untuk melakukan klasifikasi data.

A. Naïve Bayes

Tahapan berikutnya adalah melakukan perhitungan manual menggunakan algoritma *Naïve Bayes* yang akan diterapkan pada tiga dokumen.

Tabel 5. Data

No	Text Processing	Label
1	eks presiden act Ibnu khajar vonis 3 tahun penjara	Fakta
2	usut tuntas investasi bodong penipu	Fakta
3	act telah menjadi organisasi yang dana terorisme	Hoaks
4	polri usut duga seleweng dana manusia act	?

Tabel 6. Term Fakta

Fakta				
Word	nk	N	Vocab	P(wk vj)
eks	1	14	21	0,05714
presiden	1	14	21	0,05714
act	1	14	21	0,05714
ibnu	1	14	21	0,05714

khajar	1	14	21	0,05714
vonis	1	14	21	0,05714
3	1	14	21	0,05714
tahun	1	14	21	0,05714
penjara	1	14	21	0,05714
usut	1	14	21	0,05714
tuntas	1	14	21	0,05714
investasi	1	14	21	0,05714
bodong	1	14	21	0,05714
penipu	1	14	21	0,05714
telah	0	14	21	0,02857
jadi	0	14	21	0,02857
organisasi	0	14	21	0,02857
yang	0	14	21	0,02857
dana	0	14	21	0,02857
teroris	0	14	21	0,02857

Tabel 7. Term Hoaks

Hoaks				
Word	nk	N	Vocab	P(wk vj)
eks	0	7	21	0,035714
presiden	0	7	21	0,035714
act	1	7	21	0,071428
ibnu	0	7	21	0,035714
khajar	0	7	21	0,035714
vonis	0	7	21	0,035714
3	0	7	21	0,035714
tahun	0	7	21	0,035714
penjara	0	7	21	0,035714
usut	0	7	21	0,035714
tuntas	0	7	21	0,035714
investasi	0	7	21	0,035714
bodong	0	7	21	0,035714
penipu	0	7	21	0,035714
telah	1	7	21	0,071428
jadi	1	7	21	0,071428
organisasi	1	7	21	0,071428
yang	1	7	21	0,071428
dana	1	7	21	0,071428
teroris	1	7	21	0,071428

Berikut adalah komentar yang akan diuji “polri usut duga seleweng dana manusia act”, apakah komentar tersebut masuk kedalam kategori fakta atau hoaks.

$$\begin{aligned}
 P(\text{polri} \mid \text{fakta}) &= \frac{0+1}{14+21} = \frac{1}{35} = 0,02857 \\
 P(\text{usut} \mid \text{fakta}) &= \frac{1+1}{14+21} = \frac{2}{35} = 0,05714 \\
 P(\text{duga} \mid \text{fakta}) &= \frac{0+1}{14+21} = \frac{1}{35} = 0,02857 \\
 P(\text{seleweng} \mid \text{fakta}) &= \frac{0+1}{14+21} = \frac{1}{35} = 0,02857 \\
 P(\text{dana} \mid \text{fakta}) &= \frac{0+1}{14+21} = \frac{1}{35} = 0,02857 \\
 P(\text{manusia} \mid \text{fakta}) &= \frac{0+1}{14+21} = \frac{1}{35} = 0,02857 \\
 P(\text{act} \mid \text{fakta}) &= \frac{1+1}{14+21} = \frac{2}{35} = 0,05714 \\
 P(\text{polri} \mid \text{hoaks}) &= \frac{0+1}{7+21} = \frac{1}{28} = 0,03571 \\
 P(\text{usut} \mid \text{hoaks}) &= \frac{0+1}{7+21} = \frac{1}{28} = 0,03571 \\
 P(\text{duga} \mid \text{hoaks}) &= \frac{0+1}{7+21} = \frac{1}{28} = 0,03571 \\
 P(\text{seleweng} \mid \text{hoaks}) &= \frac{0+1}{7+21} = \frac{1}{28} = 0,03571 \\
 P(\text{dana} \mid \text{hoaks}) &= \frac{1+1}{7+21} = \frac{2}{28} = 0,07142 \\
 P(\text{manusia} \mid \text{hoaks}) &= \frac{0+1}{7+21} = \frac{1}{28} = 0,03571 \\
 P(\text{act} \mid \text{hoaks}) &= \frac{1+1}{7+21} = \frac{2}{28} = 0,07142
 \end{aligned}$$

Setelah menghitung bobot setiap term langkah selanjutnya yaitu menghitung prior probability untuk kelas fakta dan hoaks, tujuannya adalah untuk menentukan *prior probability* dari class fakta dan hoaks. Dari data latih dapat dicari sebagai berikut :

a. *Prior Probability* untuk kelas Fakta

$$P(\text{Fakta}) = \frac{2}{3} = 0,66$$

b. *Prior Probability* untuk kelas hoaks

$$P(\text{Hoaks}) = \frac{1}{3} = 0,33$$

$$P(\text{Fakta}|\text{d3}) = 0,66 * 0,02857 * 0,05714 * 0,02857 * 0,02857 * 0,02857 * 0,02857 * 0,05714 = 0,000000000036.$$

$$P(\text{Hoaks}|\text{d3}) = 0,33 * 0,03571 * 0,03571 * 0,03571 * 0,03571 * 0,07142 * 0,03571 * 0,07142 = 0,000000000087.$$

Dokumen yang diujikan “polri usut duga seleweng dana manusia act” dikategorikan dokumen dengan label hoaks. Kesimpulan ini diambil berdasarkan hasil perhitungan yang memperlihatkan *class* hoaks memiliki nilai yang lebih tinggi dibandingkan dengan nilai pada *class* fakta.

B. Support Vector Machine

Setelah dilakukan text processing maka dilakukan pembobotan dengan TF-IDF. Berikut diambil sampel dari 2 komentar yang telah di *processing* :

Komentar 1: gaji tinggi aksi cepat tanggap setara presiden indonesia

Komentar 2 : mahfud md act ajib hokum kutuk bukti salah

A. Perhitungan manual SVM

1) Proses Training

a) Melakukan kernelisasi

$$\begin{aligned}
 &K(x_1, x_1) \\
 &K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \\
 &= \exp(-0.5 \| [0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0, 0, 0, 0, 0, 0, 0] - [0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0, 0, 0, 0, 0, 0, 0] \|^2) \\
 &= \exp(-0.5 \| [0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301] \|^2) \\
 &= \exp(-0.5 \| [0, 0, 0, 0, 0] \|^2) \\
 &= \exp(0) \\
 &= 1
 \end{aligned}$$

$$\begin{aligned}
 &K(x_1, x_2) \\
 &K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \\
 &= \exp(-0.5 \| [0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0, 0, 0, 0, 0, 0, 0] - [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301] \|^2) \\
 &= \exp(-0.5 \| [0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, -0.301, -0.301, -0.301, -0.301, -0.301, -0.301, -0.301] \|^2) \\
 &= \exp(-0.5(1.449616)) \\
 &= \exp(-0.724808) \\
 &= 0.48442
 \end{aligned}$$

$$\begin{aligned}
 &K(x_2, x_1) \\
 &K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \\
 &= \exp(0.5 \| [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301] - [0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0, 0, 0, 0, 0, 0, 0] \|^2) \\
 &= \exp(0.5 \| [-0.301, -0.301, -0.301, -0.301, -0.301, -0.301, -0.301, -0.301, -0.301, -0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301] \|^2) \\
 &= \exp(-0.5(1.449616)) \\
 &= \exp(-0.724808) \\
 &= 0.48442
 \end{aligned}$$

$$\begin{aligned}
 &K(x_2, x_2) \\
 &K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \\
 &= \exp(0.5 \| [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301] - [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, 0.301] \|^2) \\
 &= \exp(0.5 \| [0, 0, 0, 0, 0, 0, 0, 0, 0, 0] \|^2) \\
 &= \exp(0) \\
 &= 1
 \end{aligned}$$

Sehingga membentuk matriks :

$$K(x_i, x_j) = \begin{bmatrix} 1 & 0.48442 \\ 0.48442 & 1 \end{bmatrix}$$

b) Melakukan perhitungan hessian matrix

$$\begin{aligned}
 x_1 x_1 &= (-1)(-1)(1 + 0.5^2) = 1.25 \\
 x_1 x_2 &= (-1)(1)(0.48442 + 0.5^2) = -0.73442 \\
 x_2 x_1 &= (1)(-1)(0.48442 + 0.5^2) = -0.73442 \\
 x_2 x_2 &= (1)(1)(1 + 0.5^2) = 1.25
 \end{aligned}$$

Sehingga membentuk matrix :

$$D_{ij} = \begin{bmatrix} 1.25 & -0.73442 \\ -0.73442 & 1.25 \end{bmatrix}$$

c) Mencari nilai error

$$\begin{aligned}
 x_1 &= (1.25 + (-0.73442)) \times 0.5 = 0.25779 \\
 x_2 &= (-0.73442 + 1.25) \times 0.5 = 0.25779
 \end{aligned}$$

Sehingga membentuk matrix :

$$E_i = \begin{bmatrix} 0.25779 \\ 0.25779 \end{bmatrix}$$

d) menghitung delta alpha

$$x_1 = \text{Min}(\text{Max}(0.5(1 - 0.25779), -0.5), 1 - 0.5) = \text{Min}(0.371105) \\ = 0.371105$$

Dilakukan sampai perhitungan X2, sehingga membentuk matrix :

$$\alpha_i = \begin{bmatrix} 0.371105 \\ 0.371105 \end{bmatrix}$$

e) Menghitung alpha baru

$$x_1 = (0.5 + 0.371105) = 0.871105$$

Lakukan sampai x2 sehingga membentukmatrix:

$$\alpha_i = \begin{bmatrix} 0.871105 \\ 0.871105 \end{bmatrix}$$

Diulangi sampai keadaan konvergen pada svm biner 1 diulangi hingga iterasi ke 2, berikut matrixnya:

$$\alpha_i = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

f) Mencari nilai $w \cdot x^+$ dan $w \cdot x^-$ untuk mencari bias

Nilai negatif terletak pada data ke-1

$$w \cdot x^- = ((-1 \times 1 \times 1) + (1 \times 1 \times 0.48442)) \\ = (-1 + (0.48442)) \\ = -0.51558$$

Nilai positif terletak pada data ke-2

$$w \cdot x^+ = ((-1) \times 1 \times 0.48442) + (1 \times 1 \times 1)) \\ = (-0.48442 + 1) = 0.51558$$

g) Menghitung nilai bias

$$b = -\frac{1}{2}(-0.51558 + 0.51558) = 0$$

2. Proses Testing

Data uji diambil dari komentar 1 dengan vektor sebagai berikut :

$$x_1 = \begin{bmatrix} 0.301, 0.301, 0.301, 0.301, 0.301, 0.301, \\ 0.301, 0.301, 0, 0, 0 \dots \end{bmatrix}$$

a) melakukan perhitungan kernel

Tabel 8. Perhitungan Kernel

K(x,x1)	K(x,x2)
1	0.48442

b) Untuk mengetahui kelas data uji maka dihitung keputusan data sign()

$$D(x) = \text{sign}((-1) \times 1 \times 1) + (1) \times 1 \times 0.48442)) + 0 \\ (x) = \text{sign}(-0.51558) \\ (x) = -1 \text{ (negatif)}$$

Dari hasil perhitungan diatas komentar yang telah dilakukan pengujian merupakan kelas negatif (hoaks).

3.4 Evaluasi Klasifikasi

Tahap Evaluasi klasifikasi dilakukan ketika klasifikasi telah selesai dilakukan. Pengujian pada tahap ini ditujukan untuk mengetahui nilai akurasi, presisi, dan *recall* dengan memanfaatkan *Confusion Matrix*.

A. Confusion Matrix Naïve Bayes

Tabel 9. Confusion Matrix NB

Kelas Klasifikasi	Kelas Klasifikasi Fakta	Kelas Klasifikasi Hoaks
Fakta	209	149
Hoaks	89	269
Presisi	70,13%	
Recall	58,38%	
Akurasi	66,76%	

B. Confusion Matrix SVM

Tabel 10. Confusion Matrix SVM

Kelas Klasifikasi	Kelas Klasifikasi Fakta	Kelas Klasifikasi Hoaks
Fakta	182	176
Hoaks	73	285
Presisi	71,37%	
Recall	50,84%	
Akurasi	65,22%	

IV. SIMPULAN

Dalam penelitian ini dilakukan beberapa tahapan proses, proses pertama yaitu mengumpulkan data dengan scraping dan dilakukan pelabelan pada data, setelah itu tahapan proses selanjutnya yaitu text processing, lalu dilakukan pengklasifikasian dengan Metode Naïve Bayes dan Support Vector Machine (SVM). Kinerja kedua algoritma dalam proses pengklasifikasian yakni Naïve Bayes dan Naïve Bayes memberikan hasil yang dapat dipercaya. Kepercayaan ini dibuktikan dengan tingkat akurasi yang diperoleh keduanya lebih dari 66,76%. Berdasarkan hasil evaluasi proses klasifikasi, algoritma Naïve Bayes menghasilkan akurasi 66,76%, presisi 70,13%, dan recall 58,38%. Sedangkan hasil evaluasi klasifikasi Support Vector Machine (SVM) menghasilkan akurasi 65,22%, presisi 71,37%, dan recall 50,84%. Tingkat akurasi yang dimiliki oleh algoritma *Naïve Bayes* mendapat nilai yang lebih tinggi, yakni diangka 66,76% dibandingkan dengan algoritma *Support Vector Machine (SVM)* yang memperoleh nilai akurasi pada angka 65,22%.

REFERENCES

- [1] Rahma Hidayat, A., Pandu Adikara, P., & Adinugroho, S. (2020). Klasifikasi Hoaks Kesehatan di Media Sosial menggunakan Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 4(6), 1702–1708. <http://j-ptiik.ub.ac.id>.
- [2] Prasetya, F., & Ferdiansyah, F. (2022). Analisis Data Mining Klasifikasi Berita Hoax COVID 19 Menggunakan Algoritma Naive Bayes. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 4(1), 132. <https://doi.org/10.30865/json.v4i1.4852>.
- [3] Ula, M., Rahmat, T., & Alfianov, M. M. (2020). ANALISA DAN DETEKSI KONTEN HOAX PADA MEDIA BERITA INDONESIA MENGGUNAKAN MACHINE LEARNING. *Jurnal Teknologi Terapan and Sains 4.0*, 1(2), 229. <https://doi.org/10.29103/tts.v1i2.3263>.

- [4] Dinata, R. K., Hasdyna, N., & Alif, M. (2021). Applied of Information Gain Algorithm for Culinary Recommendation System in Lhokseumawe. *Journal Of Informatics And Telecommunication Engineering*, 5(1), 45-52.
- [5] Widowati & Sadikin, 2021)Widowati, T. T., & Sadikin, M. (2021). Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 11(2), 626–636. <https://doi.org/10.24176/simet.v11i2.4568>.
- [6] Dinata, R. K., Akbar, H., & Hasdyna, N. (2020). Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus. *ILKOM Jurnal Ilmiah*, 12(2), 104-111.
- [7] Raharjo, R. A., Sunarya, I. G., & Divayana, D. G. H. (2022). Perbandingan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Kasus Analisis Sentimen Terhadap Data Vaksin Covid-19 Di Twitter.
- [8] Dwianto, E., & Sadikin, M. (2021). Analisis Sentimen Transportasi Online pada Twitter Menggunakan Metode Klasifikasi Naïve Bayes dan Support Vector Machine. *Format : Jurnal Ilmiah Teknik Informatika*, 10(1), 94. <https://doi.org/10.22441/format.2021.v10.i1.009>.
- [9] Nugraha, F. T., & Hendry, H. (2023). Implementasi Machine Learning Sebagai Analisis Kepuasan Pelanggan Terhadap Penggunaan Aplikasi Kai Access. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 8(4), 1248–1255.
- [10] Dinata, R. K., Fajriana, F., Zulfa, Z., & Hasdyna, N. (2020). Klasifikasi Sekolah Menengah Pertama/Sederajat Wilayah Bireuen Menggunakan Algoritma K-Nearest Neighbors Berbasis Web. *CESS (Journal of Computer Engineering, System and Science)*, 5(1), 33-37.
- [11] Martantoh, E., & Yanih, N. (2022). Implementasi Metode Naïve Bayes Untuk Klasifikasi Karakteristik Kepribadian Siswa Di Sekolah MTS Darussa'adah Menggunakan Php Mysql. *Jurnal Teknologi Sistem Informasi*, 3(2), 166–175. <https://doi.org/10.35957/jtsi.v3i2.2896>.