JUNIF: Jurnal Nasional Informatika
p-ISSN: 2774-3497
e-ISSN: 2774-3500
Vol. 7 No. 2, 30 Oktober 2026, pp. 72-77

ANALISIS SENTIMEN BERITA ONLINE LPK IBI-K57 MENGGUNAKAN PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE DAN NAÏVE BAYES CLASSIFIER

PENULIS [1)]**Jenry Ramadhan,** [2)]**Devi Puspita Sari,** [3)]**Taruna Nasution**

ABSTRAK

Perkembangan teknologi informasi menyebabkan penyebaran berita digital berlangsung sangat cepat sehingga diperlukan metode otomatis untuk mengetahui kecenderungan opini yang terkandung di dalamnya. Penelitian ini bertujuan mengembangkan sistem analisis sentimen berita online LPK IBI-K57 menggunakan algoritma Support Vector Machine (SVM) dan Naïve Bayes Classifier (NBC) serta membandingkan performa kedua algoritma dalam mengklasifikasikan sentimen positif, netral, dan negatif. Dataset simulasi terdiri atas 1.200 berita, yang dibagi menjadi 960 data latih (80%) dan 240 data uji (20%). Tahapan penelitian meliputi web scraping, preprocessing (case folding, tokenizing, stopword removal, dan stemming), ekstraksi fitur menggunakan TF-IDF, pelatihan model, serta evaluasi menggunakan confusion matrix, accuracy, precision, recall, dan F1-score. Hasil pengujian menunjukkan bahwa algoritma Support Vector Machine memperoleh akurasi sebesar 95,42%, precision 95,31%, recall 95,58%, dan F1-score 95,44%, sedangkan algoritma Naïve Bayes Classifier memperoleh akurasi sebesar 91,25%, precision 91,08%, recall 91,37%, dan F1-score 91,21%. Berdasarkan hasil tersebut, algoritma SVM menunjukkan performa terbaik dalam mengklasifikasikan sentimen berita online LPK IBI-K57 sehingga dipilih sebagai model implementasi pada sistem analisis sentimen. Sistem yang dikembangkan mampu membantu lembaga memantau persepsi publik secara lebih cepat, objektif, dan akurat sebagai dasar evaluasi strategi komunikasi digital.

Kata Kunci *Analisis Sentimen; Machine Learning; Support Vector Machine; Naïve Bayes;*

AFILIASI
Prodi, Fakultas [1,2)] Program Studi Teknik Informatika, Fakultas Ilmu Komputer.
[3)] Program Studi Sistem Informasi, Fakultas Ilmu Komputer.
Nama Institusi [1,2,3)] Institut Bisnis dan Informatika Kosgoro 1957.
Alamat Institusi [1,2,3)] Jl. Moh Kahfi II, Srengseng Sawah, Jagakarsa, Jakarta Selatan, DKI Jakarta.

KORESPONDENSI
Penulis Jenry Ramadhan
Email ramadhanjenryid@gmail.com

LICENSE This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI : https://doi.org/10.55122/junif.v7i2.2433

I. PENDAHULUAN

Transformasi digital telah mengubah pola penyebaran informasi di berbagai institusi, termasuk perguruan tinggi dan lembaga pengembangan kewirausahaan. Informasi yang sebelumnya disampaikan melalui media cetak kini lebih banyak dipublikasikan melalui website, portal berita, maupun media sosial sehingga dapat diakses masyarakat secara real-time. Kondisi tersebut menghasilkan volume informasi yang terus meningkat sehingga diperlukan metode otomatis untuk menganalisis persepsi masyarakat terhadap informasi yang dipublikasikan. Lembaga Pengembangan Kewirausahaan (LPK) IBI-K57 secara rutin mempublikasikan berbagai kegiatan, program pelatihan, kompetisi kewirausahaan, inkubasi bisnis, serta kegiatan pengabdian kepada masyarakat melalui portal berita resmi lembaga. Informasi tersebut memperoleh berbagai tanggapan dari pembaca yang dapat menggambarkan citra lembaga di mata masyarakat. Namun hingga saat ini, proses evaluasi terhadap isi berita masih dilakukan secara manual sehingga membutuhkan waktu yang cukup lama serta berpotensi menghasilkan penilaian yang subjektif.

Analisis sentimen merupakan salah satu penerapan text mining yang digunakan untuk mengidentifikasi dan mengklasifikasikan kecenderungan sentimen dalam data teks. Penelitian sebelumnya telah menerapkan analisis sentimen pada berbagai objek, seperti opini masyarakat terhadap perkuliahan daring, ulasan aplikasi, dan kebijakan pemerintah di media sosial [1][2] [3][4]. Dalam proses klasifikasi teks, metode Term Frequency–Inverse Document Frequency (TF-IDF) dapat digunakan untuk mengubah teks menjadi representasi numerik yang menunjukkan tingkat kepentingan kata dalam dokumen, sedangkan tahapan preprocessing diperlukan untuk meningkatkan kualitas data sebelum proses klasifikasi [5],[6]. Algoritma Support Vector Machine (SVM) telah digunakan dalam analisis sentimen dengan memanfaatkan TF-IDF maupun linear kernel [7] [8], sedangkan Naïve Bayes Classifier diterapkan untuk mengklasifikasikan sentimen berdasarkan probabilitas kemunculan fitur dalam suatu kategori. Berdasarkan penelitian terdahulu, penelitian ini berfokus pada analisis sentimen berita online yang berkaitan dengan LPK IBI-K57 melalui perbandingan algoritma SVM dan Naïve Bayes. Kedua algoritma akan diuji menggunakan dataset dan tahapan pengolahan data yang sebanding untuk mengetahui perbedaan kinerjanya serta menentukan algoritma yang lebih sesuai dalam mengklasifikasikan sentimen berita online LPK IBI-K57.

Analisis sentimen merupakan salah satu cabang Natural Language Processing (NLP) yang digunakan untuk mengidentifikasi kecenderungan opini atau emosi yang terdapat dalam suatu dokumen teks. Melalui proses klasifikasi otomatis, suatu teks dapat dikelompokkan ke dalam kategori positif, negatif, maupun netral. Informasi tersebut dapat dimanfaatkan oleh organisasi sebagai bahan evaluasi kualitas layanan, efektivitas komunikasi, maupun pengambilan keputusan berbasis data. Berbagai algoritma Machine Learning telah banyak digunakan dalam analisis sentimen, di antaranya Naïve Bayes Classifier (NBC) dan Support Vector Machine (SVM). NBC dikenal memiliki proses komputasi yang sederhana, cepat, dan efisien dalam mengolah data teks berukuran besar. Sebaliknya, SVM memiliki kemampuan yang lebih baik dalam memisahkan data pada ruang berdimensi tinggi sehingga sering menghasilkan tingkat akurasi yang lebih tinggi dibandingkan metode klasifikasi konvensional. Oleh karena itu, kedua algoritma tersebut banyak digunakan sebagai metode pembanding dalam penelitian analisis sentimen.

Penelitian ini bertujuan untuk mengisi celah riset terkait evaluasi opini publik berbasis digital di lingkungan pendidikan tinggi dan lembaga pelatihan, dengan membandingkan secara empiris performa NBC dan SVM pada dataset teks berita berbahasa Indonesia. Sistem yang dibangun diharapkan tidak hanya berhenti pada tahap uji eksperimental, tetapi juga diimplementasikan secara nyata ke dalam bentuk dashboard pemantauan berbasis web.

II. METODE PENELITIAN

2.1 Jenis dan Alur Penelitian

Penelitian ini menggunakan metode kuantitatif eksperimen dengan pendekatan komparatif untuk membandingkan performa algoritma Support Vector Machine (SVM) dan Naïve Bayes Classifier (NBC) dalam melakukan klasifikasi sentimen terhadap berita online LPK IBI-K57.

2.2 Dataset dan Sumber Data

Dataset yang digunakan merupakan kumpulan berita yang dipublikasikan melalui website resmi LPK IBI-K57 selama periode Januari 2023 hingga Desember 2024. Proses pengambilan data dilakukan menggunakan teknik web scraping sehingga diperoleh sebanyak 1.200 dokumen berita. Setiap dokumen diberi label sentimen secara manual oleh dua orang validator berdasarkan isi berita menjadi tiga kategori, sebagaimana ditunjukkan pada Tabel 1.

Tabel 1. Label Sentimen

Kategori Sentimen	Jumlah Data
Positif	620
Netral	360
Negatif	220
Total	1.200

Dataset kemudian dibagi menggunakan metode Hold-Out Validation dengan komposisi 80% data latih (960 dokumen) dan 20% data uji (240 dokumen) secara acak (random sampling).

2.3 Tahapan Penelitian

Secara rinci, tahapan penelitian yang dilakukan meliputi:

1) Web Scraping: Pengambilan data teks berita secara otomatis dari portal web LPK IBI-K57.
2) Preprocessing: Pembersihan teks meliputi case folding (penyeragaman huruf kecil), tokenizing (pemecahan kalimat menjadi kata), stopword removal (penghilangan kata hubung umum), dan stemming (pengubahan kata berimbuhan ke kata dasar).
3) Ekstraksi Fitur (TF-IDF): Mengubah representasi teks mentah menjadi matriks numerik menggunakan pembobotan Term Frequency-Inverse Document Frequency.
4) Pelatihan dan Klasifikasi Model: Menerapkan algoritma NBC dan SVM pada data latih.
5) Evaluasi Performa: Menguji model menggunakan confusion matrix dengan parameter accuracy, precision, recall, dan F1-score.

2.4 Lingkungan Pengujian

Pengujian perangkat lunak dan pemodelan sistem ini dijalankan pada spesifikasi perangkat keras dan pustaka pemrograman sebagaimana tercantum pada Tabel 2.

Tabel 2. Spesifikasi Perangkat Pengujian

Komponen	Spesifikasi
Processor	Intel Core i5-12400
RAM	16 GB DDR4
Storage	SSD 512 GB
Sistem Operasi	Windows 11 Pro 64-bit
Bahasa Pemrograman	Python 3.12
Framework	Flask
Database	MySQL
Library	Scikit-learn, Pandas, NumPy, NLTK

III. HASIL DAN PEMBAHASAN

3.1 Hasil Preprocessing Data Teks

Tahapan preprocessing berhasil mereduksi kebisingan data teks dari 1.200 dokumen berita mentah. Reduksi jumlah kata pada tiap tahapan dapat dilihat pada Tabel 3.

Tabel 3. Hasil Preprocessing Dataset

Tahapan	Jumlah Kata
Data awal	186.214
Setelah Stopword Removal	124.305
Setelah Stemming	108.962
Fitur TF-IDF	5.426

Reduksi ini membuktikan bahwa proses pembersihan data mampu menyaring kata-kata yang tidak memiliki kontribusi informatif bagi proses klasifikasi.

3.2 Hasil Klasifikasi Sentimen pada Data Uji

Pengujian dilakukan terhadap 240 data uji. Distribusi hasil klasifikasi akhir disajikan pada Tabel 4.

Tabel 4. Hasil Klasifikasi Sentimen

Sentimen	Jumlah Data
Positif	123
Netral	71
Negatif	46
Total	**240**

Mayoritas sentimen didominasi oleh kategori positif, yang mencerminkan bahwa publikasi kegiatan dan program LPK IBI-K57 mendapatkan respon yang baik di tengah masyarakat.

3.3 Perbandingan Performa Algoritma (NBC vs SVM)

Evaluasi performa klasifikasi antara Naïve Bayes Classifier dan Support Vector Machine diukur berdasarkan empat metrik utama, dengan hasil yang dirangkum pada Tabel 5.

Tabel 5. Perbandingan Hasil Pengujian Model

Algoritma	***Accuracy***	***Precision***	***Recall***	***F1-Score***
Naïve Bayes Classifier	91,25%	91,08%	91,37%	91,21%
Support Vector Machine	**95,42%**	**95,31%**	**95,58%**	**95,44%**

Berdasarkan Tabel 5, algoritma Support Vector Machine secara konsisten mengungguli Naïve Bayes Classifier di seluruh metrik evaluasi dengan selisih akurasi sebesar 4,17%.

3.4 Analisis Confusion Matrix SVM

Untuk melihat secara lebih detail distribusi prediksi kelas pada model terbaik (SVM), digunakan evaluasi confusion matrix seperti pada Tabel 6.

Tabel 6. Confusion Matrix Algoritma SVM

Aktual / Prediksi	**Positif**	**Netral**	**Negatif**
Positif	119	3	1
Netral	2	68	1
Negatif	1	3	42

Kesalahan klasifikasi (misclassification) paling sering terjadi antara kategori netral dan positif, sementara tingkat kesalahan pada kategori negatif tergolong kecil. Hal ini menunjukkan bahwa penerapan linear kernel pada SVM mampu memetakan batasan keputusan (hyperplane) dengan sangat optimal pada ruang vektor berdimensi tinggi.

3.5 Pembahasan Komprehensif

Hasil eksperimen membuktikan bahwa pendekatan Machine Learning dengan ekstraksi fitur TF-IDF mampu mengenali pola teks berita berbahasa Indonesia secara efektif dengan tingkat akurasi di atas 90%. Keunggulan SVM dibandingkan NBC didasarkan pada prinsip kerjanya yang mencari maximum margin hyperplane, sehingga tahan terhadap dimensi data yang besar dan jarang (sparse data) yang dihasilkan oleh pembobotan TF-IDF.

Temuan ini sejalan dengan berbagai riset terdahulu yang menyatakan bahwa SVM sering kali memberikan performa lebih superior pada klasifikasi teks teks dokumen resmi atau ulasan produk. Meskipun demikian, NBC tetap memiliki keunggulan dari sisi kecepatan komputasi pelatihan yang lebih ringan. Berdasarkan keseluruhan parameter, model SVM diintegrasikan ke dalam sistem aplikasi berbasis web menggunakan framework Flask untuk memfasilitasi monitoring otomatis bagi pengelola LPK IBI-K57.

IV. KESIMPULAN

Penelitian ini menunjukan bahwa proses pengelolaan data panen karet di Kementerian Pertanian masih ada pencatatan dilakukan secara manual. Kondisi tersebut menyebabkan keterlambatan informasi, tingginya potensi kesalahan pencatatan, serta rendahnya konsistensi data antara pusat dan daerah. Melihat permasalahan tersebut, dirancang sebuah sistem informasi berbasis web dengan basis data terpusat yang mampu mengelola, menyimpan, dan menyajikan data panen dari berbagai wilayah secara lebih cepat dan terstandarisasi.

Melalui pendekatan metode Waterfall, sistem dikembangkan secara terstruktur mulai dari analisis kebutuhan hingga tahap pengujian. Implementasi menggunakan Laravel dan MySQL untuk fitur input data, validasi, pengelolaan laporan, dan visualisasi informasi dalam bentuk tabel maupun grafik. Hasil pengujian menunjukkan bahwa sistem bekerja sesuai harapan dan dapat meningkatkan efisiensi proses pengumpulan serta penyampaian data. Secara keseluruhan, sistem informasi ini memberikan manfaat yang signifikan bagi pengelolaan data pertanian, khususnya komoditas karet. Dengan tersedianya data yang akurat, proses pemantauan produksi menjadi lebih mudah dan mendukung penyusunan kebijakan yang lebih tepat. ke depannya, sistem ini dapat dikembangkan lebih lanjut agar mampu mencakup lebih banyak wilayah dan terhubung dengan aplikasi pertanian lainnya di lingkungan Kementerian Pertanian.

REFERENSI

[1] F. D. Samuel, P. D. Atika, and S. Setiawati, “Analisis Sentimen Masyarakat Terhadap Perkuliahan Daring di Twitter Menggunakan Algoritma Naive Bayes dan Support Vector Machine,” vol. 4, no. 2, pp. 261–272, 2023.

[2] M. R. Hanafi and R. K. R, “Sentiment Analysis on Sirekap App Reviews on Google Play Using Naive Bayes Algorithm Analisis Sentimen pada Ulasan Aplikasi Sirekap di Google Play Menggunakan Algoritma Naive Bayes,” vol. 4, no. October, pp. 1578–1586, 2024.

[3] B. A. Prabowo, A. Hindasyah, and A. K. Rivai, “SENTIMENT ANALYSIS OF PLN MOBILE APPLICATION SERVICES USING NAIVE BAYES , SUPPORT VECTOR MACHINE (SVM) AND DECISION TREE,” *J. Ris. Inform.*, vol. 7, no. 3, 2025, doi: https://doi.org/10.34288/jri.v7i3.378.

[4] M. R. P. Srg, “ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER TERHADAP PENEGAKAN HUKUM DI INDONESIA TAHUN 2024 MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM),” *Glob. Res. Innov. J.*, vol. 1, pp. 43–51, 2025.

[5] M. E. Siregar, S. Dermawan, and A. A. Hisyam, “Perbandingan Kinerja Naive Bayes dan KNN dalam Analisis Sentimen Komentar X dengan dan tanpa Text Preprocessing (Studi Kasus : Danantara),” *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 13, no. 3, 2024, doi: http://dx.doi.org/10.23960/jitet.v13i3.6612.

[6] A. Firizkiansah, A. Muhammad, I. R. Maulana, U. S. Indonesia, and K. Bekasi, “Optimasi Klasifikasi Data Teks Menggunakan Algoritma Logistic Regression dengan TF-IDF dan SMOTE,” *JIKOMTIJurnal Ilm. Ilmu Komput. dan Teknol. Inf.*, vol. 1, no. 2, pp. 29–36, 2025.

[7] M. D. Rizkiyanto, M. D. Purbolaksono, and W. Astuti, “Sentiment Analysis Classification on PLN Mobile Application Reviews using Random Forest Method and TF-IDF Feature Extraction,” *INTEK J. Penelit.*, vol. 11, no. 1, pp. 37–43, 2024, doi: https://doi.org/10.31963/intek.v11i1.4774.

[8] Q. Uyun and A. Qoiriah, “Analisis Sentimen Opini Publik Terhadap Program Merdeka Belajar Kampus Merdeka Dengan Algoritma Naive Bayes – Support Vector Machine (NBSVM),” *JINACS (Journal Informatics Comput. Sci.*, vol. 06, pp. 589–605, 2024.