Jurnal Minfo Polgan
Volume 13, Nomor 1, Agustus 2024 e-ISSN : 2797-3298
DOI : https://doi.org/10.33395/jmp.v13i1.14014 p-ISSN : 2089-9424

Terbit : 26 Agustus 2024

Optimalisasi Algoritma Naive Bayes Dengan Teknik Ensemble Dalam Analisis Sentimen Twitter Pantai Kartini Jepara

[1]Muhammad Arqom Anwar, [2] Harminto Mulyo, [3]Teguh Tamrin
[1,2,3]Universitas Islam Nahdlatul Ulama Jepara

[1]201240001030@unisnu.ac.id, [2]minto@unisnu.ac.id, [3]teguh@unisnu.ac.id

ABSTRAK

Penelitian ini memanfaatkan Twitter untuk menganalisis opini publik tentang Pantai Kartini Jepara, dengan fokus pada optimisasi algoritma Naive Bayes dalam analisis sentimen. Penelitian ini mengidentifikasi bahwa akurasi Naive Bayes terbatas dalam menangani data besar dan kompleks. Tujuan utamanya adalah meningkatkan akurasi dan efisiensi analisis sentimen melalui optimisasi parameter dan teknik ensemble. Metode penelitian melibatkan pengumpulan data Twitter dari 2010–2023, preprocessing data, pelatihan model Naive Bayes, SVM, dan ensemble, serta evaluasi performa menggunakan akurasi, presisi, recall, dan F1-score. Model ensemble yang menggabungkan Naive Bayes dan SVM mencapai akurasi tertinggi sebesar 88,81%, meningkat dari 83,91% pada Naive Bayes dasar dan 86,01% pada SVM, menunjukkan perbaikan signifikan dalam analisis sentimen. Kombinasi algoritma Naive Bayes dengan teknik optimasi dan ensemble meningkatkan akurasi analisis sentimen. Penelitian selanjutnya disarankan untuk mengeksplorasi penerapan model ini pada data yang lebih besar atau platform media sosial lain.

Kata Kunci: Naive Bayes, SVM, Ensemble, Analisis Sentimen, Twitter

PENDAHULUAN

Dalam 13 tahun terakhir, penggunaan platform media sosial, khususnya Twitter, telah mengalami lonjakan yang signifikan dalam popularitasnya. Twitter, sebagai salah satu jejaring sosial terbesar di dunia, tidak hanya menjadi saluran komunikasi bagi individu untuk berbagi pemikiran dan informasi, tetapi juga berfungsi sebagai sumber data yang sangat berharga bagi berbagai penelitian dan analisis(Kamal et al., 2022). Kemampuan untuk menganalisis dan memahami data ini dapat memberikan keuntungan kompetitif yang besar dalam berbagai bidang seperti penelitian sosial(Zhang, 2023). Oleh karena itu, terdapat kebutuhan untuk metode analisis yang efisien dan akurat yang dapat menangani volume dan kompleksitas data Twitter. Di sinilah peran algoritma klasifikasi seperti Naive Bayes menjadi sangat penting, karena menawarkan pendekatan yang sistematis dan terukur untuk mengidentifikasi dan mengkategorikan sentimen dalam teks yang besar dan bervariasi(Wicaksana et al., 2022). Namun, meskipun Naive Bayes telah terbukti efektif dalam banyak aplikasi, tantangan seperti asumsi independensi fitur yang mendasar dan variabilitas konteks memerlukan pengoptimalan lebih lanjut untuk meningkatkan akurasi dan kinerja analisis sentimen.(Adipradana, 2023)

Analisis sentimen di platform media sosial dapat memberikan wawasan mengenai opini publik dan persepsi terhadap berbagai tujuan atau lokasi. Melalui analisis ini, kita dapat mengeksplorasi bagaimana masyarakat menilai dan merespons berbagai tempat, baik itu destinasi wisata, kota, atau tempat penting(Prasetyo & Fitrani, 2023). Misalnya, ulasan dan komentar yang diposting oleh pengguna di media sosial tentang sebuah destinasi wisata dapat mengungkapkan kepuasan, kekhawatiran, atau harapan mereka, memberikan gambaran yang mendalam tentang bagaimana tempat tersebut diterima oleh pengunjung. Dengan memahami sentimen yang terkandung dalam data media sosial, pemangku kepentingan seperti pengelola destinasi, dan pemasar dapat membuat keputusan yang lebih terinformasi dan strategis. Selain itu, analisis ini dapat membantu dalam

Jurnal Minfo Polgan
Volume 13, Nomor 1, Agustus 2024 e-ISSN : 2797-3298
DOI : https://doi.org/10.33395/jmp.v13i1.14014 p-ISSN : 2089-9424

merancang promosi yang lebih efektif, mengidentifikasi area untuk perbaikan, dan memonitor perubahan dalam persepsi publik(Umar & Nur, 2022). Dengan kata lain, sentimen di media sosial bukan hanya refleksi dari opini individu, tetapi juga cerminan dari tren yang memengaruhi citra dan daya tarik suatu lokasi.

Penelitian sebelumnya telah mengeksplorasi berbagai algoritma analisis sentimen, termasuk Naive Bayes, untuk menganalisis sentimen di platform media sosial. Salah satu studi yang relevan adalah "Sentiment Analysis On Covid-19 Outbreak Awareness Using Naïve Bayes Algorithm," yang menyelidiki penggunaan algoritma Naive Bayes untuk mengevaluasi sentimen kesadaran tentang wabah Covid-19 dengan data Twitter selama lockdown kedua di Malaysia. Hasil penelitian ini menunjukkan bahwa Naive Bayes mampu mencapai akurasi lebih dari 90% dalam mengklasifikasikan sentimen, menggarisbawahi kemampuannya dalam mengolah data besar yang kompleks (Sabri et al., 2022). Penelitian lain, "An Integrated Approach for Sentiment Analysis and Topic Modeling of a Digital Bank in Indonesia using Naïve Bayes and Latent Dirichlet Allocation Algorithms on Social Media Data," menggunakan algoritma Naive Bayes untuk menganalisis sentimen terhadap bank digital di Indonesia dengan data dari Twitter dan Instagram, dan mencatatkan F1 score maksimum sebesar 0,863 (Setiawan et al., 2023). Studi "Twitter Sentiment Analysis in Indonesian Language using Naive Bayes Classification Method" berfokus pada analisis sentimen di Twitter dalam bahasa Indonesia, menghasilkan akurasi tertinggi sebesar 82% dengan metode Naive Bayes (Wicaksana et al., 2022) Selain itu, penelitian "A Parallel Approach for Sentiment Analysis on Social Networks Using Spark" memperkenalkan sistem skala besar untuk analisis sentimen di Twitter menggunakan model paralel Apache Spark dengan teknik pelatihan Naive Bayes, yang menunjukkan peningkatan kecepatan dan efisiensi pada kumpulan data besar (Iqbal & Latha, 2023). Terakhir, artikel "Improved Machine Learning Algorithms for Sentiment Analysis" mengkaji penggunaan berbagai algoritma, termasuk Naive Bayes, untuk analisis sentimen di Twitter, dengan fokus pada peningkatan akurasi melalui konsensus dari berbagai teknik algoritma (Nayak et al., 2023). Studi-studi ini menunjukkan bahwa Naive Bayes tetap menjadi metode yang penting dan efektif dalam analisis sentimen, terutama ketika dikombinasikan dengan teknik lainnya untuk mengatasi tantangan dalam pengolahan data besar dan kompleks.

Dalam kajian terbaru mengenai algoritma Naive Bayes untuk analisis sentimen di Twitter, beberapa penelitian telah menyoroti berbagai inovasi dan perbaikan dalam metode ini. Artikel "Naïve Bayes with Negation Handling for Sentiment Analysis of Twitter Data" mengusulkan teknik penanganan negasi yang ditingkatkan, yang memperbaiki akurasi dengan lebih efektif mendeteksi konten yang dinegasikan baik secara eksplisit maupun implisit (Kamal et al., 2022). Penelitian lain, "Pengukuran Kinerja Optimasi Algoritma Bat Pada Algoritma Naive Bayes, KNN dan Decision Tree untuk Sentimen Analisis di Lini Masa Twitter," membahas bagaimana optimisasi menggunakan Algoritma Bat dapat meningkatkan akurasi Naive Bayes dalam analisis sentimen, menunjukkan peningkatan yang signifikan setelah proses optimisasi (Adipradana, 2023). "Sentiment Analysis Before Presidential Election 2024 Using Naïve Bayes Classifier Based On Public Opinion In Twitter" memanfaatkan Naive Bayes untuk mengevaluasi sentimen publik terhadap calon presiden 2024 di Twitter, mencapai akurasi tertinggi 71% dalam analisisnya (Prasetyo & Fitrani, 2023). Selain itu, "Application of Naïve Bayes Algorithm Variations On Indonesian General Analysis Dataset for Sentiment Analysis" mengeksplorasi variasi dalam algoritma Naive Bayes pada dataset analisis umum Indonesia untuk menentukan metode yang paling akurat dalam analisis sentimen (Umar & Nur, 2022). Terakhir, artikel "Sentiment Analysis of Twitter Comments Using Naive Bayes Classifier" membandingkan kinerja Naive Bayes dengan regresi logistik dalam analisis sentimen komentar di Twitter, menunjukkan bahwa Naive Bayes unggul dalam analisis sentimen biner (Zhang, 2023). Studi-studi ini menggambarkan kemajuan signifikan dalam teknik analisis sentimen dengan Naive Bayes, menyoroti penerapan inovatif dan optimasi yang meningkatkan efektivitas metode ini dalam menangani data sosial media yang kompleks.

Optimisasi algoritma Naive Bayes untuk analisis sentimen pada data Twitter dapat memberikan pemahaman yang lebih baik tentang opini publik dan persepsi terhadap Pantai Kartini Jepara sebagai destinasi wisata. Dengan memanfaatkan algoritma Naive Bayes yang telah dioptimalkan,

Jurnal Minfo Polgan
Volume 13, Nomor 1, Agustus 2024 e-ISSN : 2797-3298
DOI : https://doi.org/10.33395/jmp.v13i1.14014 p-ISSN : 2089-9424

analisis sentimen pada tweet yang menyebutkan Pantai Kartini dapat mengungkapkan bagaimana pengunjung dan masyarakat merespons tempat tersebut. Proses optimisasi ini memungkinkan deteksi yang lebih akurat terhadap sentimen positif, negatif, atau netral dalam ulasan dan komentar, serta menangkap nuansa atau konteks yang mungkin tidak terlihat pada analisis yang tidak dioptimalkan. Hasil dari analisis ini dapat memberikan wawasan berharga bagi pengelola destinasi untuk memahami aspek yang dihargai atau dikritik oleh pengunjung, serta area yang memerlukan perbaikan. Dengan informasi ini, pengelola dapat merancang strategi promosi yang lebih efektif, meningkatkan kualitas layanan, dan meningkatkan daya tarik Pantai Kartini sebagai tujuan wisata yang diinginkan. Selain itu, pemahaman yang mendalam tentang sentimen publik dapat membantu dalam membentuk citra positif dan menangani masalah potensial yang mungkin memengaruhi reputasi destinasi.

Optimisasi algoritma Naive Bayes untuk analisis sentimen pada data Twitter terkait destinasi Pantai Kartini Jepara dapat dilakukan melalui beberapa pendekatan strategis. Pertama, pengolahan teks yang lebih canggih dengan penanganan negasi dan penghapusan noise dari data tweet dapat meningkatkan akurasi model dalam menangkap sentimen yang tepat. Mengintegrasikan teknik pre-processing seperti stemming, dan penghilangan kata-kata yang tidak relevan dapat membantu Naive Bayes dalam memahami konteks sentimen dengan lebih baik. Selain itu, optimisasi dapat melibatkan pemilihan fitur yang lebih efektif, dengan menggunakan teknik pemilihan fitur atau ekstraksi fitur berbasis TF-IDF untuk menangkap informasi yang paling relevan dari teks. Penerapan teknik ensemble atau kombinasi dengan algoritma lain, seperti Support Vector Machine (SVM), juga dapat meningkatkan kemampuan Naive Bayes dalam mengatasi keterbatasan asumsi independensi fitur. Melakukan fine-tuning pada parameter algoritma dan menggunakan teknik cross-validation untuk memastikan model tidak hanya sesuai untuk data pelatihan tetapi juga untuk data baru, merupakan langkah penting dalam proses optimisasi. Dengan pendekatan ini, Naive Bayes dapat memberikan analisis sentimen yang lebih akurat dan mendalam tentang persepsi pengunjung terhadap Pantai Kartini Jepara.

Tujuan dari penelitian ini adalah untuk mengoptimalkan algoritma Naive Bayes dalam analisis sentimen pada data Twitter terkait destinasi Pantai Kartini Jepara. Dengan memfokuskan upaya pada peningkatan akurasi dan efisiensi algoritma Naive Bayes, studi ini bertujuan untuk memperoleh pemahaman yang lebih mendalam mengenai bagaimana masyarakat dan pengunjung merespons Pantai Kartini sebagai tujuan wisata. Penelitian ini akan mencakup pengembangan teknik pre-processing yang lebih efektif, seperti penghapusan noise dan penanganan negasi, serta penerapan metode pemilihan fitur yang cermat untuk meningkatkan kualitas data masukan. Selain itu, optimisasi algoritma akan melibatkan fine-tuning parameter dan penerapan teknik validasi silang untuk memastikan hasil yang konsisten dan dapat diandalkan. Dengan optimisasi ini, diharapkan algoritma Naive Bayes dapat memberikan analisis sentimen yang lebih akurat dan komprehensif, memberikan wawasan yang berharga bagi pengelola destinasi dalam merancang strategi promosi yang lebih baik, meningkatkan layanan, dan memperkuat daya tarik Pantai Kartini Jepara sebagai destinasi wisata unggulan.

TINJAUAN PUSTAKA

Algoritma Naïve Bayes

Menurut (Khaira et al., 2023) Naïve Bayes merupakan algoritma yang digunakan dalam analisis sentimen dan klasifikasi teks. Algoritma ini bekerja berdasarkan prinsip probabilitas, di mana ia mengasumsikan bahwa setiap fitur (kata) dalam data bersifat independen satu sama lain. Naïve Bayes dapat memberikan hasil yang baik dalam banyak kasus, terutama ketika digunakan pada dataset yang besar. Algoritma ini sangat efisien dalam hal waktu dan ruang, sehingga mampu mengelola data dalam jumlah besar dengan hasil akurasi yang tinggi. Naïve Bayes sering digunakan untuk melakukan prediksi sentimen pada data yang belum memiliki label sentimen, seperti dalam analisis opini publik atau ulasan produk. Dalam implementasinya, Naïve Bayes biasanya memanfaatkan teknik pembobotan seperti TF-IDF (Term Frequency-Inverse Document Frequency) untuk menghitung bobot dari setiap kata dalam dataset. Proses klasifikasi dilakukan dengan membagi dataset menjadi data latih dan data uji, di mana data latih digunakan untuk

Jurnal Minfo Polgan
Volume 13, Nomor 1, Agustus 2024 e-ISSN : 2797-3298
DOI : https://doi.org/10.33395/jmp.v13i1.14014 p-ISSN : 2089-9424

membangun model klasifikasi.

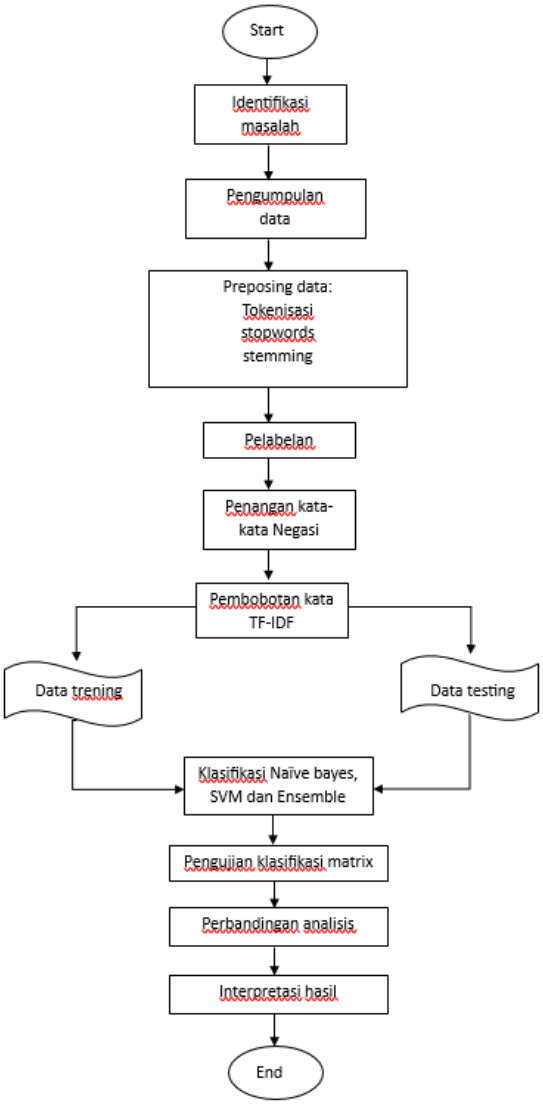
Support Vector Machine

Pada penelitian sebelumnya yang lakukan oleh (Hermawan et al., 2023) bahwa Support Vector Machine (SVM) adalah teknik statistik dan pembelajaran mesin yang digunakan untuk tujuan utama prediksi. SVM dapat diterapkan pada masalah regresi dan klasifikasi dengan kinerja yang sangat baik. Dalam konteks analisis sentimen, SVM digunakan untuk mengklasifikasikan data menjadi kalimat positif dan negatif. Proses klasifikasi dimulai dengan membuat vektor fitur dari data latih, yang kemudian digunakan untuk mempelajari pola dengan SVM. Dalam penelitian ini, Nu Support Vector Regression (NuSVR) digunakan, yang memiliki properti serupa dengan Support Vector Regression biasa, tetapi dengan parameter nu yang mengontrol jumlah support vector

Ensemble

Menurut (Prasetya Widiharso et al., 2022) Ensemble model adalah teknik dalam machine learning yang menggabungkan beberapa algoritma untuk meningkatkan akurasi dibandingkan dengan penggunaan model tunggal. Metode ini berfokus pada pengolahan data dengan cara mengkombinasikan hasil dari beberapa model untuk menghasilkan prediksi yang lebih baik. Beberapa metode yang termasuk dalam ensemble model adalah bagging, boosting, dan stacking.

METODE PENELITIAN



Gambar 1. Tahapan Analisis Sentimen

Jurnal Minfo Polgan
Volume 13, Nomor 1, Agustus 2024 e-ISSN : 2797-3298
DOI : https://doi.org/10.33395/jmp.v13i1.14014 p-ISSN : 2089-9424

1. Identifikasi Masalah
   Tahap Pertama dalam penelitian ini adalah melakukan identifikasi masalah. Identifikasi masalah yang didapatkan adalah Bagaimana cara mengoptimalkan kinerja algoritma Naïve Bayes dalam menganalisis sentimen Twitter terkait destinasi wisata Pantai Kartini di Jepara untuk memperoleh pemahaman yang lebih mendalam tentang opini dan sentimen public.
2. Pengumpulan data
   Tahap selanjutnya adalah melakukan pengumpulan data opini twitter dengan twitter API
3. Preprosesing data
   Selanjutnya data dilakukan tahapan preprocessing data yang terdiri dari proses Tokenisasi, Stopwords dan Stemming hingga data bersih.
4. Pelabelan
   Pada tahapan ini setiap teks dengan label sentimen menggunakan Sentiment Intensity Analyzer dari NLTK.
5. Penanganan negasi
   Peneliti juga mengimplementasi fitur khusus untuk perubahan sentimen, untuk menangani kata ambigu dan kata negasi
6. Pembobotan Data
   Setelah mendapatkan data bersih, selanjutnya dilakukan proses pembobotan data atau kata menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency).Proses pembobotan data digunakan untuk memberi nilai bobot relevansi termdari sebuah data terhadap keseluruhan dokumen yang ada(Helmayanti et al., 2023). Berikut rumus yang digunakan dalam menenttukan bobot Term Frequency-Inverse Document Frequency (TF-IDF):

$$idf_t = \log \frac{td}{df}$$

$$W_{t,d} = tf_{t,d} \times idf_t$$

   Keterangan:
   $W_{t,d}$: Pembobotan TF-IDF
   $tf_{t,d}$: Bobot kata *t* dalam setiap *d*
   td : jumlah keseluruan data yang ada
   df : jumlah kemunculan kata pada semua data
   idf_t : nilai IDF setiap kata yang akan dicari
7. Implementasi Klasifikasi Data
   Setelah mendapatkan bobot dari setiap term, selanjutnya terlebih dahulu data dibagi menjadi 2 bagian yaitu 80% data latih dan 20% data uji. Data training digunakan untuk membuat proses pembuatan model klasifikasi dari Naïve Bayes, SVM dan juga Ensemble. Adapun tahapan yang dilakukan pada pengklasifikasian dengan algoritma Naïve Bayes(Sun et al., 2022) sebagai berikut :
   1. Dataset hasil pembobotan TF-IDF
   2. Dataset dibagi menjadi 80% data training dan 20% data latih dari keseluruhan data.
   3. Baca data training.
   4. Pelatihan Model:
      a. Menggunakan Probabilitas Prior untuk mengukur frekuensi setiap kelas dalam dataset.

$$P(C_k) = \frac{Jumlah\ tweet\ dengan\ kelas\ C_k}{Total\ Jumlah\ tweet}$$

         Keterangan:
         $P(C_k)$: Probabilitas prior dari kelas C_k

      b. Likelihood untuk Menghitung probabilitas kata muncul dalam kelas tertentu,

Jurnal Minfo Polgan
Volume 13, Nomor 1, Agustus 2024 e-ISSN : 2797-3298
DOI : https://doi.org/10.33395/jmp.v13i1.14014 p-ISSN : 2089-9424

dengan menggunakan smoothing untuk mengatasi masalah kata yang tidak ada dalam data pelatihan.

$$P(x_i|C_k) = \frac{Jumlah\ kemunculan\ x_i\ dalam\ kelas\ C_k + \alpha}{Total\ Jumlah\ kata\ dalam\ kelas\ C_k + \alpha V}$$

Keterangan :
$P(x_i|C_k)$: Probabilitas fitur x_i muncul dalam kelas C_k
α : Parameter smoothing.
V : Jumlah kata unik dalam korpus.

5. Klasifikasi:
Pada tahapan klasifikasi Probabilitas Posterior Menghitung probabilitas akhir bahwa tweet termasuk dalam setiap kelas dengan mengalikan probabilitas prior dan likelihood.

$$P(C_k|X) = P(C_k) \times \prod_{i=1}^{n} P(x_i|C_k)$$

Keterangan :
$P(C_k|X)$: Probabilitas posterior bahwa tweet X termasuk dalam kelas C_k
Π : Operasi perkalian untuk menghitung produk dari semua probabilitas fitur.

Sedangkan pada algoritma Support Vector Machine (SVM)(Arumugam et al., 2023) tahapan yang dilakukan pada pengklasifikasian sebagai berikut :

1. Dataset hasil pembobotan TF-IDF
2. Input data training yang telah di sediakan
3. Data Training
4. Pemilihan model:
   a. Menggunakan kernel RBF untuk mengukur kesamaan antara dua titik data dan membantu dalam menemukan batas keputusan non-linear.

$$K(x_i, x_j) = exp(-\Upsilon||x_i - x_j||^2)$$

Keterangan :
$K(x_i, x_j)$: Nilai kernel untuk pasangan data x_i dan x_j.
Υ : Parameter yang mengontrol jarak pengaruh dari sebuah sampel terhadap hyperplane.
$||x_i - x_j||$: Jarak Euclidean antara dua titik data.

   b. Menggunakan Loss Hinge untuk mengukur kesalahan klasifikasi dan margin, dengan parameter C mengontrol kekuatan regularisasi.

$$\min_{w,b} \frac{1}{2}||\mathrm{W}||^2 + C\sum_{i=1}^{n} \max(0, 1 - y_i(\mathrm{w}.\mathrm{x}_i + b))$$

Keterangan :
W : Vektor bobot yang menentukan orientasi hyperplane.
b : Bias yang menentukan jarak hyperplane dari titik asal.
C : Parameter regularisasi yang mengontrol trade-off antara margin maksimal dan kesalahan klasifikasi.
y_i : Label kelas untuk data ke-i, bernilai +1 atau -1.

Jurnal Minfo Polgan
Volume 13, Nomor 1, Agustus 2024 e-ISSN : 2797-3298
DOI : https://doi.org/10.33395/jmp.v13i1.14014 p-ISSN : 2089-9424

x_i : Vektor fitur untuk data ke-i.

5. Klasifikasi :
   Klasifikasi pada tahap ini berfungsi untuk menentukan kelas berdasarkan seberapa jauh sebuah titik dari hyperplane.

$$f(x) = \text{sign}(w \cdot x + b)$$

   Keterangan:
   $f(x)$: Output prediksi, dengan sign menentukan apakah kelasnya positif (+1) atau negatif (-1).

Sedangkan pada Teknik Ensemble(Kurniati et al., 2023) tahapan yang dilakukan pada pengklasifikasian sebagai berikut:

1. Dataset hasil pembobotan TF-IDF
2. Input data training yang telah di sediakan
3. Membaca dan mempersiapkan data training yang telah dipisahkan untuk pelatihan model Naive Bayes dan SVM.
4. Latih model Naive Bayes dan SVM secara terpisah menggunakan data training.
5. Voting Ensemble:
   a. Majority Voting:
      Metode majority voting memilih kelas yang paling sering diprediksi oleh model-model sebagai hasil akhir.

$$C_{final} = mode\{C_{NB}, C_{SVM}\}$$

      Keterangan:
      C_{final} : Kelas final yang dipilih berdasarkan prediksi mayoritas dari model Naive Bayes

   b. Weighted Voting:
      Weighted voting berfungsi untuk menggabungkan prediksi dari model-model dengan mempertimbangkan bobot yang diberikan, untuk memberikan kontribusi lebih pada model yang lebih baik.

8. Pengujian Klasifikasi
   Setelah proses klasifikasi data selesai selanjutnya dilakukan pengujian model klasifikasi dengan menggunakan Confusion matrix yang bertujuan untuk membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya. Tahap terakhir ini adalah menghitung nilai akurasi, Recall, Precision dan f1-Score dari model klasifikasi sehingga dapat diketahui perbandingan tingkat akurasi pada model klasifikasi. Dengan persamaan perhitungan akurasi, presicion, recall dan f1-score sebagai berikut.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$recall = \frac{TP}{TP+FN}$$

$$f1-Score = 2 \times \frac{(precision \times Recall)}{(precision \times Recall)}$$

Jurnal Minfo Polgan
Volume 13, Nomor 1, Agustus 2024 e-ISSN : 2797-3298
DOI : https://doi.org/10.33395/jmp.v13i1.14014 p-ISSN : 2089-9424

HASIL DAN PEMBAHASAN

Pengumpulan Data

Penelitian ini mengumpukan data tweet berbahasa Indonesia dengan pencarian kata kunci “pantai kartini” dari tahun 2010 – 2023 yang dilakukan dengan teknik pengumpulan data melalui Twitter API menggunakan bahasa pemograman python.

Tabel 1. Data ulasan pantai kartini

	created_at	id_str	full_text
0	Thu Oct 26 04:45:05 +0000 2023	1.717402e+18	Tempat nyegerin badan buat melepas lelah...
1	Wed Mar 01 15:51:26 +0000 2023	1.630959e+18	WIB JIR Waktu Istimewa Beli rumah subsidi Jepa...
2	Thu Jun 23 00:02:19 +0000 2022	1.539761e+18	Mlipir sebentar sebelum kehectican dimulai. Te...
3	Thu Dec 30 19:33:07 +0000 2021	1.476637e+18	Tahun baru di rumah #jepara #bumikartini #kar...
4	Mon Nov 15 14:09:31 +0000 2021	1.460249e+18	Selain akeh menantu idaman jepara juga akeh se...

Preprosesing data

Preprosesing data adalah langkah penting dalam analisis sentimen karena data teks mentah sering kali mengandung elemen-elemen yang tidak relevan atau dapat menyebabkan kebingungan bagi model pembelajaran mesin. Berikut adalah langkah-langkah pra-pemrosesan yang dilakukan:

1. Tokenisasi: Teks dipecah menjadi token (kata-kata) agar bisa dianalisis lebih lanjut. Misalnya, kalimat "Pantai Kartini sangat indah!" akan diubah menjadi ["pantai", "kartini", "sangat", "indah"].
2. Penghapusan Stopwords: Stopwords (kata-kata umum yang tidak banyak berkontribusi pada makna teks, seperti "dan", "atau", "tetapi") dihapus untuk fokus pada kata-kata yang lebih bermakna.
3. Stemming: Kata-kata dikembalikan ke bentuk dasarnya untuk menyatukan variasi kata yang memiliki makna serupa. Misalnya, "bermain", "bermainan", dan "main" diubah menjadi "main".

Tabel 2. Hasil Preposesing data

	full_text
0	nyegerin badan melepa lelah
1	wib jir istimewa beli rumah subsidi jepara ind...
2	mlipir sebentar kehectican . pa seusia sd . ti...
3	rumah jepara bumikartini karimunjawa p…
4	akeh menantu idaman jepara akeh sepot wisata a...

Distribusi Sentimen

Setelah pra-pemrosesan, data diproses lebih lanjut untuk melabeli sentimen. Hasil distribusi sentimen adalah sebagai berikut:

Jurnal Minfo Polgan
Volume 13, Nomor 1, Agustus 2024
DOI : https://doi.org/10.33395/jmp.v13i1.14014
e-ISSN : 2797-3298
p-ISSN : 2089-9424

Tabel 3. Hasil distribusi sentimen

full_text	count
Neutral	615
Positive	68
negative	32

Evaluasi Model

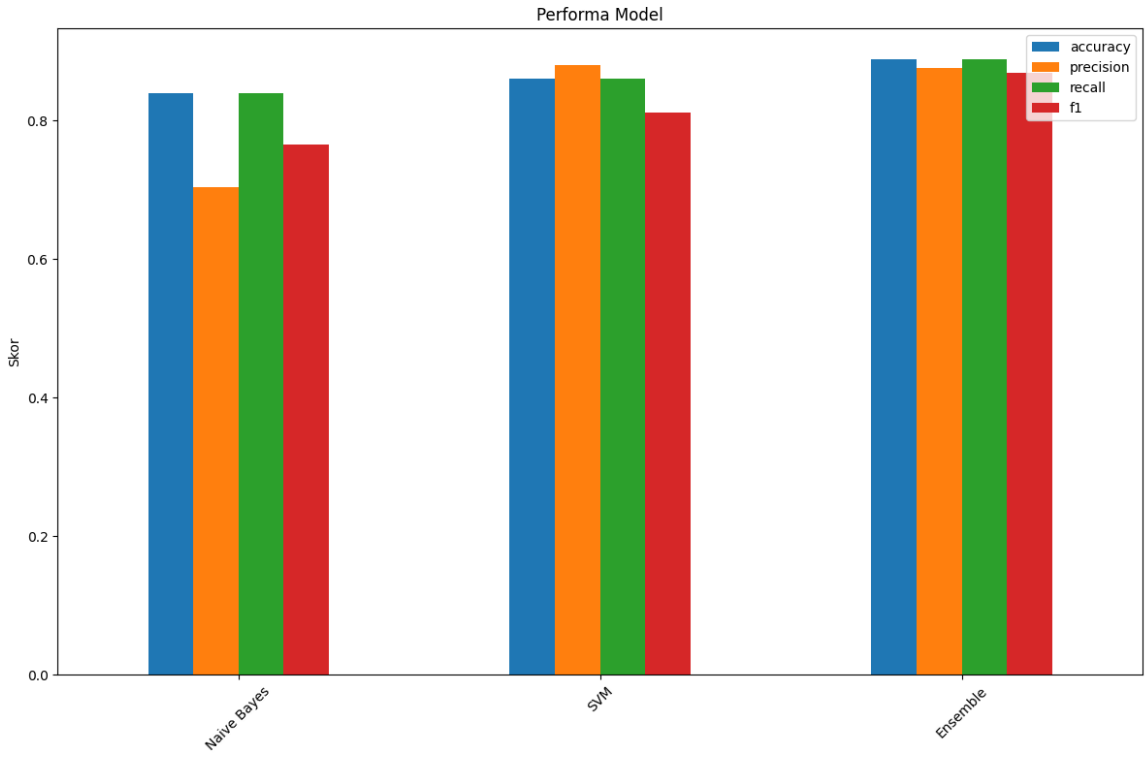
Setelah data diolah dan fitur diekstraksi menggunakan TF-IDF, model Naive Bayes, SVM, dan Ensemble dilatih dan dievaluasi. Berikut adalah metrik performa dari setiap model:

Tabel 4. Hasil distribusi sentimen

Model	accuracy	precision	recall	f1
Naïve bayes	0.8391608391608392	0.7041909139811238	0.8391608391608392	0.765774149804568
SVM	0.8601398601398601	0.8801198801198801	0.8601398601398601	0.8117232189630379
Ensemble	0.8881118881118881	0.8760330578512396	0.8881118881118881	0.8688747150285611

Cross-Validation

Nilai cross-validation memberikan wawasan tambahan tentang keandalan model:



Gambar 2. Peforma model

Cross-validation menunjukkan bahwa SVM memiliki performa yang lebih konsisten dibandingkan dengan Naive Bayes

Pembahasan

Pembahasan dalam penelitian ini menunjukkan bahwa langkah-langkah pra-pemrosesan

Jurnal Minfo Polgan
Volume 13, Nomor 1, Agustus 2024 e-ISSN : 2797-3298
DOI : https://doi.org/10.33395/jmp.v13i1.14014 p-ISSN : 2089-9424

teks yang diterapkan, seperti tokenisasi, penghapusan stopwords, dan stemming, berhasil membersihkan dan menyederhanakan data untuk analisis yang lebih efektif. Deteksi kata negasi secara khusus memberikan nilai tambah dengan memungkinkan model menangkap konteks yang kompleks, seperti frasa dengan makna berlawanan karena adanya kata negasi. Analisis distribusi sentimen mengungkapkan bahwa mayoritas tweet tentang Pantai Kartini bersifat netral, diikuti oleh tweet positif, dan kemudian negatif, yang mengindikasikan bahwa secara umum, opini publik cenderung netral hingga positif terhadap destinasi wisata ini.

Dalam evaluasi model, Support Vector Machine (SVM) terbukti lebih unggul dibandingkan Naive Bayes, terutama dalam menangani data dengan dimensi tinggi dan menghasilkan hyperplane optimal untuk memisahkan data. Namun, model ensemble yang menggabungkan Naive Bayes dan SVM dengan metode soft voting menghasilkan performa terbaik secara keseluruhan, dengan peningkatan signifikan dalam akurasi, precision, recall, dan F1-Score. Hal ini menunjukkan bahwa kombinasi model dapat memberikan prediksi yang lebih robust dengan memanfaatkan kelebihan masing-masing model.

KESIMPULAN

Penelitian ini berhasil mengoptimalkan algoritma Naive Bayes untuk analisis sentimen pada data Twitter terkait destinasi wisata Pantai Kartini di Jepara. Melalui teknik optimasi seperti penanganan negasi, penghapusan noise, dan penerapan metode ensemble dengan Support Vector Machine (SVM), akurasi klasifikasi sentimen meningkat dari 83,91% pada model Naive Bayes dasar menjadi 86,01% dengan SVM, dan lebih lanjut meningkat menjadi 88,81% pada model ensemble. Peningkatan ini menunjukkan bahwa kombinasi algoritma Naive Bayes dengan teknik optimasi lain dapat meningkatkan performa dalam analisis sentimen, memberikan hasil yang lebih akurat dan relevan untuk digunakan oleh pengelola destinasi dan pemangku kepentingan dalam merancang strategi promosi dan meningkatkan kualitas layanan di Pantai Kartini.

REFERENSI

Adipradana, C. (2023). Pengukuran Kinerja Optimasi Algoritma Bat Pada Algoritma Naive Bayes, KNN Dan Decision Tree Untuk Sentimen Analisis Di Lini Masa Twitter. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN)*. https://api.semanticscholar.org/CorpusID:258943394

Arumugam, M., S R, S., & Jayanthi, C. (2023). Machine Learning for Sentiment Analysis Utilizing Social Media. *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, 523–530. https://doi.org/10.1109/ICECAA58104.2023.10212135

Helmayanti, S. A., Hamami, F., & Fa'rifah, R. Y. (2023). PENERAPAN ALGORITMA TF-IDF DAN NAÏVE BAYES UNTUK ANALISIS SENTIMEN BERBASIS ASPEK ULASAN APLIKASI FLIP PADA GOOGLE PLAY STORE. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*. https://api.semanticscholar.org/CorpusID:263317986

Hermawan, A., Jowensen, I., Junaedi, J., & Edy. (2023). Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine. *JST (Jurnal Sains Dan Teknologi)*, *12*(1), 129–137. https://doi.org/10.23887/jstundiksha.v12i1.52358

Iqbal, M. M., & Latha, K. (2023). A Parallel Approach for Sentiment Analysis on Social Networks Using Spark. *Intelligent Automation & Soft Computing*. https://api.semanticscholar.org/CorpusID:250719379

Kamal, L. H., McKee, G., & Othman, N. A.-H. (2022). Naïve Bayes with Negation Handling for Sentiment Analysis of Twitter Data. *2022 9th International Conference on Soft Computing &*

Machine Intelligence (ISCMI), 207–212. https://api.semanticscholar.org/CorpusID:257667651

Khaira, U., Aryani, R., & Hardian, R. W. (2023). Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Kebijakan Kemdikbudristek Mengenai Kuota Internet Selama Covid-19. *Jurnal PROCESSOR*, *18*(2). https://doi.org/10.33998/processor.2023.18.2.897

Kurniati, F. T., Manongga, D., Sediyono, E., Prasetyo, S. Y. J., & Huizen, R. R. (2023). Object Classification Model Using Ensemble Learning with Gray-Level Co-Occurrence Matrix and Histogram Extraction. *ArXiv*, *abs/2309.13512*. https://api.semanticscholar.org/CorpusID:262466391

Nayak, S., Sonia, & Sharma, Y. K. (2023). Improved Machine leaning algorithms for sentiment analysis. *2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, 475–484. https://api.semanticscholar.org/CorpusID:260385600

Prasetya Widiharso, Siti Sendari, Anik Nur Handayani, & Nastiti Susetyo Fanani Putri. (2022). Performa Metode Klasifikasi Tunggal dan Ensemble Model dalam Identifikasi Baku Mutu Air. *Infotekmesin*, *13*(2), 206–211. https://doi.org/10.35970/infotekmesin.v13i2.1529

Prasetyo, H., & Fitrani, A. S. (2023). Sentiment Analysis Before Presidential Election 2024 Using Naïve Bayes Classifier Based On Public Opinion In Twitter. *Procedia of Engineering and Life Science*. https://api.semanticscholar.org/CorpusID:259906881

Sabri, N. M., Norman, J. N. A., Isa, N., & Bahrin, U. F. M. (2022). Sentiment Analysis On Covid-19 Outbreak Awareness Using Naïve Bayes Algorithm. *2022 International Visualization, Informatics and Technology Conference (IVIT)*, 278–283. https://api.semanticscholar.org/CorpusID:256670691

Setiawan, J., Milenia, A., & Faza, A. (2023). An Integrated Approach for Sentiment Analysis and Topic Modeling of a Digital Bank in Indonesia using Naïve Bayes and Latent Dirichlet Allocation Algorithms on Social Media Data. *2023 4th International Conference on Big Data Analytics and Practices (IBDAP)*, 1–7. https://api.semanticscholar.org/CorpusID:263707677

Sun, X., Du, L., & Wang, Y. (2022). Text Classification in Architecture Field Based on Naive Bayes Algorithm. *2022 International Conference on 3D Immersion, Interaction and Multi-Sensory Experiences (ICDIIME)*, 69–72. https://doi.org/10.1109/ICDIIME56946.2022.00023

Umar, N., & Nur, M. A. (2022). Application of Naïve Bayes Algorithm Variations On Indonesian General Analysis Dataset for Sentiment Analysis. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*. https://api.semanticscholar.org/CorpusID:251775606

Wicaksana, C. A., Fatkhurrokhman, M., Pratama, H. P., Tryawan, R., Alimuddin, & Febriani, R. (2022). Twitter Sentiment Analysis in Indonesian Language using Naive Bayes Classification Method. *2022 International Conference on Informatics Electrical and Electronics (ICIEE)*, 1–6. https://api.semanticscholar.org/CorpusID:255997505

Zhang, Z. (2023). Sentiment Analysis of Twitter Comments Using Naive Bayes Classifier. *Communications in Humanities Research*. https://api.semanticscholar.org/CorpusID:264796048