Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 7 Nomor 1, Maret 2023, pp. 1-12
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

Optimasi Sentimen Analisis Informatif dan Tidak Informatif dari Tweet di BMKG Menggunakan Algoritma Naive Bayes dan Metode Teknik Pengambilan Sampel Minoritas Sintetis

Muhammad Yusuf Hidayatulloh[1], Anto Sunanto[2], Armansyah[3], Muhammad Farrell Afelino Gevin[4], Dedi Dwi Saputra[5]
[1,2,3,4,5]Universitas Nusa Mandiri, Indonesia
e-mail: 11213207@nusamandiri.ac.id, 11213173@nusamandiri.ac.id, 11213242@nusamandiri.ac.id, 11213235@nusamandiri.ac.id, dedi.eis@nusamandiri.ac.id

Abstract

The emergence of computer-based and mobile-based social networks seems to have received high attention from the public. Evidenced by the increasing number of social networks that appear. Friendster, Facebook, Twitter, Linkd In and many others. Twitter is one of the social media used to find information, Twitter users generally report every activity. They are even more helped by the existence of increasingly sophisticated cellphones. The system created in this study to optimize the analysis of informative and uninformative sentiment using a rapid miner application with the Naïve Bayes, Naïve Bayes + Adaboost, SVM, and SVM PSO methods using data taken from twitter @infoBMKG. The research method used is the collection of tweet data from twitter taken by the Crawling method. The data taken is tweets in Indonesian with a total of 1,000 tweets from the @infoBMKG twitter account. The results of the nave Bayes algorithm test carried out in this study were to measure the performance of accuracy, precision, recall, AUC from the results of the training and submission of datasets that had gone through the data preprocessing process. From the results of the research that has been done, it is proven that the optimization of informative and uninformative sentiment analysis from tweets on BMKG's twitter gets good results using the Support Machine Vector method with higher Accuracy, Recall, and AUC values than other methods.

Keywords*: Data Mining, Sentiment Analyst, Naïve Bayes, Support Machine Vector, BMKG*

Abstrak

Kemunculan jejaring sosial berbasis komputer dan handphone nampaknya mendapat perhatian tinggi dari masyarakat. Terbukti dengan semakin banyaknya jejaring sosial yang muncul. Friendster, Facebook, Twitter, Linkd In dan masih banyak lainnya. Twitter merupakan salah satu sosial media yang digunakan dalam mencari informasi, Para pengguna twitter umumnya melaporkan setiap aktivitasnya. Mereka bahkan semakin terbantu dengan adanya handphone yang semakin canggih. Sistem yang dibuat pada penelitian ini untuk optimasi analisis sentiment informatif dan tidak informatif menggunakan aplikasi rapid miner dengan metode Naïve Bayes, Naïve Bayes + Adaboost, SVM, dan SVM PSO dengan menggunakan data yang di ambil dari twitter @infoBMKG. Metode penelitian yang digunakan adalah pengumpulan data tweet dari twitter yang diambil dengan metode Crawling. Data yang diambil adalah tweet dalam bahasa Indonesia dengan jumlah 1.000 tweet dari akun twitter @infoBMKG. Hasil dari tes algoritma metode naïve bayes yang dilakukan dalam penelitian ini adalah untuk mengukur kinerja akurasi, presisi, recall, AUC dari hasil pelatihan dan pengajuan dataset yang telah melalui proses pra pengolahan data. Dari hasil penelitian yang telah dilakukan terbukti bahwa optimasi sentimen analisa informatif dan tidak informatif dari tweet pada twitter BMKG mendapatkan hasil yang baik menggunakan metode Support Machine Vector dengan nilai Akurasi, Recall, dan AUC yang lebih tinggi dibandingkan metode lainnya.

Kata kunci*: Data Mining, Sentimen Analis, Naïve Bayes, Support Machine Vector, BMKG*

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 7 Nomor 1, Maret 2023, pp. 1-12
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

1. PENDAHULUAN

Kemunculan jejaring sosial berbasis komputer dan *handphone* nampaknya mendapat perhatian tinggi dari masyarakat. Terbukti dengan semakin banyaknya jejaring sosial yang muncul. Friendster, Facebook, Twitter, Linkd In dan masih banyak lainnya. Masayarakat yang mendapatkan kemudahan dari jejaring sosial tersebut semakin getol dalam menggunakannya [14]. Twitter merupakan salah satu sosial media yang digunakan dalam mencari informasi, Para pengguna twitter umumnya melaporkan setiap aktivitasnya. Mereka bahkan semakin terbantu dengan adanya handphone yang semakin canggih. Fitur Twitter sendiri sudah dapat dinikmati melalui telepon genggam yang artinya semakin tidak dapat lepas dari penggunanya. Inilah yang membuat mereka tidak berhenti melaporkan aktivitasya. Tidak sedikit bahkan yang lebih aktif di Twitter ketimbang di pergaulan nyata [14]. Data dari twitter ini memiliki karakteristik yang tidak terstruktur dan banyak memuat noise sehingga dibutuhkan teks mining yang memiliki peran penting dalam bidang data *mining* [5].

Teks mining merupakan proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah data tak terstruktur yang nantinya akan diperoleh pola-pola data, tren dan ekstraksi pengetahuan yang potensial dari data teks. Masukan untuk penambangan teks adalah data yang tidak (atau kurang) terstruktur, seperti dokumen, word, PDF, kutipan teks dll sedangkan masukan untuk penambangan data adalah data yang terstruktur. Salah satu tujuan penggunaan teks mining adalah analisis sentimen [5].

Penelitian dalam bidang analisis sentimen dewasa ini banyak dilakukan karena topik ini sangat menarik untuk dibahas salah satunya adalah penelitian Ratnawati, dkk (2017) tentang pengklasifikasian opini film yang data komentarnya diambil dari komentar yang dituliskan di twitter berdasarkan sentimen positif, sentimen negatif dan sentimen netral pada level kalimat. Analisis sentimen dilakukan dengan menggunakan metode deep learning dengan akurasi yang didapat 80,99% [5].

Sistem yang dibuat pada penelitian ini untuk optimasi analisis sentiment informatif dan tidak informatif menggunakan aplikasi *rapid miner* dengan metode *Naïve Bayes, Naïve Bayes + Adaboost, SVM, dan SVM PSO* dengan menggunakan data yang di ambil dari twitter @infoBMKG.

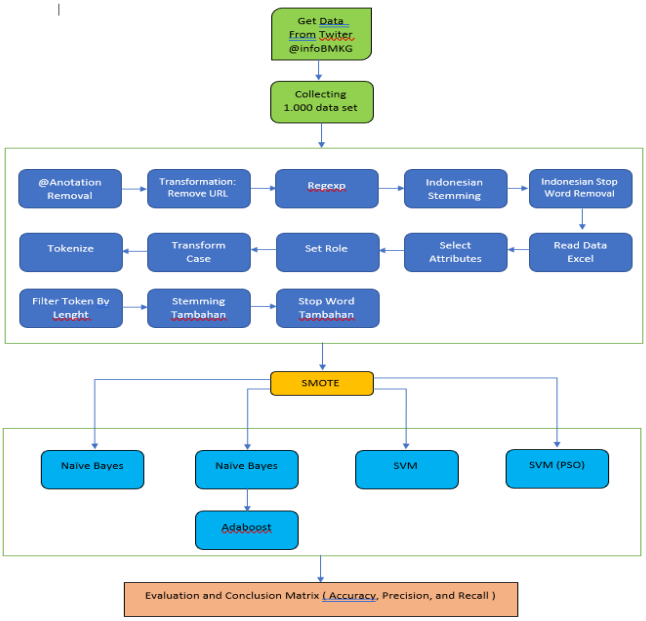
2. METODOLOGI PENELITIAN

Metode penelitian yang digunakan adalah pengumpulan data tweet dari twitter yang diambil dengan metode *Crawling*. Data yang diambil adalah tweet dalam bahasa Indonesia dengan jumlah 1.000 tweet dari akun twitter @infoBMKG. Data diambil secara acak baik dari akun pengguna biasa, akun stasiun klimatologi dari berbagai daerah maupun akun media online lain di Twitter. Metode yang digunakan melalui beberapa tahapan. Riset data yang kami gunakan sekitar 4 perbandingan, yaitu :

a) Naïve Bayes
b) Naïve Bayes + Adaboost
c) Support Vector Machine

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 7 Nomor 1, Maret 2023, pp. 1-12
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

d) *Support Vector Machine* (PSO)

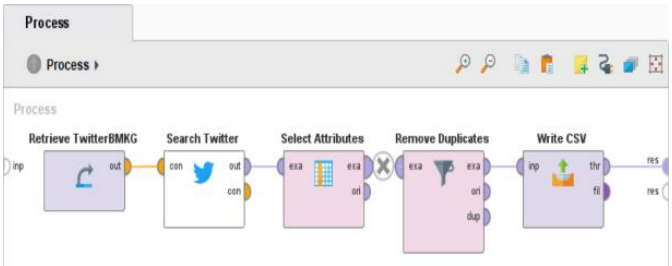


Gambar 1. Alur Penelitian

Dapat dilihat dari alur penelitian tersebut, di mulai dengan mengambil 1000 data dari twitter, dan selanjutnya kita bisa melanjutkan *pre processing* GataFramework dan menggunakan metode Naïve Bayes, Naïve Bayes + Adaboost, SVM dan SVM (PSO) yang akan dijelaskan dalam bagian ini.

a) Data Set

Data yang digunakan dalam penelitian ini diambil dari Twitter BMKG (@infoBMKG). Dari data yang diambil merupakan semua tweet yang berhubungan dengan BMKG, baik berisi informasi mengenai prakiraan cuaca, informasi bencana yang terjadi di setiap daerah, maupun komentar dari akun pengguna biasa yang tidak berisikan informasi. Aplikasi yang kami gunakan adalah Rapidminer versi 9.10.01. Alur data yang di ambil dapat dilihat pada gambar berikut :



Gambar 2. Koneksi Data Set

Dalam proses pengambilan data tersebut, yang pertama kali dilakukan adalah menyambungkan koneksi dengan twitter yang akan diambil datanya, akun tersebut adalah @infoBMKG, setelah itu di isi bagian kolom limit untuk

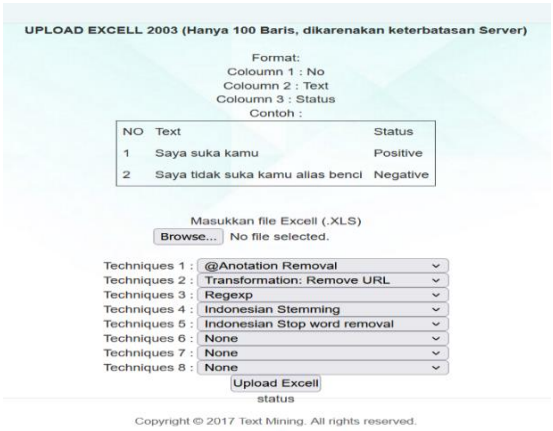
Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 7 Nomor 1, Maret 2023, pp. 1-12
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

menentukan berapa banyak batas data yang akan diambil dan di atur tanggal saat ini. Dilanjutkan proses memilih atribut dan atur atribut, lalu memilih proses pengapusan duplikat apabila ada data yang sama, maka akan dihilangkan duplikat tersebut. Pada tahap akhir memilih proses *Write CSV* untuk menentukan dimana file yang akan diambil dan disaring tersebut disimpan. Setelah semua data yang terkumpul makan akan dilakukan pelabelan kalimat jika mengandung informasi atau tidak mengandung informasi pada data tersebut. Jika proses pelabelan data selesai, maka bisa dilanjutkan ke bagian proses *GataFramework*.

Tabel 1. Data Set Twitter BMKG

No	Text	Status
188	RT @infoBMKG: Weekly Weather Outlook ep. 77	Not_Informatif
189	@infoBMKG Semoga BMKG lbh bisa memprediksi kapan perkiraan gempa nya	Not_Informatif
190	RT @infoBMKG: #Gempa Mag:5.2, 10-Apr-22 04:22:13 WIB, Lok:3.12 LS,139.98 BT (20 km Tenggara KAB-JAYAPURA-PAPUA), Kedlmn:53 Km, tdk berpoten	Informatif
191	Gempa Mag:5.2, 10-Apr-22 04:22:13 WIB, Lok:3.12 LS,139.98 BT (20 km Tenggara KAB-JAYAPURA-PAPUA), Kedlmn:53 Km, tdk berpotensi tsunami. via @infoBMKG https://t.co/EgiFVYjIfG	Informatif
192	#RekanCommuters Berikut kami informasikan Prakiraan Cuaca pada Minggu, 10 April 2022. Ingat selalu untuk mempersiapkan perlengkapan tambahan agar kamu terlindung dari hujan saat bepergian.	Informatif

b) Proses Gataframework



Gambar 3. Tampilan Gataframework

Pada proses ini data yang telah didapatkan dari proses pengambilan data pada *rapidminer* dan dilakukan *preprocessing* dengan menggunakan *Gataframwork*. *Gataframework* adalah *framework* untuk *preprocessing text mining* yang menyediakan *Indosenian stop removal*, *Indonesian stemming*, *annotation removal*, *transformation* : *remove url*, dan *regexp* [10].

Langkah pertama yang dilakukan adalah membagi dataset menjadi 50 atau maksimal 100 data, karena server terbatas dan menggunakan file dengan format excel. Berikut ini adalah tabel BMKG yang telah di proses oleh *Gataframework*.

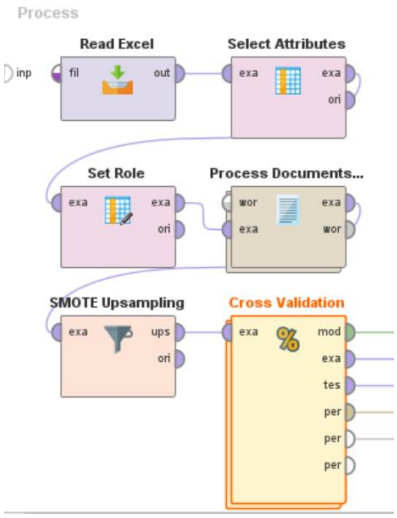
Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 7 Nomor 1, Maret 2023, pp. 1-12
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

Tabel 2. Dataset Twitter BMKG setelah Proses Gataframework

Text	Indonesian Stemming	Indonesian Stop word removal
Beberapa provinsi di Indonesia esok hari berada dalam kategori 'Waspada' terhadap bencana hidrometeorologi (Genangan, Banjir, Banjir Bandang, dll) dampak dari potensi hujan lebat.	beberapa provinsi di Indonesia esok hari ada dalam kategori waspada hadap bencana hidrometeorologi genang banjir banjir bandang dll dampak dari potensi hujan lebat	provinsi indonesia esok kategori waspada hadap bencana hidrometeorologi genang banjir banjir bandang dll dampak potensi hujan lebat
@infoBMKG Kok banyak gempa ya??	kok banyak gempa ya	gempa
Prakiraan Cuaca, 09 April 2022.	prakira cuaca April	prakira cuaca april
@infoBMKG Sekarang kok sering gempa ya	sekarang kok sering gempa ya	gempa
@infoBMKG update yang terbaru dong min,kita merasakan gempa di wamena?	update yang baru dong min kita rasa gempa di wamena	update min gempa wamena

c) Pra Pemrosesan Pada *Rapidminer*

Rapidminer adalah perangkat lunak untuk pengolahan data. Menggunakan prinsip dan algoritma penambangan data, *Rapidminer* mengekstrak pola dari kumpulan data besar dengan menggabungkan metode statistik, buatan intelijen dan *database* [10]. Berikut adalah gambar pada saat dilakukan pra pemrosesan pada rapid miner.

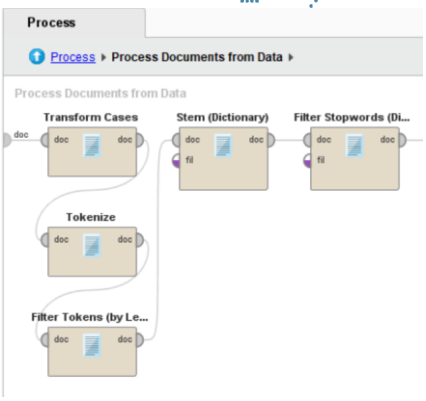


Gambar 4. Pra Pemrosesan Pada *Rapidminer*

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 7 Nomor 1, Maret 2023, pp. 1-12
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

d) Proses Pengolahan Data

Pada tahapan ini dilakukan *preprocessing* seperti *transform case*, *tokenize*, *filter token by (length)*, *stem* dan juga *stopword*.



Gambar 5. Proses Pengolahan Data

1) *Transform Case*

Pada tahapan *transform case* ini bertujuan untuk mengubah semua huruf pada tweet menjadi huruf kecil semua atau menjadi huruf kecil semua. Pada penelitian ini semua teks pada tweet diubah menjadi huruf kecil karena mayoritas teks berupa pendapat maupun pertanyaan yang sebagian besar menggunakan huruf kecil [12].

2) *Tokenize*

Pada tahapan *tokenize* ini bertujuan untuk menghilangkan tanda baca, simbol dan karakter yang bukan merupakan huruf biasa pada setiap dokumen *review.* Semua karakter yang tidak diperlukan akan dibuang, termasuk *white space* yang berlebihan dan semua tanda baca [12].

3) *Filter token by (length)*

Pada tahapan ini bertujuan untuk mengambil kata-kata penting dari token. Dalam proses ini kata-kata yang memiliki panjang tertentu akan dihapus atau dihilangkan dari teks [10].

4) *Stemming*

Pada tahapan ini dilakukan pengelompokan kata kedalam beberapa kelompok yang memiliki kata dasar yang sama dan melakukan perubahan untuk proses pembobotan dengan melakukan perhitungan terhadap ada atau tidak adanya sebuah kata didalam dokumen. Dengan tujuan semua kata yang telah dipilih pada tahap sebelumnya, akan diubah menjadi kata dasar dan menghilangkan imbuhan yang tidak diperlukan [10].

```
apa:ap
update:apdet
apa:ape
aplikasi:apknya
april:apr
```

Gambar 6. *Stemming*

5) *Stopword Removal*

Pada tahapan ini dilakukan penghapusan kata-kata yang tidak relevan, seperti kata konjungsi, kata kasar, dan kata yang dianggap tidak terkait ke dalam penelitian yang dilakukan. Apabila kata-kata tersebut dihapus maka bobot kata pada teks akan lebih bernilai, dan tahap ini akan menyempurnakan dataset sebelumnya [10].

e) SMOTE

SMOTE merupakan teknik untuk menyeimbangkan jumlah distribusi data sampel pada kelas minoritas dengan cara menyeleksi data sampel tersebut hingga jumlah data sampel menjadi seimbang dengan jumlah sampel pada kelas mayoritas [1]. SMOTE meminimalisasi ketidak seimbangan kelas pada dataset dengan membangkitkan data sintesis [8].

f) Validasi Silang (*Cross Validation*)

Validasi silang merupakan metode yang dilakukan untuk mendapatkan model terbaik. Metode yang digunakan di dalam penelitian ini adalah sebagai berikut :

1) *Naïve Bayes*

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan [9]. Dasar dari teorema Naïve Bayes yang digunakan adalah rumus Bayes sebagai berikut [6]:

$$P(C|X) = \frac{P(C).P(X|C)}{P(X)} \quad (1)$$

Deskripsi :

X : Atribut
C : Kelas
P(*C*|*X*) : Probabilitas hipotesis berdasarkan C dengan syarat X
P(X|C) : Probabilitas X berdasarkan kondisi pada hipotesis C
P(X) : Probabilitas X

2) *Adaboost*

Algoritma *AdaBoost* sendiri merupakan singkatan dari *Adaptive Boosting*, algoritma ini banyak diterapkan pada model prediksi dalam data *mining*. Inti dari algoritma *AdaBoost* adalah memberi bobot lebih pada pengamatan yang tidak tepat [10]. *Adaboost* dan variannya telah berhasil diterapkan ke beberapa bidang (*domain*) karena dasar teorinya yang kuat, prediksi yang akurat, dan kesederhanaan yang luar biasa. Langkah-langkah dalam algoritma *Adaboost* adalah sebagai berikut [2]:

a) Input: Suatu kumpulan sample penelitian dengan label $\{(x_i, y_i)\}, \ldots, (x_n, y_n)\}$ sebuah algoritma pembelajaran komponen, jumlah siklus T.
b) Inisialisasi : Bobot sampel latihan $W_i^1 = 1/N$, for all $i = 1, \ldots, N$
   Do for : t = 1,...,T
   Gunakan algoritma pembelajaran komponen untuk melatih komponen

klasifikasi h_t pada sampel bobot pelatihan.

Hitung kesalahan pelatihannya pada h_t: $\varepsilon_t = \sum_i^n = 1\, W_i^t, y_i \neq h_t(x_i)$

Tetapkan bobot untuk komponen pengklasifikasi $h_t = a_t = \frac{1}{2} ln\, (\frac{1-\varepsilon_t}{\varepsilon_t})$

Pembaruan berat sampel pelatihan $W_i^t + 1\ \frac{w_i^t\ exp\ \{a_t y_i h_t(x_i)\}}{C_t}, i$

Keluaran :

$$f(x) = sign\ (\sum_t^T{}_{=1} a_t h_t(x)) \quad (2)$$

3) *Support Vector Machine* (SVM)

Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian harmonis konsep-konsep unggulan dalam bidang pattern recognition. Sebagai salah satu metode pattern recognition, usia SVM terbilang masih relatif muda. SVM adalah metode machine learning yang bekerja atas prinsip Structural Risk Minimization (SRM) dengan tujuan menemukan hyperplane terbaik yang memisahkan dua buah class pada input space [7]. Secara prinsip SVM adalah linear classifier bidang dalam komputer sains, yang memetakan suatu data ke dalam konsep tertentu yang telah didefinisikan sebelumnya (*Pattern Recognition)* [3].

Fungsi deskriptif dari SVM dapat dilihat sebagai berikut [11]:

$$h(X) = z^x\ \emptyset(X) + c \quad (3)$$

Dimana X merupakan fitur *vector*, z merupaka *vector* dari bobot yang berbeda, ∅ merupakan fungsi pemetaan non-linier, dan c adalah *vector* bias. z dan c diperoleh secara otomatis dari dataset *training* [11].

4) SVM (*Particle Swam Optimization*)

Particle Swam Optimization adalah algoritma iteratif berbasis populasi. Populasi terdiri dari banyak partikel, di mana diinisialisasi dengan populasi solusi acak dan digunakan untuk memecahkan masalah optimasi [3]. Salah satu metode optimasi paling sederhana untuk memodifikasi beberapa parameter. Optimasi pada PSO dapat dilakukan dengan cara menyeleksi atribut (attribute selection) dan feature selection, serta meningkatkan bobot atribut (attribute weight) pada semua atribut atau variabel yang digunakan [4].

g) Evaluasi

Penelitian ini membandingkan tingkat akurasi antara *Naïve Bayes*, *Naïve Bayes* + *Adaboost*, SVM, SVM (PSO), pendekatan yang dievaluasi menggunakan akurasi, presisi, recall dan AUC. Penjelasan dari evaluasi akan dijelaskan sebagai berikut :

1) Akurasi
   Akurasi adalah evaluasi berdasarkan jumlah proporsi prediksi yang benar.
2) Presisi
   Presisi adalah tingkat akurasi antara informasi yang diminta oleh pengguna dan jawaban yang diberikan oleh sistem.

3) Recall

Recall adalah tingkat keberhasilan sistem dalam mengambil informasi.

4) AUC

AUC digunakan untuk mengukur diskriminatif kinerja dengan memperkirakan probabilitas keluaran yang diperoleh dari sampel yang dipilih secara acak dari informatif dan tidak informatif, semakin besar nilai *AUC*, semakin kuat klasifikasi yang dihasilkan. Sejak *AUC* adalah bagian dari satuan luas persegi, nilai yang dihasilkan akan selalu sama dengan itu menghasilkan, antara 0,0 dan 1,0 [13].

Berikut ini panduan untuk klasifikasikan keakuratan diagnosa menggunakan AUC [12]:

1. 0,90-1,00= klasifikasi sangat baik.
2. 0.80-0.90 = klasifikasi baik.
3. 0,70-0,80 = klasifikasi wajar.
4. 0.60-0.70 = klasifikasi buruk.
5. 0.50-0.60 = klasifikasi kegagalan.

3. HASIL DAN PEMBAHASAN

3.1. Pembahasan

1. *Naïve Bayes*

Hasil dari tes algoritma metode *naïve bayes* yang dilakukan dalam penelitian ini adalah untuk mengukur kinerja akurasi, presisi, *recall*, *AUC* dari hasil pelatihan dan pengajuan dataset yang telah melalui proses pra pengolahan data. Berikut adalah hasil pengujian algoritma *Naïve Bayes* :

Tabel 3. Akurasi *Naïve Bayes*

	true Informatif	**true Not_Informatif**	**class precision**
pred. Informatif	413	98	80.82%
pred. Not_Informatif	134	449	77.02%
class recall	75.50%	82.08%	

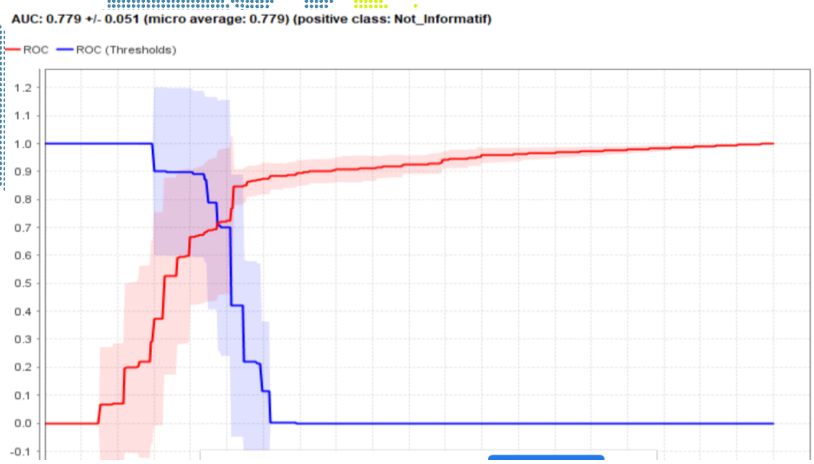
Tabel 4. Presisi *Naïve Bayes*

	true Informatif	**true Not_Informatif**	**class precision**
pred. Informatif	413	98	80.82%
pred. Not_Informatif	134	449	77.02%
class recall	75.50%	82.08%	

Tabel 5. *Recall Naïve Bayes*

	true Informatif	**true Not_Informatif**	**class precision**
pred. Informatif	413	98	80.82%
pred. Not_Informatif	134	449	77.02%
class recall	75.50%	82.08%	

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 7 Nomor 1, Maret 2023, pp. 1-12
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti



Gambar 7. *AUC Naïve Bayes*

2. *Support Machine Vector (SVM)*

Hasil dari tes algoritma metode *support machine vector (SVM)* yang dilakukan dalam penelitian ini adalah untuk mengukur kinerja akurasi, presisi, *recall*, *AUC* dari hasil pelatihan dan pengajuan dataset yang telah melalui proses pra pengolahan data. Berikut adalah hasil pengujian algoritma *Support Machine Vector*:

Tabel 6. Akurasi *Support Machine Vector*

	true Informatif	**true Not_Informatif**	**class precision**
pred. Informatif	378	58	86.70%
pred. Not_Informatif	169	489	74.32%
class recall	69.10%	89.40%	

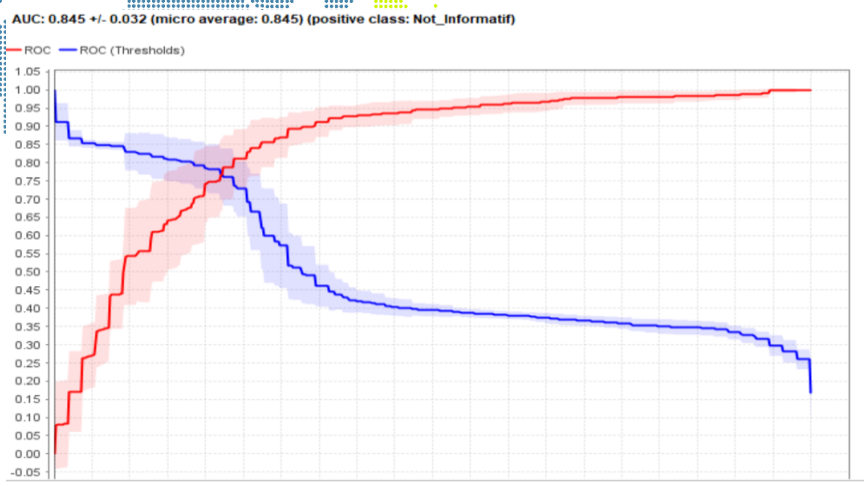
Tabel 7. Presisi *Support Machine Vector*

	true Informatif	**true Not_Informatif**	**class precision**
pred. Informatif	378	58	86.70%
pred. Not_Informatif	169	489	74.32%
class recall	69.10%	89.40%	

Tabel 8. *Recall Support Machine Vector*

	true Informatif	**true Not_Informatif**	**class precision**
pred. Informatif	378	58	86.70%
pred. Not_Informatif	169	489	74.32%
class recall	69.10%	89.40%	

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 7 Nomor 1, Maret 2023, pp. 1-12
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti



Gambar 8. *AUC Support Machine Vector*

3.2. Pengujian

Dari hasil pengujian yang telah didapatkan pada penelitian sentiment analis Informatif dan Tidak Informatif dari tweet di BMKG menggunakan metode *Naïve Bayes, Naïve Bayes + Adaboost, Support Machine Vectore (SVM), dan SVM Particle Swam Optimization (PSO),* maka didapatkan hasil sebagai berikut :

Tabel 9. Pengujian

Algoritma	Akurasi	Presisi	*Recall*	AUC
Naïve Bayes	78.79%	76.93%	82.06%	0.779
Naive Bayes + Adaboost	76.60%	79.00%	72.72%	0.827
Support Machine Vector	79.25%	74.35%	89.38%	0.845
SVM Particle Swam Optimization	78.52%	76.72%	82.44%	0.835

Dari data perbandingan dari pengujian eksperimen menggunakan *Naïve Bayes, Naïve Bayes + Adaboost, Support Machine Vector, SVM Particle Swam Optimization*. Diketahui bahwa nilai Akurasi, *Recall*, AUC dari algoritma *Support Machine Vector (SVM)* lebih tinggi dari yang lain algoritma, yaitu Akurasi 79.25%, *Recall* 89.38%, AUC 0.845. Hasil perbandingan dari algoritma menunjukkan bahwa algoritma dalam menentukan analisis sentimen informatif dan tidak informatif lebih baik dari algoritma lain.

4. SIMPULAN

Dalam penelitian ini telah dilakukan analisa terhadap tweet pada twitter BMKG untuk mencari optimasi sentiment analisa informatif dan tidak informatif dengan menggunakan rapidminer dan dilanjutkan pra pemrosesan pada *Gataframework* untuk lebih mengoptimalkan hasil yang di inginkan. Dari hasil penelitian yang telah dilakukan terbukti bahwa optimasi sentimen analisa informatif dan tidak informatif dari tweet pada twitter BMKG mendapatkan hasil yang baik menggunakan metode *Support Machine Vector* dengan nilai Akurasi, *Recall*, dan AUC yang lebih tinggi dibandingkan metode lainnya.

DAFTAR PUSTAKA

[1] Kasanah, A. N., Muladi, M., & Pujianto, U. (2019). Penerapan Teknik Smote Untuk Mengatasi Imbalance Class Dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma Knn. *Jurnal Resti (Rekayasa Sistem Dan Teknologi Informasi)*, *3*(2), 196–201. Https://Doi.Org/10.29207/Resti.V3i2.945

[2] Laila Qadrini, Andi Seppewali, & Asra Aina. (2021). Decision Treedan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial. *Jurnal Inovasi Penelitian*, *2*(7), 1959–1965.

[3] Nilawati, L., & Achyani, Y. E. (2019). Optimasi Metode Particle Swarm Optimization (Pso) Pada Prediksi Penilaian Apartemen. *Paradigma - Jurnal Komputer Dan Informatika*, *21*(2), 227–234. Https://Doi.Org/10.31294/P.V21i2.6159

[4] Que, V. K. S., Iriani, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, *9*(2), 162–170. Https://Doi.Org/10.22146/Jnteti.V9i2.102

[5] Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *Inovtek Polbeng - Seri Informatika*, *3*(1), 50. Https://Doi.Org/10.35314/Isi.V3i1.335

[6] Salmi, N., & Rustam, Z. (2019). Naïve Bayes Classifier Models For Predicting The Colon Cancer. *Iop Conference Series: Materials Science And Engineering*, *546*(5). Https://Doi.Org/10.1088/1757-899x/546/5/052068

[7] Savitri, N. L. P. C., Rahman, R. A., Venyutzky, R., & Rakhmawati, N. A. (2021). Analisis Klasifikasi Sentimen Terhadap Sekolah Daring Pada Twitter Menggunakan Supervised Machine Learning. *Jurnal Teknik Informatika Dan Sistem Informasi*, *7*(1), 47–58. Https://Doi.Org/10.28932/Jutisi.V7i1.3216

[8] Siringoringo, R. (2018). Klasifikasi Data Tidak Seimbang Menggunakan Algoritma Smote Dan K-Nearest Neighbor. *Jurnal Isd*, *3*(1), 44–49.

[9] Supriyadi, R., Maulidah, N., Fauzi, A., Nalatissifa, H., & Diantika, S. (2022). Penerapan Algoritma Naive Bayes Dan Support Vector Machine Dalam Memprediksi Autisme. *Swabumi*, *10*(1), 55–59. Https://Doi.Org/10.31294/Swabumi.V10i1.12294

[10] Technology, I. (2022). *Sentiment Analysis On Twitter Shopeecare Using Naive Bayes , Adaboost , And Svm (Evolution) Algorithm Comparative Methods*. *19*(1), 21–30.

[11] Tuhuteru, H., & Iriani, A. (2018). Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine Dan Naive Bayes Classifier. *Jurnal Informatika: Jurnal Pengembangan It*, *3*(3), 394–401. Https://Doi.Org/10.30591/Jpit.V3i3.977

[12] Wahyudi, I., Bahri, S., & Handayani, P. (2019). *Aplikasi Pembelajaran Pengenalan Budaya Indonesia*. *V*(1), 135–138. Https://Doi.Org/10.31294/Jtk.V4i2

[13] Wisdayani, D. S. (2019). *Perbandingan Algoritma K-Nearest Neighbor Dan Naive Bayes Untuk Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas Di Kabupaten Pati Jawa Tengah*. 9–25.

[14] Zukhrufillah, I. (2018). Gejala Media Sosial Twitter Sebagai Media Sosial Alternatif. *Al-I'lam: Jurnal Komunikasi Dan Penyiaran Islam*, *1*(2), 102. Https://Doi.Org/10.31764/Jail.V1i2.235