ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER MENGENAI KEBIJAKAN SELEKSI JALUR ZONASI MENGGUNAKAN METODE KLASIFIKASI NAÏVE BAYES DAN SELEKSI FITUR PARTICLE SWARM OPTIMIZATION

Yudhistira [1], Didit Widiyanto[2], Mayanda Mega Santoni[3]
Informatika / Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jakarta
Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok, Daerah Khusus Ibukota Jakarta 12450
yudhistira@upnvj.ac.id[1], didit.widiyanto@upnvj.ac.id[2], megasantoni@upnvj.ac.id[3]

Abstrak. Media sosial menjadi wadah untuk menampung opini atau sentimen masyarakat. Contoh aplikasi yang sering digunakan untuk membahas sentimen tersebut adalah Twitter. Pengguna Twitter sering menyampaikan pendapatnya tentang beberapa topik termasuk pendidikan khususnya kebijakan pemerintah mengenai seleksi sekolah jalur zonasi yang dapat dilihat di Twitter. Berdasarkan latar belakang kondisi tersebut, maka dibutuhkan penelitian tentang sentimen masyarakat mengenai kebijakan seleksi sekolah jalur zonasi. Penelitian ini dilakukan dengan menggunakan metode Machine Learning yaitu klasifikasi Naïve Bayes dan seleksi fitur menggunakan Particle Swarm Optimization untuk mengklasifikasikan tweet positif atau tweet negatif yang masyarakat lontarkan khususnya pelajar dan orang tuanya yang memperebutkan kursi di sekolah-sekolah negeri. Hasil dari penelitian ini membuktikan bahwa klasifikasi Naïve Bayes dengan seleksi fitur Particle Swarm Optimization mendapatkan nilai model evaluasi pada akurasi lebih besar dibandingkan tanpa menggunakan Particle Swarm Optimization dimana akurasi sebesar 78%, presisi sebesar 23%, recall sebesar 45%, dan specificity sebesar 82%. Sementara itu, nilai model evaluasi klasifikasi Naïve Bayes tanpa menggunakan seleksi fitur Particle Swarm Optimization lebih kecil, dimana akurasi sebesar 75%, presisi sebesar 37%, recall sebesar 28%, dan specificity sebesar 87%. Terjadi kenaikan performa pada akurasi dan recall, serta terjadi penurunan performa pada presisi dan specificity. Nilai evaluasi pada presisi dan recall mendapatkan nilai yang rendah, karena data yang tidak seimbang, dimana perbandingan antara label positif dan negatif adalah 1:4.
Kata Kunci: Twitter, Analisis Sentimen, Sekolah Negeri, Zonasi, *Naïve Bayes*, *Particle Swarm Optimization*

1 Pendahuluan

Pendidikan merupakan komponen yang sangat penting dalam menunjang kualitas sumber daya manusia supaya menjadi seseorang yang kompeten dan memiliki daya saing yang tinggi. Melalui sekolah dan dan pendidikan, masyarakat berharap dapat merubah kehidupan dan kesejahteraan sosial menjadi lebih baik serta menggapai masa depan yang cemerlang dengan mengikuti pendidikan setinggi-tingginya. Salah satu hal yang menjadi standar pendidikan sekarang adalah tahapan seleksi seperti seleksi universitas ada SNMPTN, SBMPTN, dan mandiri, begitu juga dengan sekolah menengah salah satunya yaitu jalur zonasi. Prosedur zonasi pada PPDB banyak menimbulkan permasalahan pada pelaksanaanya di berbagai provinsi. Sebagai contoh satu kasus di daerah Kabupaten Banyumas. Dengan kuota jalur zonasi 80 %, prestasi 15%, dan jalur perpindahan tugas orang tua/wali paling banyak 5% dari daya tampung sekolah. Hal ini menyebabkan keresahan bagi masyarakat yang ingin mendaftar ke sekolah negeri. Artinya sistem seleksi zonasi hanya dapat mengisi maksimal 15% untuk siswa berprestasi, sehingga kasus yang ada di daerah Kabupaten Banyumas membuat masyarakat setempat kecewa karena dengan peraturan zonasi. Dapat disimpulkan bahwa sistem seleksi zonasi masih tidak bisa mengakomodasi semua calon peserta didik yang baru dalam sekolah negeri. Bahkan calon peserta didik yang tinggal di area yang tidak terjangkau zona sekolah akan kesulitan masuk sekolah negeri jika tidak memiliki prestasi [1].

hasil survey APJII yang disampaikan oleh Sekretaris Jenderal APJII Henri Kasyfi Soemartono menyatakan hasil utama dari survei pengguna Internet di Indonesia 2019 sampai 2020. Saat ini presentase pengguna internet di Indonesia berjumlah 73,7 persen, presentasenya naik yang awalnya 64,8 persen pada tahun 2018. Jika digabungkan dengan angka dari proyeksi Badan Pusat Statistik (BPS) maka populasi Indonesia tahun 2019 berjumlah 266.911.900 juta jiwa, sehingga perkiraan pengguna internet di Indonesia sebanyak 196,7 juta pengguna. Salah satu penggunaan internet yang sering digunakan masyarakat adalah media sosial seperti Instagram, Twitter, Facebook, dll. Twitter adalah media sosial yang paling banyak digunakan di Indonesia [2].

Dari latar belakang tersebut, maka dibutuhkan penelitian tentang opini masyarakat mengenai kebijakan seleksi sekolah jalur zonasi. Cara yang tepat yaitu melakukan analisis sentimen mengenai tweet yang membahas kebijakan seleksi sekolah jalur zonasi di jejaring sosial Twitter. Penelitian ini memakai metode seleksi fitur, dan klasifikasi Naïve Bayes untuk mengklasifikasi tweet positif dan tweet negatif yang masyarakat lontarkan tentang kebijakan pemerintah mengenai seleksi sekolah jalur zonasi.

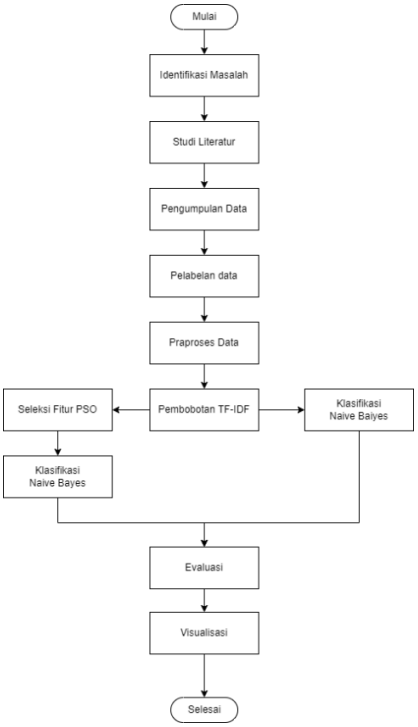
Penggunaan metode klasifikasi Naïve Bayes bisa memproses data yang jumlahnya banyak serta mempunyai akurasi yang terbilang tinggi. Menurut penelitian tentang analisis sentiment terhadap performa timnas sepakk bola Indonesia pada media sosial twitter mendapatkan nilai akurasi sebanyak 87% [3]. Dari hasil penelitian diatas dapat dibuktikan bahwa tingkat ketepatan akurasi yang tinggi didapatkan oleh metode Naïve Bayes. Tetapi masih bisa dimaksimalkan dengan seleksi fitur menggunakan Particle Swarm Optimization, dengan begitu hasil akurasi yang dihasilkan akan lebih maksimal lagi. Pada penelitian yang dilakukan terhadap analisis sentimen dewan perwakilan rakyat dengan algoritma berbagai klasifikasi seperti naïve bayes dan support vector machine berbasis Particle Swarm Optimization, mendapatkan tingkat akurasi yang lebih besar dari penggunaan klasifikasi saja. Tingkat akurasi untuk penggunaan klasifikasi Naïve Bayes 70,69% dengan menggunakan seleksi fitur Particle Swarm Optimization 73,49%. Sedangkan tingkat akurasi menggunakan klasifikasi Support Vector Machine 71,04% dengan seleksi fitur mencapai 75,03% [4]. Hal ini membuktikan bahwa penggunaan Particle Swarm Optimization sangat optimal untuk memaksimalkan performa model evaluasi dari klasifikasi.

Berdasarkan penjelasan latar belakang diatas, maka penulis tertarik untuk memilih judul "Analisis Sentimen Pada Media Sosial Twitter Mengenai Kebijakan Seleksi Sekolah Jalur Zonasi Menggunakan Metode Klasifikasi Naïve Bayes Dan Seleksi Fitur Particle Swarm Optimization" untuk mengetahui hasil performa model evaluasi dari perbandingan metode klasifikasi Naïve Bayes dengan menggunakan seleksi fitur Particle Swarm Optimization dan tanpa seleksi fitur.

2 Metodologi Penelitian

2.1 Tahap Penelitian

Dalam melakukan penelitian, ada beberapa tahapan penelitian yang dilakukan sebagai berikut:



Gambar 1. Tahap Penelitian

2.2 Identifikasi Masalah

Tahapan penelitian ini merupakan tahapan untuk menjelaskan masalah yang akan dibahas pada suatu penelitian. Dari penelitian ini yaitu permasalahan dalam mengklasifikasikan tweet yang mengandung opini atau sentimen tentang kebijakan pemerintah mengenai seleksi sekolah jalur zonasi dengan metode klasifikasi Naïve Bayes dan seleksi fitur Particle Swarm Optimization.

2.3 Studi Literatur

Studi literatur pada penelitian ini dipakai sebagai sumber pustaka dengan mengumpulkan berbagai jurnal tentang masalah analisis sentimen, text mining, metode klasifikasi Naïve Bayes, dan seleksi fitur Particle Swarm Optimization yang dibahas oleh peneliti pada penelitian dan sebagainya. Kumpulan pembahasan dicari menggunkan beberapa macam literatur, website, e-book, dan jurnal yang saling berkaitan untuk sarana pendukung penelitian. Setelah itu, studi pustaka dijadikan dasar untuk memecahkan permasalahan dalam penelitian ini.

2.4 Pengumpulan Data

Akuisisi data dilakukan untuk mencari serta memperoleh data yang dibutuhkan untuk penelitian. Data yang dimaksud pada penelitian ini adalah tweet pengguna twitter yang mention atau hastag zonasi yang akan digunakan sebagai data training dan data testing. Data twitter dapat diambil dengan menggunakan API twitter yang sudah tersedia pada twitter.

2.5 Pelabelan Data

Pelabelan data dikerjakan secara manual oleh 3 orang penilai ke dalam 2 kategori, label positif dan negatif. Berikut pada Tabel 1 adalah contoh hasil pelabelan secara manual oleh 3 orang penilai :

Tabel 1. Contoh Pelabelan Data *Tweet*

Data *Tweet*	Penilai 1	Penilai 2	Penilai 3	Label Akhir
bagus sih semenjak ada zonasi kumpulan anak orang kaya jadi hilang	Positif	Positif	Positif	Positif
waktu itu zaman zonasi, nangis ovt gabakalan dpt, untungnya dapet	Positif	Positif	Negatif	Positif
kayanya gara gara zonasi, karena sekolah pembangunan ga merata tiba tiba ada zonasi yang jaraknya tuh minim banget, jï¿½	Positif	Negatif	Negatif	Negatif

Dari hasil penilaian label tersebut masih terdapat perbedaan pendapat antara masing-masing penilai dalam mengkategorikan label tweet, maka diperlukan hasil yang menunjukan persetujuan antar penilai dengan perhitungan *Kappa Value*. Berikut hasil perhitungan *kappa value* untuk data tweet yang sudah diberi nilai oleh penilai :

Rumus Persamaan (1) dengan *Kappa Value*

$$Kappa = \frac{P_0 - P_e}{1 - P_e} \quad \textbf{(1)}$$

Keterangan :
Kappa : Koefisien dari nilai kesepakatan dimana 0 untuk persetujuan secara kebetulan tidak satu penilaian, dan 1 untuk persetujuan yang mutlak.
P_0 : Proporsi frekuensi penilai yang penilaian sama
P_e : Peluang kesepakatan antar penilai yang penilaiannya berbeda.

Dimana persamaan (2) dengan P_e:

$$P_e = P(positif)^2 + P(negatif)^2 \quad \textbf{(2)}$$

Hasil kappa value yang digunakan dapat dikatakan objektif untuk menilai sebuah kesepakatan [5]. Hasil kesepakatan dari 3 orang penilai dengan pengukuran kappa value dapat dikategorikan pada tabel 2 dimana moderate adalah batas minimal untuk melanjutkan ke tahap selanjutnya [6] :

Tabel 2. Nilai Kesepakatan *Kappa Value* :

Kesepakatan	**Nilai k**
None	0.00 - 0.20
Minimal	0.21 – 0.39
Weak	0.40 – 0.59
Moderate	0.60 – 0.79
Strong	0.80 – 0.90
Almost Perfect	k > 0.90

2.6 Praproses Data

Sebelum data tweet diklasifikasi, perlu dilakukan pra proses terlebih dahulu karena data belum berstruktur dan memiliki banyak noise. Tahapan dari pra proses ini terdiri dari case folding, cleansing, tokenizing, normalizing, filtering dan stemming.

2.5.1 Case Folding

Tahap pertama pada pra proses data yaitu Case Folding, dimana seluruh data tweet yang didapat dari yang memiliki huruf kapital (uppercase) dikonversi menjadi huruf non-kapital (lowercase), bertujuan untuk mencegah terjadinya case sensitive.

2.5.2 Cleansing

Sebelum dilakukan tokenisasi, perlu dilakukan pembersihan pada data tweet yang tidak memiliki nilai dengan menghapus username, tag @, hashtag #, link URL, dan juga emoji serta juga menghapus beberapa karakter tanda baca atau whitespace untuk mengurangi noise.

2.5.3 Tokenizing

Setelah pembersihan data, data dilakukan tokenisasi, yaitu metode pemenggalan dokumen menjadi potongan-potongan kata yang disebut token

2.5.4 Normalizing

Setelah token-token terbuat, data tweet yang diperoleh perlu dilakukan normalisasi, dimana kata tidak baku diubah menjadi kata baku, dan kata yang sesuai dengan kamus besar Bahasa Indonesia (KBBI).

2.5.5 Filtering

Setelah data sudah di melewat tahap normalisasi bahasa, selanjutnya adalah penghilangan stopword, yaitu penghapusan kata yang berulang kali muncul sehingga tidak penting dan tidak berpengaruh pada performa proses klasifikasi.

2.5.6 Stemming

Tahap terakhir pada pra proses data, data-data tweet akan dilakukan stemming sebagai proses untuk menghilangkan imbuhan pada kata-kata yang dikembalikan ke kata dasar.

2.7 Pembobotan Kata (TF-IDF)

Setelah pra proses data dilakukan, maka perhitungan bobot term dengan mengubah kata-kata pada data tweet menjadi sebuah angka sehingga kata/term dapat dikenali sebagai fitur untuk klasifikasi nanti. Metode perhitungan Term Frequency (TF) dan Inverse Document Frequency (IDF) menggunakan rumus :

$$wi,j = tfi,j\ log\ \left(\frac{N}{dfj}\right) \quad \textbf{(3)}$$

Keterangan :

wi.j	:bobot dokumen ke-i untuk kata ke-j
tfi.j	:banyak kata j yang dicari pada dokumen i
N	:total dokumen
Dfj	:banyak dokumen yang mengandung kata ke-j

2.8 Seleksi Fitur dengan Particle Swarm Optimization

Tahap selanjutnya adalah seleksi fitur menggunakan Particle Swarm Optimization, dengan langkah awal yaitu menentukan posisi awal partikel dan kecepatan awal secara random, lalu menentukan Local Best dan Global Best. Local Best pertama, yaitu posisi awal partikel. Untuk menentukan local best selanjutnya adalah dengan membandingkan nilai fitness partikel sekarang dengan local best. Jika nilai fitness pada partikel sekarang lebih kecil dari nilai fitness di local best, maka local best akan diperbarui dengan nilai posisi sekarang. Setelah mendapat local best, ditentukan partikel global best dengan membandingkan nilai fitness dari setiap local best yang memiliki nilai fitness terkecil. Selanjutnya algoritma tersebut akan berhenti jika nilai fitness terbaik sudah tercapai atau iterasi sudah maksimal. Menghitung kecepatan dan posisi menggunakan persamaan :

$$v_j = v_j(i-1) + c_1 r_1 \left(P_{best,j} - X_j(i-1)\right) + c_2 r_2 \left(G_{best} - X_j(i-1)\right) \quad \textbf{(4)}$$

$$x_j(i) = v_j(i) + x_j(i-1) \quad \textbf{(5)}$$

Keterangan :

v_j	: Kecepatan partikel
x_j	: Posisi partikel
j	: 1, 2, … , N mempresentasikan jumlah partikel
$P_{best,j}$	: posisi terbaik dari partikel ke j
G_{best}	: posisi terbaik global
$c1$ dan $c2$	: learning factor
$r1$ dan $r2$	: bilangan acak antara 0 sampai 1

2.9 Klasifikasi dengan Naïve Bayes

Klasifikasi menggunakan Naïve Bayes akan dilakukan dengan Multinomial Naïve Bayes. Data akan dibagi menjadi data latih sebesar 80% dan data uji sebesar 20%. Proses klasifikasi menggunakan algoritma Naïve Bayes dihitung dengan persamaan:

Persamaan Naïve Bayes (6)

$$P(x|y) = \frac{P(x)\times P(y|x)}{P(y)} \quad \textbf{(6)}$$

Keterangan :

x : Hipotesa data y merupakan suatu kelas spesifik.

y : Data dengan kelas yang belum diketahui.

P(x|y) : Peluang dari hipotesa x bila kondisi y (posterior).

P(x) : Peluang dari hipotesa x.

P(y|x) : Peluang dari hipotesa y bila kondisi x.

P(y) : Peluang dari data sampel yang diamati.

Persamaan untuk menghitung peluang pad amasing-masing kelas (7)

$$P(x) = \frac{|\, doc\ x\,|}{|document|} \quad \textbf{(7)}$$

Keterangan :

$P(x)$: Peluang hipotesa x.

$doc\ x$: Jumlah dokumen dari kategori x.

$|document|$: Jumlah dokumen dari setiap kategori

Persamaan untuk mengukur peluang setiap kata yang berasal dari dokumen yang ada berdasarkan kategori (8)

$$P(Wi|x) = \frac{Count(Wi,x)+1}{|x|+|V|} \quad \textbf{(8)}$$

Keterangan :

$P(Wi|x)$: Peluang kata Wi pada kelas x.

$Count(Wi,x)$: Total kemunculan kata Wi pada kelas x.

$|x|$: Total kata pada kelas x.

$/V/$: Total semua kata

Persamaan untuk mengklasifikasi data testing (9)

$$x_{MAP} = argmax\ P(x)\ \Pi_{i=1}^{n} P(W_i|x) \quad \textbf{(9)}$$

Keterangan :

x_{MAP} : Kategori yang memiliki probabilitas paling tinggi.

$argmax$: Nilai terbesar dari fungsi.

$P(x)$: Peluang kemunculan suatu dokumen yang memiliki kelas x.

$P(W_i|x)$: Peluang kemunculan *Wi* pada kelas x.

$\Pi_{i=1}^{n}$: Perkalian rating antar atribut.

2.10Evaluasi

Pada tahap akhir penelitian, model klasifikasi akan melewati tahap evaluasi menggunakan metode confusion matrix yang sudah dijelaskan pada tabel untuk menganalisis hasil performanya. Berikut rumus yang digunakan untuk menghitung evaluasi pada penelitian ini :

Persamaan (10) untuk menghitung nilai akurasi :

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN}\, x\ 100\% \quad \textbf{(10)}$$

Persamaan (11) untuk menghitung nilai *recall* (*sensitivity*):

$$Recall = \frac{TP}{TP+FN} x\ 100\% \quad \textbf{(11)}$$

Persamaan (12) untuk menghitung nilai presisi :

$$Presisi = \frac{TP}{TP+FP} x\ 100\% \quad \textbf{(12)}$$

Persamaan (13) untuk menghitung nilai *Specificity* :

$$Specificity = \frac{TN}{TN+FP} x\ 100\% \quad \textbf{(13)}$$

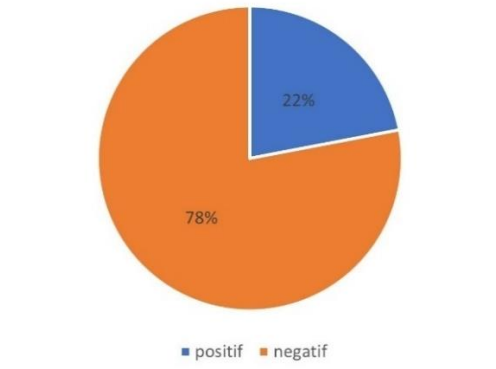
2.11 Visualisasi

Hasil sentimen masyarakat terhadap seleksi sekolah jalur zonasi berdasarkan data tweet yang didapat akan divisualisasi dalam bentuk wordcloud yang berisi frekuensi kemunculan kata terbanyak pada masing-masing sentimen positif ataupun negatif.

3 Hasil Pembahasan

Penelitian ini menggunakan data tweet yang diperoleh dengan mengaplikasikan metode crawling dengan bahasa Python memanfaatkan API (Application Programming Interface) yang disediakan oleh twitter. Data hasil crawling disimpan dalam bentuk file .csv yang berisikan record tweet. Data tweet yang diambil mulai dari tanggal 18 November 2021 sampai 12 Januari 2022 dengan kata kunci "zonasi", dan "#zonasi" dari hasil crawling data tersebut melalui proses penyaringan sehingga terkumpul sebanyak 500 data tweet yang berisikan sentimen opini publik.

Pelabelan data dikerjakan secara manual oleh 3 orang penilai ke dalam 2 kategori, label positif dan negatif. Dari hasil penilaian label tersebut masih terdapat perbedaan pendapat antara masing-masing penilai dalam mengkategorikan label tweet yang hasil yang perlu perhitungan Kappa Value untuk menunjukan tingkat persetujuan antar penilai. Setelah data tweet dilakukan voting hasil penilaian labelnya, hasil dari pelabelan data tweet sebanyak 500 tweet adalah 110 tweet berlabel positif dan 390 tweet berlabel negatif yang dipersentasekan pada Gambar 2 berikut :



Gambar 2. Presentase Hasil Pelabelan Data

Dari hasil visualisasi tersebut dapat diketahui bahwa persentase sentimen positif masyarakat terhadap seleksi sekolah jalur zonasi sebesar 22% sedangkan sentimen negatif masyarakat terhadap seleksi sekolah jalur zonasi sebesar 78%. Maka dapat dilihat bahwa sentimen negatif masyarakat terhadap seleksi sekolah jalur zonasi lebih banyak daripada sentimen positif.

Setelah data yang diperoleh diberi label, dilakukan beberapa tahapan praproses data, yaitu case folding, cleansing, tokenizing, normalizing, filtering, dan stemming. Hasil dari praproses data dapat dilihat pada Tabel 3

Tabel 3. Hasil praproses data

Data *tweet*
['bagus', 'semenjak', 'zonasi', 'kumpul', 'anak', 'orang', 'kaya', 'hilang']
['pikirin', 'jalur', 'zonasi', 'prestasi', 'korban', 'lempar']
['jalur', 'zonasi', 'sampah', 'murid', 'jiwa', 'kompetitif']
['zaman', 'zonasi', 'menang', 'pikir', 'gabakal', 'untung']

Kemudian data diatas dilakukan perhitungan pembobotan kata dengan variabel kata atau term yang diperoleh dari hasil praproses data dari jumlah data tweet sebanyak 500 tweet menggunakan metode perkalian Term Frequency - Inverse Document Frequency (TF-IDF) seperti pada Tabel 4 :

Tabel 4. Perhitungan Pembobotan Kata (TF-IDF)

Term	Dokumen				DF	IDF	TF-IDF			
	D1	D2	D3	D4			D1	D2	D3	D4
bagus	1	0	0	0	1	0,477	0,477	0	0	0
semenjak	1	0	0	0	1	0,477	0,477	0	0	0
zonasi	1	1	1	1	4	0	0	0	0	0
kumpul	1	0	0	0	1	0,477	0,477	0	0	0
Anak	1	0	0	0	1	0,477	0,477	0	0	0
orang	1	0	0	0	1	0,477	0,477	0	0	0
Kaya	1	0	0	0	1	0,477	0,477	0	0	0
hilang	1	0	0	0	1	0,477	0,477	0	0	0
pikirin	0	1	0	0	1	0,477	0	0,477	0	0
Jalur	0	1	1	0	2	0,176	0	0,176	0,176	0
prestasi	0	1	0	0	1	0,477	0	0,477	0	0
korban	0	1	0	0	1	0,477	0	0,477	0	0
lempar	0	1	0	0	1	0,477	0	0,477	0	0
sampah	0	0	1	0	1	0,477	0	0	0,477	0
murid	0	0	1	0	1	0,477	0	0	0,477	0
Jiwa	0	0	1	0	1	0,477	0	0	0,477	0
kompetitif	0	0	1	0	1	0,477	0	0	0,477	0
zaman	0	0	0	1	1	0,477	0	0	0	0,477
menang	0	0	0	1	1	0,477	0	0	0	0,477
Piker	0	0	0	1	1	0,477	0	0	0	0,477
gabakal	0	0	0	1	1	0,477	0	0	0	0,477
untung	0	0	0	1	1	0,477	0	0	0	0,477

Data tweet yang sudah diberi bobot dengan TF-IDF dibagi menjadi data latih (training) yang diambil 80% secara acak, sedangkan data uji (testing) diambil 20% yang berisi hasil sisa dari proses pembagian data latih (training). Perbandingan pembagian data latih (training) dan data uji (testing) secara bebas dan imbang seperti pada Tabel 5 berikut :

Tabel 5. Pembagian Data

	Label Positif	Label Negatif	Total
Data Latih (Training)	89	311	400
Data Uji (Testing)	21	79	100
Total	110	390	500

Setelah dilakukan pembagian data, maka masuk ke tahapan klasifikasi. Proses klasifikasi akan diterapkan menggunakan Naïve Bayes (NB) dengan menemukan nilai probabilitas tertinggi pada sentimen postif dan sentimen negatif dan diperoleh hasil nilai uji terbaik pada sentimen postif dan negatif untuk membuat model klasifikasi. Hasil evaluasi model NB dengan confusion matrix dapat dilihat pada Tabel 7 :

Tabel 6. Confusion Matrix dari Model Klasifikasi NB

Aktual	**Prediksi**	
	Positif	**Negatif**
Positif	6 (TP)	10 (FN)
Negatif	15 (FP)	69 (TN)

Maka dari Tabel 6 Confusion Matrix dapat dihitung hasil evaluasi model NB menggunakan rumus (10), (11), (12), dan (13) sebagai berikut :

Akurasi	$= \frac{TP+TN}{TP+FN+FP+TN} x\ 100\%$	$= \frac{6+69}{6+15+10+69} x\ 100\%$	= 75%
Recall	$= \frac{TP}{TP+FN} x\ 100\%$	$= \frac{6}{6+10} x\ 100\%$	= 28%
Presisi	$= \frac{TP}{TP+FP} x\ 100\%$	$= \frac{6}{6+15} x\ 100\%$	= 37%
Specificity	$= \frac{TN}{TN+FP} x\ 100\%$	$= \frac{69}{69+10} x\ 100\%$	= 87%

Selanjutnya dilakukan klasifikasi kedua akan memanfaatkan seleksi fitur dengan algoritma Particle Swarm Optimization (PSO), dengan parameter nilai learning factor (c1 dan c2), inersia (w), jumlah tetangga (k), p-norm Minkowski (p), dan jumlah partikel (n_particles), dimana pada penelitian ini yang digunakan adalah parameter default atau bawaan, yaitu untuk nilai (c1) sebesar 0.5, (c2) sebesar 0.5, (w) sebesar 0.9, (p) sebesar 9, dan (n_particles) sebanyak 30. Proses seleksi fitur dengan algoritma Particle Swarm Optimization (PSO) menghasilkan pengurangan fitur. Hasil pengurangan fitur tersebut akan diaplikasikan kembali pada proses pemodelan klasifikasi dengan algoritma Support Vector Machine (SVM) berkernel Radial Basis Function (RBF) dengan C = 10 dan γ = 0.01. Maka dilakukan percobaan sebanyak 4 skenario dengan iterasi PSO yang berbeda, yaitu 50, 100, 250, 500, dan 800 dengan parameter yang sama yang dapat dilihat pada Tabel 7 berikut :

Tabel 7. Hasil Akurasi Percobaan 5 Skenario Iterasi PSO

Iterasi PSO	**Banyak Fitur yang digunakan**	**Akurasi SVM+PSO**
50	523	77%
100	534	74%
250	545	78%
500	520	76%
800	590	76%

Dari hasil 5 skenario iterasi PSO dan diperoleh hasil nilai akurasi terbaik pada 250 iterasi dengan jumlah fitur sebanyak 545 dari 1069 fitur, yang selanjutnya hasil jumlah fitur tersebut digunakan untuk membuat model klasifikasi NB. Hasil evaluasi model NB dengan seleksi fitur PSO dengan confusion matrix dapat dilihat pada Tabel 8 :

Tabel 8. Confusion Matrix dari Model Klasifikasi NB + PSO

Aktual	**Prediksi**	
	Positif	**Negatif**
Positif	5 (TP)	6 (FN)
Negatif	16 (FP)	73 (TN)

Maka dari Tabel 9 Confusion Matrix dapat dihitung hasil evaluasi model NB + PSO menggunakan rumus (10), (11), (12), dan (13) sebagai berikut :

Akurasi $= \frac{TP+TN}{TP+FN+FP+TN} x\ 100\% = \frac{5+73}{5+6+16+73} x\ 100\% = 78\%$

Recall $= \frac{TP}{TP+FN} x\ 100\% = \frac{5}{5+6} x\ 100\% = 45\%$

Presisi $= \frac{TP}{TP+FP} x\ 100\% = \frac{5}{5+16} x\ 100\% = 23\%$

Specificity $= \frac{TN}{TN+FP} x\ 100\% = \frac{73}{73+16} x\ 100\% = 82\%$

Maka dari 2 tahapan klasifikasi yang dilakukan, didapatkan hasil confusion matrix dan hasil evaluasi berupa akurasi, recall, presisi, dan specificity berdasarkan kedua percobaan yang telah dilakukan, dapat dibandingkan kedua performa model tersebut pada Tabel 9 berikut :

Tabel 9. Hasil Perbandingan Performa Model NB dan NB + PSO

	Akurasi	**Recall (Sensitivity)**	**Presisi**	**Specificity**
SVM	75%	28%	37%	87%
SVM + PSO (500 ITERASI)	78%	45%	23%	82%

Performa model terbaik untuk klasifikasi sentimen analisis data tweet aplikasi PeduliLindungi adalah model klasifikasi Naïve Bayes (NB) dengan seleksi fitur menggunakan Particle Swarm Optimization (PSO) sebanyak 250 iterasi.

Kemudian tahapan terakhir yaitu visualisasi berdasarkan sentimen data tweet dengan label positif dan negatif secara manual, dengan tujuan menggambarkan hasil penelitian sentimen terhadap seleksi sekolah jalur zonasi. Visualisasi akan digambarkan oleh wordcloud.

Gambar 3. Wordcloud Sentimen Positif

Berdasarkan Gambar 3 tersebut, kata yang sering muncul pada tweet sentimen positif yaitu "zonasi", "sekolah", "masuk", "pakai", dan "prestasi". Dari kata-kata tersebut dapat disimpulkan bahwa masuk sekolah tidak hanya menggunakan jalur zonasi, tapi bisa juga menggunakan prestasi.

Gambar 4. Wordcloud Sentimen Negatif

Berdasarkan Gambar 4, kata yang sering muncul pada tweet sentimen negatif yaitu "zonasi", "sekolah", "masuk", "nilai", dan "rumah". Dari kata-kata tersebut dapat disimpulkan bahwa jalur zonasi menjadi masalah untuk siswa yang memiliki nilai bagus tetapi jarak rumahnya jauh.

4 Kesimpulan dan Saran

4.1 Kesimpulan

Analisis sentimen mengenai seleksi sekolah jalur zonasi menggunakan data yang merupakan tweet dari pengguna twitter dengan rentang waktu bulan November 2021 sampai Januari 2022 sebanyak 500 tweet. Data tersebut dilabeli manual oleh tiga anotator yang akan dihitung validitas pelabelannya menggunakan kappa value, yang didapatkan 110 tweet sentimen positif, dan 390 tweet sentimen negatif. Kemudia dilakukan pembersihan data pada tahap pra proses dengan beberapa tahapan, setelah data dibersihkan dilakukan pembobotan kata menggunakan TF-IDF, dimana setiap kata yang sudah dibobotkan akan menjadi fitur. Lalu fitur yang didapatkan akan diseleksi menggunakan metode optimasi particle swarm optimization, fitur yang sudah diseleksi akan dilakukan pembagian data training dan data testing dengan perbandingan 80:20 menggunakan metode klasifikasi naïve bayes. Hasil pemodelan klasifikasi menggunakan metode naïve bayes mendapatkan nilai akurasi sebesar 75%, presisi sebesar 37%, recall sebesar 28%, dan specificity sebesar 87%. Kemudian menggunakan metode klasifikasi naïve bayes dengan metode seleksi fiturnya menggunakan particle swarm optimization didapatkan nilai akurasi sebesar 78%, presisi sebesar 23%, recall sebesar 45%, dan specificity sebesar 82%. Model evaluasi yang digunakan tidak terklasifikasi dengan baik, disebabkan data yang tidak seimbang dengan perbandingan antara data label positif dengan label negatif adalah 1:4.

Seleksi fitur particle swarm optimization dapat mengoptimasi metode klasifikasi naïve bayes dalam analisis sentimen dengan dibuktikan pada penelitian ini mampu meningkatkan nilai akurasi sebesar 3%, meningkatkan nilai recall sebesar 17%, tetapi menurunkan nilai presisi sebesar 14% dan menurunkan nilai specificity sebesar 5%. Waktu pengujian menggunakan particle swarm optimization lebih lama karena harus melakukan pengujian berdasarkan iterasi yang sudah ditentukan dan sampai mendapatkan nilai fitness terbaik.

4.2 Saran

Berdasarkan hasil pembahasan dan analisis yang sudah dilakukan pada penelitian ini, dapat diberikan saran untuk pengembangan penelitian ini sebagai bahan evaluasi bagi penelitian selanjutnya sebagai berikut:

- Disarankan untuk penelitian selanjutnya untuk menggunakan data yang lebih banyak supaya dapat meningkatkan lagi performa dari klasifikasi.Untuk penelitian pada masa yang akan datang, pada pelabelan data diharapkan menggunakan ahli tata Bahasa sehingga tidak terjadi bias pada hasil label masing-masing data tweet, dan tidak memerlukan 3 orang penilai.
- Disarankan untuk penelitian selanjutnya untuk meningkatkan pada tahap pra proses serta pemodelan

data yang lebih baik sehingga bisa mendapatkan evaluasi model yang baik.Untuk penelitian pada masa yang akan datang, dapat menggunakan algoritma seleksi fitur yang lain seperti Chi-Square atau Information Gain.

- Penelitian selanjutnya dapat meggunakan metode klasifikasi atau seleksi fitur yang lainnya supaya dapat membandingkan hasil evaluasi yang didapatkan.
- Disarankan untuk penelitian selanjutnya untuk melakukan penyeimbangan data terlebih dahulu jika data belum seimbang supaya model evaluasi bisa terklasifikasikan dengan baik

Referensi

[1] Mulyani, T., & Muryati, D. T. (2020). ANALISIS YURIDIS MENGENAI SISTEM ZONASI DALAM PENERIMAAN PESERTA DIDIK BARU. JURNAL USM LAW REVIEW, 3(1), 34-58.

[2] KOMINFO. "Survei Penetrasi Pengguna Internet di Indonesia Bagian Penting dari Transformasi Digital". Kominfo.go.id. www.kominfo.go.id/content/detail/30653 (Diakses Oktober 28, 2021).

[3] Pradana, Y. R. Y., Astiningrum, M., & Hani'ah, M. (2020, October). Analisis Sentimen Tentang Opini Terhadap Performa Timnas Sepak Bola Indonesia Pada Twitter. In Seminar Informatika Aplikatif Polinema (pp. 35-39).

[4] FAISAL, Anas, et al. (2020). Analisis Sentimen Dewan Perwakilan Rakyat Dengan Algoritma Klasifikasi Berbasis Particle Swarm Optimization. JOINTECS (Journal of Information Technology and Computer Science), 5(2), 61-70.

[5] Feizizadeh, B., Darabi, S., Blaschke, T., & Lakes, T. (2022). QADI as a New Method and Alternative to Kappa for Accuracy Assessment of Remote Sensing-Based Image Classification. Sensors, 22(12), 4506.

[6] Researchgate. “Interpretatition of Cohens Kappa”. Researchgate.net. https://www.researchgate.net/figure/Interpretation-of-Cohens-kappa_tbl1_347970159 (Diakses Juni 2, 2022).

[7] Karsito, K., & Taufiq, A. (2020). Analisis Sentimen Terhadap Pemindahan Ibu Kota Pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization. Jurnal SIGMA, 10(3), 173-182.

[8] Gunawan, D., Riana, D., Ardiansyah, D., Akbar, F., & Alfarizi, S. (2020). Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023. V (1), 135–138.

[9] Krisdiyanto, T. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers. Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi, 7(1), 32-37.

[10] Mahesh, B. (2020). Machine Learning Algorithms-A Review. International Journal of Science and Research (IJSR), 381-386.