



Analisis Sentimen Pembangunan IKN (Ibu Kota Nusantara) Pada Twitter Menggunakan Metode K- Nearest Neighbor, Naive Bayes Dan Support Vector Machines

*Yossy Veiebrian Fitri Prasmono ¹, Arya Adhyaksa Waskita ²

^{1,2,)}Teknik Informatika, Teknik, Universitas Pamulang, Tangerang Selatan, Banten

Email: ¹yossyveif@gmail.com, ² aawaskita@unpam.ac.id

ABSTRACT

This research investigates the Nusantara Capital City (IKN) relocation, which has generated diverse opinions, including concerns over the chosen location and the swift ratification of related laws. Recently, the Indonesian government has called on the public to support IKN's development. To assess public sentiment regarding this relocation, sentiment analysis was performed on a dataset of tweets. After data cleaning, 502 tweets were analyzed, yielding 337 positive and 163 negative comments. The analysis utilized Support Vector Machine (SVM), Naive Bayes (NB), and K-Nearest Neighbor (K-NN) algorithms, incorporating feature selection through Particle Swarm Optimization (PSO). This study compares the performance of Naive Bayes, SVM, and K-NN without feature selection against those methods with feature selection, specifically analyzing their Area Under Curve (AUC) values to identify the most effective algorithm. The results indicate that the PSO-based SVM algorithm achieved the highest performance, with an accuracy of 97.63% and an AUC of 0.997. This research successfully identifies an optimal algorithm for classifying positive and negative comments regarding the relocation of the Nusantara Capital City, contributing valuable insights to public sentiment analysis in this context.

Keywords: Nusantara Capital City, IKN, Sentiment Analysis, Naive Bayes, K-Nearest Neighbor, Support Vector Machines, Twitter.

ABSTRAK

Penelitian ini menyelidiki relokasi Ibu Kota Nusantara (IKN), yang telah menghasilkan beragam opini, termasuk kekhawatiran mengenai lokasi yang dipilih dan pengesahan cepat undang-undang terkait. Baru-baru ini, pemerintah Indonesia mengajak masyarakat untuk mendukung pembangunan IKN. Untuk menilai sentimen publik mengenai relokasi ini, dilakukan analisis sentimen pada dataset tweet. Setelah proses pembersihan data, 502 tweet dianalisis, menghasilkan 337 komentar positif dan 163 komentar negatif. Analisis ini menggunakan algoritma Support Vector Machine (SVM), Naive Bayes (NB), dan K-Nearest Neighbor (K-NN), dengan pemilihan fitur melalui Particle Swarm Optimization (PSO). Penelitian ini membandingkan kinerja Naive Bayes, SVM, dan K-NN tanpa pemilihan fitur dengan metode yang menggunakan pemilihan fitur, dengan menganalisis nilai Area Under Curve (AUC) untuk mengidentifikasi algoritma yang paling efektif. Hasilnya menunjukkan bahwa algoritma SVM berbasis PSO mencapai kinerja tertinggi, dengan akurasi 97,63% dan AUC 0,997. Penelitian ini berhasil mengidentifikasi algoritma optimal untuk mengklasifikasikan komentar positif dan negatif terkait relokasi Ibu Kota Nusantara, memberikan wawasan berharga untuk analisis sentimen publik dalam konteks ini.

Kata kunci: Ibu Kota Nusantara, IKN, Analisis Sentimen, Naive Bayes, K-Nearest Neighbor, Support Vector Machines, Twitter.

1. PENDAHULUAN

Pemindahan ibu kota negara bukanlah fenomena baru dalam sejarah, karena banyak negara telah melakukannya dengan berbagai pertimbangan, seperti alasan politik, ekonomi, atau sosial. Di Indonesia, wacana pemindahan ibu kota yang sudah muncul sejak era Presiden Soeharto kini semakin menarik perhatian publik, terutama dengan berkembangnya teknologi informasi dan komunikasi. Melalui media sosial seperti Twitter, masyarakat dengan mudah menyuarakan opini mereka, menjadikan platform ini salah satu ruang diskusi utama untuk membahas berbagai isu, termasuk rencana pemindahan ibu kota.

Metode yang digunakan dalam penelitian ini adalah *K-Nearest Neighbor*, *Naive Bayes*, dan *Support Vector Machines*, dengan tambahan *feature selection* dengan *Particle Swarm Optimization* (PSO). Penelitian akan membandingkan metode *K-Nearest Neighbor*, *Naive Bayes*, dan *Support Vector Machines* tanpa *Particle Swarm Optimization* dibandingkan dengan metode *K-Nearest Neighbor*, *Naive Bayes*, dan *Support Vector Machines* dengan *Particle Swarm Optimization* (PSO). Hal ini untuk menguji pengaruh penggunaan *feature selection Particle Swarm Optimization* (PSO) dengan membandingkan nilai AUC (*Area Under Curve*) dari metode tersebut.

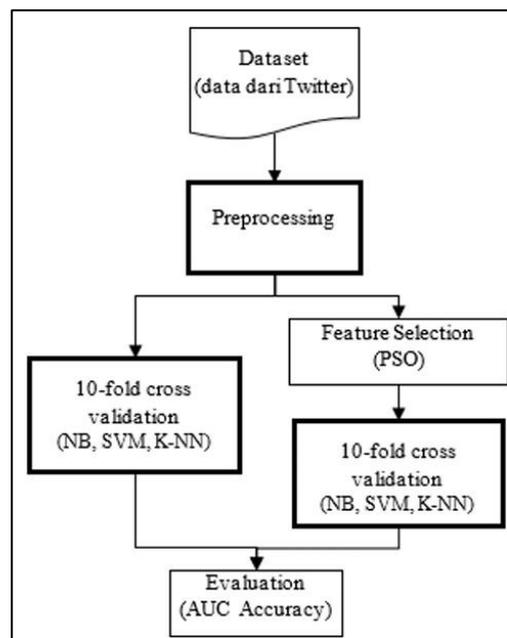
Penelitian oleh Sang Intan Rizqi Adi, Bintang Bakkara, Kurniaman Andreas Zega, Femi Nabila Vielita, dan Nur Aini Rakhmawati (2024) didapatkan bahwa melakukan penelitian terhadap progress IKN menggunakan model Decision Tree. Dari dataset yang digunakan dalam penelitian, jumlah twitter yang mengandung sentimen positif, negatif, dan netral terkait pemindahan ibu kota Indonesia memiliki perbandingan persentase 42,3% untuk sentimen positif, 27,6% untuk sentimen negatif, dan 30,1% untuk sentimen netral. Sehingga antara sentimen positif, negatif, dan netral, besar persentase yang paling besar merupakan sentimen positif yaitu 42,3%, kemudian diikuti dengan sentimen netral, dan terakhir sentimen negatif. Dari hasil tersebut dapat dilihat bahwa mayoritas masyarakat Indonesia mendukung dan memiliki harapan terhadap kesuksesan pemindahan IKN yang sedang berjalan [1].

Berdasarkan penjelasan sebelumnya maka penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap pemindahan ibu kota Indonesia. Penelitian ini akan melakukan komparasi algoritma *K-Nearest Neighbor*, *Naive Bayes Classifier* dan *Support Vector Machines* dalam melakukan klasifikasi. Kinerja algoritma

tersebut dapat bekerja dengan baik dalam melakukan analisis sentimen. Penelitian ini diharapkan dapat menyajikan hasil analisis yang tepat dan akurat dalam analisis sentimen masyarakat terhadap terhadap pemindahan ibu kota Indonesia.

2. METODE

Penelitian analisis kombinasi algoritma *K-Nearest Neighbor*, *Naive Bayes Classifier*, dan *Support Vector Machines* untuk klasifikasi pembangunan Ibu Kota Nusantara (IKN) akan dilakukan dengan beberapa tahapan atau langkah - langkah seperti gambar 3.1 dibawah ini:



Penjelasan tahapan - tahapan penelitian analisis sentimen pembangunan Ibu Kota Nusantara (IKN) sebagai berikut:

1. *Business Understanding*

Proses memahami objek penelitian yang akan dilakukan untuk menghasilkan hasil sesuai dengan yang diharapkan. Pemahaman terhadap objek penelitian dapat dilakukan dengan mencari dan menggali informasi dari sumber objek. Objek yang digunakan yaitu Ibu Kota Nusantara (IKN) yang didapatkan dari media sosial di Twitter. Maka, yang harus dilakukan yaitu membuat akun media sosial di Twitter dan buat aplikasi untuk mendapatkan *API keys* dan *access tokens*. Berikut langkah-langkah membuat akun media sosial Twitter :

- 1) Buka twitter.com/signup.
- 2) Klik tombol daftar.

- 3) Kotak *pop-up* "Buat akun Anda" akan muncul, dan kalian akan dipandu dalam proses pendaftaran. Kalian akan diminta untuk memasukkan informasi seperti nama dan nomor ponsel atau alamat email.
- 4) Jika mencantumkan alamat email saat mendaftar, kami akan langsung mengirim email berisi petunjuk agar kami dapat memverifikasi alamat *email*.
- 5) Jika mencantumkan nomor ponsel saat mendaftar, kami akan langsung mengirim pesan teks berisi kode agar kami dapat memverifikasi nomor ponsel.
- 6) Setelah memasukkan informasi, klik Berikutnya.
- 7) Di kotak *pop-up* Sesuaikan pengalaman kalian, centang jika ingin melacak di mana saja, kalian dapat melihat konten Twitter di internet, lalu klik Berikutnya.
- 8) Pelajari cara menyesuaikan pengaturan untuk akun baru kalian.

Berikut cara mendapatkan *API Keys* dan *access token* pada Twitter:

- 1) Setelah aplikasi dibuat, kalian akan melihat *API Key* dan *access token* di halaman aplikasi tersebut.
- 2) Untuk mendapatkan *API Key* dan *access token*, klik kanan pada aplikasi Twitter, lalu pilih *inspect*, kemudian pilih tab *application* dan dapatkan *auth_token*.
- 3) Simpan *access token* untuk digunakan *crawling* data.

2. *Data Understanding*

Dalam penelitian ini proses yang dilakukan pada tahap ini yaitu pengambilan data mentah yang dilakukan sesuai objek penelitian dengan atribut yang diperlukan. Berikut langkah-langkah *crawling* data Twitter untuk mendapatkan data sentimennya:

1) Mendaftar dan Mendapatkan Akses API Twitter

Membuat akun media sosial di Twitter dan dapatkan *API keys* dan *access tokens*. Kemudian, pastikan memahami aturan penggunaan API Twitter untuk menghindari pelanggaran kebijakan. Pada tahap *business understanding* telah dilakukan pendaftaran akun Twitter dan mendapatkan *access token*.

2) Menentukan Tujuan, Kriteria Data dan Limit Data

Tentukan jenis data yang ingin diambil, sesuai objek yang diteliti akan

mengambil data *tweet* dengan *keyword* “IKN” dengan range tanggal 02 Februari 2023 hingga 02 Februari 2024 dan filter bahasa yang digunakan adalah Bahasa Indonesia dengan limit 500 data *tweet* sentimen.

3) Membangun atau Menggunakan *Crawler*

Untuk melakukan *crawling* data menggunakan *node.js*. Dengan *Node.js* kita dapat menjalankan kode *JavaScript* di mana pun, tidak hanya terbatas pada lingkungan *browser*.

4) Pengolahan dan Analisis Data

Setelah *crawling* data dan data diperoleh sesuai kriteria dan limit yang telah ditentukan, kemudian dilakukan proses pembersihan data (*data cleaning*). Data yang diperoleh dari hasil data *cleaning* tersebut akan digunakan sebagai bahan penelitian.

3. *Data Preparation (Preprocessing)*

Data preparation dilakukan sebelum dataset dimasukkan kedalam model. Proses tersebut meliputi beberapa tahap, yaitu:

(1) *Transform case*

Berikut contoh dari *transform case*:

Text yang didapat: Hello World

Hasil *transform case*: hello world

(2) *Remove http*

Berikut contoh dari *remove http*:

Text yang didapat: <http://www.ikn.com/page>

Hasil *remove http*: www.ikn.com/page

(3) *Remove @*

Berikut contoh dari *remove @*:

Text yang didapat: @ikn

Hasil *remove @*: ikn

(4) *Tokenize*

Berikut contoh dari *Tokenize*:

Text yang didapat: Ibu Kota Nusantara.

Hasil *Tokenize*: ['Ibu', 'Kota', 'Nusantara', '.']

(5) *Filter tokens by length*

Berikut contoh dari *Filter tokens by length*:

Filter Token yang Panjangnya Lebih dari 3 Karakter

Text yang didapat: ['Ia', 'menetap', 'di', 'IKN']

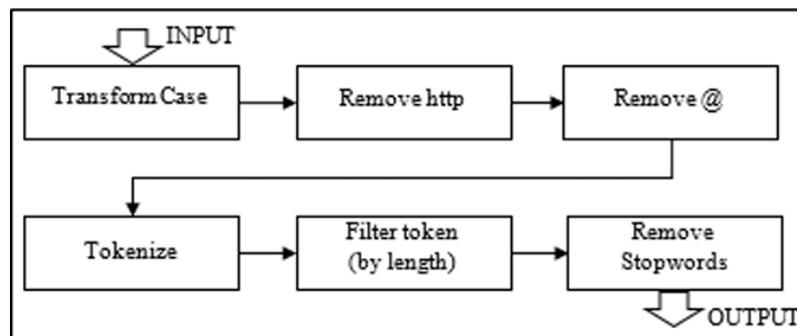
Hasil *Filter tokens by length*: ['menetap', 'IKN']

(6) *Remove Stopwords*

Berikut contoh dari *Remove Stopwords*:

Text yang didapat: Ini adalah contoh menghapus stopwords dari sebuah kalimat.

Hasil *Remove Stopwords*: contoh menghapus stopwords kalimat. Alur kerja data *preparation* dapat dilihat pada gambar 2 berikut.



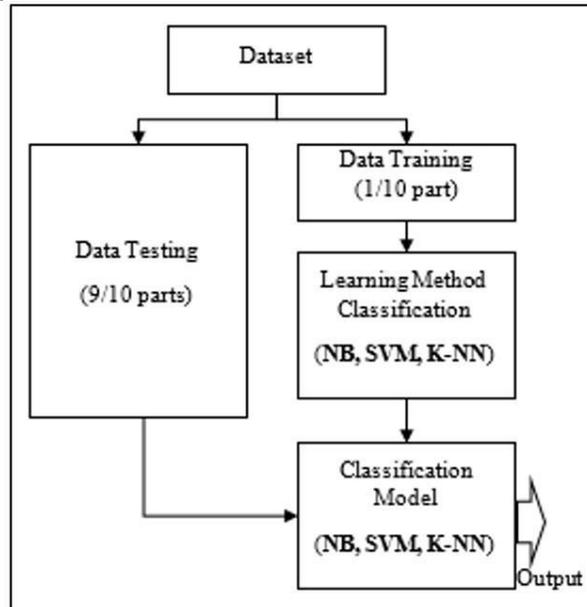
Gambar 3.2 Alur proses Data Preparation (Preprocessing)

4. Pelabelan Data Sentimen

Data yang telah di *preprocessing* akan diexport menjadi CSV dan dilakukan pelabelan secara manual. Dengan proses mengklasifikasikan teks *tweet* ke dalam kategori sentimen “positif” dan “negatif”. Ini adalah langkah penting dalam analisis sentimen, yang digunakan untuk memahami isi teks *tweet*.

5. *Modelling*

Proses ini dilakukan menggunakan metode *10-fold cross validation*, yaitu metode membagi dataset menjadi 10 bagian yang mana 1 bagian diantaranya menjadi data uji (testing) sedangkan bagian lainnya menjadi data latih (*training*). Kemudian data tersebut dimasukkan ke dalam model algoritma yang diuji yaitu *K-Nearest Neighbor*, *Naive Bayes*, dan *Support Vector Machines*. Hal ini dilakukan bergantian pada setiap bagian data hingga didapatkan nilai terbaik dari model tersebut. Alur proses *10-fold cross validation* dapat dilihat pada gambar 3 berikut.



Gambar 3.3 Alur proses 10-fold cross validation

6. Evaluation

Tahap ini melakukan proses evaluasi terhadap hasil dari pemodelan dengan membandingkan hasil dari pemodelan yang berupa *accuracy*, *precision*, *recall* dan AUC antara model yang menggunakan *feature selection* PSO dengan model yang tidak menggunakan *feature selection* PSO. Tingkat akurasi dan ROC Curve dari masing-masing algoritma diklasifikasikan berdasarkan nilai AUC sebagai berikut: 0,90-1,00 = sangat baik. 0,80-0,90 = baik. 0,70-0,80 = cukup. 0,60-0,70 = buruk. 0,50-0,60 = gagal.

3. HASIL DAN PEMBAHASAN

Untuk mendapatkan hasil dari penelitian ini harus melalui beberapa tahapan antara lain: tahapan persiapan data dan implementasi RapidMiner. Berikut langkah- langkah yang diterapkan dalam penelitian ini terdiri dari beberapa proses, antara lain:

3.1. Business Understanding

Pada penelitian ini sumber informasi yang digunakan yaitu media sosial Twitter. Penelitian dilakukan untuk menggali informasi mengenai analisis sentimen mengenai pembangunan Ibu Kota Nusantara (IKN) berdasarkan data Twitter tersebut.

3.2. Data Understanding

Pengambilan data Twitter dilakukan menggunakan tools Google Colab dan program *tweet harvest*. Tweet Harvest merupakan proses mengumpulkan data dari

platform Twitter, biasanya berupa *tweet*, untuk keperluan analisis lebih lanjut. Proses ini sering digunakan dalam penelitian, analisis sentimen, pemantauan media sosial, analisis tren, dan banyak lagi. Dengan menggunakan *keyword* “IKN”, dan “Ibu Kota Nusantara” berdasarkan *range* tanggal yang ditentukan, kemudian data tersebut di-*export* ke dalam format file CSV. Data yang diperoleh sebanyak 502 *tweet*. Data tersebut kemudian dilakukan pembersihan (*data cleaning*) untuk menghilangkan duplikasi data. Data bersih yang diperoleh dari hasil data *cleaning* sebanyak 500 *tweet*, yang terdiri dari 337 *tweet* positif dan 163 *tweet* negatif.

3.3. Data Preparation (Preprocessing)

Proses data *preparation* dilakukan terhadap data 500 *tweet* tersebut dalam 6 tahapan, sebagaimana disebutkan dalam metodologi, menggunakan RapidMiner Studio Educational 9.10 dengan contoh hasil dapat dilihat pada tabel 4.1 sampai dengan tabel 4.6 berikut:

Tabel 4.1 Hasil dari *Transform Case*

<i>Text</i>	<i>Transform Case</i>
@detikcom Biarin aja doi pergi kan Ibu Kota Nusantara Dan biarkan kami tetap di Ibu Kota Indonesia Jangan balik lagi yak ??	@detikcom biarin aja doi pergi kan ibu kota nusantara dan biarkan kami tetap di ibu kota indonesia jangan balik lagi yak??

Tabel 4.2 Hasil dari *Remove HTTP*

<i>Text</i>	Hasil <i>Remove HTTP</i>
kegiatan apel gelar pasukan dalam rangka operasi Pengamanan vvip terkait kunjungan kerja presiden ri ke wilayah ibu kota nusantara di kecamatan sepaku kabupaten penajam paser utara kalimantan timur dilaksanakan dengan sukses. https://t.co/TusceKMCDy	kegiatan apel gelar pasukan dalam rangka operasi Pengamanan vvip terkait kunjungan kerja presiden ri ke wilayah ibu kota nusantara di kecamatan sepaku kabupaten penajam paser utara kalimantan timur dilaksanakan dengan sukses.

Tabel 4.3 Hasil dari *Remove @*

<i>Text</i>	Hasil dari <i>Remove @</i>
@Adiprayitno_20 ok resmi namanya berubah jd ikan.. ibu kota artis nusantara	ok resmi namanya berubah jd ikan.. ibu kota artis nusantara

Tabel 4.4 Hasil dari *Tokenize*

<i>Text</i>	Hasil dari <i>Tokenize</i>
persiapan sempurna untuk hut ri di ibu kota nusantara! mari rayakan bersama.	persiapan sempurna untuk hut ri di ibu kota nusantara mari rayakan bersama

Tabel 4.5 Hasil dari *Filter by Length*

<i>Text</i>	Hasil dari <i>Filter by Length</i>
boediantar4 islam nusantara itu kiblatnya ibu kota nusantara (ikn) jadi	boediantar4 islam nusantara itu kiblatnya ibu kota nusantara (ikn) jadi

haji gak harus ke mekkah cukup ke ikn	haji gak harus mekkah cukup ikn
--	--

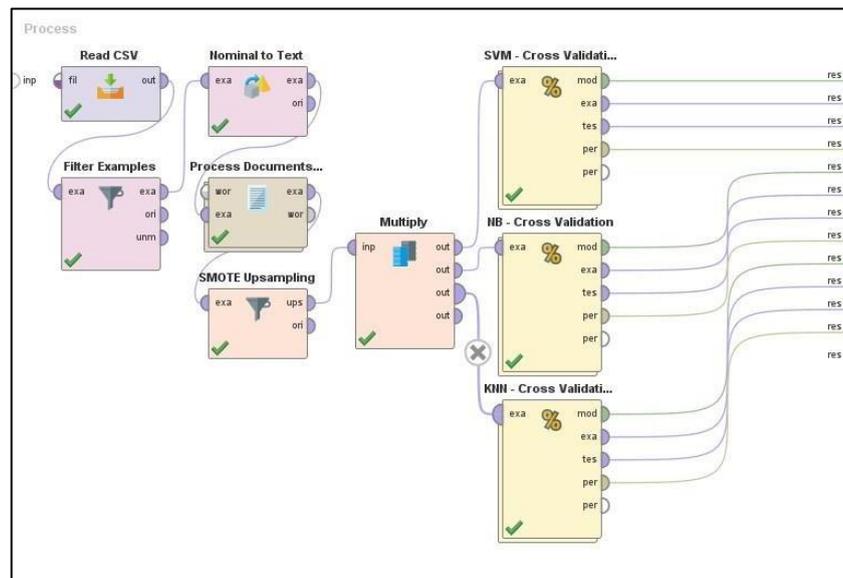
Tabel 4.6 Hasil dari *Remove Stopwords*

<i>Text</i>	Hasil dari <i>Remove Stopwords</i>
boediantar4 islam nusantara itu kiblatnya ibu kota nusantara (ikn) jadi haji gak harus ke mekkah cukup ke ikn	boediantar4 islam nusantara kiblatnya ibu kota nusantara (ikn) haji gak harus mekkah cukup ikn.

3.4. Modelling

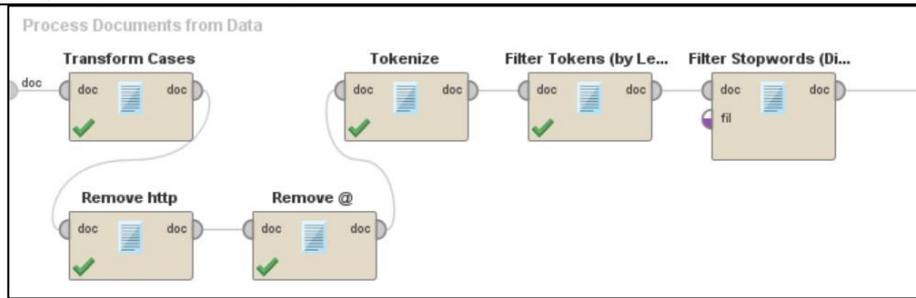
Proses Pada tahap ini melibatkan teknik data *mining* dengan algoritma klasifikasi menggunakan RapidMiner Studio Educational 9.10. Hasil pengujian model yang dilakukan adalah mengklasifikasi *tweet* positif dan *tweet* negatif menggunakan algoritma *K-Nearest Neighbor*, *Naive Bayes*, dan *Support Vector Machines* dan *feature selection* PSO.

Pertama, dilakukan perbandingan algoritma algoritma K-Nearest Neighbor, Naive Bayes, dan Support Vector Machines tanpa *feature selection* PSO. Desain model tersebut pada aplikasi yang digunakan dapat dilihat pada Gambar 4.2. Pada gambar tersebut dapat dijelaskan bahwa data yang berhasil diperoleh telah disimpan dalam sebuah file CSV yang akan dibaca oleh operator *Read Excel*. Sebelum diproses lebih lanjut, terhadap data tersebut dilakukan operasi SMOTE Upsampling untuk menyeimbangkan kelas data dengan menambah jumlah data kelas minor agar setara dengan kelas mayor. (Angelina Puput Giovanni, 2020) [5].



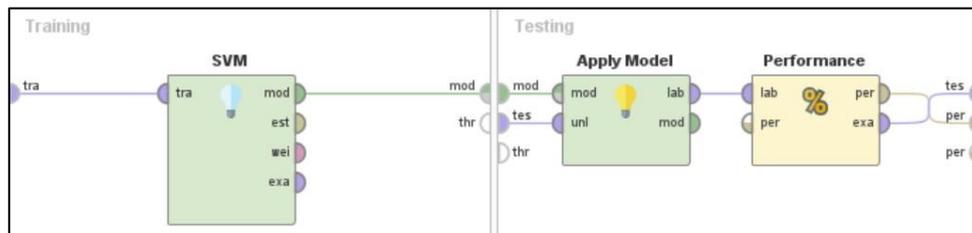
Gambar 4.2 Desain model perbandingan K-Nearest Neighbor, Naive Bayes, dan Support Vector Machines

Setelah data diseimbangkan, selanjutnya dilakukan operasi *Process Document*. Proses ini adalah tahap data *preparation* yang di dalamnya terdiri dari operator *transform cases*, *remove http*, *remove @*, *tokenize*, *filter tokens by length*, dan *Stopword Filter*, sebagaimana dapat dilihat pada Gambar 4.3 berikut.

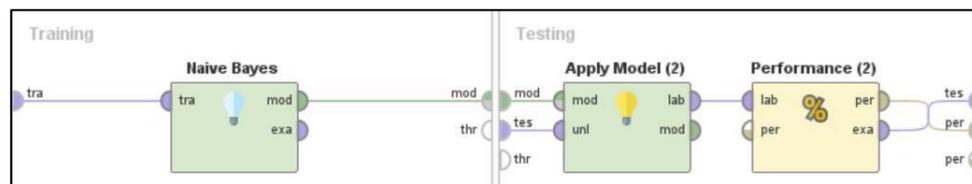


Gambar 4.3 Proses dokumen

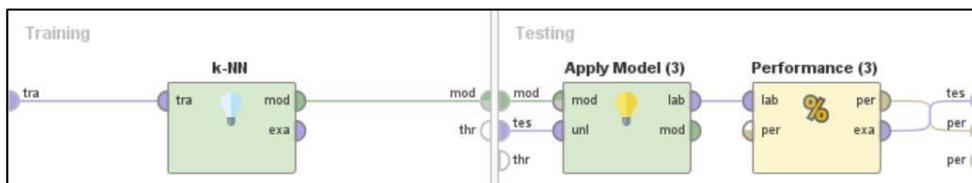
Selanjutnya, hasil proses tersebut digandakan menggunakan operator *Multiply* agar dapat digunakan dalam 3 algoritma pemodelan. Masing-masing pemodelan dilakukan menggunakan metode *10 fold cross validation* dengan operator *Cross Validation* sebagaimana dapat dilihat pada gambar 6 (SVM), gambar 7 (NB), dan gambar 8 (K-NN).



Gambar 4.4 Proses didalam Cross Validation- SVM



Gambar 4.5 Proses didalam Cross Validation -NB



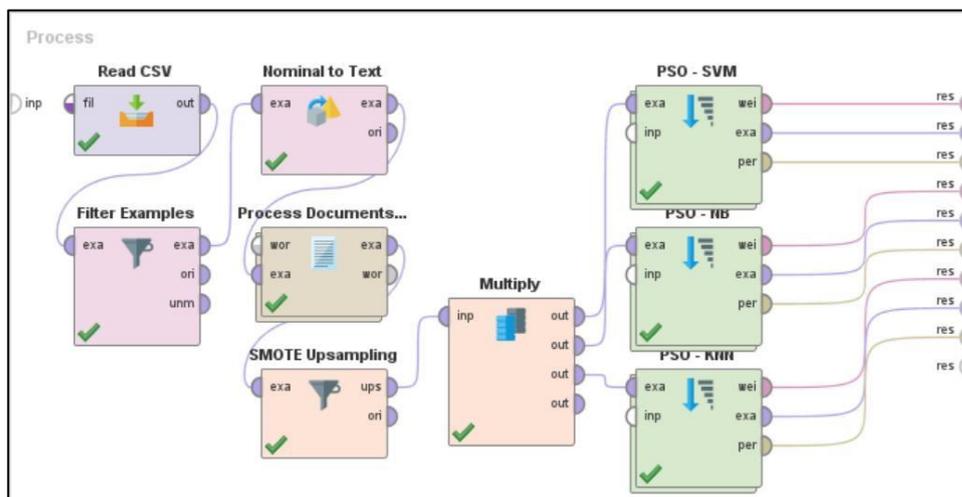
Gambar 4.6 Proses didalam Cross Validation-K-NN

3.5. Evaluation

Tahap evaluasi bertujuan untuk menentukan nilai kegunaan model yang telah berhasil dibuat pada langkah sebelumnya. Metode evaluasi yang digunakan dalam penelitian ini adalah metode *10-fold cross validation*. Tingkat akurasi dan *ROC Curve*

dari masing-masing algoritma diklasifikasikan berdasarkan nilai AUC sebagai berikut: 0,90-1,00 = sangat baik. 0,80-0,90 = baik. 0,70-0,80 = cukup. 0,60-0,70 = buruk. 0,50-0,60 = gagal. Hasil pengujian akurasi dari algoritma NB, SVM, K-NN tanpa menggunakan *feature selection* PSO dapat dilihat pada gambar 4.8 sampai dengan gambar 4.10 berikut.

Berikutnya, dilakukan perbandingan algoritma K-Nearest Neighbor, Naive Bayes, dan Support Vector Machines yang menggunakan *feature selection* PSO. Proses diawali dengan membaca data dari *file* dengan operator *read* Excel hingga penggandaan data dengan operator *Multiply*, memiliki proses yang sama seperti model desain sebelumnya. Sementara operator PSO ditempatkan setelah operator *Multiply* sebagaimana dapat dilihat pada gambar 4.7 berikut.



Gambar 4.7 Desain Model Perbandingan NB (PSO), SVM (PSO), dan K-NN (PSO)

Setelah dilakukan *feature selection* PSO, langkah selanjutnya adalah menjalankan masing-masing algoritma pemodelan menggunakan metode *10 fold cross validation* dengan operator *Cross Validation* sebagaimana proses sebelumnya pada gambar 4.8, gambar 4.9, dan gambar 4.10.

accuracy: 85.75% +/- 4.56% (micro average: 85.76%)			
	true Negatif	true Positif	class precision
pred. Negatif	324	83	79.61%
pred. Positif	13	254	95.13%
class recall	96.14%	75.37%	

Gambar 4.8 Tingkat accuracy algoritma NB

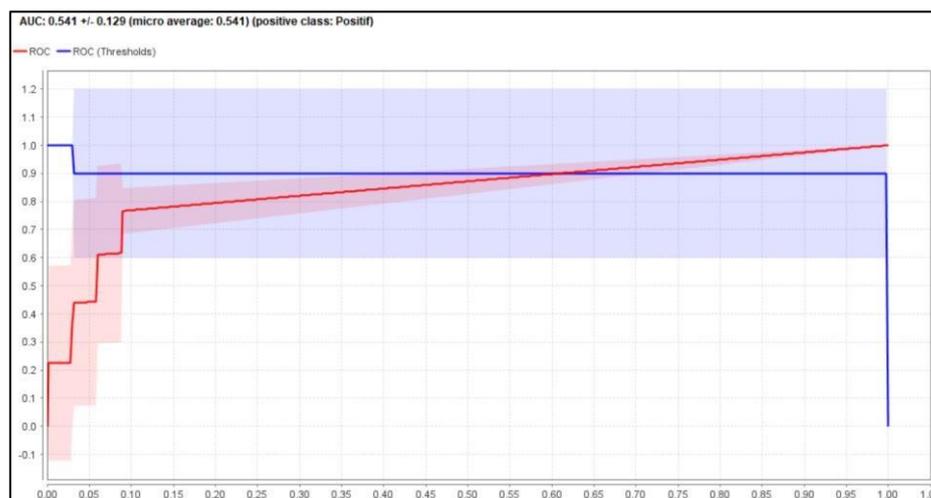
accuracy: 75.53% +/- 4.58% (micro average: 75.52%)			
	true Negatif	true Positif	class precision
pred. Negatif	337	165	67.13%
pred. Positif	0	172	100.00%
class recall	100.00%	51.04%	

Gambar 4.9 Tingkat accuracy algoritma SVM

accuracy: 81.18% +/- 6.84% (micro average: 81.16%)			
	true Negatif	true Positif	class precision
pred. Negatif	223	13	94.49%
pred. Positif	114	324	73.97%
class recall	66.17%	96.14%	

Gambar 4.10 Tingkat accuracy algoritma K-NN

Hasil pengujian performa dalam bentuk *ROC Curve* dari algoritma NB, SVM, K-NN tanpa menggunakan *feature selection* PSO dapat dilihat pada gambar 4.11 sampai dengan gambar 4.13 berikut. Pada gambar *ROC Curve* algoritma NB menampilkan hasil AUC 0,541 yang artinya masuk ke kategori klasifikasi Gagal.



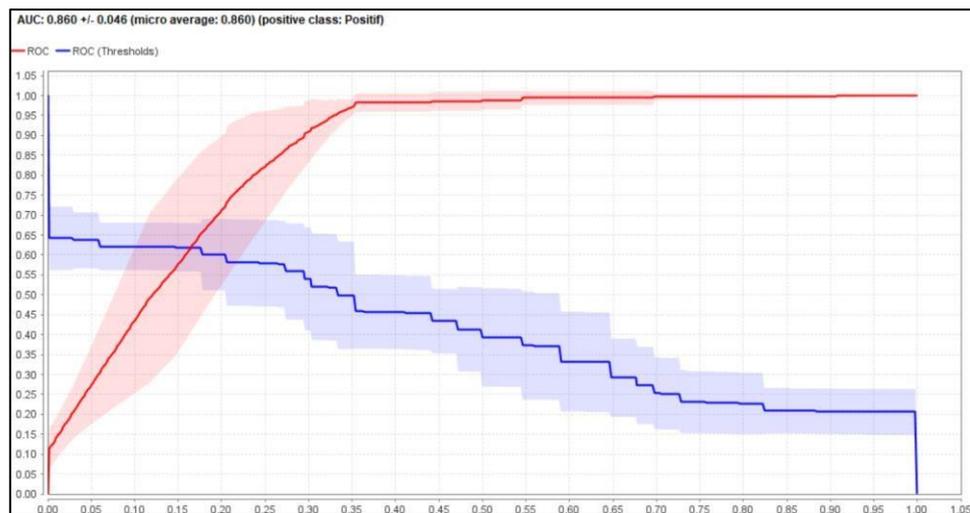
Gambar 4.11 ROC Curve algoritma NB

Pada gambar *ROC Curve* algoritma SVM menampilkan hasil AUC 0,966 yang artinya masuk ke kategori klasifikasi Sangat Baik.



Gambar 4.12 ROC Curve algoritma SVM

Pada gambar *ROC Curve* algoritma K-NN menampilkan hasil AUC 0,860 yang artinya masuk ke kategori klasifikasi Baik.



Gambar 4.13 ROC Curve algoritma KNN

Dengan demikian, hasil akurasi dan nilai AUC dari algoritma NB, SVM, K-NN tanpa menggunakan *feature selection* PSO dapat dilihat pada tabel 4.7 berikut.

Tabel 4.7 Akurasi dan AUC algoritma NB, SVM, dan K-NN tanpa PSO

Algoritma	Accuracy	AUC	Klasifikasi
NB	85,76%	0,541	Gagal
SVM	75,52%	0,966	Sangat Baik
K-NN	81,16%	0,860	Baik

Berdasarkan hasil tersebut, dapat diketahui bahwa tanpa menggunakan *feature selection* Particle Swarm Optimization, algoritma Support Vector Machines memiliki performa yang terbaik jika dibandingkan algoritma Naive Bayes dan K-Nearest Neighbor dalam penelitian ini.

Hasil pengujian akurasi dari algoritma Naive Bayes, Support Vector Machines, dan K-Nearest Neighbor dengan menggunakan *feature selection* Particle Swarm Optimization dapat dilihat pada gambar 4.14 sampai dengan gambar 4.16 berikut.

accuracy: 98.52%			
	true Negatif	true Positif	class precision
pred. Negatif	337	10	97.12%
pred. Positif	0	327	100.00%
class recall	100.00%	97.03%	

Gambar 4.14 Tingkat accuracy algoritma NB (PSO)

accuracy: 97.63%			
	true Negatif	true Positif	class precision
pred. Negatif	336	15	95.73%
pred. Positif	1	322	99.69%
class recall	99.70%	95.55%	

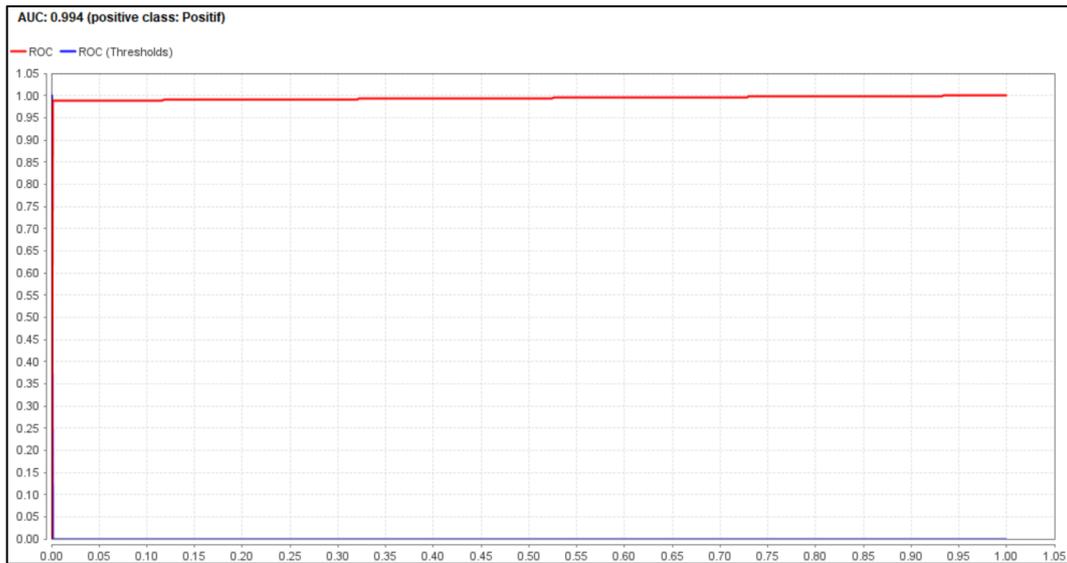
Gambar 4.15 Tingkat accuracy algoritma SVM (PSO)

accuracy: 97.18%			
	true Negatif	true Positif	class precision
pred. Negatif	330	12	96.49%
pred. Positif	7	325	97.89%
class recall	97.92%	96.44%	

Gambar 4.16 Tingkat accuracy algoritma KNN (PSO)

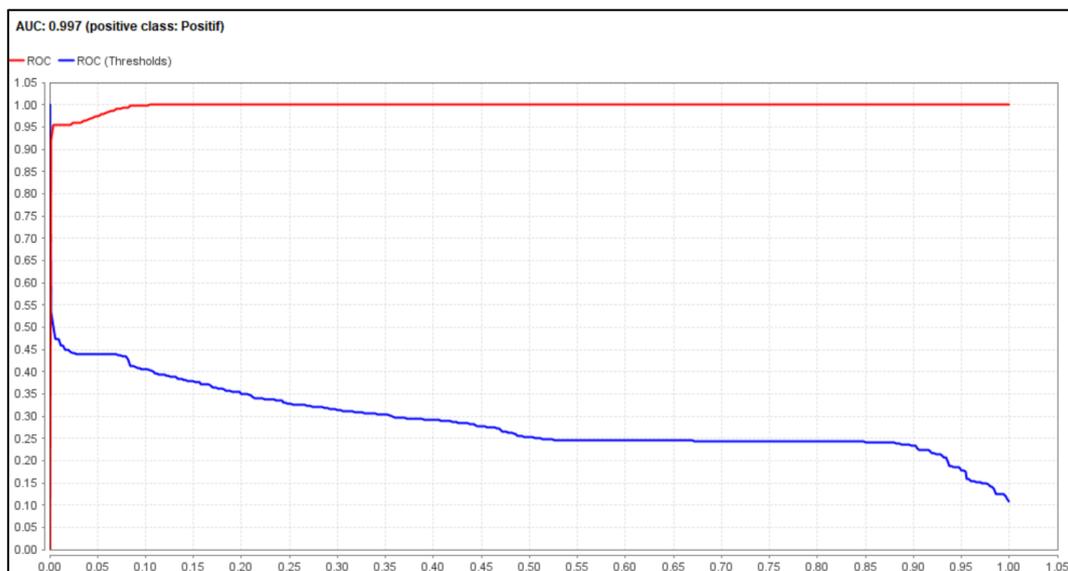
Hasil pengujian performa dalam bentuk *ROC Curve* dari algoritma NB, SVM, K- NN tanpa menggunakan *feature selection* PSO dapat dilihat pada gambar 4.17 sampai dengan gambar 4.19 berikut.

Pada gambar *ROC Curve* algoritma NB menggunakan PSO menampilkan hasil AUC 0,994 yang artinya masuk ke kategori klasifikasi Sangat Baik.



Gambar 4.17 ROC Curve algoritma NB (PSO)

Pada gambar *ROC Curve* algoritma SVM menggunakan PSO menampilkan hasil AUC 0,997 yang artinya masuk ke kategori klasifikasi Sangat Baik.



Gambar 4.18 ROC Curve algoritma SVM (PSO)

Pada gambar *ROC Curve* algoritma K-NN menggunakan PSO menampilkan hasil AUC 0,995 yang artinya masuk ke kategori klasifikasi Sangat Baik.



Gambar 4.19 ROC Curve algoritma KNN (PSO)

Dengan demikian, hasil akurasi dan nilai AUC dari algoritma NB, SVM, K-NN dengan menggunakan *feature selection* PSO dapat dilihat pada tabel 4.8 dan gambar 4.21 berikut.

Tabel 4.8 Akurasi dan AUC algoritma NB, SVM, dan K-NN menggunakan PSO

Algoritma	Accuracy	AUC	Klasifikasi
NB (PSO)	98,52%	0,994	Sangat Baik
SVM (PSO)	97,63%	0,997	Sangat Baik
K-NN (PSO)	97,18%	0,995	Sangat Baik

Gambar 4.21 Compare ROC menggunakan PSO

Berdasarkan hasil tersebut, dapat diketahui bahwa jika menggunakan *feature selection* PSO, maka algoritma Support Vector Machines memiliki performa yang terbaik jika dibandingkan algoritma Naive Bayes dan K-Nearest Neighbor dalam penelitian ini.

4. KESIMPULAN

Penelitian yang dilakukan terhadap 500 data *tweet* dengan hasil *crawling* data menggunakan *tweet harvest*. Dengan mengacu pada pertanyaan dari penelitian yang dirumuskan pada Bab 1 dalam tesis ini, maka mendapatkan kesimpulan sebagai berikut:

1. Penelitian dilakukan dari hasil *crawling* data sentimen twitter menggunakan *tweet harvest*. Terdapat 502 data tweet yang didapat dan dilakukan

data

preprocessing sejumlah 500 valid. Dari jumlah data tersebut didapatkan pelabelan dengan jumlah 337 tweet positif dan 163 *tweet* negatif.

2. Berdasarkan data yang didapat dengan jumlah 502 data sentimen tentang Ibu Kota Nusantara selanjutnya dilakukan *preprocessing* data, dan dilakukan pembobotan menggunakan TF-IDF. Kemudian diuji menggunakan *Cross Validation* dan *Particle Swarm Optimization*. Sehingga didapatkan nilai *accuracy* dan *AUC*nya.
3. Berdasarkan data yang sudah diuji menggunakan aplikasi Rapidminer dapat disimpulkan bahwa penggunaan *feature selection* PSO dalam algoritma klasifikasi dapat meningkatkan performa dan akurasi. Hasil dari perbandingan tanpa PSO dapat dilihat dari bentuk *confusion matrix* dan ROC dengan algoritma Naive Bayes mendapatkan *accuracy* 98,52% dan *AUC* 0,541, algoritma Support Vector Machine mendapatkan *accuracy* 75,52% dan *AUC* 0,966, dan algoritma K-Nearest Neighbor mendapatkan *accuracy* 81,16% dan *AUC* 0,860. Sedangkan perbandingan menggunakan PSO dapat dilihat dari bentuk *confusion matrix* dan ROC dengan algoritma Naive Bayes mendapatkan *accuracy* 98,52% dan *AUC* 0,994, algoritma Support Vector Machine mendapatkan *accuracy* 97,63% dan *AUC* 0,997, dan algoritma K-Nearest Neighbor mendapatkan *accuracy* 97,18% dan *AUC* 0,995. Sehingga algoritma Support Vector Machine menggunakan PSO memiliki performa yang terbaik jika dibandingkan dengan NB, SVM, K-NN, NB (PSO) dan K-NN (PSO).

5. DAFTAR PUSTAKA

- [1] Sang Intan Risqi Adi, B. B. (2024). Analisis Sentimen Masyarakat Terhadap Progress IKN Menggunakan Model Decision Tree. Tangerang, Januari 2024, Vol 8 No.1, , 57- 65.
- [2] Liu, H. T. (2013). Electrical Power and Energy Systems An experimental investigation of two Wavelet-MLP hybrid frameworks for wind speed prediction using GA and PSO optimization. International Journal of Electrical Power & Energy Systems, 161-173.

-
- [3] Saleh, A. (2015). Implementasi Naive Bayes . Journal of Informatics, Information System, Software Engineering and Applications (INISTA), 73-81.
 - [4] Naik, A. &. (2016). Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime. Procedia Computer Science, 662–668.
 - [5] Angelina Puput Giovani, A. T. (2020). ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI . Jurnal TEKNOINFO, 116-124.