Jurnal KomtekInfo

https://jkomtekinfo.org/ojs

2023 Vol. 10 No. 1 Hal: 1-7 p-ISSN: 2356-0010, e-ISSN: 2502-8758

Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN

Syahril Dwi Prasetyo[1]✉, Shofa Shofiah Hilabi[2], Fitri Nurapriani[3]

[1,2,3] Prodi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Buana Perjuangan Karawang

Si19.syahrilprasetyo@mhs.ubpkarawang.ac.id

Abstract

The relocation and construction of the Archipelago National Capital during the President's era, Mr. Ir. H. Joko Widodo will be relocated in stages from 2024 to 2045. With this, it becomes a conversation and invites a lot of reactions, especially for the Indonesian people. The issue of relocating the national capital is a sensitive matter, so it is widely discussed on social media, including Twitter. Social media is used to convey opinions or expression. Utilization of social media is of course a service and facility that can be utilized for a political issue or matters being discussed. Therefore this study aims to analyze the sentiments of the Indonesian people toward the relocation of the Archipelago's Capital City. In this study, the methods used were Naïve Bayes (NB) and K-Nearest Neighbor (KNN). The results of the study present the results of the comparative performance of these methods that the Naïve Bayes method provides an accuracy level of sentiment analysis of 82.27%, a Precision value of 86.36%, and a Recall value of 76.93%. The performance of the KNN method also presents analysis results with an accuracy rate of 88.12%, a Precision of 93.98, and a recall value of 81.53%. Based on the results of this analysis, the analysis process using the KKN method outperforms the NB method in measuring sentiment toward the relocation of the Archipelago's Capital City.

Keywords: Sentiment Analysis, Capital City of the Archipelago, K-Nearest Neighbor (KNN), Naive Bayes (NB), Comparison

Abstrak

Pemindahan serta pembangunan Ibu Kota Negara Nusantara pada masa Presiden Bapak Ir. H. Joko Widodo akan direlokasi secara bertahap dari tahun 2024 hingga 2045. Dengan hal ini tersebut maka menjadi sebuah perbincangan dan mengundang banyak reaksi, terutama bagi masyarakat Indonesia. Persoalan dalam Pemindahann Ibu Kota Negara merupakan hal yang sensitif sehingga ramai diperbincangkan di media sosial termasuk Twitter. Pada dasarnya media sosial digunakan untuk menyampaikan pendapat atau sebuah ekspresi. Pemanfaatan media sosial ini tentunya menjadi layanan dan fasilitas yang dapat dimanfaatkan untuk sebuah isu politik atau hal-hal yang sedang dibahas. Maka dari itu penelitian ini bertujuan untuk melakukan analisis sentimen masyarakat Indonesia terhadap pemindahan Ibu Kota Nusantara. Dalam penelitian ini metode yang digunakan adalah Naïve Bayes (NB) dan K-Nearest Neighbor (KNN). Hasil penelitian menyajikan hasil komparasi kinerja metode tersebut bahwa metode Naïve Bayes memberikan tingkat akurasi analisis sentimen sebesar 82.27%, nilai Precision sebesar 86.36% dan nilai Recall sebesar 76.93%. Kinerja metode KNN juga menyajikan hasil analisis dengan tingkat akurasi sebesar 88,12%, Precision sebesar 93.%98 dan nilai recall sebesar 81.53%. Berdasarkan hasil analisis tersebut maka proses analisis menggunakan metode KKN mengungguli metode NB dalam mengukur sentimen terhadap pemindahan Ibu Kota Nusantara.

Kata kunci: Sentimen Analisis, Ibu Kota Nusantara, K-Nearest Neighbor (KNN), Naive Bayes (NB), Komparasi

1. Pendahuluan

Dalam upaya pemindahan ibu kota Nusantara telah lama menjadi wacana pemerintah, Tetapi tepatnya pada tahun 2017 upaya tersebut kembali dimunculkan oleh kementrian [1]. Mengenai pemindahan Ibu Kota Negara Nusantara, yakni pada masa Presiden Bapak Ir. H. Joko Widodo, Pembangunan Ibu Kota Nusantara pada pertengahan Maret 2022 akan direlokasi secara bertahap dari tahun 2024 hingga 2045 [2]. Tepat pada Senin, 26 Agustus 2019, Melalui siaran pers, lokasi baru IKN berada di Kabupaten Penajam Paser Utara Provinsi Kalimantan Timur, tepatnya di sebagian Kabupaten Kutai Kartanegara [3]. Dalam hal tersebut adanya pemindahan ibu kota negara Indonesia, tentu mengundang berbagai reaksi, terutama bagi masyarakat Indonesia. Mengingat ibu kota negara baru di Indonesia merupakan hal yang sensitif sehingga ramai diperbincangkan di media sosial termasuk Twitter.

Twitter adalah suatu media dan layanan sosial microblogging yang populer di kalangan pengguna internet. Penggunanya dapat mengekspresikan apa yang dipikirkan dalam pesan waktu nyata [4]. Pengguna Twitter bebas menyampaikan pendapat atau ekspresi mereka tentang layanan, fasilitas atau isu politik atau hal-hal yang sedang dibahas [5].

Dalam topik analisis sentimen bertujuan untuk mengklasifikasikan teks dalam sebuah kalimat [6].

Diterima: 05-01-2023 | Revisi: 25-01-2023 | Diterbitkan: 30-03-2023 | doi: 10.35134/komtekinfo.v10i1.330

Aspek, atau data dokumen digunakan untuk menentukan opini yang terkandung dalam kalimat atau dokumen memiliki sentien positif atau negatif [7]. Proses tersebut dapat dilakukan dengan menggunakan model analisis teks mining.

Pada sisi lain teks mining dapat bekerja dalam komputer dengan tujuan memproses informasi lama secara eksplisit untuk menghasilkan penemuan informasi baru [8]. Hal ini sejalan dengan definisi data mining yang menambang sebuah sumber data berupa teks tertentu. Biasanya data didapatkan dari dokumen sehingga bisa menganalisa keterhubungan antar dokumen [9]. Penelitian Analisis Sentimen ini tentunya semakin berkembang di bidang text mining dalam berbagai publikasi jurnal menggunakan Naïve Bayes classifier dan KNN yang telah dilakukan oleh berbagai peneliti sebelumnya [10].

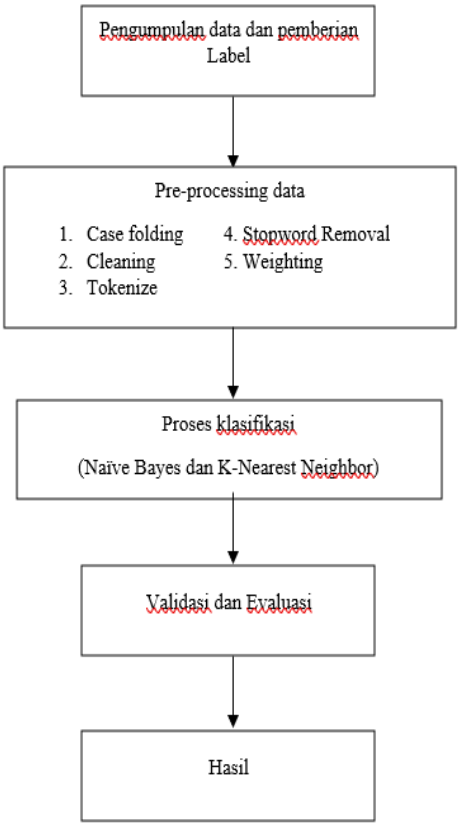
Penelitian terdahulu dengan topik yang terkait diantaranya analisis dari twitter terhadap calon Presiden Indonesia 2019 dengan menggunakan data berjumlah 1500 dan menggunakan teknik pembelajaran mesin yaitu Naïve Bayes Classifier yang memperoleh hasil akurasi sebesar 64.6% pada paslon 1, dan 58% untuk paslon 2 [11]. Penelitian yang lainnya dalam pembahasan opini publik Covid-19 pada Twitter menggunakan metode Naïve Bayes dan KNN. Dengan menggunakan metode NB, Mendapatkan nilai akurasi sebesar 63.21% sedangkan nilai KNN sebesar 58.10% yang berarti metode NB paling akurat dibandingkan metode KNN [12]. Penelitian tentang Sistem Analisis Sentimen pada ulasan produk dengan menggunakan total 1500 data dari femaledaily.com dan nilai akhir sebesar 77.78%, data tersebut kasifikasi menggunakan Naive Bayes [13],[14].

Penelitian dalam proses analisis sentimen masyarakat Indonesia terhadap pemindahan Ibu Kota Nusantara menggunakan teks mining dengan mengadopsi algoritma metode Naïve Bayes dan KNN. Algoritma tersebut merupakan metode analisis klasifikasi yang telah banyak digunakan [15]. Algoritma memiliki keunggulan dan kelemahan dalam mengkasifikasikan data masing-masing, diantara Naïve Bayes Classifier dan Metode K-Nearest Neighbor algoritma mana yang paling tepat dalam mengklasifikasikan data [16].

Berdasarkan penjelasan sebelumnya maka penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap pemindahan ibu kota Indonesia. Penelitian ini akan melakukan komparasi algoritma NB dan KNN dalam melakukan klasigikasi. Kinerja algoritma tersebut dapak 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. Metodologi Penelitian

Dalam penelitian ini metode klasifikasi yang digunakan yaitu Naïve Bayes dan K-Nearest Neighbor, Sebenarnya masih banyak metode klasifikasi tetapi pada penelitian ini hanya fokus pada dua metode, Untuk mendapatkan hasil yang terbaik, diperlukan beberapa tahapan, yang meliputi pengumpulan data dan pelabelan data, dilanjutkan dengan preprocessing [17],[18]. Data yang telah diproses sebelumnya akan diklasifikasikan menggunakan algoritma Naïve Bayes dan KNN [19]. Berikut tahap yang dapat dilihat:



Gambar 1. Tahapan Penelitian

Pada gambar 1 merupakan tahap dalam penelitian yang dimana tahap awal yaitu pengumpulan data dan pemberian label pada data set. Kemudian pada tahap selanjutnya yaitu memproses/membersihkan data untuk digunakan dalam penelitian. Selanjutnya proses klasifikasi, evaluasi dan validasi merupakan tahap akhir untuk mendapatkan nilai perbandingan.

2.1 Naïve Bayes

Naive Bayes merupakan metode klasifikasi sederhana yang menghitung semua probabilitas berdasarkan teorema Bayes yang dikombinasikan dengan kombinasi nilai frekuensi database. Dalam tugas klasifikasi, tujuannya adalah untuk memprediksi label kelas dari sampel yang diberikan berdasarkan sekumpulan fitur atau karakteristik. Misalnya, berdasarkan data tentang usia, jenis kelamin, dan

pendapatan seseorang, tujuannya mungkin untuk memprediksi apakah mereka cenderung memiliki pekerjaan bergaji tinggi atau tidak [20].

2.2 K-Nearest Neighbor

Algoritma KNN adalah algoritma pembelajaran terawasi yang dapat digunakan untuk tugas klasifikasi dan regresi. Ini bekerja dengan menemukan K titik data terdekat ke sampel yang diberikan, dan menggunakan label kelas atau nilai dari titik data ini untuk membuat prediksi tentang label kelas atau nilai sampel. KNN dapat melakukan prosedur berbasis matematika untuk mengevaluasi nilai-nilai kriteria tersebut ke dalam klasifikasi. Dalam tugas klasifikasi, KNN bekerja dengan mencari titik data K dalam set pelatihan yang paling dekat dengan sampel yang sedang diklasifikasikan, berdasarkan ukuran jarak seperti jarak Euclidean. Ini kemudian menetapkan sampel ke label kelas yang paling umum di antara K tetangga terdekat [21].

2.3 Pengumpulan Data

Pada langkah awal penelitian, para peneliti mengidentifikasi query tweet untuk kumpulan data. Query tweet yang digunakan dalam penelitian ini adalah ibukota atau ikn nusantara. Dalam penelitian ini, data berasal dari layanan microblogging Twitter. Para peneliti menggores data menggunakan metode pengumpulan tweet yang disediakan oleh API Twitter dan menyimpannya dalam format csv [22].

2.4. Pre-Processing

Pra-pemrosesan adalah proses menyiapkan data mentah untuk dianalisis atau digunakan dalam model pembelajaran mesin. Ini adalah langkah penting dalam proses pembelajaran mesin, karena kualitas dan karakteristik data dapat memengaruhi performa model secara signifikan. Proses tahap processing dilakukan guna pengolahan data mentah menjadi set data yang sudah jadi untuk memilih data yang akan diproses dalam dokumen. Untuk memaksimalkan output preprocessing dari penelitian sebelumnya, penelitian ini melakukan beberapa tahapan pre-processing yaitu [23]:

a. Case folding : Sebuah proses untuk mengedit teks dokumen ke dalam bentuk lower case
b. Cleaning : Data yang digunakan perlu adanya proses pembersihan data seperti symbol tautan URL, angka. Data twitter sendiri tentunya banyak data kotor seperti tagar, angka, nama pengguna, URL dan teks retweet
c. Tokenize : Proses data yang sebelumnya kalimat kemudian dipecah menjadi kata perkata
d. Stopword Removal : Kata yang terdapat di dalam stoplist akan melalui tahap pembersihan
e. Weighting : dalam proses ini untuk pembobotan kata dengan TF-IDF.

2.5. Klasifikasi

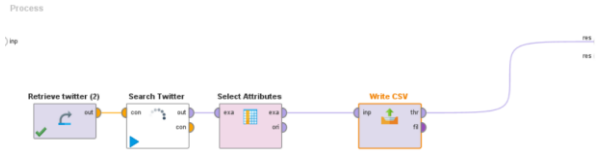
Dalam proses pembelajaran mesin, klasifikasi adalah tugas memprediksi label kelas dari sampel yang diberikan berdasarkan sekumpulan fitur atau karakteristik. Misalnya, diberikan data tentang usia, jenis kelamin, dan pendapatan seseorang, tujuan tugas klasifikasi mungkin untuk memprediksi apakah mereka cenderung memiliki pekerjaan bergaji tinggi atau tidak.Ada banyak algoritma berbeda yang dapat digunakan untuk klasifikasi, termasuk algoritma NB dan KNN, Algoritme ini menggunakan pendekatan berbeda untuk belajar dari data pelatihan dan membuat prediksi tentang label kelas dari sampel baru. Untuk mengevaluasi performa algoritme klasifikasi, pendekatan umum adalah membagi data menjadi set pelatihan dan set pengujian, dan menggunakan set pelatihan untuk melatih model dan set pengujian untuk mengevaluasi kinerjanya. Ini dapat dilakukan dengan menggunakan metrik seperti akurasi, presisi, dan daya ingat, yang mengukur proporsi prediksi yang benar yang dibuat oleh model.

2.6. Validasi dan Evaluasi

Pada tahap ini data yang telah dilakukan pembobotan akan divalidasi menggunakan K-fold Cross Validation. Dengan melakukan pengelompokan antara data uji dan data latih kemudian data akan dilakukan pengujian sebanyak mungkin. Lalu Pada tahap evaluasi bertujuan untuk mengetahui nilai dari Accuracy, Precision dan Recall. Tahap ini menggunakan metode Cross Validation dengan nilai k-folds [24].

3. Hasil dan Pembahasan

Pada penelitian analisis sentimen ini menggunakan data tweet hasil crawling menggunakan Rapid Miner dengan jumlah 800 data yang sudah melalui proses pembersihan data. Dalam mengenai pemindahan Ibu Kota Negara yaitu Nusantara, dengan menggunakan metode Naive Bayes Classifier yang akan dibandingkan dengan Algoritma K-Nearest Neighbors (KNN). Tujuannya untuk mengetahui hasil perbandingan keberhasilan nilai accuracy, precision, dan recall. dengan beberapa tahap Preprocessing, Validasi dan Evaluasi pada proses preprocessing serta untuk memberikan label sentimen mengenai data tweet yang digolongkan menjadi label sentimen positif dan negatif. Berikut proses pengambilan data pada twitter pada Gambar 2.



Gambar 2. Proses Pengambilan Data

Gambar 2. Merupakan proses pengambilan data yang dimana pada tahap pertama membuat retrieve koneksi

Twitter, lalu kemudian disambungkan ke operator Search Twitter, Dalam parameter Search Twitter masukan kata kunci dan jumkah data yang ingin diambil mengenai topik tweet yaitu pemindahan ibu kota dengan data yangg diambil sebanyak 2000 data. Selanjutnya yaitu pada operator Select Attributes untuk memilih atribut sebagai data.

3.1. Preprocessing

Pra-pemrosesan dapat menjadi proses berulang, karena pilihan langkah pra-pemrosesan dapat bergantung pada karakteristik data dan persyaratan khusus dari algoritme pembelajaran mesin. Penting untuk mempertimbangkan dengan hati-hati langkah-langkah pra-pemrosesan yang sesuai untuk data dan masalah yang dihadapi, karena langkah-langkah ini dapat berdampak signifikan terhadap performa model. Tujuan dari tahap pengolahan data ini adalah untuk memudahkan analisis data pada tahap selanjutnya. Pada fase ini dilakukan pembersihan dan persiapan data agar data yang akan dianalisis lebih terorganisir dan siap dilakukan penelitian.

a. Case Folding

Case Folding adalah langkah pra-pemrosesan yang melibatkan konversi semua karakter dalam string ke huruf besar-kecil, seperti huruf kecil atau huruf besar. Langkah ini sering digunakan saat bekerja dengan data teks, karena dapat membantu mengurangi dimensi data dengan mengurangi jumlah kata atau token unik. Pada Case Folding menggunakan Transform Case dari Rapid Miner, Operator ini merubah semua karakter dalam teks menjadi huruf kecil pada dokumen teks yang akan digunakan.

Tabel 1. Proses Case Folding

Sebelum Case Folding	Sesudah Case Folding
Kami Mahasiswa Jawa Timur sangat dukung Pembangunan IKN karena manfaat nya.	kami mahasiswa jawa timur sangat dukung pembangunan ikn karena manfaat nya.
Untuk pak sekedar saran dari Rakyatmu sebaiknya isu Resesi jangan terlalu dibesarkan sehingga menakutnakuti Rakyatmu	untuk pak sekedar saran dari rakyatmu sebaiknya isu resesi jangan terlalu dibesarkan sehingga menakutnakuti rakyatmu

Berikut merupakan sebuah hasil sebelum dan sesudah dilakukan case folding. Yang dimana karakter pada tabel sebelum case folding merupakan text asli. Kemudian menghasilkan karakter dalam string ke huruf besar-kecil. Kasus lipat dapat berguna dalam situasi di mana kasus teks tidak menyampaikan makna apapun, seperti dalam tugas pengolahan bahasa alami. Namun, itu mungkin tidak sesuai dalam situasi di mana kasus teks itu penting, seperti ketika menganalisis kata benda atau akronim yang tepat.

b. Tokenizing

Tokenisasi adalah proses memecah sepotong teks menjadi token individu, yang biasanya berupa kata-kata atau tanda baca. Tokenisasi adalah langkah pra-pemrosesan umum saat bekerja dengan data teks, karena memungkinkan teks diproses dan dianalisis dengan lebih mudah. Tahap Tokenizing menggunakan operator tokenize dari Rapid Miner yang digunakan untuk memotongan proses data yang sebelumnya kalimat kemudian dipecah menjadi kata perkata.

Tabel 2. Proses Tokenizing

Sebelum Tokenizing	Sesudah Tokenizing
kami mahasiswa jawa timur sangat dukung pembangunan ikn karena manfaat nya	‘kami’ ’mahasiswa ’ ’jawa ’ ’timur ’ ’sangat ’ ’dukung ’ ’pembangunan ’ ’ikn ’ ’karena ’ ’manfaat ’ ’nya ’
untuk pak sekedar saran dari rakyatmu sebaiknya isu resesi jangan terlalu dibesarkan sehingga menakutnakuti rakyatmu	‘untuk’ ’pak ’ ’sekedar ’ ’saran ’ ’dari’ ’rakyatmu ’ ’sebaiknya’ ’isu ’ ’resesi ’ ’jangan ’ ’terlalu ’ ’dibesarkan’ ’sehingga’ ’menakutnakuti’ ’rakyatmu’

Berikut merupakan sebuah hasil sebelum dan sesudah dilakukan tokenizing. Yang dimana pada tabel 2 merupakan hasil pemecahan teks menjadi kata-kata individual. Ini adalah bentuk tokenisasi yang paling umum, dan sering digunakan sebagai titik awal untuk tugas seperti pemodelan bahasa atau klasifikasi teks.

c. Stopword Removal

Stopword Removal adalah langkah pra-pemrosesan yang melibatkan penghapusan kata-kata umum dari sepotong teks yang tidak dianggap penting untuk makna teks. Kata-kata ini, yang dikenal sebagai stopwords, mencakup kata-kata seperti "a", "and", "the", dan "but", dan sering digunakan untuk menghubungkan klausa atau kalimat. Stopwords biasanya dihapus dari data teks sebelum diproses atau dianalisis, karena kata-kata tersebut tidak memberikan banyak makna pada konten teks dan dapat meningkatkan dimensi data secara tidak perlu. Menghapus stopword dapat membantu mengurangi beban komputasi pemrosesan teks dan dapat meningkatkan performa beberapa algoritma pembelajaran mesin.

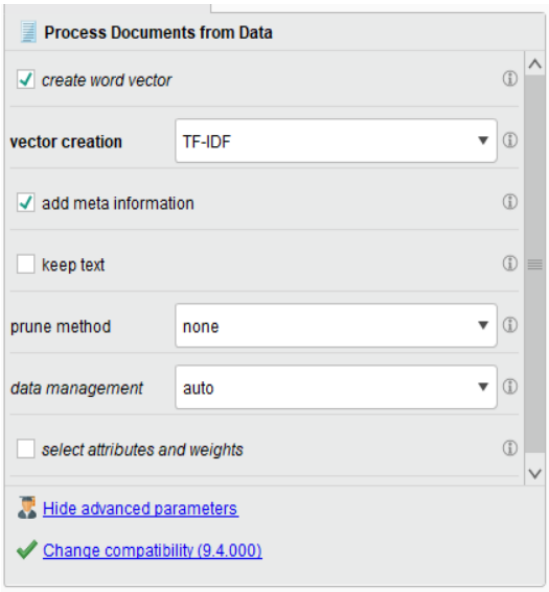
Tabel 3. Stopword Removal

Sebelum Stopword Removal	Sesudah Stopword Removal
kami mahasiswa jawa timur sangat dukung pembangunan ikn karena manfaat nya.	mahasiswa jawa timur dukung pembangunan ikn manfaat
untuk pak sekedar saran dari rakyatmu sebaiknya isu resesi jangan terlalu dibesarkan sehingga menakutnakuti rakyatmu	sekedar saran rakyatmu isu resesi dibesarkan menakutnakuti rakyatmu

Berikut merupakan sebuah hasil sebelum dan sesudah dilakukan Stopword Removal. Yang dimana pada tabel 3 merupakan hasil penghapusan kata-kata yang dikenal tidak penting. Penting untuk diperhatikan bahwa stopword mungkin penting dalam konteks tertentu, seperti saat menganalisis struktur gramatikal kalimat atau saat frekuensi stopword digunakan sebagai fitur dalam model pembelajaran mesin. Dalam kasus ini, penghapusan stopword mungkin.

d. Weighting

Pembobotan adalah teknik yang memberikan bobot atau kepentingan untuk setiap fitur dalam dataset. Ini dapat berguna dalam pembelajaran mesin, karena memungkinkan model untuk memberikan lebih banyak atau lebih sedikit penekanan pada fitur tertentu saat membuat prediksi. Fitur pembobotan dapat berguna dalam situasi di mana fitur tertentu lebih penting atau relevan untuk tugas yang sedang dikerjakan. Ini juga dapat membantu meningkatkan kinerja beberapa algoritma pembelajaran mesin, dengan memberi model lebih banyak informasi tentang kepentingan relatif setiap fitur. Namun, penting untuk mempertimbangkan dengan hati-hati skema pembobotan yang sesuai untuk data dan masalah yang dihadapi, karena pilihan bobot dapat berdampak signifikan terhadap performa model terlihat pada Gambar 3.

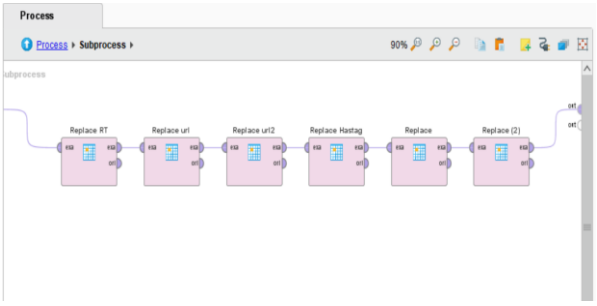


Gambar 3. Proses TF-IDF

TF-IDF (term frequency-inverse document frequency) adalah skema pembobotan yang biasa digunakan dalam pemrosesan bahasa alami untuk menetapkan bobot pada term dalam dokumen berdasarkan kepentingan atau keinformatifannya. Ini dihitung sebagai produk dari term frequency (TF) dan inverse document frequency (IDF). Frekuensi istilah adalah ukuran pentingnya suatu istilah dalam satu dokumen, dan dihitung sebagai berapa kali suatu istilah muncul dalam dokumen dibagi dengan jumlah total istilah dalam dokumen. Frekuensi dokumen terbalik adalah ukuran pentingnya suatu istilah di seluruh kumpulan dokumen, dan dihitung sebagai logaritma dari jumlah total dokumen dibagi dengan jumlah dokumen yang berisi istilah tersebut. Berikut merupakan proses tahap TF-IDF menggunakan procces document.

e. Cleanning

Pembersihan data adalah proses mengidentifikasi dan mengoreksi atau menghapus data yang tidak valid atau tidak akurat. dari kumpulan data Ini merupakan langkah penting dalam pra-pemrosesan data, karena kualitas dan karakteristik data dapat memengaruhi kinerja model pembelajaran mesin secara signifikan. Proses cleaning yaitu sebuah proses membersihkan/menghilangkan karakter Dalam proses RT, URL, tagar (#), pengguna dan simbol lainnya. Proses cleaning pada rapidminer terlihat pada Gambar 4.

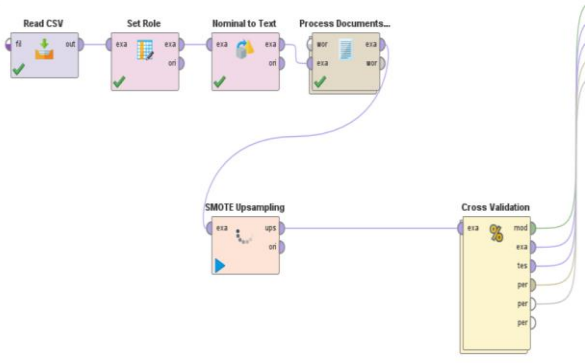


Gambar 4. Proses Cleaning

Berikut merupakan sebuah proses pembersihan. Yang dimana pada tahap ini menggunakan operator replace. Pembersihan data dapat menjadi proses yang memakan waktu dan membosankan, tetapi penting untuk memastikan bahwa data tersebut akurat dan bebas dari kesalahan sebelum digunakan untuk analisis atau pemodelan. Ini juga merupakan proses berulang. karena pilihan langkah pembersihan data dapat bergantung pada karakteristik data dan persyaratan khusus dari algoritme pembelajaran mesin.

3.2 Validasi

Validasi data adalah proses pengecekan keakuratan, kualitas, dan integritas data. Ini adalah langkah penting dalam pra-pemrosesan data, karena membantu memastikan bahwa data tersebut sesuai untuk analisis atau penggunaan dalam model pembelajaran mesin. Dalam tahap ini untuk memvalidasi atau data yang digunakan evaluasi adalah cross validation untuk mendapatkan nilai akurasi yang terbaik. operator yang digunakan validation dapat dilihat pada gambar 6.



Gambar 5. Tahap Validasi

Pada Gambar 5 Read Excel digunakan untuk menampilkan hasil pengambilan data melalui media

sosial Twitter, dengan tipe file .xls, dan operator Set Role digunakan dalam menentukan peran label pada kumpulan data yang akan diolah. Selain itu, pemrosesan data nominal diubah menjadi string menggunakan operator Nominal ke Teks. Setelah itu, lanjutkan mengerjakan dokumen proses dengan beberapa Langkah. Kemudian pada Teknik upsampling SMOTE bisa mengatasi ketidak seimbangan kelas dan melanjutkan proses ke operator pengujian dilakukan sebanyak 10 kali, Berdasarkan penelitian, validasi silang sebanyak 10 kali adalah pilihan terbaik untuk hasil yang akurat.

3.3 Klasifikasi dan Evaluasi

Pada pengujian ini akan menghasilkan data dari klasifikasi yang telah menentukan jumlah sentimen positif dan negatif yang diuji dalam metode Confusion Matrix. Confusion Matrix merupakan pengujian keakuratan untuk perhitungan nilai Accuracy, Recall, Precision yang di klasifikasi menggunakan persamaan yang terdapat pada Formula 1-3.

$$(1) Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$(2) Precision = \frac{TP}{TP + FP}$$

$$(3) Recall \frac{TP}{TP + FN}$$

True Positive (TP) merupakan jumlah prediksi data positif yang benar dan True Negative (TN) adalah jumlah nilai prediksi data negatif yang benar, Kemudian False Positive (FP) adalah jumlah data negatif tetapi diprediksi sebagai data yang tidak relevant, lalu False Negative (FN) merupakan jumlah data positif namun diprediksi sebagai data yang tidak relevan. Adapun pengujian confusion matrix disajikan pada Gambar 6 dan 7.

accuracy: 82.27% +/- 3.63% (micro average: 82.27%)

	true Negative	true Positive	class precision
pred. Negative	425	112	79.14%
pred. Positive	60	373	86.14%
class recall	87.63%	76.91%	

Gambar 6. Hasil Confusion Matrix Metode Naïve Bayes

accuracy: 88.12% +/- 3.83% (micro average: 88.12%)

	true Negative	true Positive	class precision
pred. Negative	554	108	83.69%
pred. Positive	31	477	93.90%
class recall	94.70%	81.54%	

Gambar 7. Hasil Confusion Matrix Metode K-Nearest Neighbors

Gambar 6 dan 7 merupakan hasil dari pengujian confusion matrix. Dalam pengujian confusion matrik pada gambar diatas mengahasilkan nilai akurasi, presisi dan recall. Berdasarkan dari analisis pengujian menggunakan metode NB dan KNN dari masing-masing metode dapat dirangkum hasilnya seperti tabel dibawah ini.

Tabel 4. Perbandingan Metode

Classifier	Accuracy	Precision	Recall
Naive Bayes	82.27%	86.36%	76.93%
KNN	88.12%	93.98%	81.53%

Penting untuk mempertimbangkan karakteristik data dan persyaratan khusus tugas dengan hati-hati saat memilih metode pembelajaran mesin, karena pilihan metode dapat memengaruhi performa model secara signifikan. Dalam penelitian ini mencoba dua metode yang membandingkan kinerjanya dengan menggunakan teknik seperti validasi silang untuk menentukan pendekatan terbaik untuk masalah yang diberikan. Berdasarkan tabel 4 hasil perhitungan evaluasi yang telah dilakukan proses pengolahan data dengan metode NB dan KNN terlihat juga bahwa metode K-Nearest Neighbor terbaik untuk menyatakan nilai tingkat akurasi yang tinggi dengan nilai akurasi 88.12% diikuti oleh algoritma NB, dengan tingkat akurasi 82.27%.

Kesimpulan

Dalam penelitian ini setelah menyelesaikan tahapan analisis data sentimen twitter terhadap pembangunan Ibu Kota Negara Nusantara, dapat disimpulkan, diantaranya data tweet yang diperoleh dianalisis menggunakan metode NB dengan tingkat Accuracy analisis sentimen sebesar 82.27% dengan nilai Precision sebe20sar 86.36% dan nilai Recall sebesar 76.93%, lalu tingkat akurasi metode KKN sebesar 88,12% dengan Precision sebesar 93.98% dan nilai recall sebesar 81.53%. Dengan nilai tersebut dapat disimpulkan bahwa metode KNN memiliki tingkat akurasi lebih unggul dibandingkan dengan metode NB. Dari hasil metode klasifikasi yang digunakan, masyarakat akan mengetahui adanya sentimen terhadap perpindahan ibu kota Nusantara.

Daftar Rujukan

[1] S. D. Saputra, T. Gabriel J, and M. Halkis, "Analisis Strategi Pemindahan Ibu Kota Negara Indonesia Ditinjau Dari Perspektif Ekonomi Pertahanan (Studi Kasus Upaya Pemindahan Ibu Kota Negara dari DKI Jakarta Ke Kutai Kartanegara Dan Penajam Paser Utara) Strategy Analysis Relocation Of The Capital Cit," *J. Ekon. Pertahanan*, vol. 7, p. 192, 2021.

[2] D. Nugroho, "Bentuk Ibu Kota Negara Nusantara Dalam Negara Kesatuan Republik Indonesia," *Indones. J. Polit. Policy* , vol. 4, no. 1, pp. 53–62, 2022, [Online]. Available: https://journal.unsika.ac.id/index.php/IJPP.

[3] A. Kodir, N. Hadi, I. K. Astina, D. Taryana, N. Ratnawati, and Idris, "The dynamics of community response to the development of the New Capital (IKN) of Indonesia," *Dev. Soc. Chang. Environ. Sustain.*, no. Nugroho 2020, pp. 57–61, 2021, doi: 10.1201/9781003178163-13.

[4] H. A. Sulaiman *et al.*, "Electronics Engineering , Computer Engineering and Information Technology."

[5] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14,

no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
[6] J. Teknika, R. K. Septiani, S. Anggraeni, and S. D. Saraswati, "Klasifikasi Sentimen Terhadap Ibu Kota Nusantara (IKN) pada Media Sosial Menggunakan Naive Bayes," *Teknika*, vol. 16, no. 2, pp. 245–254, 2022, [Online]. Available: https://jurnal.polsri.ac.id/index.php/teknika/article/view/4875.
[7] R. Y. Yanis, "Analisis Sentimen terhadap Debat Pemilihan Gubernur Jakarta Tahun 2017," *Aiti*, vol. 15, no. 2, pp. 128–134, 2018, doi: 10.24246/aiti.v15i2.128-134.
[8] S. Hilabi *et al.*, "Analysis of Drug Data Mining with Clustering Technique Using K-Means Algorithm," *J. Phys. Conf. Ser.*, vol. 1908, no. 1, 2021, doi: 10.1088/1742-6596/1908/1/012024.
[9] M. A. Djamaludin, A. Triayudi, and E. Mardiani, "Analisis Sentimen Tweet KRI Nanggala 402 di Twitter menggunakan Metode Naïve Bayes Classifier," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 2, pp. 161–166, 2022, doi: 10.35870/jtik.v6i2.398.
[10] F. Septianingrum, A. Susilo, and Y. Irawan, "Metode Seleksi Fitur Untuk Klasifikasi Sentimen Menggunakan Algoritma Naive Bayes : Sebuah Literature Review," vol. 5, pp. 799–805, 2021, doi: 10.30865/mib.v5i3.2983.
[11] S. Nurul, J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," vol. 5, no. 3, pp. 279–285, 2019, doi: 10.26418/jp.v5i3.34368.
[12] M. Syarifuddin, "Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Meggunakan Metode Naive Bayes Dan KNN," vol. 15, no. 1, pp. 23–28, 2020, doi: 10.33480/inti.v15i1.1347 VOL.
[13] J. Edukasi, B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," vol. 4, no. 2, pp. 113–118, 2018, doi: 10.26418/jp.v4i2.27526.
[14] S. Lestari, M. Mupaat, and A. Erfina, "Analisis Sentimen Masyarakat Indonesia terhadap Pemindahan Ibu Kota Negara Indonesia pada Twitter," *JUSIFO (Jurnal Sist. Informasi)*, vol. 8, no. 1, pp. 13–22, 2022, doi: 10.19109/jusifo.v8i1.12116.
[15] N. K. Sriwinarti and P. Juniarti, "Analisis Metode K-Nearest Neighbors (K-NN) Dan Naive Bayes Dalam Memprediksi Kelulusan Mahasiswa (Analysis of K-Nearest Neighbors (K-NN) and Naive Bayes Methods in Predicting Student Graduation)," vol. 3, no. 2, pp. 106–112, 2021.
[16] F. Sodik and I. Kharisudin, "Analisis Sentimen dengan SVM , NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," vol. 4, pp. 628–634, 2021.
[17] J. A. Josen Limbong, I. Sembiring, K. Dwi Hartomo, U. Kristen Satya Wacana, and P. Korespondensi, "Analisis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes Dan K-Nearest Neighbor Analysis of Review Sentiment Classification on E-Commerce Shopee Word Cloud Based With Naïve Bayes and K-Nearest Neighbor Meth," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, pp. 347–356, 2019, doi: 10.25126/jtiik.202294960.
[18] A. L. Hananto and A. Y. Rahman, "User experience measurement on go-jek mobile app in Malang City," *Proc. 3rd Int. Conf. Informatics Comput. ICIC 2018*, no. October 2018, pp. 1–6, 2018, doi: 10.1109/IAC.2018.8780423.
[19] A. Rohman, "Model Algoritma K-Nearest Neighbor (K-NN) Untuk Prediksi Kelulusan Mahasiswa," 2012.
[20] D. W. T. Putra, A. O. Utami, Minarni, and G. Y. Swara, "Accuracy Level of Diagnosis of ENT Diseases in Expert System," *J. KomtekInfo*, vol. 6, no. 2, pp. 127–134, 2019, doi: 10.35134/komtekinfo.v6i2.51.
[21] H. Yanto, "Sistem Pendukung Keputusan untuk Seleksi Usulan Pengajuan Sertifikasi Guru Menggunakan Algoritma K-Nearest Neigh Bor Berbasis Web," *J. KomtekInfo*, vol. 5, no. 2, pp. 42–50, 2018, doi: 10.35134/komtekinfo.v5i2.22.
[22] A. M. Tukino, "Klasifikasi Untuk Prediksi Cuaca Menggunakan Esemble Learning," *Petir*, vol. 13, no. 2, pp. 138–147, 2020, doi: 10.33322/petir.v13i2.998.
[23] M. K. Anam, B. N. Pikir, and M. B. Firdaus, "Penerapan Na¨ive Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen danPemeritah," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 139–150, 2021, doi: 10.30812/matrik.v21i1.1092.
[24] S. Ernawati and R. Wati, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel," *J. Khatulistiwa Inform.*, vol. 6, no. 1, pp. 64–69, 2018.