
**PENERAPAN ALGORITMA KLASIFIKASI NAÏVE BAYES UNTUK
ANALISIS SENTIMEN TENTANG PEMINDAHAN
IBU KOTA NEGARA**

Abdul Halim Hasugian^{1*}, Raissa Amanda Putri¹, Mulya Alfian Simatupang¹
Universitas Islam Negeri Sumatera Utara, Medan
e-mail: abdulhalimhasugian@uinsu.ac.id

Abstract: *The decision of the Government of the Republic of Indonesia regarding the relocation of the territory of the Indonesian state capital to the Nusantara Capital in North Penajam Paser, East Kalimantan is certainly an interesting discussion. This is a new record in Indonesian history because it has never happened before, becoming one of the big news stories that has generated a lot of opinions or reactions from the public. Twitter is one of the social media that people use to express opinions. Various kinds of opinions, both positive and negatif, are expressed by Indonesian people through their Twitter social media accounts. The data was obtained by crawling techniques from social media Twitter with the keywords "ibu kota nusantara" and "ikn" with the amount of data used as many as 606 tweets. The data is then labeled (class) using the lexicon-based method but first translated into English so that it is detected by the lexicon dictionary. From the labeling results, there were 400 tweets (66%) with positive labels and 206 tweets (34%) with negatif labels. The data preprocessing stage is carried out in preparation for the data to be processed for sentiment analysis. Of the many opinions obtained, an analysis of public opinion sentiment about the relocation of the national capital will be carried out using the Naïve Bayes classification algorithm. This study also applied the TF-IDF word weighting method with the n-gram feature used, namely bigram (n=2). The system will be evaluated using a confusion matrix. The performance of the system built in this study can be said to be good.*

Keywords: *relocation of national capitals; sentiment analysis; naïve bayes classifier.*

Abstrak: Keputusan Pemerintah Republik Indonesia tentang pemindahan wilayah ibu kota negara Indonesia menjadi Ibu Kota Nusantara di Penajam Paser Utara, Kalimantan Timur tentu menjadi pembahasan yang menarik. Ini merupakan catatan baru dalam sejarah Indonesia karena belum pernah terjadi sebelumnya, menjadi salah satu berita besar yang menimbulkan banyak opini atau reaksi dari masyarakat. Twitter merupakan salah satu media sosial yang digunakan masyarakat untuk mengungkapkan opini. Berbagai macam opini antara positif dan negatif diutarakan masyarakat Indonesia melalui akun media sosial Twiternya. Data diperoleh dengan teknik crawling dari media social Twitter dengan kata kunci "ibu kota nusantara" dan "ikn" dengan jumlah data yang digunakan sebanyak 606 tweet. Data kemudian diberi label (kelas) dengan metode lexicon-based tetapi terlebih dahulu menerjemahkannya kedalam Bahasa Inggris agar terdeteksi oleh kamus leksikon. Dari hasil pelabelan diperoleh data berlabel positif sebanyak 400 tweet (66%) dan negatif sebanyak 206 tweet (34%) . Tahap preprocessing data dilakukan sebagai persiapan untuk data agar bisa diproses untuk analisis sentimen. Dari sekian banyak opini yang diperoleh, maka akan dilakukan analisis sentimen opini masyarakat tentang pemindahan ibu kota negara dengan menggunakan algoritma klasifikasi Naïve Bayes. Pada penelitian ini juga diterapkan metode pembobotan kata TF-IDF dengan fitur n-gram yang digunakan yaitu bigram (n=2). Sistem akan dievaluasi dengan menggunakan confusion matrix. Kinerja dari sistem yang dibangun pada penelitian ini dapat dikatakan baik.

Kata kunci: pemindahan ibu kota negara; analisis sentimen; naïve bayes classifier.

PENDAHULUAN

Ibu Kota Negara merupakan pusat segala kegiatan yang berhubungan dengan kenegaraan dan diatur oleh perundang-undangan. Ibu kota sebagai pusat dari berbagai macam aktivitas tentunya tidak lepas dari permasalahan. Pindahan Ibu Kota Negara Indonesia menjadi berita terpopuler di pertengahan tahun 2019. Permasalahan kompleks yang terjadi pada ibu kota Negara Indonesia yaitu masalah demografi karena penduduk terbanyak ada di Jakarta kisaran 10.277.628 jiwa dikutip dari Bappenas (2019), kemacetan, dan kondisi lingkungan yang kian memburuk disertai banjir tahunan. Sebagaimana diketahui Pemerintah telah mencanangkan pemindahan Ibukota dari Jakarta ke sebuah kawasan di Kalimantan Timur (Purnama & Chotib, 2022). Masalah yang ada pada Jakarta sekarang sudah sesuai dengan kriteria yang mendasari pemindahan ibu kota menurut Agassi (2013) yaitu masalah sosial, ekonomi, politik, dan lingkungan. Permasalahan terus menumpuk tanpa sejalan dengan penanggulangannya menyebabkan diambil keputusan sulit untuk memindahkan ibu kota negara dari Jakarta ke Kalimantan Timur (Sakariana et al., 2020).

Topik berita mengenai pemindahan ibu kota negara banyak beredar di twitter. Oleh karena itu banyak juga penggunanya yang turut meramaikan, salah satunya dengan mengutarakan opini. Keputusan Pemerintah Indonesia dalam hal pemindahan ibu kota negara ini tentu menjadi bahasan yang menarik sehingga menimbulkan beragam reaksi komentar pro dan kontra dari para pengguna Twitter. Keterbatasan pengguna dalam menyaring informasi tersebut tentunya akan mempersulit masyarakat atau bahkan pemerintah dalam menganalisis informasi yang beredar di masyarakat (Sunesti & Putri, 2022). Twitter merupakan salah satu jejaring sosial yang paling mudah digunakan, karena hanya memerlukan waktu yang singkat tetapi informasi yang disampaikan dapat langsung menyebar

secara luas (Rezeki et al., 2020). Peningkatan jumlah pengguna Twitter di Indonesia tiap selang waktu menjadikannya sebagai media yang paling efektif untuk mengumpulkan umpan balik dan melibatkan publik dalam mengambil tindakan atas kebijakan pemerintah (Mustaqillah et al., 2023).

Analisis sentimen adalah salah satu cabang penelitian dari text mining yang berguna untuk mengklasifikasi dokumen teks berupa opini berdasarkan sentimen (Idris et al., 2023). Analisis sentimen dapat mengklasifikasikan opini-opini tersebut di antaranya menggunakan metode dalam machine learning seperti Naïve Bayes Classifier. Analisis sentimen (juga disebut sebagai analisis subjektivitas atau penambangan opini) adalah teknik Natural Language Processing (NLP) yang mengidentifikasi pola dan fitur penting informasi dari sekumpulan korpus teks. Analisis sentimen menggunakan NLP untuk mengekstraksi pola pengetahuan yang signifikan dari sejumlah besar data tekstual. Analisis sentimen mengekstraksi perasaan penulis dalam bentuk subjektivitas (objektif dan subjektif), polaritas (negatif, positif, dan netral), dan perasaan (marah, bahagia, terkejut, sedih, cemburu, dan campur aduk (Lamba & Madhusudhan, 2022).

Algoritma secara umum dapat diartikan sebagai urutan dari langkah - langkah yang disusun secara logis dan sistematis dan memiliki peranan penting dalam menyelesaikan suatu masalah (Retta et al., 2019). Naïve Bayes Classifier merupakan salah satu algoritma dalam teknik data mining yang menerapkan teori bayes dalam klasifikasi, suatu algoritma memanfaatkan teorema bayes (Prakoso & Sutanto, 2019). Naïve Bayes Classifier melakukan klasifikasi berdasarkan hasil perhitungan dari peluang terjadinya distribusi kata-kata yang terdapat dalam dokumen. Seluruh variabel yang terdapat dalam persamaan Naïve Bayes Classifier dianggap sebagai variabel bebas dan tidak memiliki hubungan dengan variabel lainnya.

Performa Naïve Bayes memiliki waktu klasifikasi yang singkat sehingga mempercepat proses analisis sentimen (Gunawan et al., 2018).

Tujuan penelitian yaitu analisis sentimen pada data komentar pengguna media sosial khususnya Twitter terkait pemindahan Ibu kota negara dengan metode Naïve Bayes dengan feature extraction TF-IDF yang dikombinasikan dengan teknik n-gram sebagai keterbaruan penelitian ini. Tujuan yang ingin dicapai dari penelitian ini untuk mengetahui tanggapan masyarakat terhadap pemindahan ibu kota negara melalui media sosial. Hasil penelitian ini adalah model klasifikasi untuk analisis sentimen yang dapat digunakan untuk mengetahui persepsi publik positif atau negatif terhadap pemindahan Ibu Kota.

METODE

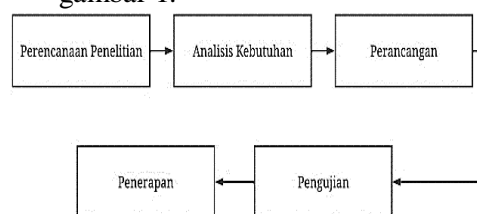
Lokasi penelitian ini dilakukan di Laboratorium Multimedia Fakultas Sains dan Teknologi UINSU Medan, Jl. Lapangan Golf No. 120, Desa Durian Jangak, Kecamatan Pancur Batu, Kabupaten Deli Serdang, Sumatera Utara. Dalam penelitian ini digunakan berupa sekumpulan data opini tweet berbahasa Indonesia yang mengandung salah satu atau beberapa dari kata kunci berikut yaitu “ibu kota nusantara” atau “ikn” hasil dari crawling data akan disimpan dan diproses dalam format berkas .csv. Data yang diperoleh atau dikumpulkan oleh peneliti secara langsung dari sumber datanya. Data primer disebut juga sebagai data asli atau data baru yang memiliki sifat up to date. Untuk mendapatkan data primer, peneliti harus mengumpulkannya secara langsung (Mustafa et al., 2020).

Adapun teknik pengumpulan data yang digunakan dalam mengumpulkan informasi:

1. Penelitian Kepustakaan, data yang diperoleh berupa buku cetak, jurnal penelitian, e-book dan situs web yang berhubungan dengan penelitian

ini, baik itu objek ataupun metodenya.

2. Observasi, mengambil data opini masyarakat mengenai kebijakan pemerintah tentang pemindahan ibu kota negara pada media sosial Twitter. Data tersebut diperoleh dengan menggunakan teknik web crawling dan memanfaatkan library sncrape (Hanifah & Nurhasanah, 2018). Tahapan awal penelitian ini dilakukan sesuai dengan rencana kegiatan dengan melalui beberapa tahapan yang dapat dilihat pada gambar 1.



Gambar 1. Diagram Perencanaan Penelitian

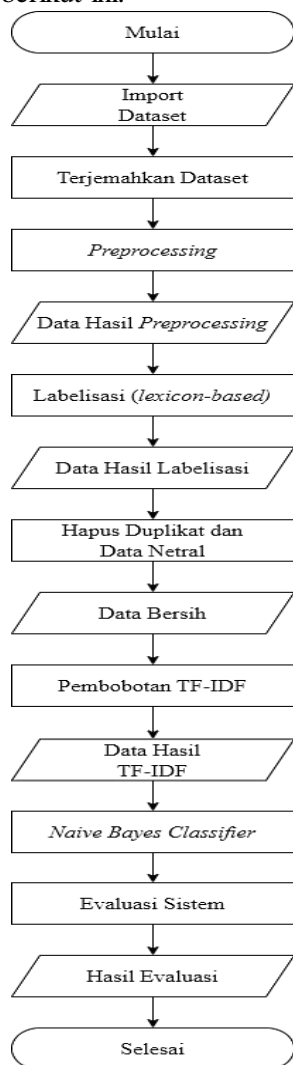
Analisa Kebutuhan Proses

Rincian mengenai analisis kebutuhan pada penelitian ini guna melakukan pengembangan pada penelitian yaitu , Analisis deskriptif, bertujuan untuk menunjukkan gambaran opini masyarakat Indonesia pada Twitter tentang peminahan ibu kota negara. Analisis sentimen, bertujuan untuk melabeli data menjadi kelas sentimen positif atau negatif dengan metode lexicon-based. Visualisasi, bertujuan untuk memberikan identifikasi mengenai pola kata yang terdapat dalam dataset. Metode pengujian, pada penelitian dataset diproses menggunakan algoritma Naïve Bayes Classifier dengan perbandingan data latih:data uji sebesar 9:1. Confusion Matrix, bertujuan untuk mengevaluasi hasil dari kinerja metode pengkla sifikasian yang digunakan sehingga didapat kesimpulan mengenai penelitian yang dilakukan.

Perancangan

Proses perancangan sistem pada penelitian ini bertujuan agar sistem yang dihasilkan sesuai dengan kebutuhan dan

dapat membantu pada saat penerapan rancangan ini menjadi sebuah sistem sehingga dapat menghemat waktu pembuatan. Proses perancangan dari sistem analisis sentimen yang akan dibangun terdiri dari perancangan flowchart. Adapaun sistem tersebut dimodelkan dalam bentuk flowchart dari perancangan sistem ini dapat dilihat gambar berikut ini:



Gambar 2. Flowchart Sistem Analisis Sentimen

Pengujian

Proses pengujian pada sistem yang dibangun memiliki tujuan untuk memastikan bahwa sistem yang dibangun telah berjalan seperti semestinya. Proses pengujian pada sistem yang dibangun untuk melakukan analisis sentimen masyarakat dilakukan dengan mengumpulkan semua data yang diperlukan,

dilanjutkan tahap preprocessing lalu di simpan dalam format csv. Setelah proses preprocessing selesai semua data akan melewati proses dengan pemberian label (nilai) dengan menggunakan Teknik lexicon-based dan selanjutnya akan dilakukannya klasifikasi pada data tersebut menggunakan metode.

Penerapan

Penelitian ini digunakan untuk mengetahui tanggapan atau opini dari masyarakat Indonesia pada media social Twitter, dengan cara memasukkan data berupa opini atau sentimen yang telah diberikan label (nilai) kedalam sistem dalam bentuk file dengan format csv untuk selanjutnya melakukan proses preprocessing dan klasifikasi setelah selesai proses klasifikasi selanjutnya akan dilakukan proses evaluasi sistem untuk mengetahui nilai accuracy, recall, precision, dan f1-score sehingga dapat diketahui tingkat akurasi dari hasil klasifikasi yang dilakukan oleh sistem.

HASIL DAN PEMBAHASAN

Presentasi Data

Dataset pada penelitian ini berupa opini atau tanggapan masyarakat terhadap keputusan pemerintah Republik Indonesia mengenai pemindahan ibu kota negara Indonesia dari Jakarta ke Kalimantan Timur atau dengan nama lain yaitu Ibu Kota Nusantara. Data berasal dari media sosial *Twitter* yang diperoleh dengan Teknik *crawling* dengan menggunakan *library Python* yaitu *snsrape*.

Proses ini menggunakan Bahasa pemrograman *Python* dengan *tools Jupyter Notebook*. Langkah pertama dengan mengunduh dan memasang *library "Snsrape"* kemudia mendeklarasikannya kedalam sistem. Kemudian data *tweet* dicrawling dengan kata kunci "ibu kota nusantara" atau "ikn".

Setelah *crawling* data dengan kata kunci tersebut selesai, maka selanjutnya list dari "*tweets_list*" diubah menjadi *dataframe* menggunakan fungsi *pd.DataFrame* dan disimpan ke dalam

berkas dengan format `.csv` dengan nama berkas “1100.csv”. Data yang diperoleh berjumlah 1100 data *tweet*. Data yang diperoleh dari proses *crawling* dapat dilihat dari kepala *dataframe* berturut-turut dengan kolom yaitu tanggal *tweet* dipublikasikan, id *tweet*, username, dan text yang berupa *tweet* itu sendiri.

Terjemahkan Data

Pada penelitian ini akan diterapkan metode labelisasi dataset dengan *lexicon-based*, kamus leksikon yang digunakan merupakan sekumpulan kata dalam Bahasa Inggris. Oleh sebab itu, tahap terjemahkan *dataset* dilakukan agar nantinya pelabelan sentimen bisa diterapkan dengan metode ini. Untuk memudahkan penerjemahan dari Bahasa Indonesia ke dalam Bahasa Inggris penulis memanfaatkan *library Googletrans* pada *Python*. Berikut disajikan tampilan hasil dari implementasi proses ini pada *Jupyter Notebook*.

	text
0	@puspuuus @Teguhwidiarto @soundBroKen @jokowi ...
1	The Minister of Public Works and Public Housin...
2	The President met Foxconn bosses, investment o...
3	The President hopes that IKN will create jobs ...
4	Asking BWT to work hard to complete various IK...

Gambar 3. Hasil Terjemahan Data

Preprocessing Data

Selanjutnya akan dilakukan proses *preprocessing* data. Proses ini perlu dilakukan karena data tersebut masih memiliki banyak elemen yang tidak diperlukan (*noise*) dalam proses analisis sentimen. Oleh karena itu data tersebut perlu dibersihkan dari segala elemen yang mengganggu terlebih dahulu (Rizaldi et al., 2020).

Dataset yang telah diperoleh disimpan dan akan diproses dalam berkas dengan format `.csv`. Sistem dari proses *preprocessing* data yang meliputi *cleaning*, *case folding*, *tokenizing*, *normalization*, *stopword removal* dan *stemming*.

Cleaning

Tahap pertama dalam yaitu *cleaning* dengan tujuan elemen yang tidak diperlukan dihilangkan atau dihapuskan agar data minim dari gangguan. Elemen tersebut berupa sebutan (*mention*), tagar, nama pengguna, tautan (*link*), alamat surel, angka, tanda baca, dan emotikon. Berikut disajikan tampilan hasil dari implementasi proses ini pada *Jupyter Notebook*.

	text
0	looking for ikn funds
1	the minister of public works and public housin...
2	the president met foxconn bosses investment of...
3	the president hopes that ikn will create jobs ...
4	asking bwt to work hard to complete various ik...

Gambar 4. Hasil Cleaning

Case Folding

Case folding adalah langkah selanjutnya dalam *preprocessing* data. Pada bagian ini akan dilakukan penyeragaman semua huruf yang terdapat dalam *dataset* menjadi huruf kecil (*lower case*). Berikut disajikan tampilan hasil dari implementasi proses ini pada *Jupyter Notebook*.

	text
0	@puspuuus @teguhWidiarto @soundbroken @jokowi ...
1	the minister of public works and public housin...
2	the president met foxconn bosses, investment o...
3	the president hopes that ikn will create jobs ...
4	asking bwt to work hard to complete various ik...

Gambar 5. Hasil Case Folding

Tokenizing

Tahap *tokenizing* dilakukan untuk memisahkan dokumen dalam dataset menjadi per kata agar tahap selanjutnya dapat dilakukan. Berikut disajikan tampilan hasil dari implementasi proses ini pada *Jupyter Notebook*.

	text
0	[looking, for, ikn, funds]
1	[the, minister, of, public, works, and, public...
2	[the, president, met, foxconn, bosses, investm...
3	[the, president, hopes, that, ikn, will, creat...
4	[asking, bwt, to, work, hard, to, complete, va...

Gambar 6. Hasil *Tokenizing*

Normalization

Tahap *normalization* dilakukan untuk mengubah, menghapus, memperbaiki kata atau istilah seperti singkatan, kata tidak baku, dan bentuk penulisan kata tidak baku yang sering digunakan saat bersosial media di dalam dokumen yang tidak terdeteksi ke dalam Bahasa Indonesia yang baku. Berikut disajikan tampilan hasil dari implementasi proses ini pada *Jupyter Notebook*.

	text
0	looking for ikn funds
1	the minister of public works and public housin...
2	the president met bosses investment of inclu...
3	the president hopes that ikn will create jobs ...
4	asking to work hard to complete various ikn a...

Gambar 7. Hasil *Normalization*

Stopword Removal

Tahap *stopword removal* dilakukan untuk menghapus kata dalam dokumen yang tidak memiliki arti seperti kata ganti, kata hubung, dan lainnya. Berikut disajikan tampilan hasil dari implementasi proses ini pada *Jupyter Notebook*.

	text
0	[looking, ikn, funds]
1	[minister, public, works, public, housing, req...
2	[president, met, bosses, investment, including...
3	[president, hopes, ikn, create, jobs, future, ...
4	[asking, work, hard, complete, various, ikn, a...

Gambar 8. Hasil *Stopword Removal*

Stemming

Tahap *stemming* dilakukan untuk menghapus seluruh imbuhan pada kata

yang terdapat di dalam dokumen seperti awalan,akhiran, dan penjamakan kata yang dalam Bahasa Inggris contohnya *-ing, -ed, -ment* dan lainnya (Findawati & Rosid, 2020). Berikut disajikan tampilan hasil dari implementasi proses ini pada *Jupyter Notebook*.

	text
0	look ikn fund
1	minist public work public hous request constru...
2	presid met boss invest includ ikn
3	presid hope ikn creat job futur capabl
4	ask work hard complet variou ikn access toll road

Gambar 9. Hasil *Stemming*

Labelisasi Data

Dataset yang telah diperoleh sebelumnya, selanjutnya akan dilakukan proses pemberian label (kelas) terhadap data tersebut. Kelas sentimen dikategorikan menjadi dua kelas yaitu positif dan negatif. Pada tahap ini, data akan diproses secara otomatis yaitu melakukan perhitungan nilai skor sentimen dengan kamus leksikon (*lexicon-based*).Kamus *lexicon* merupakan metode pengklasifikasian yang menggunakan kamus yang berisi kata opini untuk menentukan suatu sentimen kelas positif dan negatif dari suatu data teks. Kamus yang digunakan pada penelitian ini memakai kamus opini positif dan negatif dari penelitian oleh Liu et al., (2015) (Furqan et al., 2022).

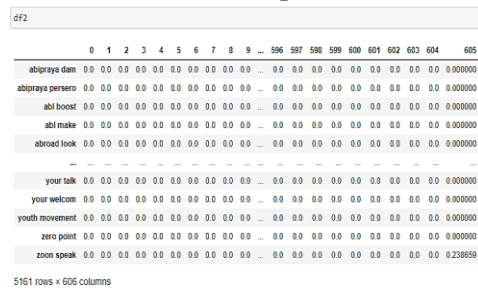
Formula yang umum digunakan untuk menghitung skor sentimen (*StSc*) adalah (Aline Bessa, 2022):

$$StSc = \frac{\text{jumlah kata positif} - \text{jumlah kata negatif}}{\text{total jumlah kata}} \quad (1)$$

Pembobotan *TF-IDF*

Tahap perhitungan bobot dari semua *term* (kata) berdasarkan seberapa besar intensitas kemunculan *term* tersebut dalam dokumen yang terdapat pada *dataset*. Pada penelitian ini dihasilkan sebanyak 5161 *term*. Berikut ini

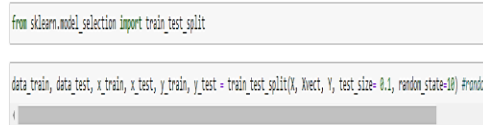
adalahhasil dari proses *TF-IDF*.



Gambar 10. Document Term Matrix Proses TF-IDF

Split Dataset

Split data merupakan tahapan untuk membagi *dataset* menjadi dua bagian yang terdiri dari data latih dan data uji dimana komposisi pembagiannya akan lebih banyak jumlah data latih dibandingkan data uji. Dalam penelitian ini akan diterapkan *split* data dengan perbandingan 9:1 yang artinya 90% dari total dataset akan digunakan sebagai data latih dan 10% lainnya akan digunakan sebagai data uji yang merujuk pada penelitian sebelumnya.



Gambar 11. Split Dataset

Klasifikasi Naïve Bayes Classifier

Setelah tahap pembobotan kata selesai, maka akan dilakukan tahap pengklasifikasian terhadap data uji dengan metode *Naïve Bayes Classifier*. Berikut disajikan contoh perhitungan klasifikasi terhadap data uji dengan data latih yang sebelumnya sudah dilakukan tahap preprocessing dan pelabelan. Tabel 1. Contoh Data Latih.

Tabel 1. Contoh Data Latih

Dokumen	Data Opini	Label
D1	construct nusantara ikn access toll road acceler good grate god add favor	Positive
D2	kpu call impact	Negative

Dokumen	Data Opini	Label
	papua new guinea ikn rais elect problem	
D3	hope ikn develop process safe smooth victori sake great Indonesia	Positive
D4	debt incur legal plan debt legal like debt build high speed train debt build ikn pleas explain pleas	Negative
D5	ikn lazi listen anymor govern debt count anymore	Negative

Tabel 2. Contoh Data Uji

Data tweet	Hasil Preprocessin g	Labe l
@OposisiCerdas Nggak usah bangun IKN kalau duitnya ngutang Pak	build ikn owe money	?

Tahapan dalam proses perhitungan klasifikasi kelas pada data uji dimulai dari perhitungan nilai *prior probability*, *conditional probability*, dan *posterior probability*. Berikut disajikan tahap perhitungan metode *Naïve Bayes Classifier* terhadap data uji tersebut.

- Perhitungan nilai *prior probability*

$$P(Positive) = \frac{2}{5} = 0,4$$

$$P(Negative) = \frac{3}{5} = 0,6$$
- Perhitungan nilai *conditional probability* dengan *Laplacian smoothing*

- Positif

$$P(\text{build ikn}|\text{Positive}) = \frac{0+1}{41,97+109,13} = 0,006$$

$$P(\text{ikn owe}|\text{Positive}) = \frac{0+1}{41,97+109,13} = 0,006$$

$$P(\text{owe money}|\text{Positif}) = \frac{0+1}{41,97+109,13} = 0,006$$

- Negatif

$$P(\text{build ikn}|\text{Negative}) = \frac{1+1}{69,25+109,13} = 0,011$$

$$P(\text{ikn owe}|\text{Negative}) = \frac{0+1}{69,25+109,13} = 0,005$$

$$P(\text{owe money}|\text{Negative}) = \frac{0+1}{69,25+109,13} = 0,005$$

3. Perhitungan nilai *posterior probability*

$$P(\text{Opini}|\text{Positif}) = 0,4 * 0,006^3 = 8,64 * 10^{-8}$$

$$P(\text{Opini}|\text{Negatif}) = 0,6 * 0,011 * 0,005^2 = 1,65 * 10^{-5}$$

Dari contoh perhitungan didapatkan hasil dengan label yang paling tinggi adalah **Negative** dengan nilai $1,65 * 10^{-5}$. Sehingga hasil klasifikasi untuk data uji adalah **Negative**. Berikut ini adalah tahapan untuk melakukan prediksi klasifikasi menggunakan metode *Naïve Bayes Classifier*.

```
# objek naive bayes classifier
mnb = MultinomialNB()
mnb.fit(x_train, y_train)
mnb_pred = mnb.predict(x_test)
mnb_acc = accuracy_score(y_test, mnb_pred)
print("Test accuracy: {:.2f}%".format(mnb_acc*100))
```

Gambar 12. Proses Klasifikasi Dataset

Evaluasi Hasil

Data yang digunakan merupakan opini berlabel positif hasil dari tahap pelabelan dengan teknik *lexicon-based* secara otomatis. Data positif berjumlah 400 dari 606 data opini. Data yang digunakan merupakan opini berlabel positif hasil dari tahap pelabelan dengan teknik *lexicon-based* secara otomatis. Data positif berjumlah 206 dari 606 data opini. Pada penelitian ini, *confusion*

matrix digunakan sebagai evaluasi hasil dari sistem yang dibangun. *Confusion matrix* menyajikan informasi berupa angka dengan bentuk matriks berordo tertentu yang pada penelitian ini berordo 2x2. Dari *confusion matrix* dapat dihitung nilai *accuracy*, *precision*, *recall*, dan *f1-score* sebagai hasil dari sistem analisis sentimen dan pengklasifikasian menggunakan metode *Naïve Bayes Classifier*. Berikut disajikan Tabel *confusion matrix* pada penelitian ini.

Tabel 3. Confusion Matrix Pada Model Klasifikasi

<i>ij</i>		Kelas Prediksi (<i>j</i>)	
		Positif	Negatif
Kelas Aktual (<i>i</i>)	Positif	47	0
	Negatif	14	0

Maka dapat dihitung seberapa besar nilai *accuracy*, *precision*, *recall*, dan *f1-score*, yaitu:

$$Accuracy = \frac{47+0}{47+0+14+0} \times 100\% = 77.05\%$$

$$Precision = \frac{47}{47+14} \times 100\% = 77.05\%$$

$$Recall = \frac{47}{47+0} \times 100\% = 100\%$$

$$f1score = \frac{2 \times 100 \times 77.05}{100 + 77.05} \times 100\% = 87.04\%$$

SIMPULAN

Pada penelitian ini digunakan dataset yang diperoleh dengan teknik *crawling* didapat sejumlah 1100 data tweet mengenai keputusan pemerintah Indonesia terkait pemindahan ibu kota negara dari Jakarta ke Penajam Paser Utara, Kalimantan Timur. Dengan menerapkan preprocessing data, labelisasi dan penghapusan duplikat maka didapat sebanyak 606 data tweet (opini) dengan penyeragaman kata kunci “ibu kota nusantara” atau “ikn”. Pelabelan data dalam penelitian ini menggunakan Teknik *lexicon-based* dengan kamus berbahasa Inggris. Teknik

ini akan menghitung seberapa banyak kata yang terdeteksi dalam kamus terhadap dokumen (opini) dan menghitung skor sentimennya. Dengan menerapkan Teknik ini didapat rincian data dalam kelas positif sebanyak 66% yaitu 400 data, dan negatif sebanyak 34% yaitu 206 data.

Tingkat akurasi yang dihasilkan dengan menggunakan metode Naïve Bayes Classifier pada sistem klasifikasi sentimen dapat dikatakan baik, hal ini dapat dilihat dari hasil akurasi pada dataset yang berjumlah 606 data dengan perbandingan data latih dan data uji sebesar 9:1 dan pembobotan kata dengan TF-IDF dan n-gram (n=2) diperoleh nilai accuracy sebesar 77,05%, precision sebesar 77,05%, recall sebesar 100%, dan f1-score sebesar 87,04%.

DAFTAR PUSTAKA

- Findawati, Y., & Rosid, M. A. (2020). *Buku Ajar Text Mining* (R. Dijaya (ed.); 1st ed.). Umsida Press. <https://doi.org/10.21070/2020/978-623-6833-19-3>
- Furqan, M., Sriani, S., & Sari, S. M. (2022). Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia. *Techno.Com*, 21(1), 52–61. <https://doi.org/10.33633/tc.v21i1.5446>
- Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, 4(2), 17–29.
- Hanifah, R., & Nurhasanah, I. S. (2018). Implementasi Web Crawling untuk Mengumpulkan Informasi Wisata Kuliner di Bandar Lampung. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(5), 531–535. <https://doi.org/10.25126/jtiik.201855842>
- Idris, I. S. K., Mustofa, Y. A., & Salihi, I. A. (2023). Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM). *Jambura Journal of Electrical and Electronics Engineering*, 5(1), 32–35. <https://doi.org/10.37905/jjee.v5i1.16830>
- Lamba, M., & Madhusudhan, M. (2022). Sentiment Analysis. In *Text Mining for Information Professionals* (1st ed., pp. 191–211). Springer International Publishing. https://doi.org/10.1007/978-3-030-85085-2_7
- Mustafa, P. S., Gusdiyanto, H., Victoria, A., Masgumelar, N. K., Lestariningsih, N. D., Maslacha, H., Ardiyanto, D., Utama, H. A., Boru, M. J., Fachrozi, I., Rodriquez, E. I. S., Prasetyo, T. B., & Romadhana, S. (2020). Metodologi Penelitian Kuantitatif, Kualitatif, Dan Penelitian Tindakan Kelas Dalam Pendidikan Olahraga. Fakultas Ilmu Keolahragaan Universitas Negeri Malang.
- Mustaqlilah, R., Widyaningtyas, O., & Wantoro, T. (2023). Efektivitas Penggunaan Twitter Sebagai Sarana Peningkatan Berpikir Kritis Mahasiswa Ilmu Komunikasi. *MUKASI: Jurnal Ilmu Komunikasi*, 2(1), 18–28. <https://doi.org/10.54259/mukasi.v2i1.1346>
- Prakoso, B. S., & Sutanto, G D. (2019). Penerapan Metode Decision Tree Dan Naïve Bayes Untuk. *Jurnal IKRA-ITH Informatika*, 3(2), 27–32.
- Purnama, S. J., & Chotib. (2022). Analisis Kebijakan Publik Pemindahan Ibu Kota Negara. *Jurnal Ekonomi Dan Kebijakan Publik*, 13(1), 57–70. <https://doi.org/10.22212/jekp.v13i1.2155>
- Retta, A. M., Isroqmi, A., & Nopriyanti, T. D. (2019). Pengaruh Penerapan Algoritma Pembelajaran Pemrograman Komputer. *Jurnal Inovasi Pendidikan Matematika*, 2(2), 126–135.
- Rezeki, S. R. I., Restiviani, Y., & Zahara,

- R. (2020). Penggunaan sosial media twitter dalam komunikasi organisasi (studi kasus pemerintah provinsi DKI Jakarta dalam penanganan covid-19). *Journal of Islamic and Law Studies*, 04(02), 63–78.
- Rizaldi, M. R., Adiwijaya, A., & Al Faraby, S. (2020). Klasifikasi Argument Pada Teks dengan Menggunakan Metode Multinomial Logistic Regression Terhadap Kasus Pemandangan Ibu Kota Indonesia di Twitter. *Jurnal Media Informatika Budidarma*, 4(4), 904–913. <https://doi.org/10.30865/mib.v4i4.23>
- Sakariana, M. I. D., Indriati, & Dewi, C. (2020). Analisis Sentimen Pemandangan Ibu Kota Indonesia Dengan Pembobotan Term BM25 Dan Klasifikasi Neighbor Weighted K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 4(3), 748–755.
- Sunesti, Y., & Putri, A. K. (2022). Narasi Ibukota Negara Baru Di Twitter: Dari Isu Kearifan Lokal Hingga Isu Anak Muda. *Konferensi Nasional Sosiologi IX APSSI 2022 Balikpapan*, 1–3.