

**KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP PROSES
PEMINDAHAN IBU KOTA NEGARA (IKN) INDONESIA PADA MEDIA
SOSIAL *TWITTER* MENGGUNAKAN METODE *NAÏVE BAYES***

SKRIPSI

Oleh:

MOCH. REINALDY DESTRA FACHREZA

NIM. 17650053



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP PROSES
PEMINDAHAN IBU KOTA NEGARA (IKN) INDONESIA PADA MEDIA
SOSIAL *TWITTER* MENGGUNAKAN METODE *NAÏVE BAYES***

SKRIPSI

Oleh:

MOCH. REINALDY DESTRA FACHREZA
NIM. 17650053

Diajukan kepada:

**Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

HALAMAN PERSETUJUAN

**KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP PROSES
PEMINDAHAN IBU KOTA NEGARA (IKN) INDONESIA
PADA MEDIA SOSIAL TWITTER MENGGUNAKAN
METODE NAÏVE BAYES**

SKRIPSI

Oleh:
MOCH. REINALDY DESTRA FACHREZA
NIM. 17650053

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal: 12 Juni 2023

Pembimbing I



Prof. Dr. Suhartono, M.Kom
NIP. 19680519 200312 1 001


Pembimbing II



Dr. M. Ainul Yaqin, M.Kom
NIP. 19761013 200604 1 004

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Fachrul Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP PROSES PEMINDAHAN IBU KOTA NEGARA (IKN) INDONESIA PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE NAÏVE BAYES

SKRIPSI

Oleh:

MOCH. REINALDY DESTRA FACHREZA

NIM. 17650053

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S. Kom)
Pada Tanggal: 19 Juni 2023

Susunan Dewan Penguji

Ketua Penguji : Dr. Totok Chamidy, M.Kom
NIP. 19691222 20060

Anggota Penguji I : Okta Oomaruddin Aziz, M.Kom
NIP. 19911019 201903 1 013


Anggota Penguji II : Prof. Dr. Suhartono, M.Kom
NIP. 19680519 200312 1 001

Anggota Penguji III : Dr. Ainul Yaqin, M.Kom
NIP. 19761013 200604 1 004



Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Fachrul Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Moch. Reinaldy Destra Fachreza
NIM : 17650053
Program Studi : Teknik Informatika
Fakultas : Sains dan Teknologi
Judul Skripsi : Klasifikasi Sentimen Masyarakat Terhadap Proses
Pemindahan Ibu Kota Negara (IKN) Indonesia Pada Media
Sosial *Twitter* Menggunakan Metode *Naïve Bayes*.

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar – benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 19 Juni 2023
Yang membuat pernyataan,



Moch. Reinaldy Destra Fachreza
NIM. 17650053

MOTTO

*“Unfold the pages of life with harmonious adequacy, embrace tranquility
as chapters intertwine, until the climactic verse unveils.”*

HALAMAN PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Penulis persembahkan skripsi ini kepada keluarga penulis, terutama untuk Bapak Casimin dan Ibu Sri Utami serta kakak Atika Puspita Sari yang senantiasa memberikan dukungan dan pembelajaran kepada penulis melalui arti perjuangan, kesabaran, dan ketekunan. Kepada keluarga besar penulis dari pihak Ayah maupun pihak Ibu yang senantiasa ikut mendoakan agar dapat menyelesaikan masa studi. Terakhir untuk Khotibul Umam, Apt. Reihan Khairiati, S.Farm, Ahmad Rifqi Rosadi, Rizal Sulton Aulia, Faishal Alfarisi, Arya Abimanyu, Ayu Dian Fitri Mellina, Fajar Dewantara, dan teman-teman lainnya yang belum penulis sebutkan yang selalu menjadi pendengar yang baik, bertukar pikiran, dan menjadi salah satu *support system* penulis. Semoga kasih sayang Allah *subhanahu wa ta'ala* selalu menyertai mereka.

KATA PENGANTAR

Assalamu'alaikum warahmatullahi wabarakatuh

Segala puji bagi Allah *subhanahu wa ta'ala* yang telah melimpahkan rahmat dan karunia-Nya serta shalawat beriring salam tak lupa dihanturkan kepada baginda Rasulullah *shalallahu 'alaihi wa sallam* sehingga penulis mampu merampungkan penulisan skripsi yang berjudul **“Klasifikasi Sentimen Masyarakat Terhadap Proses Pemindahan Ibu Kota Negara (IKN) Indonesia Pada Media Sosial *Twitter* Menggunakan Metode *Naïve Bayes*”** sebagai salah satu syarat kelulusan untuk mendapatkan gelar sarjana pada Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Skripsi ini tidak dapat terwujud tanpa adanya doa, bantuan, bimbingan dan motivasi dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih sedalam-dalamnya kepada:

1. Prof. Dr. H. M. Zainuddin, MA, selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Sri Harini, M. Si, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan, ST., M.MT., IPM selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Hani Nurhayati, M.T., selaku Dosen Wali yang telah bersedia meluangkan waktu dalam membimbing, dan memberikan motivasi sehingga skripsi ini dapat terselesaikan.
5. Prof. Dr. Suhartono, M,Kom selaku Dosen Pembimbing I yang telah banyak bersedia meluangkan waktunya dalam membimbing, memberi saran dan arahan kepada penulis sehingga skripsi ini dapat terselesaikan.
6. Dr. M. Ainul Yaqin, M,Kom selaku Dosen Pembimbing II yang juga bersedia meluangkan waktunya dalam membimbing dan memberi arahan kepada penulis sehingga skripsi ini dapat terselesaikan.
7. Dr. Totok Chamidy, M.Kom, dan Okta Qomaruddin Aziz, M.Kom selaku Dosen Penguji yang telah memberikan kritik dan masukan membangun kepada penulis selama proses penyelesaian skripsi ini.

8. Seluruh Dosen dan Jajaran Staf Program Studi Teknik Informatika yang telah mengajarkan ilmu yang bermanfaat kepada penulis.
9. Teman-teman Teknik Informatika Angkatan 2017 UNOCORE dan Rohmad Aji Irawan yang telah menemani dan saling bertukar pikiran saat mengerjakan skripsi, serta memberikan banyak pengalaman dan dukungan berharga.
10. Semua pihak yang secara langsung maupun tidak langsung turut membantu penulis menyelesaikan skripsi yang tidak dapat disebutkan satu persatu.
11. Penulis sendiri yang telah memberikan yang terbaik dalam penyelesaian skripsi.

Penulis menyadari bahwa dalam pengerjaan skripsi ini masih terdapat banyak kekurangan sehingga penulis terbuka terhadap kritik dan saran yang membangun dari para pembaca. Penulis berharap semoga skripsi ini dapat memberikan manfaat tidak hanya bagi penulis namun juga bagi para pembaca.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Malang, 19 Juni 2023

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	i
PERNYATAAN KEASLIAN TULISAN	ii
MOTTO	i
HALAMAN PERSEMBAHAN	ii
KATA PENGANTAR	iii
DAFTAR ISI	v
DAFTAR GAMBAR	7
DAFTAR TABEL	8
ABSTRAK	10
ABSTRACT	11
تحبلا صلختسم	12
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Pernyataan Masalah.....	5
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	5
1.6 Sistematika Penulisan.....	6
BAB II STUDI PUSTAKA	7
2.1 Klasifikasi.....	7
2.2 Pemandangan Ibu Kota Negara.....	8
2.3 Sosial Media Twitter.....	9
2.4 Pengolahan Bahasa Alami.....	11
2.5 <i>Text Mining</i>	11
2.6 <i>Naïve Bayes Classifier</i>	12
2.7 Integrasi Islam.....	12
2.8 Penelitian Terkait.....	15
BAB III DESAIN DAN IMPLEMENTASI SISTEM	19
3.1 Pengumpulan Data.....	19
3.2 Rancangan Sistem.....	22
3.3 <i>Preprocessing</i>	23
3.4 TF-IDF.....	30
3.5 Naïve Bayes.....	37
3.6 <i>Gaussian Naïve Bayes</i>	39
3.7 Skenario Uji.....	49
BAB IV UJI COBA DAN PEMBAHASAN	61
4.1 Rangkaian Pengujian.....	61
4.1.1 <i>Input Data</i>	61
4.1.2 Pembagian Data.....	62
4.1.3 Jumlah Fitur.....	63
4.1.4 Pelatihan Data.....	63
4.1.5 Pengujian Data.....	64

4.1.5 Hasil Evaluasi	64
4.2 Hasil Pengujian.....	65
4.3 Pembahasan	74
BAB V KESIMPULAN DAN SARAN	79
5.1 Kesimpulan.....	79
5.1 Saran	79
DAFTAR PUSTAKA	

DAFTAR GAMBAR

Gambar 2.1 Proses Klasifikasi	7
Gambar 3.1 Data Hasil <i>Scraping</i>	19
Gambar 3.2 Rancangan Sistem	22
Gambar 3.3 Tahapan <i>Preprocessing</i>	23
Gambar 3.4 Alur <i>Case Folding</i>	25
Gambar 3.5 Alur <i>Cleaning</i>	26
Gambar 3.6 Alur <i>Stopword Removal</i>	27
Gambar 3.7 Alur <i>Stemming</i>	28
Gambar 3.8 Alur <i>Tokenizing</i>	29
Gambar 3.9 Alur <i>Gaussian Naïve Bayes</i>	41
Gambar 3.10 Grafik Hasil Pengujian Akurasi	59
Gambar 4.1 Visualisasi Hasil Pengujian	74

DAFTAR TABEL

Tabel 2.1 Pembaharuan dengan Metode Penelitian Terkait.....	17
Tabel 3.1 Pelabelan	21
Tabel 3.2 <i>Case Folding</i>	25
Tabel 3.3 <i>Cleaning</i>	27
Tabel 3.4 Contoh <i>Stopword Removal</i>	28
Tabel 3.5 <i>Stemming</i>	29
Tabel 3.6 <i>Tokenizing</i>	30
Tabel 3.7 Tabel Nilai TF-IDF	32
Tabel 3.8 Tabel Contoh Dataset.....	33
Tabel 3.9 Tabel Nilai TF	33
Tabel 3.10 Tabel Nilai DF dan IDF	34
Tabel 3.11 Tabel Hasil Nilai TF-IDF	35
Tabel 3.12 Tabel Peringkat Nilai TF-IDF.....	35
Tabel 3.13 Tabel Fitur.....	42
Tabel 3.14 Tabel Fitur Kelas Mendukung	43
Tabel 3.15 Tabel Fitur Kelas Tidak Mendukung.....	44
Tabel 3.16 Tabel <i>Mean</i>	44
Tabel 3.17 Tabel Standar Deviasi	45
Tabel 3.18 Tabel Input Data Uji Baru.....	45
Tabel 3.19 Tabel Probabilitas Fitur.....	46
Tabel 3.20 Tabel Probabilitas Kelas	46
Tabel 3.21 Contoh dataset dengan 100% Berlabel Mendukung	49
Tabel 3.22 Contoh dataset dengan 90% Berlabel Mendukung	50
Tabel 3.23 Contoh dataset dengan 80% Berlabel Mendukung	51
Tabel 3.24 Contoh dataset dengan 70% Berlabel Mendukung.....	51
Tabel 3.25 Contoh dataset dengan 60% Berlabel Mendukung.....	52
Tabel 3.26 Contoh dataset dengan 50% Berlabel Mendukung.....	53
Tabel 3.27 Contoh dataset dengan 40% Berlabel Mendukung.....	53
Tabel 3.28 Contoh dataset dengan 30% Berlabel Mendukung.....	54
Tabel 3.29 Contoh dataset dengan 20% Berlabel Mendukung.....	55
Tabel 3.30 Contoh dataset dengan 10% Berlabel Mendukung.....	55
Tabel 3.31 Tabel Hasil Skenario Pengujian.....	56
Tabel 3.32 Pembagian Persentase Dokumen Uji	58
Tabel 3.33 Hasil Akurasi Pembagian Persentase Dokumen Uji	58
Tabel 4.1 Tabel Contoh Input Data.....	62
Tabel 4.2 Tabel Pembagian Persentase Penggunaan Fitur.....	63
Tabel 4.3 Keterangan Rumus <i>Confusion Matrix</i>	65
Tabel 4.4 Hasil Uji Penggunaan Total Fitur Sebesar 10%.....	66
Tabel 4.5 Hasil Uji Penggunaan Total Fitur Sebesar 20%.....	67
Tabel 4.6 Hasil Uji Penggunaan Total Fitur Sebesar 30%.....	68

Tabel 4.7 Hasil Uji Penggunaan Total Fitur Sebesar 40%.....	69
Tabel 4.8 Hasil Uji Penggunaan Total Fitur Sebesar 50%.....	70
Tabel 4.9 Hasil Uji Penggunaan Total Fitur Sebesar 60%	70
Tabel 4.10 Hasil Uji Penggunaan Total Fitur Sebesar 70%	71
Tabel 4.11 Hasil Uji Penggunaan Total Fitur Sebesar 80%	72
Tabel 4.12 Hasil Uji Penggunaan Total Fitur Sebesar 90%.....	73
Tabel 4.13 Hasil Uji Pada <i>Confussion Matrix</i>	74

ABSTRAK

Fachreza, Moch Reinaldy Destra. 2023. **Klasifikasi Sentimen Masyarakat Terhadap Proses Pemindahan Ibu Kota Negara (IKN) Indonesia Pada Media Sosial Twitter Menggunakan Metode Naïve Bayes**. Skripsi. Program Studi Teknik Informatika. Fakultas Sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (1) Prof. Dr. Suhartono, M.Kom (2) Dr. M. Ainul Yaqin, M.Kom

Kata kunci: IKN, Klasifikasi Sentimen, Naïve Bayes, Twitter

Beberapa waktu lalu, DPR mengesahkan Undang-Undang (UU) Nomor 3 Tahun 2022 tentang IKN pada 18 Januari 2022 dan kemudian Presiden Joko Widodo resmi telah menandatangani UU IKN pada 15 Februari 2022. Dengan demikian, Ibu kota Indonesia IKN akan dipindahkan ke Kabupaten Penajam Paser Utara dan Kabupaten Kutai Kartanegara, Provinsi Kalimantan Timur. Respon yang masyarakat terkait keputusan tersebut beraneka ragam, Banyak yang menanggapi dengan sentimen mendukung, namun ada pula yang menanggapi dengan sentimen tidak mendukung. Dewasa ini, banyak cara untuk mengamati informasi yang terhimpun pada media sosial. Berbagai tanggapan yang disampaikan melalui media sosial dapat digunakan sebagai data penelitian klasifikasi sentimen. Metode Naïve Bayes lumrah digunakan untuk penelitian jenis ini. Penelitian ini menggunakan jenis *Gaussian Naïve Bayes* karena asumsi independensi yang dilakukan oleh metode ini, fitur-fitur yang tidak memberikan kontribusi signifikan pada klasifikasi dapat diabaikan, sehingga mengurangi dampak dari fitur-fitur yang tidak relevan. Penelitian ini bertujuan untuk mengukur tingkat akurasi klasifikasi sentimen masyarakat di media sosial twitter terhadap proses pemindahan ibu kota negara. Sistem yang dibuat memberikan hasil uji coba terbaik pada penggunaan fitur sebanyak 80% dengan accuracy 82,0%, precision 76,9%, dan recall 100%

ABSTRACT

Fachreza, Moch Reinaldy Destra. 2023. **Sentiment Classification of Society Towards the Relocation Process of the National Capital (IKN) of Indonesia on Twitter Social Media Using Naïve Bayes Method.** Undergraduate Thesis. Department of Informatics Engineering, Faculty of Science and Technology, State Islamic University of Maulana Malik Ibrahim Malang. Supervisor: (I) Prof. Dr, Suhartono, M.Kom (II) Dr. M. Ainul Yaqin, M.Kom

Earlier, the Indonesian House of Representatives (DPR) passed Law Number 3 of 2022 regarding the National Capital Integrated Coastal Development (IKN) on January 18, 2022, and later on February 15, 2022, President Joko Widodo officially signed the IKN Law. Consequently, the capital city of Indonesia will be relocated to Penajam Paser Utara Regency and Kutai Kartanegara Regency, East Kalimantan Province. The public's response to this decision varies. Many have expressed their support, while others have expressed their opposition. Nowadays, there are many ways to observe the information gathered on social media. Various responses conveyed through social media can be used as data for sentiment classification research. The Naïve Bayes method is commonly used for this type of research. This study uses the Gaussian Naïve Bayes classifier due to its assumption of independence. Features that do not significantly contribute to the classification can be ignored, reducing the impact of irrelevant features. This study aims to measure the accuracy, precision, and recall of sentiment classification among the public on Twitter regarding the capital relocation process. The developed system yielded the best results during testing with 80% of the features, achieving an accuracy of 82.0%, precision of 76.9%, and recall of 100%.

Keywords: IKN, Naïve Bayes, Sentiments Classification, Twitter

ثجبالا صلختسم

في إندونيسيا على وسائل (IKN) فاشريزا، موك رينالدي ديسترا ٢٠٢٣. تصنيف مشاعر المجتمع تجاه عملية نقل عاصمة الدولة
التواصل الاجتماعي تويتر باستخدام طريقة نايف بايز. سكريسي. برنامج دراسة هندسة المعلوماتية. كلية العلوم والتكنولوجيا
جامعة الإسلامية النيجيرية مولانا مالك إبراهيم مالانغ. المشرفون: ١: (أ.د. سوهارتونو، ماجستير في الحاسوب) ٢ (د. م.
عين اليقين، ماجستير في الحاسوب

الكلمات الرئيسية: تصنيف المشاعر، تويتر، نايف بايز، IKN

في ١٨ يناير ٢٠٢٢، ومن ثم وقع IKN بعض الوقت السابق، أقر مجلس النواب قانون رقم ٣ لعام ٢٠٢٢ بشأن
إلى مقاطعة IKN في ١٥ فبراير ٢٠٢٢. وبذلك، ستنقل عاصمة إندونيسيا IKN الرئيس جوكو ويدودو رسمياً على قانون
،بيناجام باسر أوتارا ومقاطعة كوتاي كارتانيجارا في محافظة كاليمانتان تيمور. التفاعل الذي أبداه الناس بشأن هذا القرار متنوع
حيث أبدى العديد منهم مشاعر داعمة، ولكن هناك أيضاً من رد بمشاعر غير داعمة. في الوقت الحالي، هناك العديد من الطرق
لمراقبة المعلومات التي تجمع على وسائل التواصل الاجتماعي. يمكن استخدام مختلف الردود التي يتم نشرها عبر وسائل التواصل
تستخدم عادة في هذا النوع من الأبحاث. يستخدم هذا Naïve Bayes الاجتماعي كبيانات لبحث تصنيف المشاعر. طريقة
بسبب الافتراضات المستقلة التي تجربها هذه الطريقة، حيث يمكن تجاهل الميزات التي Gaussian Naïve Bayes البحث نوع
لا تساهم بشكل كبير في التصنيف، مما يقلل من تأثير الميزات غير ذات الصلة. يهدف هذا البحث إلى قياس مستوى دقة تصنيف
مشاعر الجمهور على وسائل التواصل الاجتماعي تويتر تجاه عملية نقل عاصمة البلاد. يعطي النظام الذي تم إنشاؤه أفضل نتائج
للاختبار عند استخدام ٨٠٪ من الميزات بدقة ٨٢٪، دقة ٧٦.٩٪، واسترجاع ١

BAB I

PENDAHULUAN

1.1 Latar Belakang

Ibu kota adalah wilayah letak lokasi pusat pemerintahan suatu negara, tempat himpunan segala administratif, yaitu lembaga eksekutif, lembaga legislatif, dan Lembaga yudikatif. (Tim Pustaka Phoenix, 2013). Ibu kota negara diambil dari bahasa latin *caput* yang bermakna kepala (*head*), dan terpaut dengan tata *capitol* yang berarti pusat pokok pemerintahan. Jakarta merupakan Ibu Kota Indonesia, mempunyai kedudukan vital sebagai inti sebuah negara dan pusat pemerintahan yang diatur dalam Undang - Undang. Jakarta sebagai ibu kota pastinya berbeda dari daerah lainnya meskipun mempunyai kewenangan yang serupa sebagai sebuah provinsi, pastinya Kota Jakarta sebagai ibu kota memiliki kedudukan eksklusif sebagai ibukota, yang mempunyai tugas sebagai tempat penguasaan politik, ekonomi, pembangunan dan bidang krusial lainnya kepada tiap-tiap provinsi yang lain. Jakarta dijadikan sebagai Ibu Kota negara diatur dalam Undang - Undang Republik Indonesia Nomor 10 tahun 1964 Tentang Pernyataan Daerah Khusus Ibukota Jakarta Raya Tetap Sebagai Ibu Kota Negara Republik Indonesia. (Hutasoit, 2018).

Dalam asal - usulnya, Ibu kota Indonesia sempat beberapa kali beralih ke Yogyakarta, serta Bukittinggi sebagai ibu kota pemerintahan darurat kala itu. Pertama, pemindahan pusat pemerintahan dari Jakarta ke Yogyakarta dilakukan ketika adanya Agresi Militer ke- I yang dilakukan oleh Belanda tepat pada tanggal 29 September 1945, berjarak lima bulan setelah pengumuman kemerdekaan.

Alurnya, tanggal 2 Januari 1946 Sultan Hamengkubuwana IX menugaskan utusan ke Jakarta agar ibu kota NKRI dialihkan ke Yogyakarta. Kemudian Presiden Soekarno mengalihkan ibu kota ke Yogyakarta pada tanggal 4 Januari 1946 yang merupakan awal perpindahan ibu kota yang pernah dilakukan dalam sejarah. Latar belakang yang mendasari peristiwa tersebut dikarenakan Belanda telah menguasai Jakarta. Disisi lain, Yogyakarta mempunyai kesiapan dari bidang ekonomi, bidang politik, dan bidang keamanan. Kemudian Presiden Soekarno membuat surat kuasa yang diberikan kepada Safruddin Prawiranegara di Bukit Tinggi guna membangun pemerintahan darurat pada waktu tersebut. Syarifudin Parwiranegara mendeklarasikan pembentukan Pemerintahan Darurat Republik Indonesia pada tanggal 22 Desember 1948 di Bukit Tinggi. Selepas Bukit Tinggi, ibu kota pernah beralih pada tempat yang dirahasiakan selama pemerintahan darurat Republik Indonesia (PDRI). Pemerintahan kembali ke Jakarta pada saat revolusi telah usai, tidak ditemukan urgensi lain mengenai rencana penunjukan ibu kota negara yang baru hingga dikemudian hari. (Ridho, 2020)

Pada masa kepemimpinan Presiden Joko Widodo, Usulan Pemindahan ibu kota baru mulai dibahas. Menurut Hadi & Ristawati, (2020), Wacana terkait pemindahan ibu kota di Indonesia dapat diketahui pada unsur:

1. Unsur efisiensi dan efektivitas, dijelaskan bahwa Jakarta dirasa kurang dalam evaluasi pada tingkat efektif dan efisien sebagai ibu kota, disebabkan oleh kepadatan di Jakarta, baik di transportasi darat, laut dan udara. Oleh karena hal tersebut, Presiden RI menetapkan pemindahan ibu kota yang awalnya akan direalisasikan secara berangsur pada tahun 2024.

2. Unsur pemerataan ekonomi. Pulau Jawa menjadi pusat pembangunan sejak Indonesia mendeklarasikan kemerdekaannya. Sehingga terbentuknya sifat sentralistik yang mengakibatkan ketimpangan apabila disandingkan dengan daerah lain di Indonesia. Pulau Jawa menjadi pusat Lembaga dan institusi besar. 54 persen dari seluruh penduduk Indonesia (150 juta penduduk) dan 58 persen PDB ekonomi Indonesia bertempat di Pulau Jawa. Harapan pemindahan ibu kota adalah terjadinya siklus perekonomian yang lebih luas sehingga tidak berpusat di Pulau Jawa.

DPR mengesahkan Undang - Undang (UU) Nomor 3 Tahun 2022 pada 18 Januari 2022 tentang IKN, lalu Presiden Joko Widodo telah menandatangani secara resmi UU tentang IKN pada tanggal 15 Februari 2022. Dengan demikian, Ibu kota Indonesia IKN direlokasikan ke Kabupaten Penajam Paser Utara dan Kabupaten Kutai Kartanegara, Provinsi Kalimantan Timur. Barometer dipilihnya Kalimantan Timur untuk dibangun IKN yang baru adalah akses lokasi yang mudah, lebih dekat dengan kota besar di sekitar wilayah tersebut, penduduk yang terbuka dan heterogen, rendahnya potensi terjadinya konflik, Didukung oleh Tri Matra Darat, Laut, Udara, Infrastruktur yang mumpuni, Air alami didapatkan dari 3 waduk, 2 waduk yang diproyeksikan akan dibangun, 4 sungai strategis, dan 4 Daerah Aliran Sungai. (Saraswati et al., 2022)

Keputusan pemindahan ibu kota negara oleh pemerintah menimbulkan perbincangan dalam masyarakat, khususnya pada sosial media. Media sosial yang memiliki banyak pengguna khususnya di Indonesia adalah Twitter. Twitter merupakan media sosial yang dikembangkan oleh Jack Dorsey yang berfungsi

untuk mengunggah pesan yang dikenal sebagai kicauan (*tweet*). Twitter digunakan oleh kalangan masyarakat sebanyak 59% dari total *media social* yang diakses serta menjadi peringkat ke-5 media sosial yang sering diakses pada tahun 2020. Maka dari itu, Twitter menjadi platform media sosial yang berpengaruh bagi para pengguna aktif di Indonesia (Krisdiyanto et al., 2021). Menurut (Malik Zuhdi et al., 2019), data yang dihimpun pada Twitter dapat bermanfaat jika dianalisis dikarenakan data yang diolah dari Twitter dapat dimanfaatkan sebagai informasi penting melalui *opinion mining*.

Kumpulan *tweet* yang ada pada Twitter menjadikannya sumber data dalam bentuk opini, maka diperlukan teknik *text mining* untuk mengumpulkan data. *Text mining* adalah teknik yang dimanfaatkan sebagai proses klasifikasi, *text mining* adalah ragam *data mining* yang berfokus untuk mencari pola menarik dari kumpulan data bersifat tekstual dalam jumlah yang besar (Ratniasih et al., 2017).

Data dari *tweet* yang berupa teks kemudian diolah menggunakan klasifikasi. Salah satu Algoritma klasifikasi adalah *Naïve Bayes*. Apabila disejajarkan dengan metode klasifikasi lainnya, *Naive Bayes* dapat mengolah data yang cukup sedikit untuk proses latihnya. Performa data latih yang dilakukan pada informasi yang dikumpulkan relatif cepat, dan menggunakan sedikit *storage* pada proses latih, dan klasifikasi yang cenderung ringan untuk digunakan serta mempunyai lebih sedikit kriteria dari metode lain seperti Jaringan Neural dan *Support Vector Machine*. (Al-Aidaros et al., 2010)

1.2 Pernyataan Masalah

Seberapa besar tingkat sentimen masyarakat dalam pengklasifikasian teks sentimen masyarakat terhadap proses pemindahan ibu kota negara (IKN) Indonesia pada media sosial *twitter* menggunakan metode *Gaussian Naïve Bayes Classifier*?

1.3 Tujuan Penelitian

Mengukur tingkat sentimen masyarakat dalam pengklasifikasian teks sentimen masyarakat terhadap proses pemindahan ibu kota negara (IKN) Indonesia pada media sosial *twitter* menggunakan metode *Gaussian Naïve Bayes Classifier*

1.4 Manfaat Penelitian

1. *Output* dari penelitian dapat mengidentifikasi sentimen masyarakat terhadap proses pemindahan Ibu Kota Negara (IKN) pada media sosial Twitter.
2. Bermanfaat sebagai sumber informasi bagi peneliti selanjutnya dalam penelitian pada bidang yang sesuai atau sejenis.

1.5 Batasan Masalah

1. Data yang dianalisis merupakan *tweet* dengan Bahasa Indonesia.
2. Data *Tweet* yang dikumpulkan tidak mengandung gambar.
3. Pengambilan data dilakukan pada periode 15 Februari 2023 – 25 Februari 2023 dengan kata kunci ibu kota pindah, ibu kota negara, pemindahan ibu kota, #ikn, dan #ibukotabaru

1.6 Sistematika Penulisan

Sistematika penulisan dalam penyusunan laporan penelitian adalah sebagai berikut:

1. Bab I Pendahuluan

Bab ini menjelaskan tentang latar belakang penelitian, pernyataan masalah, tujuan penelitian, batasan masalah, manfaat penelitian dan sistematika penulisan.

2. Bab II Kajian Pustaka

Bab ini menjelaskan tentang pembahasan teori dari berbagai literatur terkait dengan pembahasan dalam penelitian.

3. Bab III Desain dan Implementasi Sistem

Bab ini berisikan tentang proses pengambilan data, pembangunan sistem dan skenario pengujian.

4. Bab VI Uji Coba dan Pembahasan

Bab ini menjelaskan tentang hasil klasifikasi data teks menggunakan algoritma *Naïve Bayes*.

5. Bab V Kesimpulan dan Saran

Bab ini menjelaskan tentang kesimpulan dan saran untuk pengembangan penelitian terkait selanjutnya.

6. Daftar Pustaka

Daftar Pustaka berisi tentang referensi penelitian yang dirujuk dalam pembuatan penelitian ini.

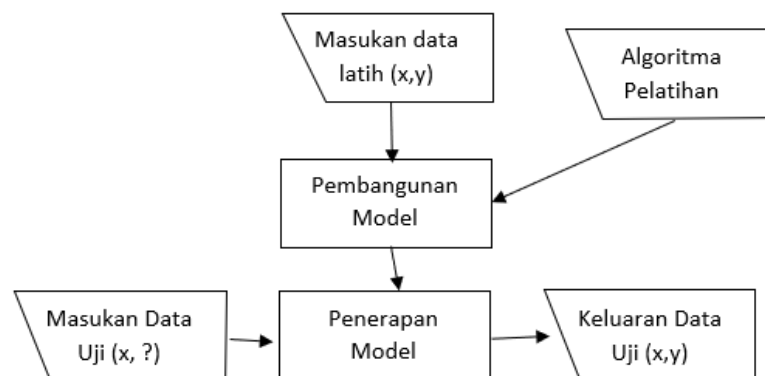
BAB II

STUDI PUSTAKA

2.1 Klasifikasi

Pertama kali, klasifikasi diterapkan dalam konteks tanaman dengan tujuan mengelompokkan spesies tertentu. Carolus von Linne, sebagai orang yang memprakarsai pengklasifikasian ini, melakukan pengelompokan pertama kali berdasarkan ciri-ciri fisik spesies. Akibatnya, ia dikenal sebagai pelopor sistem klasifikasi dan dihormati sebagai "bapak klasifikasi" (Yuli Mardi, 2019).

Menurut (Imron, 2019), konsep model dalam klasifikasi dapat disamakan dengan kotak hitam, di mana sebuah model menerima *input* dan secara internal memproses informasi tersebut untuk menghasilkan *output* berupa jawaban. Ilustrasi kerangka kerja klasifikasi dapat ditemukan dalam Gambar 2.1. Gambar tersebut menunjukkan adanya sekumpulan data latih (x, y) yang digunakan untuk melatih model. Setelah dilatih, model tersebut digunakan untuk memprediksi kelas dari data uji ($x, ?$), sehingga kelas sebenarnya y dapat diketahui.



Gambar 2.1 Proses Klasifikasi

Menurut (Annur, 2018), Proses klasifikasi melibatkan penciptaan model atau fungsi yang mampu menggambarkan dan membedakan kelas-kelas data atau konsep. Tujuan utamanya adalah untuk menggunakan model tersebut dalam memprediksi kelas dari objek yang tidak memiliki label kelas. Ada berbagai jenis klasifikasi yang umum digunakan, seperti pohon keputusan/klasifikasi, klasifikasi Bayesian (*Naïve Bayes*), jaringan saraf (*Neural Networks*), metode *fuzzy*, dan Mesin Vector Dukungan (*Support Vector Machines*) (Wibawa et al., 2018). Menurut (Malik Zuhdi et al., 2019), Dalam rangka melakukan pengklasifikasian, terdapat dua tahapan utama yang dilakukan, yaitu:

1. Tahap pelatihan (*training*) di mana model dibangun menggunakan kumpulan data pelatihan yang telah diberi label.
2. Tahap pengujian (*testing*) digunakan untuk mengevaluasi akurasi model yang telah dibangun pada tahap pelatihan. Biasanya, data yang disebut sebagai set data uji (*test set*) digunakan untuk memprediksi label.

2.2 Pemindahan Ibu Kota Negara

Definisi resmi mengenai ibu kota di Indonesia tidak secara tegas diatur dalam peraturan perundang-undangan. Namun, jika ditafsirkan secara implisit berdasarkan Undang - Undang Nomor 29 Tahun 2007, terdapat penjelasan dalam Pasal 5 yang menyebutkan bahwa ibu kota dimaksudkan sebagai tempat kedudukan lembaga pusat yang mencakup pemerintahan eksekutif, legislatif, dan yudikatif. Selain itu, ibu kota juga merupakan tempat kedudukan perwakilan negara asing dan kantor perwakilan lembaga internasional (Saraswati et al., 2022).

Setelah Undang - Undang Nomor 3 Tahun 2022 tentang Ibu Kota Negara diterbitkan dan dimasukkan ke dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN) Tahun 2020 hingga 2024, proyek pemindahan Ibu Kota Negara (IKN) mendapat respons yang beragam dari masyarakat, seperti yang sering terjadi dalam negara-negara demokrasi. Kebijakan pemindahan IKN didasarkan pada kebutuhan untuk mengembangkan wilayah ekonomi baru, mengurangi kesenjangan antar wilayah, dan meningkatkan kesejahteraan masyarakat. Kebijakan pemindahan ini juga diambil sebagai respons terhadap potensi konflik sosial dan degradasi lingkungan yang dapat terjadi (Saraswati et al., 2022).

2.3 Sosial Media Twitter

Twitter adalah platform *microblogging* yang merupakan bentuk blog sederhana yang memungkinkan pengguna untuk memposting pesan-pesan singkat. Dengan menggunakan Twitter, pengguna dapat berbagi berbagai informasi dalam berbagai format, seperti teks, gambar, grafik, audio, dan video (Mardiana & Zi'ni, 2020). Menurut (Zuhdi et al., 2019), terdapat beberapa istilah yang dipakai dalam twitter, antara lain:

1. *Mention* merupakan penyebutan *user* Twitter lainnya yang sedang berinteraksi atau hendak berinteraksi dengan pemilik nama pengguna Twitter yang bersangkutan. Penulisan *mention* menggunakan '@' dengan diikuti nama pengguna yang hendak disebut.
2. *Hashtag* adalah penanda sebuah topik pembicaraan di Twitter. *Hashtag* biasanya digunakan untuk mengidentifikasi topik yang dimuat dalam *tweet* yang di dalamnya terdapat label *hashtag*. *hashtag* ditulis

menggunakan ‘#’ kemudian penulisan nama topik terkait. *Hashtag* sering kali dimanfaatkan untuk menaikkan tingkat visibilitas suatu topik atau tweet agar dapat dimunculkan dengan topik terkait yang sedang banyak diperbincangkan pada sosial media Twitter atau sering disebut *trending topics*.

3. *Trending Topics* adalah himpunan topik yang dibicarakan yang mengalami peningkatan atensi dalam Twitter.

Untuk mengumpulkan data berupa tweet, Twitter menyediakan API yang diperuntukkan untuk *developer* yang terdapat pada <https://developer.twitter.com>. Menurut Zuhdi et al., (2019), API (*Application Programming Interface*) adalah sekumpulan perintah, fungsi, dan protokol yang digunakan dalam pengembangan perangkat lunak untuk sistem tertentu. API juga berfungsi sebagai dokumentasi yang menggambarkan antarmuka, fungsi, kelas, dan struktur yang digunakan dalam pengembangan perangkat lunak. Dengan adanya API, seorang programmer dapat lebih mudah mengintegrasikan perangkat lunak yang sedang dikembangkan dengan perangkat lunak lainnya. Dalam penelitian terkait, Zuhdi et al., (2019), juga menjelaskan dua tipe Twitter API yang tersedia :

1. *Twitter REST API*

Termuat dari *Twitter REST* dan *Twitter Search*. Data inti dan objek inti Twitter diambil dari *Twitter REST*, sedangkan proses pencarian tentang suatu objek maupun *trending topics* yang sedang diperbincangkan diambil dari *Twitter Search*.

2. *Twitter Streaming API*

Twitter Streaming API dimanfaatkan sebagai alat penggali data. Informasi didapatkan secara *realtime* dengan volume yang tinggi yang berasal dari *Streaming* API.

2.4 Pengolahan Bahasa Alami

Pengolahan Bahasa Alami (PBA) atau *Natural Language Processing* (NLP) merupakan komponen penting dalam *text mining* dan merupakan subbidang dari kecerdasan buatan atau *Artificial Intelligence*. PBA berfokus pada pengembangan komputer yang memiliki kemampuan untuk memahami dan menafsirkan makna bahasa manusia. Ini dilakukan dengan mengubah bahasa manusia menjadi bentuk dokumen atau teks yang lebih terstruktur, agar lebih mudah untuk diproses oleh program komputer dan memberikan respons yang sesuai. Tujuan utama PBA adalah melampaui manipulasi teks yang hanya berfokus pada sintaksis (dikenal juga sebagai '*wordcounting*') untuk mencapai pemahaman yang lebih baik dan memproses bahasa alami dengan mempertimbangkan batasan semantik, gramatikal, dan kontekstual (Imron, 2019)

2.5 Text Mining

Text mining adalah suatu proses yang melibatkan pemahaman mendalam terhadap sejumlah dokumen dari waktu ke waktu dengan menggunakan rangkaian alat analisis. *Text mining* ini bertujuan untuk menganalisis dan mengidentifikasi pola menarik, termasuk tren, data yang berbeda dari yang diharapkan, serta konsep pencarian yang tidak memiliki konten atau bahkan yang relevan. (Setianingrum et al., 2018)

2.6 *Naïve Bayes Classifier*

Naïve Bayes adalah salah satu algoritma yang memiliki kemampuan untuk melakukan pengklasifikasian dengan kecepatan tinggi. Selain itu, *Naïve Bayes* juga dikenal sebagai algoritma yang sangat efisien dan efektif bahkan ketika digunakan untuk menganalisis data dalam skala besar (Ahmad et al., 2017). Menurut (Annur, 2018), *Naïve Bayes* merupakan metode statistik dalam pengklasifikasian yang dapat digunakan untuk memperkirakan probabilitas suatu objek menjadi anggota dari kelas tertentu. Penelitian tersebut memperoleh nilai akurasi sebanyak 73%.

Pada penelitian terkait, mendapatkan tingkat akurasi sebesar 89% pada penelitian Klasifikasi Surat Keluar (Setianingrum et al., 2018). Dalam penerapan algoritma *Naïve Bayes* untuk pengkategorian, terdapat langkah-langkah yang melibatkan pencarian probabilitas yang relevan untuk melakukan prediksi di masa depan berdasarkan data historis. Untuk melaksanakan langkah-langkah tersebut, *dataset* diperlukan sebagai sumber data. Dalam algoritma *Naïve Bayes*, *dataset* tersebut dibagi menjadi dua *subset*, yaitu data latih dan data uji. (Sidiq & Christanto, 2020).

2.7 Integrasi Islam

1. Musyawarah dan pengambilan keputusan

Ada ayat yang terdapat dalam Al-Qur'an yang membahas tentang musyawarah dan pentingnya konsultasi dalam pengambilan keputusan . QS.

Asy-Syura: 38 dalam Al-Qur'an menyatakan,

وَالَّذِينَ اسْتَجَابُوا لِرَبِّهِمْ وَأَقَامُوا الصَّلَاةَ وَأَمْرُهُمْ شُورَىٰ بَيْنَهُمْ وَمِمَّا رَزَقْنَاهُمْ يُنفِقُونَ ؕ

"Dan mereka (orang-orang yang menjalankan urusan mereka) melaksanakan urusan mereka dengan musyawarah di antara mereka."

Ayat ini dapat diintegrasikan dalam konteks pemindahan ibukota negara dengan beberapa cara:

1. **Pentingnya Musyawarah:** Ayat ini menegaskan pentingnya musyawarah dalam mengambil keputusan. Dalam konteks pemindahan ibukota negara, musyawarah dapat menjadi mekanisme yang diterapkan untuk melibatkan berbagai pihak yang terkait, seperti pemerintah, ahli, masyarakat sipil, dan pemangku kepentingan lainnya. Melalui musyawarah, pendapat dan perspektif dari berbagai pihak dapat didengar dan dipertimbangkan sebelum pengambilan keputusan.
2. **Partisipasi Aktif:** Ayat ini juga menunjukkan bahwa mereka yang terlibat dalam menjalankan urusan mereka ikut terlibat dalam musyawarah. Dalam konteks pemindahan ibukota negara, partisipasi aktif dari berbagai pihak yang terlibat, seperti warga masyarakat, organisasi masyarakat, dan pemimpin lokal, sangat penting. Melalui musyawarah, semua pihak dapat berkontribusi dengan pendapat, masukan, dan saran mereka dalam proses pengambilan keputusan.
3. **Pembentukan Keputusan yang Konsensus:** Musyawarah bertujuan untuk mencapai kesepakatan yang berbasis pada konsensus. Dalam konteks pemindahan ibukota negara, musyawarah harus diarahkan pada mencapai kesepakatan yang melibatkan banyak pihak yang terkait.

Proses musyawarah ini akan memungkinkan diskusi terbuka, pertukaran pandangan, dan upaya untuk mencapai kesepakatan bersama sebelum mengambil keputusan terkait pemindahan ibukota.

4. Keputusan yang Dilaksanakan dengan Persetujuan Bersama: Setelah melalui proses musyawarah, keputusan tentang pemindahan ibukota negara harus diambil dengan persetujuan bersama dari pihak-pihak yang terlibat. Persetujuan ini mencerminkan semangat musyawarah dan konsensus yang dicari dalam mencapai keputusan tersebut. Keputusan yang diambil berdasarkan musyawarah dan persetujuan bersama memiliki potensi lebih kuat untuk mendapatkan dukungan dan implementasi yang lebih baik dari masyarakat.

Dengan mempertimbangkan ayat QS. Asy-Syura: 38, musyawarah tentang pemindahan ibukota negara dapat dilakukan dengan dasar - dasar musyawarah yang ditunjukkan pada Al-Qur'an, yaitu menghormati pendapat semua pihak terkait, mencari kesepakatan melalui diskusi terbuka, dan mencapai keputusan dengan persetujuan bersama.

2. Perpindahan Tempat

Dalam Al-Qur'an, terkandung beberapa ayat yang mengkaji terkait hijrah (migrasi) dan perpindahan tempat. Meskipun tidak secara langsung mengacu pada pemindahan ibukota negara, konsep hijrah dapat dihubungkan dengan konteks penelitian tentang proses pemindahan ibukota negara. Dalam konteks pemindahan ibukota negara, penelitian dapat mengkaji aspek penggantian dan

peralihan dalam infrastruktur, pemerintahan, dan kehidupan sosial ekonomi ketika ibukota negara dipindahkan.

وَمَنْ يُهَاجِرْ فِي سَبِيلِ اللَّهِ يَجِدْ فِي الْأَرْضِ مُرَاعًا كَثِيرًا وَسَعَةً ۚ وَمَنْ يَخْرُجْ مِنْ بَيْتِهِ مُهَاجِرًا إِلَى اللَّهِ وَرَسُولِهِ ثُمَّ يُدْرِكْهُ الْمَوْتُ فَقَدْ وَقَعَ أَجْرُهُ عَلَى اللَّهِ ۗ وَكَانَ اللَّهُ غَفُورًا رَحِيمًا

"Barangsiapa yang berhijrah di jalan Allah, niscaya mereka mendapati di muka bumi ini tempat hijrah yang luas dan rezeki yang banyak." (QS. An-Nisa: 100)

Ayat tersebut menunjukkan janji Allah SWT kepada orang-orang yang melakukan hijrah di jalan-Nya, bahwa mereka akan menemukan tempat hijrah yang lapang dan rezeki yang melimpah. Dalam konteks pemindahan ibukota negara, penelitian dapat mengkaji potensi manfaat ekonomi, sosial, dan lingkungan yang dapat diperoleh dengan pemindahan ibukota ke lokasi baru.

Meskipun ayat ini tidak secara langsung membahas pemindahan ibukota negara, konsep hijrah dalam Al-Qur'an dapat memberikan pemahaman dan perspektif tentang perpindahan tempat dan perubahan yang signifikan. Penelitian tentang sentimen masyarakat terhadap proses pemindahan ibukota negara dapat mengeksplorasi pandangan Al-Qur'an tentang hijrah serta nilai-nilai yang dapat diterapkan dalam konteks pemindahan tersebut, seperti komitmen, ketekunan, penyesuaian, dan manfaat yang diperoleh dari perubahan tersebut.

2.8 Penelitian Terkait

Terdapat penelitian terkait sejenis yang membahas klasifikasi sentimen menggunakan *Naïve Bayes* yang dapat dilihat pada tabel 2.1. Pada uraian tabel

tersebut dijelaskan Alur yang diterapkan dalam penelitian terkait. Pembaharuan dalam penelitian ini menggunakan berbagai macam proses tambahan untuk mendapatkan nilai akurasi yang maksimal, di antaranya:

1. *Preprocessing*

Menambahkan tambahan Langkah-langkah *preprocessing* dokumen dengan menggunakan *slangword remover* dan *emoji remover*. Dua langkah ini belum diterapkan pada penelitian terkait.

2. Utama

Dalam penelitian ini, digunakan metode *Gaussian Naïve Bayes* yang didasarkan pada asumsi bahwa data dari setiap label berasal dari distribusi *Gaussian* yang sederhana.

3. *Post-processing*

Menggunakan data latih dan data uji yang paling optimal dengan cara dihitung satu persatu, kemudian dihitung juga pembagian persentase fitur yang digunakan untuk mencari akurasi terbaik.

Tabel. 2.1 Pembaharuan dengan Metode Penelitian Terkait

No.	Riset	Metode		
		Preprocessing	Utama	Post
1.	Algoritma <i>Naïve Bayes</i> untuk Penentuan PKH (Program Keluarga Harapan) Berbasis Sistem Pendukung Keputusan (Studi Kasus: Kelurahan Karanganyar Gunung Semarang)	Menggunakan metode <i>Fishbone</i> Analisis dalam penulisannya	<i>Naïve Bayes</i>	perhitungan secara independen dan sistem melalui <i>Confution Matrix</i> menggunakan <i>RapidMiner</i> .
2.	Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan <i>Naïve Bayes Clasifiers</i>	Menggunakan metode <i>Casefolding, Tokenizing, Filtering, Cleansing, Stemming.</i>	<i>Naive Bayes Classifier</i>	Tidak dijelaskan pembagian secara rinci tetapi menyebutkan jumlah dokumen sebesar 1000 untuk mencari akurasi.
3.	Analisa Sentimen Terhadap Tempat Wisata di Kabupaten Rembang Menggunakan Metode <i>Naïve Bayes Classifier</i>	<i>Cleansing, Tokenizing, Stopword Removal, Stemming,</i>	<i>Naive Bayes Classifier</i>	Menghasilkan akurasi, presisi dan <i>recall</i> cukup baik, sehingga dapat diaplikasikan ke dalam sistem

No.	Riset	Metode		
		Preprocessing	Utama	Post
4.	Implementasi Algoritma <i>Multinomial Naïve Bayes</i> Sistem Klasifikasi Surat Keluar (Studi Kasus : DISKOMINFO Kabupaten Tangerang)	<i>Casefolding, Tokenizing, Stopword Removal, Stemming, seleksi fitur TF-IDF</i>	<i>Multinomial Naïve Bayes</i>	Pembagian 23 <i>document training</i> dan 1 dokumen testing untuk menemukan akurasi
5.	Sentiment Analysis Menggunakan <i>Naïve Bayes Classifier</i> Pada <i>Tweet</i> Tentang Zakat	<i>Remove duplicate, casefolding, filtering, tokenizing, Normalization, Stopword Removal, Stemming</i>	<i>Naive Bayes Classifier</i>	Pembagian data latih 950 dokumen training dan 50 dokumen testing untuk menemukan akurasi
6.	Klasifikasi Sentimen Masyarakat Terhadap Proses Pindahan Ibu Kota Negara Indonesia Pada Media Sosial <i>Twitter</i> Menggunakan Metode <i>Naïve Bayes</i>	<i>Slangword normalization, emoji remover, Casefolding, Tokenizing, Stopword Removal, Stemming, seleksi fitur TF-IDF</i>	<i>Gaussian Naïve Bayes</i>	Jumlah total terdapat 500 dokumen dengan sentimen positif sebanyak 317, dan sentimen negatif sebanyak 183. Pembagian data latih 350 dokumen <i>training</i> dan 150 dokumen <i>testing</i> serta pengujian dengan berbagai macam persentase fitur untuk menemukan akurasi.

BAB III

DESAIN DAN IMPLEMENTASI SISTEM

3.1 Pengumpulan Data

Data yang digunakan sebagai objek penelitian dikumpulkan dari Sosial Media Twitter. Data Objek penelitian ini adalah *tweet* dari pengguna aktif Twitter. *Tweet* yang mengandung opini atau komentar warga negara Indonesia dikumpulkan menggunakan teknik *scraping*. Data dikumpulkan dengan menggunakan kata kunci “ibu kota pindah, ibu kota negara, pemindahan ibu kota” serta tagar #ikn, dan #ibukotabaru yang menggunakan bahasa Indonesia. Keseluruhan data yang telah dikumpulkan melalui proses *scraping* pada periode tanggal 15 Februari 2023 hingga 25 Februari 2023 sebanyak 500 *tweet*. Data yang berhasil dikumpulkan kemudian disimpan menjadi *output* di dalam berkas *.csv*. Data yang telah disimpan kemudian diolah dengan tahapan pelabelan yang bertujuan menentukan label pada tiap kelas *tweet* yang telah melalui proses *scraping*.

105	ririnurin	@kemalugraha441 @maryshelparaiso IKN mensejahterakan warga lokal #bukotaNegara #IKN #SemangatNusantara https://t.co/hk7nictov
106	stikumiyarti	Gus A Muhamin Iskandar memberikan tiga nilai tambah jika Ibu Kota Nusantara (IKN) terealisasi. #gusmuhamin #muhaminiskandar #ikn #stikumiyarti https://t.co/S30XoYmIml
107	jpnconm	Direktur Pengelolaan Gedung dan Kawasan IKN Raih Gelar Doktor di ITB, Selamat #IKN https://t.co/swDNBLVMT
108	tribunkaltim	Poin Draft Revisi UU IKN Nusantara, Presiden Terpilih 2024 Wajib Melanjutkan, Ada Pengecualian? #IKN #IbuKotaNegara #IKN #Nusantara #PenajamPaserUtara #PPU #Kaltim #KalimantanTimur #revisiuiKN #UIKN https://t.co/c
109	indopos_id	Kementerian PUPR Bangun Ekosistem Digital Untuk Pekerja di IKN https://t.co/kGeOCyYvmg #IKN #KementerianPUPR #PekerjaKonstruksi
110	MKaruniasa	Perancang dan penghuninya juga harus berkelas dunia#Lingkungan #IKN https://t.co/8s85g8ahm7
111	ikn_id	Untuk itu, terdapat target yang harus dicapai agar IKN dapat menjadi Ibu Kota kelas dunia 🇮🇩🇨🇸 simak dalam infografik berikut! #KotaDuniauntukSemua #Nusantara #IbuKotaNegara #IKN #Indonesia https://t.co/mQ6UoiMG6
112	alfino805	Pemerintah Susun Studi Jalur Kereta Api Rute Balikpapan-IKN #IKN #Kaltim #PBB13 #Pemilu2024 https://t.co/PqgHmDrZg
113	KristyRives	Presiden Jokowi optimistis dapat menyelenggarakan upacara kemerdekaan Republik Indonesia di Ibu Kota Nusantara (IKN) pada tahun 2024 nanti. #IKN #pemerintahhebat #Indonesia #jokowi #jokowipresident
114	LoriannePerdomo	Pembangunan Ibu Kota Negara (IKN) Nusantara di Kalimantan Timur telah mencapai 22 persen hingga akhir Februari 2023 ini. #IKN #pemerintahhebat #Indonesia #jokowi #jokowipresident #RI1 #nusantara
115	BagusPrasasta	Saat ini sedang dilakukan finalisasi UU IKN untuk menambahkan poin salah satunya mengenai kewajiban bagi presiden terpilih selanjutnya untuk melanjutkan pembangunan IKN. #IKN #pemerintahhebat #Indonesia #jokowi #jokowipresident
116	BethVigil17	Dari sisi investasi, pembangunan IKN tidak menemui kendala yang berarti. Sekretaris OIKN Achamd Jaka Santos Adiwijaya mengatakan. #IKN #pemerintahhebat #Indonesia #jokowi #jokowipresident #RI1 #
117	gantar_mandala	Presiden Jokowi optimistis dapat menyelenggarakan upacara kemerdekaan Republik Indonesia di Ibu Kota Nusantara (IKN) pada tahun 2024 nanti. #IKN #pemerintahhebat #Indonesia #jokowi #jokowipresident
118	Margare6347595	Pembangunan Ibu Kota Negara (IKN) Nusantara di Kalimantan Timur telah mencapai 22 persen hingga akhir Februari 2023 ini. #IKN #pemerintahhebat #Indonesia #jokowi #jokowipresident #RI1 #nusantara
119	gaman_pradana	Untuk menjaga keberlanjutan pembangunan IKN dan adanya kepastian bagi investor. #IKN #pemerintahhebat #Indonesia #jokowi #jokowipresident #RI1 #nusantara #fyp #viral #2023 #2024 https://t.co/4cif
120	Shannon5556167	Sekretaris OIKN Achamd Jaka Santos Adiwijaya mengatakan, saat ini sudah ada lebih dari 150 letter of intent (LoI) dari investor yang diterima Otorita IKN. #IKN #pemerintahhebat #Indonesia #jokowi #jokowip
121	yulia_purwanti0	Presiden Jokowi optimistis dapat menyelenggarakan upacara kemerdekaan Republik Indonesia di Ibu Kota Nusantara (IKN) pada tahun 2024 nanti. #IKN #pemerintahhebat #Indonesia #jokowi #jokowipresident
122	azalea3737	Pembangunan Ibu Kota Negara (IKN) Nusantara di Kalimantan Timur telah mencapai 22 persen hingga akhir Februari 2023 ini. #IKN #pemerintahhebat #Indonesia #jokowi #jokowipresident
123	FMB9iD	Apa saja fokus tahap pertama pembangunan di IKN? Cek infografis berikut ini! #FMB9iD #NUSANTARA #IKN https://t.co/Eh8xhriQY1
124	jiva_tertukar	@ririnurin @maryshelparaiso IKN pasti membawa manfaat untuk seluruh wilayah Indonesia #bukotaNegara #IKN #SemangatNusantara https://t.co/LalEqMCTE
125	EisseeJea	Pembangunan IKN berdampak baik pemerataan nasional #IKN #bukotaBaru #mendunia #jokowi #presidenku https://t.co/6WwTbNUNti
126	tribunkaltim	Anies Baswedan sebut akan lanjutkan IKN Nusantara bila jadi Presiden, ini alasannya! https://t.co/yjx9Zv18IKN #IKN #Nusantara #bukotabaru #beritaterkini #beritaterbaru
127	HippySanto	@VIVAbola pembangunan IKN harus didukung sih #IKN #bukota #bukotabaru #bukotaNegara #bukotaNegaraNusantara #dukungin https://t.co/w7AXbDYBSS
128	WMandalaka	@ArnoldPoernomo pembangunan IKN harus didukung sih #IKN #bukota #bukotabaru #bukotaNegara #bukotaNegaraNusantara #dukungin https://t.co/d0WMY4SLQ5
129		

Gambar 3.1 Data Hasil Scraping

Gambar 3.1 menunjukkan data yang telah di *scraping* menggunakan API *twitter* menggunakan kata kunci #IKN. Terdapat 500 data yang berhasil disimpan ke dalam file berformat .csv.

```

!pip install tweepy

from google.colab import drive
import tweepy
import csv

consumer_key = '4buP4RyjdWgCcxzTIkdaprBEfq'
consumer_secret =
'5urxstKNpbkAsFeOodLO3sg0vWnbTL4BA5l7pMsxUNrNSYkj14'
access_token = '548170515-
fZ07iDhJLYFh12aeVsegDGS5yYAvBxwo45JzDtB6'
access_token_secret =
'FKC8db04dZxoGGqYUoBCIMzxxwuL0qEwjZKwW1Hnyera'

from google.colab import drive
drive.mount('/content/drive')

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)

api = tweepy.API(auth)

csvFile = open('/content/drive/My Drive/Colab Notebooks/data.2-
25.csv', 'a')
#Use csv writer
csvWriter = csv.writer(csvFile)

for tweet in tweepy.Cursor(api.search_tweets, lang="id",
q="IKN",until="2023-2-25", tweet_mode="extended").items(200):
    if not (hasattr(tweet, 'retweeted_status')): # to avoid
retweet
        csvWriter.writerow([tweet.user.screen_name,
tweet.id, # saving only user name and tweet
tweet.full_text.replace("\n", "")])
csvFile.close()

```

Source code di atas digunakan untuk mengambil data yang berupa *tweet* menggunakan *API Twitter*. Ditentukan juga kata kunci dan tanggal yang dibutuhkan dalam proses pencarian *tweet*. Kemudian data hasil *scraping* dimasukkan ke dalam file yang berformat .csv.

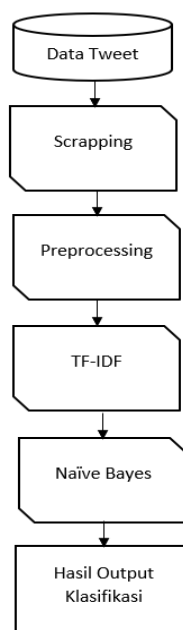
Pada tahap ini, data yang telah dikumpulkan melalui proses *scraping* akan dilakukan pelabelan secara manual. Tidak semua data akan melewati tahap pelabelan, namun dilakukan pemilihan sampel secara acak untuk kemudian diberi label, seperti yang ditunjukkan dalam Tabel 3.1. Pelabelan dilakukan dengan mengamati opini dalam data. Metode pelabelan ini melibatkan tinjauan opini secara individu, di mana opini tersebut akan dikategorikan sebagai sentimen positif atau negatif. Keuntungan dari metode pelabelan ini adalah hasil yang diperoleh akan lebih akurat dengan sentimen asli, dan tentunya tidak memerlukan tahapan komputasi. Proses pelabelan dilakukan oleh Ibu Purwati, SE. selaku Kepala Bidang Ideologi Wawasan Kebangsaan dan Kewaspadaan Nasional pada Badan Kesatuan bangsa dan Politik Kota Pasuruan.

Tabel 3.1 Pelabelan

Dokumen	Tweet	Label
D1	@dragon3mpror @maryshelparaiso warga dukung kota masa depan ikn #IbukotaNegara	Mendukung
D2	@edyaldo1 @maryshelparaiso ih keren bgt desain IKN ini #IbukotaNegar	Mendukung
D3	Potensi IKN MANGKRAK Menghamburkan Uang Negara, Kereta Cepat Yang Katanya Tidak Menggunakan APBN dll Apabila Di Usut Tuntas.	Tidak Mendukung

3.2 Rancangan Sistem

Rancangan pada sistem yang diterapkan secara sistematis dalam penelitian ini akan dibangun menggunakan bagan seperti pada gambar 3.1. Rancangan sistem menjelaskan bagaimana proses penelitian ini berlangsung, diawali dari tahap *collecting data* hingga terdapat *output* klasifikasi sehingga terjadi proses kalkulasi pada data sesuai dengan tujuan penelitian.

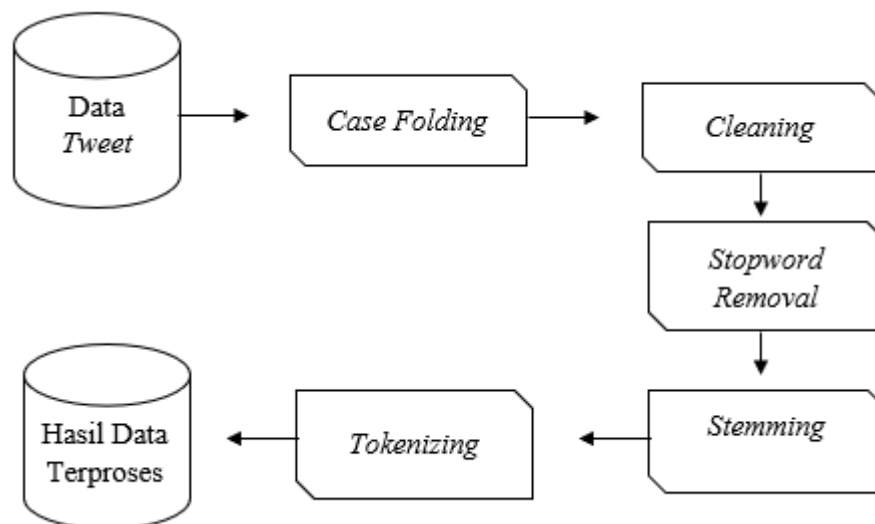


Gambar 3.2 Rancangan Sistem

Rancangan pada sistem diawali melalui Tahap pencarian data pada Twitter dengan menggunakan kata kunci yang telah ditentukan, kemudian data dikumpulkan menggunakan teknik *scraping*. Setelah data disimpan, tahap berikutnya adalah proses *preprocessing* dan proses TF-IDF. Setelah melalui tahap pengolahan data, kemudian data tersebut di melalui proses klasifikasi dengan menggunakan *Naïve Bayes* untuk selanjutnya menuju tahap akhir berupa hasil klasifikasi.

3.3 *Preprocessing*

Preprocessing merupakan proses untuk mengubah sebuah suatu dokumen abstrak menjadi terstruktur dengan cara menghilangkan atribut yang tidak dibutuhkan sehingga data menjadi sistematis dan mengurangi *noise*. Penggunaan dari *preprocessing* diperlukan karena data yang diambil berupa opini pengguna *Twitter* yang tidak ada batasan dalam penggunaan kata dalam kalimat sehingga mempengaruhi hasil dari proses klasifikasi. Data abstrak akan diolah sehingga terkonversi menjadi data yang lebih terstruktur dengan *preprocessing* melalui tahapan yang divisualisasikan pada gambar 3.2.



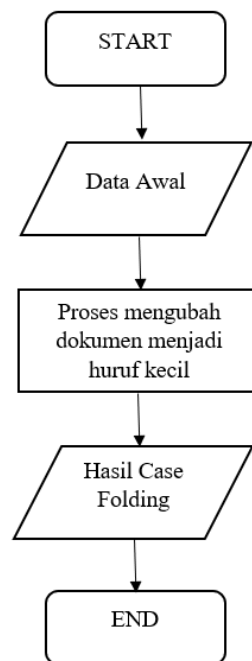
Gambar 3.3 Alur Preprocessing

Implementasi Preprocessing dalam penelitian ini dapat dilihat pada *source code* berikut mulai dari tahap cleaning hingga *tokenizing*.

```
def cleaningText(text):
    text = re.sub(r'@[A-Za-z0-9]+', '', text) # remove
    mentions
    text = re.sub(r'#[A-Za-z0-9]+', '', text) # remove
    hashtag
    text = re.sub(r'RT[\s]', '', text) # remove RT
    text = re.sub(r"http\S+", '', text) # remove link
    text = re.sub(r'[0-9]+', '', text) # remove numbers
    text = re.sub(r'[\^2]+', '', text) # remove repetition
    text = text.replace('\n', ' ') # replace new line
    text = text.translate(str.maketrans('', '',
string.punctuation)) # remove all punctuations
    text = text.strip(' ') # remove characters space from
    both left and right text
    return text
def casefoldingText(text): # Converting all the characters
    in a text into lower case
    text = text.lower()
    return text
def tokenizingText(text): # Tokenizing or splitting a
    string, text into a list of tokens
    text = word_tokenize(text)
    return text
def filteringText(text): # Remove stopwords in a text
    listStopwords = set(stopwords.words('indonesian'))
    filtered = []
    for txt in text:
        if txt not in listStopwords:
            filtered.append(txt)
    text = filtered
    return text
def stemmingText(text): # Reducing a word to its word stem
    that affixes to suffixes and prefixes
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    text = [stemmer.stem(word) for word in text]
    return text
def toSentence(list_words): # Convert list of words into
    sentence
    sentence = ' '.join(word for word in list_words)
    return sentence
```

a. *Case Folding*

Langkah pertama dalam tahapan *preprocessing* adalah melakukan proses *case folding*. *Case folding* merupakan proses untuk mengubah semua karakter dalam dokumen menjadi huruf kecil (*lowercase*). *Case folding* bertujuan untuk menyeragamkan karakter yang tidak sesuai. Alur kerja *case folding* dimulai dengan memasukkan data awal kemudian diproses dengan *case folding* sehingga menghasilkan data hasil yang diinginkan seperti yang terdapat dalam gambar 3.3.



Gambar 3.4 Alur *Case Folding*

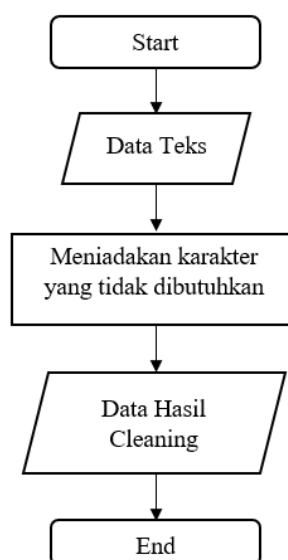
Tabel 3.2 *Case Folding*

Data Awal	Data Hasil <i>Case Folding</i>
Dukung Penuh IKN Nusantara untuk Indonesia lebih baik #IKN	dukung penuh ikn nusantara untuk indonesia lebih baik #ikn

Implementasi *case folding* ditunjukkan pada Tabel 3.1 dengan membandingkan dokumen sebelum menggunakan *case folding* dan sesudah diseragamkan menjadi huruf kecil (*lowercase*) dengan memanfaatkan fitur tersebut.

b. *Cleaning*

Proses *Cleaning* adalah proses yang dilakukan untuk menghilangkan karakter yang tidak relevan dalam klasifikasi serta tidak ada kaitannya dengan hasil klasifikasi serta tidak diperlukan dalam tahapan klasifikasi. nantinya, seperti penggunaan “#, @, angka, tautan URL, karakter spesial, dan tanda baca lainnya”. Proses *cleaning* dijelaskan secara visual melalui gambar 3.4 yang menggambarkan langkah-langkah dalam alur kerjanya. Implementasi *cleaning* juga disajikan dalam tabel 3.2 untuk mempermudah pengolahan data awal agar sesuai dengan metode klasifikasi yang digunakan.



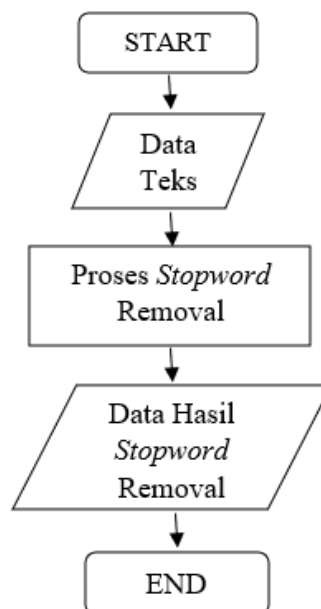
Gambar 3.5 Alur *Cleaning*

Tabel 3.3 Cleaning

Data Awal	Data Hasil Cleaning
dukung penuh ikn nusantara untuk indonesia lebih baik #ikn	dukung penuh ikn nusantara untuk indonesia lebih baik

c. *Stopword Removal*

Tahap *stopword removal* adalah langkah untuk menghilangkan kata-kata yang sering muncul dalam jumlah yang besar pada dokumen namun tidak memiliki makna penting. Penghapusan kata-kata ini tidak mempengaruhi hasil klasifikasi.. Tujuan *stopword removal* untuk mempermudah proses klasifikasi. Penelitian ini menggunakan daftar *Stopword* yang disesuaikan dengan kebutuhan penelitian. Alur kerja *stopword* dapat dilihat dari visualisasi gambar 3.5. yang menjelaskan runtutan proses yang dilalui.



Gambar 3.6 Gambar Alur *Stopword Removal*

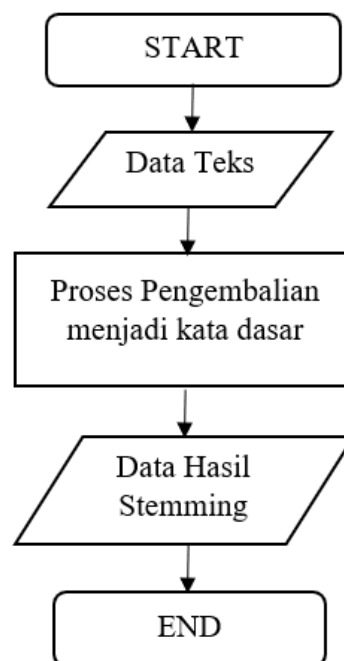
Tabel 3.4 Contoh *Stopword Removal*

Data Awal	Data Hasil <i>Stopword Removal</i>
dukung penuh ikn nusantara untuk indonesia lebih baik	dukung penuh ikn nusantara indonesia

Implementasi *stopword removal* dengan menghilangkan kata yang tidak relevan dalam proses klasifikasi dapat dilihat pada tabel 3.3.

d. *Stemming*

Tahap *Stemming* merupakan langkah dalam mengembalikan kata-kata dalam data menjadi bentuk kata dasarnya. Proses ini diperlukan untuk menghapus sufiks, prefiks, dan konfiks dalam teks Bahasa Indonesia yang dapat mengubah kata dasar menjadi bentuk yang berbeda dan menyulitkan pencarian kata dasar.

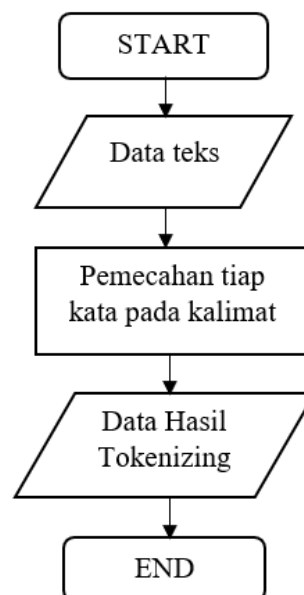
Gambar 3.7 Alur *Stemming*

Tabel 3.5 *Stemming*

Data Awal	Data Hasil <i>Stemming</i>
dukung penuh ikn nusantara indonesia	dukung penuh ikn nusantara indonesia

e. *Tokenizing*

Tahap *tokenizing* merupakan proses yang membagi kata pada kalimat menjadi kata terpisah, sehingga terdapat segmentasi kata yang membentuk kalimat. Proses *tokenizing* didasarkan pada pemisahan kalimat menggunakan spasi. Kata-kata yang telah dipisahkan tersebut disebut sebagai *token*. *Token - token* yang telah dibuat selanjutnya akan digunakan untuk proses pembobotan. Penjelasan visual mengenai langkah-langkah *tokenizing* dapat ditemukan dalam Gambar 3.7.

Gambar 3.8 Alur *Tokenizing*

Tabel 3.6 *Tokenizing*

Data Awal	Data Hasil <i>Tokenizing</i>
dukung penuh ikn nusantara indonesia	['dukung', 'penuh', 'ikn', 'nusantara', 'indonesia']

Implementasi *tokenizing* yang terdapat pada dokumen tersaji pada tabel

3.4. Sebuah kalimat akan dipisahkan menjadi kata tunggal.

3.4 TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan sebuah algoritma statistik numerik yang menghasilkan bobot nilai untuk setiap atribut (atau kata) dalam sebuah dokumen teks. Metode ini sering digunakan dalam pemrosesan bahasa alami (NLP), pencarian informasi, dan penambangan teks. Metode ini bertujuan untuk menentukan bobot kata dalam teks dengan mengukur tingkat kepentingan suatu kata dalam kumpulan dokumen. Tingkat kepentingan kata tersebut meningkat sejalan dengan jumlah total kemunculannya dalam dokumen teks. Algoritma TF-IDF digunakan untuk memberikan bobot pada hubungan antara kata (term) dan sebuah dokumen. TF-IDF adalah ukuran statistik yang digunakan untuk mengevaluasi seberapa penting sebuah kata dalam satu dokumen atau kumpulan kata. Tingkat kemunculan kata dalam sebuah dokumen menunjukkan tingkat kepentingan kata tersebut dalam dokumen tersebut. Jumlah dokumen yang mengandung kata tersebut menentukan seberapa umum kata tersebut dalam kumpulan dokumen. Bobot kata akan tinggi jika kata tersebut sering muncul dalam satu dokumen, dan bobotnya rendah jika kata tersebut sering muncul dalam beberapa dokumen. Algoritma TF-IDF ini digunakan dalam penelitian ini

karena dikenal sebagai algoritma pembobotan yang efektif, mudah diimplementasikan, dan menghasilkan akurasi yang baik. Algoritma ini mampu menghasilkan hasil yang akurat dengan waktu pemrosesan yang cepat, bahkan tanpa membutuhkan komputasi yang kompleks. Selain itu, konsep TF-IDF memberikan nilai rendah pada kata yang sering muncul dalam semua dokumen yang digunakan, sehingga metode ini juga dapat digunakan untuk menentukan kata-kata pada *StopList*. *StopList* merupakan daftar kata yang tidak relevan dalam penelitian, terutama kata-kata hubung dalam bahasa Indonesia. Langkah selanjutnya dalam penelitian ini adalah tahap seleksi fitur, di mana penulis menggunakan algoritma TF-IDF (*Term Frequency-Inverse Document Frequency*). TF merupakan jumlah total kata yang muncul dalam sebuah dokumen. Besar kecilnya nilai TF tergantung pada seberapa sering kata tersebut muncul dalam dokumen, semakin sering kemunculan kata tersebut, semakin besar nilai TF-nya. Sementara itu, IDF merupakan perhitungan seberapa banyak sebuah kata didistribusikan dalam semua dokumen yang tersedia. Nilai IDF akan semakin tinggi jika frekuensi kemunculan kata semakin sedikit, dan nilainya berbanding terbalik dengan TF.

a. Pembobotan Fitur

Tahap selanjutnya melibatkan perhitungan nilai bobot TF dan IDF untuk setiap kata menggunakan persamaan TF-IDF yang diberikan di bawah ini :

$$TF_{ij} = \frac{f_{i,j}}{\sum_{\kappa} f_{\kappa,j}} \quad (3.1)$$

Keterangan:

$f_{i,j}$ = jumlah eksistensi (i) term pada dokumen (j)

$\sum_{\kappa} f_{\kappa,j}$ = jumlah (i) term yang terdapat dalam dokumen (j).

$$IDF_i = \log\left(\frac{N}{ni}\right) + 1 \quad (3.2)$$

Keterangan:

N = Jumlah dokumen

ni = Jumlah dokumen (i) term atau kata tersebut.

Perhitungan TF-IDF dapat dilakukan dengan menggunakan persamaan (3.3)

$$TFIDF_{ij} = TF_{ij} \times IDF_i \quad (3.3)$$

Berikut contoh tabel TF-IDF.

Tabel 3.7 Tabel Nilai TF-IDF

TERM	D ₁	D ₂	...	D ₃	Bobot Total
Term ₁	TF-IDF _{1,1}	TF-IDF _{2,1}	...	TF-IDF _{n,1}	TF-IDF _{total,1}
Term ₂	TF-IDF _{1,2}	TF-IDF _{2,2}	...	TF-IDF _{n,2}	TF-IDF _{total,2}
Term ₃	TF-IDF _{1,3}	TF-IDF _{2,3}	...	TF-IDF _{n,3}	TF-IDF _{total,3}
Term ₄	TF-IDF _{1,4}	TF-IDF _{2,4}	...	TF-IDF _{n,4}	TF-IDF _{total,4}
Term ₅	TF-IDF _{1,5}	TF-IDF _{2,5}	...	TF-IDF _{n,5}	TF-IDF _{total,5}
...
Term _m	TF-IDF _{1,m}	TF-IDF _{2,m}	...	TF-IDF _{n,m}	TF-IDF _{total,m}

Tabel 3.7. Menjelaskan tentang Nilai TF-IDF dimana term merupakan kolom yang berisi daftar istilah pada dokumen, bobot total adalah kolom yang berisi bobot total untuk setiap istilah dalam semua dokumen.

b. Peringkat Fitur

Tahap berikutnya dalam seleksi fitur adalah mengurutkan fitur berdasarkan bobot total dari yang tertinggi hingga yang terendah. Bobot total dihitung dengan menjumlahkan bobot dari setiap term di seluruh dokumen yang ada.

Untuk memberikan gambaran, penerapan nilai TF-IDF dapat dilihat dalam

Tabel 3.6 sebagai referensi. Berikut contoh peringkat yang digunakan dalam implementasi TF-IDF.

Tabel 3.8 Tabel Contoh Dataset

Dokumen	Teks	Kelas
D1	Mari dukung IKN #ikn	Mendukung
D2	Pemindahan IKN Memberikan Kesempatan Revitalisasi Kota Untuk Menjadi Lebih Baik	Mendukung
D3	BIKIN KAGET! PEJABAT OTORITA IKN MENGUNDURKAN DIRI?	Tidak Mendukung
D4	Kenapa IKN harus di Kaltim?	Tidak Mendukung
D5	Realisasi Investasi di IKN Belum Ada, Menteri Basuki: Semua Masih dari APBN	Tidak Mendukung

Tabel 3.8 merupakan contoh *dataset* yang digunakan dalam menentukan nilai TF-IDF yang berisi lima dokumen dengan dua dokumen mendukung dan tiga dokumen tidak mendukung.

Tabel 3.9 Tabel Nilai TF

No.	Kata	TF				
		D1	D2	D3	D4	D5
1	mari	0.333	0	0	0	0
2	dukung	0.3333	0	0	0	0
3	ikn	0.3333	0.1666	0.1428	0.5	0.1666
4	heru	0	0.1666	0	0	0
5	perpindahan	0	0.1666	0	0	0
6	peluang	0	0.1666	0	0	0
7	revitalisasi	0	0.1666	0	0	0
8	kota	0	0.1666	0	0	0
9	bikin	0	0	0.1428	0	0
10	kaget	0	0	0.1428	0	0
11	pejabat	0	0	0.1428	0	0
12	otorita	0	0	0.1428	0	0
13	ramerame	0	0	0.1428	0	0
14	mengundurkan	0	0	0.1428	0	0
15	kaltim	0	0	0	0.5	0
16	realisasi	0	0	0	0	0.1666

No.	Kata	TF				
		D1	D2	D3	D4	D5
17	investasi	0	0	0	0	0.1666
18	menteri	0	0	0	0	0.1666
19	basuki	0	0	0	0	0.1666
20	apbn	0	0	0	0	0.1666

Dokumen tersebut kemudian diolah menggunakan rumus (3.1) tentang menghitung nilai TF, sehingga menghasilkan data pada tabel 3.9.

Tabel 3.10 Tabel Nilai DF dan IDF

No.	Kata	DF	IDF
1	mari	1	1.6989
2	dukung	1	1.6989
3	ikn	5	1
4	heru	1	1.6989
5	perpindahan	1	1.6989
6	peluang	1	1.6989
7	revitalisasi	1	1.6989
8	kota	1	1.6989
9	bikin	1	1.6989
10	kaget	1	1.6989
11	pejabat	1	1.6989
12	otorita	1	1.6989
13	ramerame	1	1.6989
14	mengundurkan	1	1.6989
15	kaltim	1	1.6989
16	realisasi	1	1.6989
17	investasi	1	1.6989
18	menteri	1	1.6989
19	basuki	1	1.6989
20	apbn	1	1.6989

Selanjutnya dilakukan perhitungan DF kata pada seluruh dokumen dan IDF menggunakan rumus (3.2). Hasil dari perhitungan termuat dalam tabel 3.10.

Tabel 3.11 Tabel Hasil Nilai TF-IDF

No.	Kata	TF-IDF				
		D1	D2	D3	D4	D5
1	mari	0.5663	0	0	0	0
2	dukung	0.5663	0	0	0	0
3	ikn	0.3333	0.1666	0.1428	0.5	0.1666
4	heru	0	0.2831	0	0	0
5	perpindahan	0	0.2831	0	0	0
6	peluang	0	0.2831	0	0	0
7	revitalisasi	0	0.2831	0	0	0
8	kota	0	0.2831	0	0	0
9	bikin	0	0	0.2427	0	0
10	kaget	0	0	0.2427	0	0
11	pejabat	0	0	0.2427	0	0
12	otorita	0	0	0.2427	0	0
13	ramerame	0	0	0.2427	0	0
14	mengundurkan	0	0	0.2427	0	0
15	kaltim	0	0	0	0.8494	0
16	realisasi	0	0	0	0	0.2831
17	investasi	0	0	0	0	0.2831
18	menteri	0	0	0	0	0.2831
19	basuki	0	0	0	0	0.2831
20	apbn	0	0	0	0	0.2831

Tabel 3.11 menunjukkan perhitungan TF IDF menggunakan rumus (3.3)

Tabel 3.12 Tabel Peringkat Nilai TF-IDF

No.	Kata	Skor	No	Kata	Skor
1	ikn	1.30952	11	pejabat	0.24271
2	mari	0.56632	12	otorita	0.24271
3	dukung	0.56632	13	ramerame	0.24271
4	heru	0.28316	14	mengundurkan	0.24271
5	perpindahan	0.28316	15	kaltim	0.84948
6	peluang	0.28316	16	realisasi	0.28316
7	revitalisasi	0.28316	17	investasi	0.28316
8	kota	0.28316	18	menteri	0.28316
9	bikin	0.24271	19	basuki	0.28316
10	kaget	0.24271	20	apbn	0.28316

Pada langkah awal dalam proses TF-IDF dilakukan dengan mencari nilai TF masing-masing fitur dari dokumen pada tabel 3.8, sehingga nilai TF fitur dapat diketahui melalui tabel 3.9. Setelah mengetahui nilai TF, langkah selanjutnya adalah mencari nilai DF dan IDF pada masing-masing fitur, lalu hasilnya terdapat pada tabel 3.10. Setelah mengetahui nilai TF, DF dan IDF, tahap terakhir adalah mencari nilai TF-IDF pada masing-masing fitur. Hasil nilai dari TF-IDF terdapat pada tabel 3.11. Dalam proses TF-IDF terdapat proses peringkatan fitur-fitur yang disajikan dalam tabel 3.12. Tabel tersebut berisi nilai dari hasil peringkat dari yang terbesar hingga terkecil dari semua fitur yang termuat dalam semua dokumen. Berikut merupakan *source code* yang digunakan untuk mencari nilai TF-IDF.

```
def calc_TF(document):
    # Counts the number of times the word appears
    in review
    TF_dict = {}
    for word in document:
        if word in TF_dict:
            TF_dict[word] += 1
        else:
            TF_dict[word] = 1
    # Computes tf for each word
    for word in TF_dict:
        TF_dict[word] = TF_dict[word] /
len(document)
    return TF_dict
text_TF = x.apply(calc_TF)
text_TF
def Calc_DF(tfDict):
    count_DF = {}
    # Run through each document's tf dictionary
    and increment countDict's (term, doc) pair
    for document in tfDict:
        for word in document:
            if word in count_DF:
                count_DF[word] += 1
            else:
                count_DF[word] = 1
    return count_DF
DF = calc_DF(text_TF)
```

```

#TF-IDF
import math
n_document = len(x)
def calc_IDF(__n_document, __DF):
    IDF_dict = {}
    for word in __DF:
        IDF_dict[word] =
math.log10((__n_document)/(__DF[word]))+1
    return IDF_dict
IDF = calc_IDF(n_document, DF)
IDF
#calc TF-IDF
def calc_TF_IDF(TF):
    TF_IDF_Dict = {}
    #For each word in the review, we multiply its
tf and its idf.
    for key in TF:
        TF_IDF_Dict[key] = TF[key] * IDF[key]
    return TF_IDF_Dict
#Stores the TF-IDF Series
text_TFIDF = text_TF.apply(calc_TF_IDF)

```

3.5 Naïve Bayes

Langkah berikutnya, yang merupakan langkah inti dari penelitian ini, adalah langkah klasifikasi. Dalam penelitian ini, digunakan algoritma *Naïve Bayes Classifier* sebagai algoritma utama untuk melakukan pengklasifikasian. Sistem dibangun dengan pendekatan statistik yang disebut teori *Bayes* yang menggunakan probabilitas bersyarat. Menurut (Suhartono et al., 2018), probabilitas bersyarat adalah perhitungan probabilitas suatu peristiwa, A, Ketika peristiwa lain, B, telah terjadi, yang dicatat sebagai $P(A|B)$ yang menggabungkan probabilitas A dan B. Keunggulan utama metode ini terletak pada asumsi yang kuat tentang ketidak terikatan antar variabel. Proses pengklasifikasian melibatkan dua tahap, yaitu pelatihan data dan pengujian data. Tahap pelatihan digunakan untuk memperkenalkan dan melatih *dataset* pada sistem, sedangkan tahap pengujian digunakan untuk menguji *dataset* dengan mengacu pada perhitungan probabilitas yang diperoleh dari tahap pelatihan. Secara umum, teorema *Bayes* direpresentasikan dengan persamaan (3.4).

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (3.4)$$

Keterangan:

$P(A|B)$ = Probabilitas Posterior

$P(B|A)$ = Estimasi Probabilitas

$P(A)$ = Probabilitas prior dari kelas

$P(B)$ = Probabilitas prior dari prediktor

Proses untuk menentukan probabilitas data dimulai dengan menghitung nilai *likelihood* berdasarkan *dataset* yang telah disiapkan. Nilai *likelihood* ini kemudian dikalikan dengan setiap nilai probabilitas data yang digunakan. Hasil perhitungan tersebut digunakan sebagai panduan untuk mengklasifikasikan data uji baru. Dalam praktiknya, sering kali nilai $P(B)$ diabaikan karena nilainya tetap.

Pada tahap klasifikasi menggunakan *Naïve Bayes Classifier*, setiap dokumen direpresentasikan dengan atribut " $X_1, X_2, X_3, \dots, X_n$ " yang mewakili kata-kata dalam dokumen tersebut. Kumpulan dokumen *tweet* direpresentasikan dengan variabel V . Saat melakukan klasifikasi, *Naïve Bayes Classifier* akan menentukan probabilitas tertinggi dari setiap kelas dokumen uji (V_{MAP}) dengan memasukkan atribut ($a_1, a_2, a_3, \dots, a_n$). Persamaan untuk V_{MAP} ditulis dalam bentuk persamaan (3.5) berikut ini:

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} P(V_j | a_1, a_2, a_3, \dots, a_n) \quad (3.5)$$

Dengan memanfaatkan teorema Bayes, persamaan (3.5) dapat diubah menjadi:

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} \frac{P(a_1, a_2, a_3, \dots, a_n | V_j) P(V_j)}{P(a_1, a_2, a_3, \dots, a_n)} \quad (3.6)$$

Nilai $P(a_1, a_2, a_3, \dots, a_n)$ tetap konstan untuk setiap dokumen V_j , sehingga persamaan (3.6) dapat dirumuskan sebagai berikut:

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} P(a_1, a_2, a_3, \dots, a_n | V_j) P(V_j) \quad (3.7)$$

Naïve Bayes Classifier mempermudah perhitungan ini dengan mengasumsikan bahwa dalam setiap kelas, setiap variabel tidak terkait satu sama lain. Dengan demikian, dapat diinterpretasikan sebagai berikut:

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} P(V_j) \prod_i P(a_i | V_j) \quad (3.8)$$

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} P(V_j) \prod_i P(a_i | V_j) \quad (3.9)$$

P Probabilitas $P(V_j)$ dan probabilitas nilai kata a_i pada setiap kelas $P(a_i | V_j)$ dihitung menggunakan rumus sebagai berikut,

$$P(V_j) = \frac{\text{docs}_j}{\text{training}} \quad (3.10)$$

$$P(a_i | V_j) = \frac{n_i + 1}{n + \text{kosa kata}} \quad (3.11)$$

Keterangan :

docs_j = Jumlah total dokumen kelas j , dan training jumlah total dokumen.

Variabel n_i = Jumlah kata a_i yang muncul dalam kelas V_j .

n = Jumlah kata yang muncul dalam kelas V_j .

Kosakata = Jumlah kata yang berbeda dalam semua dokumen.

Untuk mengatasi kemungkinan probabilitas nol, nilai 1 ditambahkan pada setiap perkiraan perhitungan probabilitas. Teknik ini dikenal sebagai teknik *Laplace Smoothing*.

3.6 *Gaussian Naïve Bayes*

Sistem menggunakan algoritma *Gaussian Naïve Bayes*, yang sering digunakan untuk mengklasifikasikan data dengan atribut yang bersifat kontinu. Algoritma ini mengandalkan distribusi normal *Gaussian* sebagai fungsi dasarnya. Algoritma ini memiliki dua parameter utama, yaitu *mean* (μ) dan *varians* (σ^2). *Mean*

adalah nilai rata-rata dari atribut kontinu, yang dapat dihitung menggunakan rumus pada persamaan (3.12). *Varians* adalah kuadrat dari simpangan nilai data pada atribut kontinu, yang dihitung menggunakan rumus pada persamaan (3.13). Algoritma ini memiliki ciri khas dalam menggunakan parameter-parameter tersebut.

$$\mu = \frac{\sum X}{N} \quad (3.12)$$

Keterangan :

μ = Nilai rata-rata (mean)
 $\sum X$ = Total nilai TF-IDF
 N = Total data

$$\sigma = \frac{\sum (X - \mu)^2}{N - 1} \quad (3.13)$$

Keterangan :

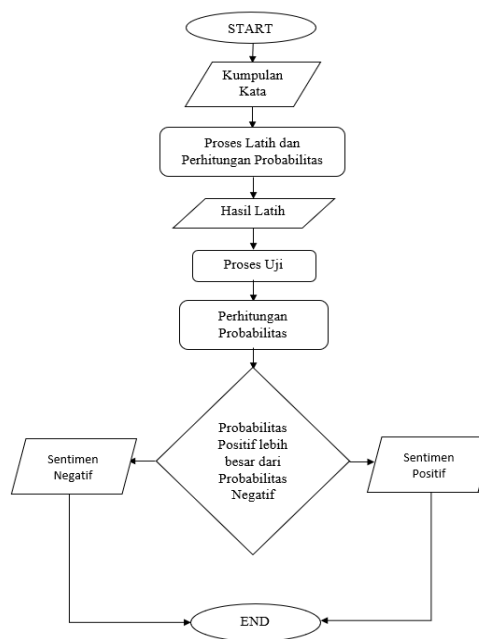
σ = varians dari data TF-IDF
 X = nilai data TF-IDF
 μ = nilai rata-rata data TF-IDF
 N = jumlah data

Apabila menghadapi data kontinu, diasumsikan secara umum bahwa nilai kontinu yang berkorelasi dengan setiap kelas memiliki distribusi *Gaussian*. Perhitungan pada data pelatihan dilakukan dengan memisahkan data berdasarkan kelas yang ada, dan kemudian dihitung nilai rata-rata dan standar deviasi untuk setiap kelas. Selanjutnya, dilakukan proses untuk menghitung probabilitas dari himpunan data kontinu menggunakan persamaan (3.14) berikut ini (Kamel et al., 2019) :

$$P(B|A) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \exp \frac{-(x-\mu)^2}{2\sigma^2} \quad (3.14)$$

Keterangan :

- $P(B|A)$ = Estimasi Probabilitas
- σ = varians dari TF-IDF
- $\pi = 3.14$
- μ = nilai rata-rata data TF-IDF



Gambar 3. 9 Gambar Alur *Gaussian Naïve Bayes*

Sebagai ilustrasi, berikut disajikan sebuah diagram alur yang menunjukkan tahapan klasifikasi menggunakan algoritma *Gaussian Naïve Bayes* pada Gambar 3.8. Proses dimulai dengan memasukkan kata-kata yang telah diberi bobot pada tahap sebelumnya, kemudian data tersebut digunakan dalam proses pelatihan. Tahap awal pelatihan melibatkan pembagian dokumen latih berdasarkan kelasnya, yaitu kelas positif dan negatif. Langkah berikutnya adalah menghitung nilai *mean* dan *varians* untuk setiap kata pada setiap kelas menggunakan persamaan (3.12) dan persamaan (3.13). Tahapan selanjutnya dalam proses ini adalah menghitung probabilitas *prior* dari masing-masing kelas menggunakan persamaan (3.10). Pada tahap ini, perhitungan probabilitas menggunakan persamaan (3.14) dengan memanfaatkan nilai *mean* dan *varians* yang diperoleh dari data latih sebelumnya. Namun, jika terdapat kata yang tidak memiliki nilai *mean* dan *varians* dalam data

latih atau memiliki nilai 0, maka probabilitas kata tersebut dapat diabaikan atau dianggap sebagai 1. Selanjutnya, mengalikan hasil probabilitas dari setiap kata dalam kelas yang sama dengan probabilitas *prior* kelas tersebut. Proses ini akan menentukan kelas hasil klasifikasi dari dokumen uji, di mana kelas tersebut ditentukan berdasarkan hasil perhitungan dengan total nilai probabilitas tertinggi.

langkah pertama klasifikasi adalah menghitung frekuensi kemunculan fitur pada setiap kelas sentimen. Fitur yang akan dihitung jumlah kemunculannya diambil dari hasil peringkat TF-IDF. Sebagai contoh, dalam perhitungan ini hanya menggunakan 80% data dari Tabel 3.12, sehingga hanya diambil 16 kata dari total 20 kata yang akan digunakan sebagai fitur, seperti yang terlihat pada Tabel 3.13.

Tabel 3.13 Tabel Fitur

No.	Kata	Mendukung	Tidak Mnedukung
1	mari	1	0
2	dukung	1	0
3	ikn	2	3
4	heru	1	0
5	perpindahan	1	0
6	peluang	1	0
7	revitalisasi	1	0
8	kota	1	0
9	bikin	0	1
10	kaget	0	1
11	pejabat	0	1
12	otorita	0	1
13	rامرة	0	1
14	mengundurkan	0	1
15	kaltim	0	1
16	realisasi	0	1

Langkah berikutnya adalah menentukan kemunculan fitur di kelas mendukung dan tidak mendukung pada tiap dokumen. Lalu kemudian mencari *mean* serta standar deviasi tiap fitur.

Tabel 3.14 Tabel Fitur Kelas Mendukung

No.	Kata	Kelas Mendukung	
		Dokumen 1	Dokumen 2
1	mari	0.5663	0
2	dukung	0.5663	0
3	ikn	0.3333	0.1666
4	heru	0	0.2831
5	perpindahan	0	0.2831
6	peluang	0	0.2831
7	Revitalisasi	0	0.2831
8	kota	0	0.2831
9	bikin	0	0
10	kaget	0	0
11	pejabat	0	0
12	otorita	0	0
13	ramerame	0	0
14	mengundurkan	0	0
15	kaltim	0	0
16	realisasi	0	0

Tabel 3.14 berisi tentang nilai fitur dua dokumen pada kelas mendukung Sejumlah 16 fitur yang digunakan. Hasil tersebut diketahui dari proses perhitungan.

Tabel 3.15 Tabel Fitur Kelas Tidak Mendukung

No.	Kata	Kelas Tidak Mendukung		
		Dokumen 1	Dokumen 2	Dokumen 3
1	mari	0	0	0
2	dukung	0	0	0
3	ikn	0.1428	0.5	0.1666
4	heru	0	0	0
5	perpindahan	0	0	0
6	peluang	0	0	0
7	Revitalisasi	0	0	0
8	kota	0	0	0
9	bikin	0.2427	0	0
10	kaget	0.2427	0	0
11	pejabat	0.2427	0	0
12	otorita	0.2427	0	0
13	ramerame	0.2427	0	0
14	mengundurkan	0.2427	0	0

Tabel 3.15 berisi tentang nilai fitur dua dokumen pada kelas tidak mendukung.. Kemudian menghitung *mean* dari tiap fitur pada kelas mendukung dan tidak mendukung.

Tabel 3.16 Tabel *Mean*

No.	Kata	Mean	
		Positif	Negatif
1	mari	0.28315	0
2	dukung	0.28315	0
3	ikn	0.4999	0.8094
4	heru	0.14155	0
5	perpindahan	0.14155	0
6	peluang	0.14155	0
7	revitalisasi	0.14155	0
8	kota	0.14155	0
9	bikin	0	0.0809
10	kaget	0	0.0809
11	pejabat	0	0.0809
12	otorita	0	0.0809
13	ramerame	0	0.0809
14	mengundurkan	0	0.0809
15	kaltim	0	0.28313
16	realisasi	0	0.09436

Kemudian menghitung nilai standar deviasi tiap fitur pada masing-masing kelas mendukung dan kelas tidak mendukung menggunakan rumus (3.11). menghasilkan nilai yang terhimpun di tabel 3.16.

Tabel 3.17 Tabel Standar Deviasi

No.	Kata	Standar Deviasi	
		Mendukung	Tidak Mendukung
1	mari	0.40043457	0.8094
2	dukung	0.40043457	0
3	ikn	0.1178747	0
4	heru	0.20018193	0
5	perpindahan	0.20018193	0
6	peluang	0.20018193	0
7	revitalisasi	0.20018193	0
8	Kota	0.200182	0
9	Bikin	0	0.14012291
10	Kaget	0	0.14012291
11	Pejabat	0	0.140123
12	otorita	0	0.14012291
13	rامرة	0	0.14012291
14	mengundurkan	0	0.14012291
15	kaltim	0	0.490401
16	realisasi	0	0.163447861

Probabilitas perhitungan kata-kata sebelumnya akan menjadi referensi untuk proses klasifikasi. Sebagai contoh, hasil klasifikasi pada set data uji dapat dilihat dalam Tabel 3.18.

Tabel 3.18 Tabel *Input* Data Uji Baru

Kalimat	Kelas
Memang realisasi perpindahan ikn memberi dampak di kaltim?	?

Tahap awal yang dilakukan adalah memilih kata-kata yang akan digunakan dan mengambil nilai probabilitas masing-masing kelas untuk perhitungan. Jika terdapat kata dalam dokumen uji yang tidak terdapat dalam Tabel 3.13, kata tersebut akan diabaikan dalam proses perhitungan seperti yang terlihat dalam Tabel 3.16 dan 3.17.

Tabel 3.19 Tabel Probabilitas Fitur

Kata	Mendukung	Tidak Mendukung
realisasi	1	1.523623791
perpindahan	1.523623791	1
ikn	2.082346078	1.897918402
kaltim	1	1.897918

Probabilitas suatu kata ditentukan pada masing - masing kelas mendukung dan tidak mendukung menggunakan rumus (3.14) sehingga menghasilkan data pada tabel 3.19.

Tabel 3.20 Tabel Probabilitas Kelas

Kelas	Prediksi
Mendukung	1.9036272153
Tidak Mendukung	3.2929412028

Berdasarkan hasil perhitungan yang tertera dalam Tabel 3.20, dapat diamati bahwa kelas "tidak mendukung" memiliki nilai yang lebih tinggi dibandingkan dengan nilai kelas "mendukung", maka dari itu hasil proses klasifikasi dokumen tersebut yaitu tidak mendukung. Tabel 3.20 merupakan hasil prediksi yang dilakukan oleh sistem yang dibangun, sehingga menghasilkan hasil klasifikasi yang berlandaskan nilai - nilai probabilitas kelas pada dokumen tersebut.

Berikut terdapat potongan kode yang digunakan untuk melakukan proses klasifikasi. Kode ini memulai dengan menghitung rata-rata dari fitur-fitur pada kelas yang mendukung dan kelas yang tidak mendukung. Setelah menemukan rata-rata untuk masing-masing kelas, dilakukan perhitungan standar deviasi untuk kedua kelas tersebut. Hasil perhitungan standar deviasi ini kemudian digunakan dalam suatu rumus (3.14) yang merupakan bagian dari algoritma *Gaussian Naïve Bayes* untuk mendapatkan hasil klasifikasi. Dalam implementasi sistem yang sedang dibuat, langkah-langkah ini akan memainkan peran penting dalam mengklasifikasikan data yang diberikan. Dengan mengumpulkan dan menganalisis statistik seperti *mean* dan standar deviasi dari kelas-kelas yang ada, sistem dapat mengidentifikasi pola dan mengambil keputusan klasifikasi yang akurat berdasarkan data baru yang masuk. Metode *Gaussian Naïve Bayes* yang diadopsi dalam sistem ini mengasumsikan bahwa setiap fitur yang diamati secara independen berkontribusi terhadap probabilitas kelas tertentu. Dengan memanfaatkan rumus (3.14) yang melibatkan *mean* dan standar deviasi, sistem dapat menghitung probabilitas kelas yang paling mungkin untuk data yang diberikan. Dengan kata lain, sistem dapat mengklasifikasikan data berdasarkan distribusi probabilistik dan informasi statistik yang telah diperoleh dari kelas-kelas pelatihan sebelumnya.


```

#Function Calc probability features
from math import sqrt
from math import pi
from math import exp
def calculate_probability(x, mean, stdev):
    if stdev == 0 or x == 0:
        return 1
    else: exponent = exp(-((x-mean)**2 / (2 * (stdev**2) )))
    return (1 / sqrt(2 * 3.14 * (stdev) * exponent)+1
# Fuction to Create the list for fill the probablity values that
being used
df_test_contain = []
for a in df1_test:
    for b in df3:
        if a == b:
            df_test_contain.append(a)
#Create reserved string to fill by the real values of probability
features
df_test_contain_value = [[ [0.0 for col in range(3)] for col in
range(len(df_test_contain))] for row in range(len(df1_test))]
probability_result_positif = [[0.0]*len(df_test_contain) for k in
range(len(df1_test))]
for i in range(len(df1_test)):
    for y in range(len(df_test_contain)):
        df_test_contain_value[i][y][0] =
df1_test.iloc[i][df_test_contain[y]]
        for a in range(len(meanpos)):
            if meanpos.iloc[a][0] == df_test_contain[y]:
                df_test_contain_value[i][y][1] = meanpos.iloc[a][1]
                df_test_contain_value[i][y][2] = meanpos.iloc[a][2]
                probability_result_positif[i][y] =
calculate_probability(df_test_contain_value[i][y][0],
df_test_contain_value[i][y][1], df_test_contain_value[i][y][2])
#Create reserved string to fill by the real values of probability
features
probability_result_negatif = [[0.0]*len(df_test_contain) for k in
range(len(df1_test))]
for i in range(len(df1_test)):
    for y in range(len(df_test_contain)):
        df_test_contain_value[i][y][0] =
df1_test.iloc[i][df_test_contain[y]]
        for a in range(len(meanneg)):
            if meanneg.iloc[a][0] == df_test_contain[y]:
                df_test_contain_value[i][y][1] = meanneg.iloc[a][1]
                df_test_contain_value[i][y][2] = meanneg.iloc[a][2]
                probability_result_negatif[i][y] =
calculate_probability(df_test_contain_value[i][y][0],
df_test_contain_value[i][y][1], df_test_contain_value[i][y][2])
# Calc Gaussian NB values
g_nb_pos = [0.0]* len(df1_test)
for i in range(len(df1_test)):
    g_nb_pos[i] =
np.prod(probability_result_positif[i])*list_prior[1]
# Calc Gaussian NB values
g_nb_neg = [0.0]* len(df1_test)
for i in range(len(df1_test)):
    g_nb_neg[i] =
np.prod(probability_result_negatif[i])*list_prior[0]

```

```

# Labeling the result of GNB
def max_value(pos,neg):
    if pos > neg:
        return 1
    else:
        return 0
x_test_list = np.array(x_test.to_list())
df_x_test = pd.DataFrame(x_test_list, columns=["Text"])
df_label_test = pd.DataFrame(label_test, columns=['label'])
data_test = [df_x_test, df_label_test]
df_data_test = pd.concat(data_test, axis=1, join='inner')
df_data_test
# Compare max values of GNB
g_nb_result = ["none"]* len(df1_test)
for i in range(len(df1_test)):
    g_nb_result[i] = max_value(g_nb_pos[i], g_nb_neg[i])
print(g_nb_result)
df_data_test.insert(2, "predict", g_nb_result, True)

```

3.7 Skenario Uji

Penentuan data dilakukan dengan cara mengambil *tweet* menggunakan teknik *scraping*. Data *tweet* berjumlah 500 data kemudian diolah dan diberi label oleh ahli sehingga menghasilkan 319 jumlah *tweet* berlabel mendukung dan 181 jumlah *tweet* berlabel tidak mendukung. Hal tersebut mengakibatkan ketimpangan pada perbandingan kelas yang dimiliki. Data berlabel mendukung memiliki jumlah 63.8% dari keseluruhan data sehingga kemungkinan dapat menyebabkan *overfitting* yang cenderung mengarah pada prediksi berlabel dominan pada data latih, sehingga hasil pengujian mengalami penurunan akurasi, maka dilakukan rangkaian skenario pengujian..

Tabel 3.21 Contoh *dataset* dengan 100% Berlabel Mendukung

No.	Teks	Label
1	Pembangunan IKN menarik banyak investor. #ikn #indonesia #investasi	Mendukung
2	investasi di IKN Nusantara menjadi 120.	Mendukung
3	IKN bukti kemajuan teknologi	Mendukung
4	ikn meratakan pembangunan #IbukotaNegara #IKN #SemangatNusantara	Mendukung

No.	Teks	Label
5	Ibu Kota Negara IKN investasi masa depan	Mendukung
6	Fasilitas pelayanan pembagian bahan baku pembangunan Ibu Kota Negara (IKN) makin banyak dibangun.	Mendukung
7	JERMAN TERTARIK BERKOLABORASI DALAM PEMBANGUNAN HIJAU IKN	Mendukung
8	Dukung IKN Nusantara, Untuk Kejayaan Bangsa dan Negara. #IKNNusantara	Mendukung
9	IKN Sejahterakan Indonesia, dukung proyek IKN. #IKNNusantara	Mendukung
10	Dukung terus pembangunan IKN Nusantara di Kaltim	Mendukung

Pada tabel 3.21. disajikan skenario pengujian dengan keseluruhan dokumen berlabel mendukung. Kemudian *dataset* diuji dengan metode yang digunakan dalam penelitian.

Tabel 3.22 Contoh *dataset* dengan 90% Berlabel Mendukung

No.	Teks	Label
1	Pembangunan IKN menarik banyak investor. #ikn #indonesia #investasi	Mendukung
2	investasi di IKN Nusantara menjadi 120.	Mendukung
3	IKN bukti kemajuan teknologi	Mendukung
4	ikn meratakan pembangunan #IbukotaNegara #IKN #SemangatNusantara	Mendukung
5	Ibu Kota Negara IKN investasi masa depan	Mendukung
6	Fasilitas pelayanan pembagian bahan baku pembangunan Ibu Kota Negara (IKN) makin banyak dibangun.	Mendukung
7	JERMAN TERTARIK BERKOLABORASI DALAM PEMBANGUNAN HIJAU IKN	Mendukung
8	Dukung IKN Nusantara, Untuk Kejayaan Bangsa dan Negara. #IKNNusantara	Mendukung
9	IKN Sejahterakan Indonesia, dukung proyek IKN. #IKNNusantara	Mendukung
10	Ramalan pengamat, IKN bakal sepi #IKN #IbuKotaNusantara #Kalimantan	Tidak Mendukung

Pada tabel 3.22. disajikan skenario pengujian dengan 9 dokumen mendukung dan 1 dokumen tidak mendukung. Kemudian *dataset* diuji dengan metode yang digunakan dalam penelitian.

Tabel 3.23 Contoh *dataset* dengan 80% Berlabel Mendukung

No.	Teks	Label
1	Pembangunan IKN menarik banyak investor. #ikn #indonesia #investasi	Mendukung
2	investasi di IKN Nusantara menjadi 120.	Mendukung
3	IKN bukti kemajuan teknologi	Mendukung
4	ikn meratakan pembangunan #IbukotaNegara #IKN #SemangatNusantara	Mendukung
5	Ibu Kota Negara IKN investasi masa depan	Mendukung
6	Fasilitas pelayanan pembagian bahan baku pembangunan Ibu Kota Negara (IKN) makin banyak dibangun.	Mendukung
7	JERMAN TERTARIK BERKOLABORASI DALAM PEMBANGUNAN HIJAU IKN	Mendukung
8	Dukung IKN Nusantara, Untuk Kejayaan Bangsa dan Negara. #IKNNusantara	Mendukung
9	jangan investasi ke ikn	Tidak Mendukung
10	Ramalan pengamat, IKN bakal sepi #IKN #IbuKotaNusantara #Kalimantan	Tidak Mendukung

Pada tabel 3.23. disajikan skenario pengujian dengan 8 dokumen mendukung dan 2 dokumen tidak mendukung. Kemudian *dataset* diuji dengan metode yang digunakan dalam penelitian.

Tabel 3.24 Contoh *dataset* dengan 70% Berlabel Mendukung

No.	Teks	Label
1	Pembangunan IKN menarik banyak investor. #ikn #indonesia #investasi	Mendukung
2	investasi di IKN Nusantara menjadi 120.	Mendukung
3	BIKIN KAGET! PEJABAT OTORITA IKN MENGUNDURKAN DIRI	Tidak Mendukung
4	ikn meratakan pembangunan #IbukotaNegara #IKN #SemangatNusantara	Mendukung

No.	Teks	Label
5	Ibu Kota Negara IKN investasi masa depan	Mendukung
6	Fasilitas pelayanan pembagian bahan baku pembangunan Ibu Kota Negara (IKN) makin banyak dibangun.	Mendukung
7	JERMAN TERTARIK BERKOLABORASI DALAM PEMBANGUNAN HIJAU IKN	Mendukung
8	Dukung IKN Nusantara, Untuk Kejayaan Bangsa dan Negara. #IKNNusantara	Mendukung
9	jangan investasi ke ikn	Tidak Mendukung
10	Ramalan pengamat, IKN bakal sepi #IKN #IbuKotaNusantara #Kalimantan	Tidak Mendukung

Pada tabel 3.24. disajikan skenario pengujian dengan 7 dokumen mendukung dan 3 dokumen tidak mendukung. Kemudian *dataset* diuji dengan metode yang digunakan dalam penelitian.

Tabel 3.25 Contoh *dataset* dengan 60% Berlabel Mendukung

No.	Teks	Label
1	Pembangunan IKN menarik banyak investor. #ikn #indonesia #investasi	Mendukung
2	investasi di IKN Nusantara menjadi 120.	Mendukung
3	BIKIN KAGET! PEJABAT OTORITA IKN MENGUNDURKAN DIRI	Tidak Mendukung
4	ikn meratakan pembangunan #IbukotaNegara #IKN #SemangatNusantara	Mendukung
5	Ibu Kota Negara IKN investasi masa depan	Mendukung
6	Fasilitas pelayanan pembagian bahan baku pembangunan Ibu Kota Negara (IKN) makin banyak dibangun.	Mendukung
7	ini keinginan para mafia lahan di IKN	Tidak Mendukung
8	Dukung IKN Nusantara, Untuk Kejayaan Bangsa dan Negara. #IKNNusantara	Mendukung
9	jangan investasi ke ikn	Tidak Mendukung
10	Ramalan pengamat, IKN bakal sepi #IKN #IbuKotaNusantara #Kalimantan	Tidak Mendukung

Pada tabel 3.25. disajikan skenario pengujian dengan 6 dokumen mendukung dan 4 dokumen tidak mendukung. Kemudian dataset diuji dengan metode yang digunakan dalam penelitian.

Tabel 3.26 Contoh *dataset* dengan 50% Berlabel Mendukung

No.	Teks	Label
1	Pembangunan IKN menarik banyak investor. #ikn #indonesia #investasi	Mendukung
2	investasi di IKN Nusantara menjadi 120.	Mendukung
3	BIKIN KAGET! PEJABAT OTORITA IKN MENGUNDURKAN DIRI	Tidak Mendukung
4	ikn meratakan pembangunan #IbukotaNegara #IKN #SemangatNusantara	Mendukung
5	Ibu Kota Negara IKN investasi masa depan	Mendukung
6	Sejumlah Anggota DPR Kembali Minta Untuk Menunda Proses Pemindahan IKN Nusantara	Tidak Mendukung
7	ini keinginan para mafia lahan di IKN	Tidak Mendukung
8	Dukung IKN Nusantara, Untuk Kejayaan Bangsa dan Negara. #IKNNusantara	Mendukung
9	jangan investasi ke ikn	Tidak Mendukung
10	Ramalan pengamat, IKN bakal sepi #IKN #IbuKotaNusantara #Kalimantan	Tidak Mendukung

Pada tabel 3.26. disajikan skenario pengujian dengan 5 dokumen mendukung dan 5 dokumen tidak mendukung. Kemudian *dataset* diuji dengan metode yang digunakan dalam penelitian.

Tabel 3.27 Contoh *dataset* dengan 40% Berlabel Mendukung

No.	Teks	Label
1	Pembangunan IKN menarik banyak investor. #ikn #indonesia #investasi	Mendukung
2	IKN Merusak Paru-Paru Dunia	Tidak Mendukung
3	BIKIN KAGET! PEJABAT OTORITA IKN MENGUNDURKAN DIRI	Tidak Mendukung

No.	Teks	Label
4	ikn meratakan pembangunan #IbukotaNegara #IKN #SemangatNusantara	Mendukung
5	Ibu Kota Negara IKN investasi masa depan	Mendukung
6	Sejumlah Anggota DPR Kembali Minta Untuk Menunda Proses Pemindahan IKN Nusantara	Tidak Mendukung
7	ini keinginan para mafia lahan di IKN	Tidak Mendukung
8	Dukung IKN Nusantara, Untuk Kejayaan Bangsa dan Negara. #IKNNusantara	Mendukung
9	jangan investasi ke ikn	Tidak Mendukung
10	Ramalan pengamat, IKN bakal sepi #IKN #IbuKotaNusantara #Kalimantan	Tidak Mendukung

Pada tabel 3.27. disajikan skenario pengujian dengan 4 dokumen mendukung dan 6 dokumen tidak mendukung. Kemudian *dataset* diuji dengan metode yang digunakan dalam penelitian.

Tabel 3.28 Contoh *dataset* dengan 30% Berlabel Mendukung

No.	Teks	Label
1	Pembangunan IKN menarik banyak investor. #ikn #indonesia #investasi	Mendukung
2	Realisasi Investasi di IKN Tersandung Masalah Lahan, Penjabaran dan Komitmen dari Otorita	Tidak Mendukung
3	Akan dianggap percuma kalau ibukota pindah ke IKN	Tidak Mendukung
4	masih paksakan proyek IKN? Siapa investornya? APBN sudah jebol 8T. Mau hutang lagi?	Tidak Mendukung
5	IKN maju terus	Mendukung
6	TOLAK IKN STOP infratraktur gak karuan Jokowi	Tidak Mendukung
7	Investor IKN kabur	Tidak Mendukung
8	harus seragam tolak IKN	Tidak Mendukung
9	Pembangunan IKN menarik banyak investor. #ikn #indonesia #investasi	Mendukung
10	Indonesia memang pantas pindah ke IKN.	Mendukung

Pada tabel 3.28. disajikan skenario pengujian dengan 3 dokumen mendukung dan 7 dokumen tidak mendukung. Kemudian *dataset* diuji dengan metode yang digunakan dalam penelitian.

Tabel 3.29 Contoh *dataset* dengan 20% Berlabel Mendukung

No.	Teks	Label
1	Ramalan pengamat, IKN bakal sepi #IKN #IbuKotaNusantara #Kalimantan	Tidak Mendukung
2	Realisasi Investasi di IKN Tersandung Masalah Lahan, Penjabaran dan Komitmen dari Otorita	Tidak Mendukung
3	Akan dianggap percuma kalau ibukota pindah ke IKN	Tidak Mendukung
4	masih paksakan proyek IKN? Siapa investornya? APBN sudah jebol 8T. Mau hutang lagi?	Tidak Mendukung
5	IKN maju terus	Mendukung
6	TOLAK IKN STOP infratraktur gak karuan Jokowi	Tidak Mendukung
7	Investor IKN kabur	Tidak Mendukung
8	harus seragam tolak IKN	Tidak Mendukung
9	Pembangunan IKN menarik banyak investor. #ikn #indonesia #investasi	Mendukung
10	Indonesia memang pantas pindah ke IKN.	Mendukung

Pada tabel 3.29. disajikan skenario pengujian dengan 2 dokumen mendukung dan 8 dokumen tidak mendukung. Kemudian *dataset* diuji dengan metode yang digunakan dalam penelitian.

Tabel 3.30 Contoh *Dataset* dengan 10% Berlabel Mendukung

No.	Teks	Label
1	Ramalan pengamat, IKN bakal sepi #IKN #IbuKotaNusantara #Kalimantan	Tidak Mendukung
2	Realisasi Investasi di IKN Tersandung Masalah Lahan, Penjabaran dan Komitmen dari Otorita	Tidak Mendukung
3	Akan dianggap percuma kalau ibukota pindah ke IKN	Tidak Mendukung

No.	Teks	Label
4	masih paksakan proyek IKN? Siapa investornya? APBN sudah jebol 8T. Mau hutang lagi?	Tidak Mendukung
5	IKN maju terus	Mendukung
6	TOLAK IKN STOP infratraktur gak karuan Jokowi	Tidak Mendukung
7	Investor IKN kabur	Tidak Mendukung
8	harus seragam tolak IKN	Tidak Mendukung
9	Jajaran Pimpinan Otoritas IKN mengundurkan diri	Tidak Mendukung
10	Indonesia memang pantas pindah ke IKN.	Mendukung

Pada tabel 3.30. disajikan skenario pengujian dengan 1 dokumen mendukung dan 8 dokumen tidak mendukung. Kemudian *dataset* diuji dengan metode yang digunakan dalam penelitian.

Proses rangkaian pengujian menghasilkan data pada tabel 3.31. Tabel ini menjelaskan tentang hasil akurasi yang didapat melalui proses klasifikasi oleh sistem yang dibuat menggunakan perbandingan dokumen latih dan dokumen uji sebesar 70:30. Pembagian ini digunakan karena menurut (Vrigazova, 2021) pembagian proporsi terbaik berada pada 70/30 dalam sebuah penelitian klasifikasi.

Tabel. 3.31 Tabel Hasil Skenario Pengujian

No.	Persentase Mendukung	Akurasi
1	100%	-
2	90%	-
3	80%	0.67
4	70%	0.67
5	60%	0.67
6	50%	0.33
7	40%	0.33
8	30%	0.33
9	20%	0.33
10	10%	0.67

Tabel 3.30. memiliki hasil akurasi terbaik pada persentase mendukung sebanyak 10%, 60%, 70%, dan 80% dengan nilai akurasi 0,67 atau 67% sistem dapat memprediksi kelas suatu dokumen dengan benar. Pada persentase mendukung sebanyak 90%, dan 100% tidak mendapatkan hasil akurasi karena sistem mengalami *error* ketika memproses skenario pengujian tersebut. Prediksi yang tidak sesuai dapat menurunkan tingkat akurasi sistem yang dibuat sehingga perlu adanya proporsi pembagian yang tepat antara dokumen latih dan dokumen tes yang dihimpun. Data yang digunakan sebagai model tidak bisa menghasilkan *output* yang *reliable* karena memiliki keterbatasan untuk dapat memprediksi kemungkinan *output*. Tabel 3. 31 memberikan hasil akurasi terbaik pada beberapa kategori pembagian sekaligus. Diperlukan pengujian lebih lanjut untuk menemukan proporsi terbaik, sehingga dilakukan skenario pengujian dengan berbagai proporsi yang digunakan pada penelitian guna mencari hasil akurasi terbaiknya.

Pada langkah ini, *dataset* yang tersedia dipecah menjadi dua bagian. Bagian pertama adalah data latih, yang digunakan untuk melatih model dan mengajarkan pola-pola sentimen yang terdapat dalam teks. Bagian kedua adalah data uji, yang digunakan untuk menguji seberapa baik model dapat memprediksi sentimen dengan tingkat akurasi yang tinggi. Rincian pembagian antara data uji dan data latih ditunjukkan dalam Tabel 3.32.

Tabel 3.32 Pembagian Persentase Dokumen Uji

No.	Jumlah Dokumen	Persentase Dokumen latih Dari Total Dokumen	Dokumen latih	Dokumen uji
1.	500	10%	50	450
2.	500	20%	100	400
3.	500	30%	150	350
4.	500	40%	200	300
5.	500	50%	250	250
6.	500	60%	300	200
7.	500	70%	350	150
8.	500	80%	400	100
9.	500	90%	450	50

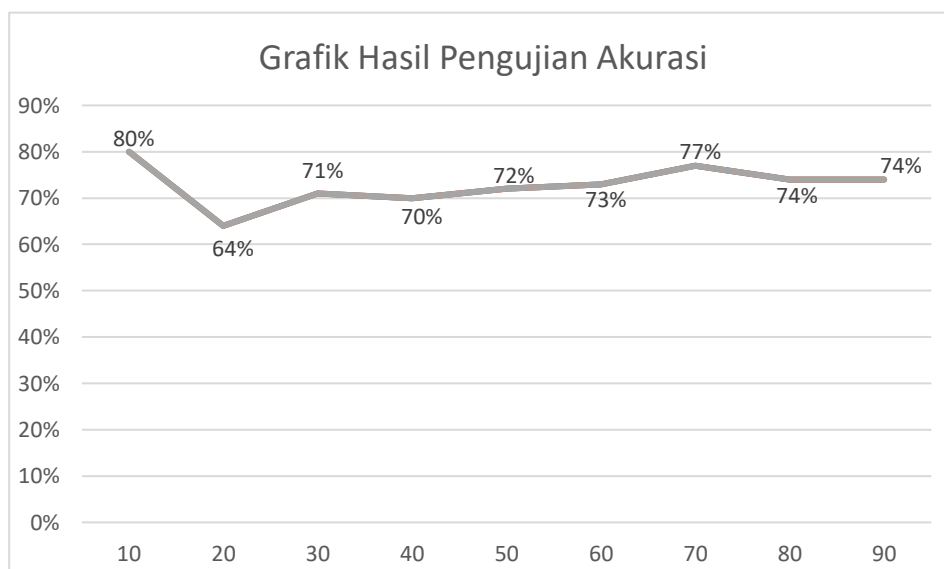
Pembagian data uji dan data latih dilakukan berulang kali hingga menemukan akurasi terbaiknya. Data hasil akurasi yang diuji dapat dilihat pada Tabel 3.33.

Tabel 3.33 Hasil Akurasi Pembagian Persentase Dokumen Uji

No.	Persentase Dokumen latih Dari Total Dokumen	Akurasi	Akurasi (%)
1.	10%	0,8	80%
2.	20%	0,64	64%
3.	30%	0,71	71%
4.	40%	0,7	70%
5.	50%	0,72	72%
6.	60%	0,73	73%
7.	70%	0,77	77%
8.	80%	0,74	74%
9.	90%	0,74	74%

Hasil akurasi pada tiap-tiap poin persentase data uji yang dimuat pada tabel 3.32. menyajikan hasil yang dihimpun pada tabel 3.33. Dari tiap poin persentase tersebut ditemukan bahwa hasil akurasi maksimal terdapat pada jumlah persentase

dokumen uji dari total dokumen keseluruhan yang berjumlah 10% dari total 500 dokumen. Sehingga dalam penelitian ini, menggunakan perbandingan 90:10 yang artinya 90% merupakan dokumen latih dan 10% merupakan dokumen uji. Berikut hasil pengujian akurasi yang ditampilkan dalam grafik.



Gambar 3.10 Grafik Hasil Pengujian Akurasi

Untuk meningkatkan akurasi dalam melakukan prediksi klasifikasi, diperlukan sebuah pendekatan yang memungkinkan untuk meningkatkan akurasi selain dengan mengatur jumlah data uji dan data latih dalam *dataset*. Cara yang diterapkan yaitu mengatur penggunaan jumlah fitur. Dengan pembagian fitur relevan serta signifikan dalam mempengaruhi kelas konkret yang telah dilabeli oleh ahli, hasil prediksi dapat menjadi lebih akurat. Dengan mengurangi fitur-fitur yang tidak relevan atau memiliki pengaruh yang minim terhadap klasifikasi, model dapat fokus pada aspek-aspek yang lebih penting dan mendapatkan hasil yang lebih sesuai dengan kelas yang diinginkan. Pengaturan jumlah fitur ini membantu memperbaiki keakuratan model klasifikasi dan mengurangi risiko adanya *overfitting* pada data.

Dengan demikian, penggunaan fitur-fitur yang tepat dapat memperkuat kemampuan model dalam mengklasifikasikan data secara lebih efektif dan akurat. Maka dari itu diperlukan pengaturan jumlah fitur yang dimuat dalam prediksi klasifikasi.

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Rangkaian Pengujian

Pengujian sistem ini dilaksanakan dengan mengikuti langkah-langkah pengujian yang telah disusun secara sistematis dan teratur. Hal ini bertujuan untuk memastikan bahwa proses pengujian pada sistem yang telah dibangun dapat berjalan dengan baik dan mudah dipahami. Tujuan utama dari pengujian ini adalah untuk mengukur tingkat akurasi, presisi, dan *recall* pada berbagai jumlah fitur yang digunakan dalam penelitian klasifikasi menggunakan metode *Gaussian Naïve Bayes* untuk mengklasifikasikan pendapat masyarakat di media sosial Twitter terkait pemindahan Ibu Kota Negara Indonesia. Sistem melakukan uji coba dengan mengikuti langkah-langkah seperti *text preprocessing*, penggunaan metode TF-IDF, dan penerapan algoritma *Gaussian Naïve Bayes*. Berikut ini adalah skenario pengujian yang digunakan dalam pengujian sistem.

4.1.1 Input Data

Pada tahap awal pengujian ini, langkah pertama yang dilakukan adalah mengimpor data ke dalam sistem. Data yang digunakan adalah hasil dari proses *scraping* pada platform media sosial Twitter, dan disimpan dalam format *.csv*. Jumlah total data yang ada adalah 500, terdiri dari 317 *tweet* yang diberi label "mendukung" dan 183 *tweet* yang diberi label "tidak mendukung". Proses pelabelan ini telah dilakukan sebelumnya oleh seorang tenaga ahli. Contoh *dataset* yang digunakan dapat dilihat pada Tabel 4.1 di bawah ini.

Tabel 4.1 Tabel Contoh *Input Data*

No.	Teks	Label
1.	Masyarakat di Kalimantan Timur khususnya Balikpapan, memiliki tekad untuk upgrading scale bisnisnya agar bisa bersaing dengan pendatang. #IKN	Mendukung
2.	Sektor UMKM mengalami pertumbuhan sejak pemerintah mulai pembangunan infrastruktur IKN. #IKN	Mendukung
3.	ikn sejahterakan ekonomi masyarakat kaltim dan buka peluang bagi UMKM #IKN	Mendukung
4.	Mari dukung IKN #ikn	Mendukung
5.	Heru: Perpindahan IKN Beri Peluang Revitalisasi Kota Lebih Baik	Mendukung
6.	BIKIN KAGET! PEJABAT OTORITA IKN MENGUNDURKAN DIRI	Tidak Mendukung
7.	Kenapa IKN di Kaltim? dr segi infrastruktur gak galah keren nya dr kota ² lain di Indonesia.	Tidak Mendukung
8.	Dulu Janji IKN Tidak Dibangun dengan APBN, Said Didu: Sekarang habiskan 1000 Triliun, Semua dari APBN!	Tidak Mendukung
9.	Pimpinan IKN Kompak Mundur Karena Takut Masuk Penjara Setelah Jokowi Lengser	Tidak Mendukung
...
500.	Realisasi Investasi di IKN Belum Ada, Menteri Basuki: Semua Masih dari APBN	Tidak Mendukung

4.1.2 Pembagian Data

Selanjutnya, dalam pengujian sistem, dilakukan pembagian *dataset* menjadi dua bagian, yaitu data latih dan data uji. Proporsi pembagiannya adalah 10% dari *dataset* untuk data latih dan 90% untuk data uji. Selain itu, *dataset* juga dibagi berdasarkan persentase penggunaan fitur yang akan dijelaskan pada poin berikutnya. Tujuan dari tahap ini adalah untuk

menentukan jumlah data yang akan digunakan dalam proses pelatihan dan pengujian, sehingga memastikan adanya data yang cukup untuk kedua proses tersebut. Pembagian 90% dan 10% dari total dokumen dipilih karena memiliki nilai akurasi tertinggi pada skenario pengujian.

4.1.3 Jumlah Fitur

Setelah itu, tindakan berikutnya adalah menentukan persentase penggunaan fitur yang akan digunakan dalam pengujian, dengan variasi persentase yang berbeda dari total fitur yang ada. Proses pembagian ini bertujuan untuk menguji tingkat akurasi sistem secara lebih komprehensif melalui beberapa variasi pengujian pada tabel 4.2.

Tabel. 4.2 Tabel Pembagian Persentase Penggunaan Fitur

Urutan Pengujian	Persentase Penggunaan Fitur
1	10 %
2	20 %
3	30 %
4	40 %
5	50 %
6	60 %
7	70 %
8	80 %
9	90 %

4.1.4 Pelatihan Data

Langkah awal dalam proses pelatihan data dimulai dengan membagi *dataset* berdasarkan jumlah kelas yang diinginkan. Dalam pengujian ini, *dataset* dibagi menjadi dua kelas, yaitu "mendukung" dan "tidak mendukung". Setelah *dataset* terbagi dalam kedua kelas tersebut, langkah

berikutnya adalah mengolah nilai rata-rata (*mean*) dan standar deviasi pada setiap kata yang terdapat dalam *dataset*.

4.1.5 Pengujian Data

Langkah ini merupakan fase terakhir dari algoritma yang diterapkan dalam sistem. Pada tahap ini, dilakukan perhitungan untuk mendapatkan hasil klasifikasi dari dokumen uji menerapkan persamaan dasar probabilitas *Gaussian Naïve Bayes* sebagaimana disajikan dalam persamaan (3.14). Variabel x mewakili bobot data uji, sementara nilai rata-rata (*mean*) dan *varians* diambil dari perhitungan pada data pelatihan. Selanjutnya, yang diperoleh untuk setiap kata dikalikan dengan nilai probabilitas *prior* dari masing-masing kelas untuk memperoleh probabilitas hasil pada setiap dokumen dalam setiap kelas. Hasil probabilitas tertinggi kemudian dijadikan sebagai klasifikasi akhir dari dokumen tersebut.

4.1.5 Hasil Evaluasi

Setelah memperoleh hasil dari data yang telah diklasifikasikan, langkah berikutnya adalah menghitung total akurasi dalam penerapan klasifikasi menggunakan metode *Naïve Bayes*. Hasil pengujian skenario akan digunakan untuk menghitung nilai akurasi, presisi, dan *recall* menggunakan Matriks Konfusi (*Confusion Matrix*). Perhitungan jumlah akurasi, presisi, dan *recall* dijalankan dengan mengacu pada persamaan (4.1), (4.2), dan (4.3).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} 100\% \quad (4.1)$$

$$Precision = \frac{TP}{TP + FP} 100\% \quad (4.2)$$

$$Recall = \frac{TP}{TP + FN} 100\% \quad (4.1)$$

Tabel. 4.3 Keterangan Rumus *Confussion Matrix*

No.	Keterangan	Definisi
1.	<i>True Positive</i> (TP)	banyaknya dokumen konkret berlabel label mendukung (1) lalu hasil dari prediksinya juga berlabel mendukung (1) atau data positif yang terprediksi positif.
2.	<i>True Negative</i> (TN)	banyaknya dokumen konkret berlabel tidak mendukung (0) namun hasil prediksinya berlabel tidak mendukung (0) atau data negatif yang terprediksi negatif.
3.	<i>False Positive</i> (FP)	banyaknya dokumen konkret berlabel tidak mendukung (0) namun hasil prediksinya berlabel mendukung (1) atau data negatif yang terprediksi positif.
4.	<i>False Negative</i> (FN)	banyaknya dokumen konkret berlabel mendukung (1) namun hasil prediksinya berlabel negatif (0). atau data positif yang terprediksi negatif.

4.2 Hasil Pengujian

Melalui proses uji yang dijalankan dengan variasi persentase fitur yang telah disebutkan, diperoleh hasil yang dapat mengindikasikan keakuratan sistem terhadap *class* konkret yang sudah ditentukan. Hasil tersebut disajikan berdasarkan setiap persentase total fitur. Dengan menggunakan data ini, dapat dievaluasi sejauh mana keselarasan dari kelas prediksi dengan kelas konkret, yang memungkinkan dilakukannya perhitungan untuk mengevaluasi kemampuan sistem melalui *Confusion Matrix*. Matriks kebingungan digunakan untuk mengukur sejauh mana klasifikasi yang dilakukan oleh sistem sesuai dengan kelas konkret, dengan memberikan informasi tentang total *true positive*, *true negative*, *false positive*, dan *false negative*. Oleh sebab itu, penggunaan metode *Gaussian Naive Bayes* dengan jumlah fitur yang telah dioptimalkan dapat memberikan hasil yang akurat dan dapat

dievaluasi secara objektif menggunakan *confusion matrix* untuk mengukur kemampuan sistem.

Dalam Tabel 4.4, terdapat ringkasan nilai hasil pengujian yang telah dilakukan dengan menggunakan 10% fitur. Nilai-nilai yang tercatat dalam tabel tersebut selanjutnya dapat diolah menggunakan rumus (4.1), (4.2), dan (4.3) untuk mengukur kemampuan sistem yang telah dikembangkan. Persamaan tersebut akan memberikan informasi yang berguna dalam mengevaluasi sejauh mana sistem mampu melakukan klasifikasi dengan akurasi yang tinggi. Dengan memanfaatkan rumus (4.1), (4.1), dan (4.3) peneliti sanggup menghitung persentase keberhasilan sistem dan menganalisis performa yang telah dicapai berdasarkan data yang terdokumentasi dalam Tabel 4.4. Ini akan memberikan pemahaman yang lebih komprehensif mengenai performa sistem dan kemampuannya dalam melakukan klasifikasi dengan tingkat akurasi yang diinginkan.

Tabel 4.4 Hasil Uji Penggunaan Total Fitur Sebesar 10%

Nomor	Konkret	Terprediksi	TP	TN	FP	FN
1	0	1	0	0	1	0
2	1	1	1	0	0	0
3	1	1	1	0	0	0
4	0	1	0	0	1	0
5	1	0	0	0	0	1
6	0	1	0	0	1	0
7	0	1	0	0	1	0
8	1	1	1	0	0	0
9	1	1	1	0	0	0
10	1	1	1	0	0	0
11	1	0	0	0	0	1
12	0	1	0	0	1	0
13	1	0	0	0	0	1
14	1	1	1	0	0	0
15	1	1	1	0	0	0
...

Nomor	Konkret	Terprediksi	TP	TN	FP	FN
50	1	1	1	0	0	0
Total			35	0	11	4

$$Accuracy = \frac{35+0}{35+0+11+4} 100\% = 70,0\%$$

$$Precision = \frac{35}{35+11} 100\% = 76,0\%$$

$$Recall = \frac{35}{35+4} 100\% = 89,7\%$$

Dalam Tabel 4.5, terdapat jumlah nilai hasil uji yang dijalankan dengan penggunaan 20% fitur. Nilai-nilai tersebut selanjutnya dapat digunakan dalam perhitungan menggunakan rumus (4.1), (4.2), dan (4.3) untuk mengukur tingkat kemampuan sistem yang telah dikembangkan.

Tabel 4.5 Hasil Uji Penggunaan Total Fitur Sebesar 20%

Nomor	Konkret	Terprediksi	TP	TN	FP	FN
1	1	1	1	0	0	0
2	1	1	1	0	0	0
3	1	1	1	0	0	0
4	1	1	1	0	0	0
5	1	1	1	0	0	0
6	1	1	1	0	0	0
7	1	1	1	0	0	0
8	1	1	1	0	0	0
9	0	1	0	0	1	0
10	1	1	1	0	0	0
11	0	1	0	0	1	0
12	0	1	0	0	1	0
13	1	1	1	0	0	0
14	1	1	1	0	0	0
15	1	1	1	0	0	0
...
50	1	1	1	0	0	0
Total			32	1	16	1

$$Accuracy = \frac{32+1}{32+1+16+1} 100\% = 66,0\%$$

$$Precision = \frac{32}{32+16} 100\% = 66,6\%$$

$$Recall = \frac{32}{32 + 1} 100\% = 96,9\%$$

Dalam Tabel 4.6, terdapat jumlah nilai hasil uji yang dijalankan dengan penggunaan 30% fitur. Nilai-nilai tersebut selanjutnya dapat digunakan dalam perhitungan menggunakan rumus (4.1), (4.2), dan (4.3) untuk mengukur tingkat kemampuan sistem yang telah dikembangkan.

Tabel 4.6 Hasil Uji Penggunaan Total Fitur Sebesar 30%

Nomor	Konkret	Terprediksi	TP	TN	FP	FN
1	1	1	1	0	0	0
2	1	1	1	0	0	0
3	1	1	1	0	0	0
4	0	1	0	0	1	0
5	1	1	1	0	0	0
6	1	1	1	0	0	0
7	0	1	0	0	1	0
8	1	1	1	0	0	0
9	1	1	1	0	0	0
10	1	1	1	0	0	0
11	0	1	0	0	1	0
12	1	1	1	0	0	0
13	1	1	1	0	0	0
14	1	1	1	0	0	0
15	1	1	1	0	0	0
...
50	1	1	1	0	0	0
Total			28	7	12	3

$$Accuracy = \frac{28 + 7}{28 + 7 + 12 + 3} 100\% = 70,0\%$$

$$Precision = \frac{28}{28 + 12} 100\% = 70,0\%$$

$$Recall = \frac{28}{28 + 3} 100\% = 90,3\%$$

Dalam Tabel 4.7, terdapat jumlah nilai hasil uji yang dijalankan dengan penggunaan 40% fitur. Nilai-nilai tersebut selanjutnya dapat digunakan dalam

perhitungan menggunakan rumus (4.1), (4.2), dan (4.3) untuk mengukur tingkat kemampuan sistem yang telah dikembangkan.

Tabel 4.7 Hasil Uji Penggunaan Total Fitur Sebesar 40%

Nomor	Konkret	Terprediksi	TP	TN	FP	FN
1	0	1	0	0	1	0
2	0	1	0	0	1	0
3	1	1	1	0	0	0
4	1	1	1	0	0	0
5	1	1	1	0	0	0
6	1	1	1	0	0	0
7	1	1	1	0	0	0
8	1	1	1	0	0	0
9	1	1	1	0	0	0
10	0	1	0	0	1	0
11	1	1	1	0	0	0
12	1	1	1	0	0	0
13	1	1	1	0	0	0
14	1	1	1	0	0	0
15	1	1	1	0	0	0
...
50	0	1	1	0	1	0
Total			27	8	12	2

$$Accuracy = \frac{27 + 8}{27 + 8 + 12 + 2} 100\% = 71,4\%$$

$$Precision = \frac{27}{27 + 12} 100\% = 69,2\%$$

$$Recall = \frac{27}{27 + 2} 100\% = 93,1\%$$

Dalam Tabel 4.8, terdapat jumlah nilai hasil uji yang dijalankan dengan penggunaan 50% fitur. Nilai-nilai tersebut selanjutnya dapat digunakan dalam perhitungan menggunakan rumus (4.1), (4.2), dan (4.3) untuk mengukur tingkat kemampuan sistem yang telah dikembangkan.

Tabel 4.8 Hasil Uji Penggunaan Total Fitur Sebesar 50%

Nomor	Konkret	Terprediksi	TP	TN	FP	FN
1	1	1	1	0	0	0
2	0	1	0	0	1	0
3	1	1	1	0	0	0
4	1	1	1	0	0	0
5	1	1	1	0	0	0
6	1	1	1	0	0	0
7	0	0	0	1	0	0
8	1	1	1	0	0	0
9	1	1	1	0	0	0
10	1	1	1	0	0	0
11	1	1	1	0	0	0
12	0	1	0	0	1	0
13	1	1	1	0	0	0
14	1	1	1	0	0	0
15	0	1	0	0	1	0
...
50	1	1	1	0	0	0
Total			31	4	15	0

$$Accuracy = \frac{31 + 4}{31 + 4 + 15 + 0} 100\% = 76\%$$

$$Precision = \frac{31}{31 + 15} 100\% = 67,3\%$$

$$Recall = \frac{31}{31 + 0} 100\% = 100$$

Dalam Tabel 4.9, terdapat jumlah nilai hasil uji yang dijalankan dengan penggunaan 60% fitur. Nilai-nilai tersebut selanjutnya dapat digunakan dalam perhitungan menggunakan rumus (4.1), (4.2), dan (4.3) untuk mengukur tingkat kemampuan sistem yang telah dikembangkan.

Tabel 4.9 Hasil Uji Penggunaan Total Fitur Sebesar 60%

Nomor	Konkret	Terprediksi	TP	TN	FP	FN
1	1	1	1	0	0	0
2	1	1	1	0	0	0
3	1	1	1	0	0	0
4	1	1	1	0	0	0
5	1	1	1	0	0	0
6	1	1	1	0	0	0

Nomor	Konkret	Terprediksi	TP	TN	FP	FN
7	0	1	0	0	1	0
8	1	1	1	0	0	0
9	1	1	1	0	0	0
10	1	1	1	0	0	0
11	0	1	0	0	1	0
12	0	0	0	1	0	0
13	1	1	1	0	0	0
14	1	1	1	0	0	0
15	1	1	1	0	0	0
...
50	1	1	1	0	0	0
Total			30	6	12	2

$$Accuracy = \frac{30 + 6}{30 + 6 + 12 + 2} 100\% = 72\%$$

$$Precision = \frac{30}{30 + 12} 100\% = 71,4\%$$

$$Recall = \frac{30}{30 + 2} 100\% = 93,7\%$$

Dalam Tabel 4.10, terdapat jumlah nilai hasil uji yang dijalankan dengan penggunaan 70% fitur. Nilai-nilai tersebut selanjutnya dapat digunakan dalam perhitungan menggunakan rumus (4.1), (4.2), dan (4.3) untuk mengukur tingkat kemampuan sistem yang telah dikembangkan.

Tabel 4.10 Hasil Uji Penggunaan Total Fitur Sebesar 70%

Nomor	Konkret	Terprediksi	TP	TN	FP	FN
1	1	1	1	0	0	0
2	1	1	1	0	0	0
3	1	1	1	0	0	0
4	0	1	0	0	1	0
5	1	1	1	0	0	0
6	1	1	1	0	0	0
7	1	1	1	0	0	0
8	1	1	1	0	0	0
9	1	1	1	0	0	0
10	1	1	1	0	0	0
11	1	1	1	0	0	0

Nomor	Konkret	Terprediksi	TP	TN	FP	FN
12	1	1	1	0	0	0
13	0	1	0	0	1	0
14	1	1	1	0	0	0
15	0	1	0	0	1	0
...
50	1	1	1	0	0	0
Total			26	6	14	4

$$Accuracy = \frac{26 + 6}{26 + 6 + 14 + 4} 100\% = 70,0\%$$

$$Precision = \frac{26}{26 + 14} 100\% = 65,0\%$$

$$Recall = \frac{26}{26 + 4} 100\% = 86,6\%$$

Dalam Tabel 4.11, terdapat jumlah nilai hasil uji yang dijalankan dengan penggunaan 80% fitur. Nilai-nilai tersebut selanjutnya dapat digunakan dalam perhitungan menggunakan rumus (4.1), (4.2), dan (4.3) untuk mengukur tingkat kemampuan sistem yang telah dikembangkan.

Tabel 4. 11 Hasil Uji Penggunaan Total Fitur Sebesar 80%

Nomor	Konkret	Terprediksi	TP	TN	FP	FN
1	1	1	1	0	0	0
2	1	1	1	0	0	0
3	1	1	1	0	0	0
4	1	1	1	0	0	0
5	1	1	1	0	0	0
6	0	1	0	0	1	0
7	1	1	1	0	0	0
8	1	1	1	0	0	0
9	1	1	1	0	0	0
10	1	1	1	0	0	0
11	1	1	1	0	0	0
12	1	1	1	0	0	0
13	1	1	1	0	0	0
14	1	1	1	0	0	0
15	1	1	1	0	0	0
...
50	0	0	0	1	0	0
Total			30	11	9	0

$$Accuracy = \frac{35 + 0}{35 + 11 + 9 + 0} 100\% = 82,0\%$$

$$Precision = \frac{35}{35 + 9} 100\% = 76,9\%$$

$$Recall = \frac{35}{35 + 0} 100\% = 100\%$$

Dalam Tabel 4.12, terdapat jumlah nilai hasil uji yang dijalankan dengan penggunaan 90% fitur. Nilai-nilai tersebut selanjutnya dapat digunakan dalam perhitungan menggunakan rumus (4.1), (4.2), dan (4.3) untuk mengukur tingkat kemampuan sistem yang telah dikembangkan.

Tabel 4.12 Hasil Uji Penggunaan Total Fitur Sebesar 90%

Nomor	Konkret	Terprediksi	TP	TN	FP	FN
1	0	1	0	0	1	0
2	1	1	1	0	0	0
3	1	1	1	0	0	0
4	0	1	1	0	0	0
5	1	0	1	0	0	0
6	0	1	1	0	0	0
7	0	1	1	0	0	0
8	1	1	1	0	0	0
9	1	1	1	0	0	0
10	1	1	0	0	1	0
11	1	0	1	0	0	0
12	0	1	0	0	1	0
13	1	0	1	0	0	0
14	1	1	1	0	0	0
15	1	1	1	0	0	0
...
50	1	1	1	0	0	0
Total			25	8	17	0

$$Accuracy = \frac{25 + 8}{25 + 8 + 17 + 0} 100\% = 77,3\%$$

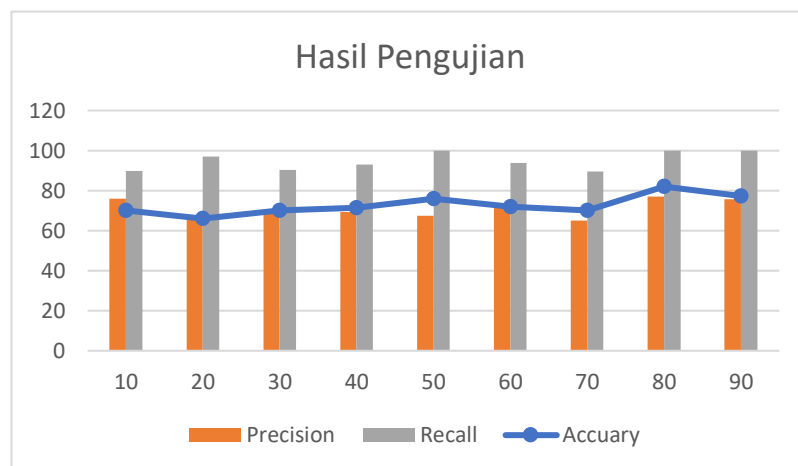
$$Precision = \frac{25}{25 + 17} 100\% = 75,7\%$$

$$Recall = \frac{25}{25 + 0} 100\% = 100\%$$

Hasil *accuracy*, *precision*, dan *recall* pada setiap persentase jumlah fitur yang dihimpun dalam penelitian disajikan pada Tabel 4.13 dan Gambar 4.1 berikut ini.

Tabel 4.13 Hasil Uji Pada *Confussion Matrix*

Total Fitur	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
10%	70,0%	76,0%	89,7%
20%	66,0%	66,6%	96,9%
30%	70,0 %	70,0%	90,3%
40%	71,4 %	69,2%	93,1%
50%	76 %	67,3%	100%
60%	72 %	71,4%	93,7%
70%	70,0 %	65,0%	89,5%
80%	82,0 %	76,9%	100%
90%	77,3 %	75,7%	100%



Gambar 4.1 Visualisasi Hasil Uji

4.3 Pembahasan

Dari proses skenario dan *outout* uji yang telah dijelaskan pada bab sebelumnya, disimpulkan bahwa setiap tahapan dalam proses uji memiliki dampak terhadap *output* klasifikasi. Skenario pengujian dengan berbagai jumlah dokumen mendukung menunjukkan hasil yang termuat pada Tabel 3.30. berisi hasil akurasi

terbaik berada pada penggunaan persentase mendukung sebanyak 10%, 60%, 70%, dan 80% dengan menggunakan metode yang diimplementasikan pada penelitian sebesar 66,7%. Setelah diketahui hasil akurasi dari beberapa jumlah persentase mendukung yang digunakan dalam skenario pengujian pertama, selanjutnya dilakukan pengujian pada skenario kedua menggunakan total dokumen sebanyak 500 dokumen. Pengujian dilakukan untuk mengetahui proporsi pembagian total dokumen latih serta total dokumen uji yang terbaik. *Output* pengujian kedua menunjukkan proporsi terbaik terdapat pada 10:90 atau 90% dokumen yang dilatih dan 10% dokumen yang diuji mencapai hasil nilai akurasi sebanyak 80% seperti penyajian dalam tabel 3.33. Kemudian hasil skenario pengujian ini dipakai untuk uji coba sistem pada bab berikutnya.

Pengujian yang dilakukan sebanyak Sembilan kali dengan penggunaan persentase fitur yang berbeda. Pada persentase 10% menghasilkan nilai akurasi sebanyak 70,0%, *precision* sebanyak 76,0%, dan *recall* sebanyak 89,7%. Pada persentase 20% menghasilkan nilai akurasi sebanyak 66,0%, *precision* sebanyak 66,6%, dan *recall* sebanyak 96,9%. Pada persentase 30% menghasilkan nilai akurasi sebanyak 70,0%, *precision* sebanyak 70,0%, dan *recall* sebanyak 90,3%. Pada persentase 40% menghasilkan nilai akurasi sebanyak 71,4%, *precision* sebanyak 69,2%, dan *recall* sebanyak 93,1%. Pada persentase 50% menghasilkan nilai akurasi sebanyak 76%, *precision* sebanyak 67,3%, dan *recall* sebanyak 100%. Pada persentase 60% menghasilkan nilai akurasi sebanyak 72%, *precision* sebanyak 71,4%, dan *recall* sebanyak 93,7%. Pada persentase 70% menghasilkan nilai akurasi sebanyak 70,0%, *precision* sebanyak 65,0%, dan *recall* sebanyak

89,5%. Pada persentase 80% menghasilkan nilai akurasi sebanyak 82,0%, *precision* sebanyak 76,9%, dan *recall* sebanyak 100%. Pengujian terakhir pada persentase 90% menghasilkan nilai akurasi sebesar 77,3%, *precision* sebanyak 75,7%, dan *recall* sebanyak 100%.

Informasi pada Gambar 4.1. Visualisasi Hasil Pengujian, dapat diketahui bahwa proses pengklasifikasi *tweet* dengan metode *Gaussian Naïve Bayes* memberikan hasil sistem maksimal ketika menggunakan fitur sebanyak 80%. Hasil pengujian menunjukkan bahwa sistem berhasil mencapai tingkat akurasi sebesar 82,0%, *precision* sebesar 76,9%, dan *recall* sebesar 100%. Angka tersebut mengindikasikan sejauh mana sistem mampu mengklasifikasikan *tweet* dengan tepat berdasarkan kelas yang telah ditentukan. Dengan menggunakan 80% fitur yang dipilih berdasarkan performa terbaik terhadap persentase lainnya, sistem dapat memberikan hasil yang memuaskan dalam pengklasifikasian *tweet* dengan hasil terbaik.

Penggunaan total fitur yang sangat sedikit dapat mengakibatkan tereliminasi fitur krusial yang dipakai untuk klasifikasi yang mengakibatkan nilai pada akurasi condong ke arah yang lebih rendah. Akan tetapi, terlalu banyak menggunakan jumlah pada fitur tidak menjanjikan nilai akurasi terbaik karena termuat banyak fitur sehingga menyebabkan kesalahan pada klasifikasi.

Penggunaan fitur sebesar 80% menghasilkan jumlah akurasi sebanyak 82,0% dimana nilai tersebut adalah jumlah akurasi tertinggi apabila disandingkan dengan menggunakan fitur yang lebih banyak karena kata fitur diseleksi sehingga menggunakan lebih sedikit fitur. Hasil penelitian menunjukkan bahwa penggunaan

lebih sedikit fitur meningkatkan hasil akurasi lebih maksimal daripada melibatkan lebih banyak fitur. Nilai akurasi ini menunjukkan berapa persentase dokumen pelatihan yang terprediksi mendukung dan tidak mendukung dari keseluruhan dokumen. Semakin tinggi *accuracy*, semakin baik kinerja dokumen pelatihan dalam mengklasifikasikan sentimen masyarakat secara keseluruhan.

Penggunaan fitur sebesar 80% memiliki nilai *precision* sebesar 76,9%. Hasil ini bukan yang terbaik, namun masih tergolong baik. Skor presisi merupakan hasil dari perhitungan yang dipengaruhi oleh jumlah *false positive*. Ketika jumlah *false positive* semakin rendah, maka skor presisi akan meningkat. Nilai dari *precision* menunjukkan berapa persentase dokumen pelatihan yang terprediksi mendukung dari total dokumen yang terprediksi mendukung. *Precision* yang tinggi menunjukkan bahwa dokumen pelatihan cenderung memberikan label tidak mendukung hanya pada *instance* yang benar-benar mengandung pendapat tidak mendukung. Hal ini penting karena memberikan keyakinan bahwa pendapat tidak mendukung yang diidentifikasi oleh dokumen pelatihan adalah valid.

Penggunaan fitur sebesar 80% memiliki nilai *recall* sebesar 100%. Hasil ini bukan juga yang terbaik, namun masih tergolong baik. Skor recall didapatkan dari perhitungan yang terpengaruh oleh jumlah *false negative*. Ketika jumlah *false negative* semakin rendah, maka skor *recall* akan meningkat. Nilai dari *recall* adalah nilai persentase dokumen yang diprediksi mendukung dibandingkan dengan total dokumen yang sebenarnya mendukung. *Recall* yang tinggi menunjukkan bahwa dokumen pelatihan memiliki kemampuan yang baik untuk menemukan *instance* mendukung, yang dalam konteks ini adalah kelas tidak mendukung.

Di samping itu, dalam tahap seleksi pada fitur, terdapat kata-kata krusial yang mestinya dapat meningkatkan kemampuan sistem. Sayangnya, kata-kata tersebut tidak digunakan karena memiliki bobot yang rendah dalam peringkat fitur. Keputusan ini dapat berdampak pada hasil pengklasifikasian yang diperoleh oleh sistem. Dalam seleksi fitur, harus dilakukan evaluasi yang lebih cermat untuk memastikan kata-kata hal tersebut tidak terabaikan.

Meskipun tidak secara langsung membahas tentang klasifikasi, sorotan dalam integrasi islam terdapat pada pentingnya menjaga dan memperbaiki interaksi sosial serta memberikan panduan moral dalam berkomunikasi dengan orang lain. Seperti yang terdapat pada QS. Al-Isra:53)

وَقُلْ لِعِبَادِي يَقُولُوا الَّتِي هِيَ أَحْسَنُ ۚ إِنَّ الشَّيْطَانَ يَنْزِعُ بَيْنَهُمْ ۗ إِنَّ الشَّيْطَانَ كَانَ لِلْإِنْسَانِ عَدُوًّا مُّبِينًا

"Dan ucapkanlah kepada hamba-hamba-Ku, bahwa mereka hendaklah mengucapkan perkataan yang lebih baik. Sesungguhnya setan itu menimbulkan permusuhan di antara mereka. Sesungguhnya setan itu adalah musuh yang nyata bagi manusia." (QS. Al-Isra:53)

Potongan Ayat ini mengingatkan kita untuk menerapkan penggunaan opini yang positif dan bermanfaat dalam berkomunikasi dengan orang lain. Apabila mereka tidak menggunakan kata yang bersifat positif, maka mengakibatkan terpaparnya pengaruh negatif yang berasal dari syaitan yang akan mempengaruhi tindakan mereka. Akibatnya, terjadi konflik dan pertentangan di antara mereka (Muhardisyah, 2017). Pada keseharian, sering kali ditemukan perbedaan pendapat, tidak terkecuali dengan kebijakan yang diterapkan oleh pemerintah. Dalam konteks ini, diharapkan dapat menjadi pengingat untuk menjaga dan memperbaiki argumen yang diungkapkan secara lisan atau tertulis.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Evaluasi kinerja sistem diterapkan dengan menggunakan beberapa metrik seperti akurasi, presisi, dan *recall*. Hasil evaluasi menunjukkan bahwa sistem mencapai tingkat accuracy terbaik sejumlah 82,0 %, *precision* sejumlah 76,9%, dan *recall* sejumlah 100%. Hasil ini menunjukkan bahwa sistem memiliki kemampuan yang cukup baik dalam mengklasifikasikan *tweet*. Untuk mencapai hasil ini, dilakukan berbagai bentuk pengujian yang telah disiapkan dalam skenario pengujian. Proses pengujian yang diterapkan menghasilkan kesimpulan informasi yang lebih baik tentang performa sistem dan memastikan bahwa algoritma *Gaussian Naïve Bayes* memberikan hasil maksimal dengan menggunakan fitur sebanyak 80% dari keseluruhan fitur yang didapatkan. Penggunaan data yang tidak seimbang mengakibatkan sistem mengalami *overfitting*, di mana prediksi cenderung condong ke arah klasifikasi positif. Hal ini berdampak pada hasil prediksi, terutama ketika adanya kekurangan data pada kelas tidak mendukung. Akibatnya, sering kali terjadi situasi di mana hasil terprediksi adalah mendukung, padahal data konkretnya adalah tidak mendukung.

5.1 Saran

Penelitian mengenai klasifikasi *tweet* menggunakan metode *Gaussian Naïve Bayes* masih terdapat kelemahan. Berdasarkan penjelasan tersebut, penelitian berikutnya diharapkan dapat mencapai hasil pengklasifikasian yang semakin

akurat. Maka dari itu, penulis menyajikan beberapa masukan guna pengembangan selanjutnya, yaitu:

1. Menambahkan jumlah data set sehingga meningkatkan jumlah penggunaan fitur pada kata dalam kamus kata di dokumen pelatihan. Harapannya, Tindakan ini dapat meningkatkan kapabilitas sistem untuk melakukan klasifikasi.
2. Memaksimalkan proses *preprocessing* dengan mengidentifikasi kata dari bahasa non formal. Ini dapat dilakukan dengan cara memperbanyak langkah lain dalam tahap pembersihan serta normalisasi untuk mengenali kata non formal seperti kata singkat, slang, dan kata berbahasa asing yang populer digunakan dalam percakapan.
3. Mengeksplorasi metode lain dalam tahap seleksi fitur, seperti *Chi-Square*, *Information Gain*, dan lain-lain. Hal ini dapat membantu meningkatkan performa sistem dalam memilih fitur-fitur yang paling relevan untuk klasifikasi.

Dengan mengimplementasikan saran-saran ini, diharapkan penelitian selanjutnya dapat mengatasi kelemahan yang ada dan menghasilkan hasil klasifikasi yang lebih akurat dan efektif.

DAFTAR PUSTAKA

- Ahmad, M., Aftab, S., Ali, I., & Hameed, N. (2017). Hybrid tools and techniques for sentiment analysis: a review. *International Journal of Multidisciplinary Sciences and Engineering*, 8(4), 28–33. www.ijmse.org
- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. In *Agustus* (Vol. 10, Issue 2).
- Borislava Vrigazova. The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems. *Business Systems Research* Vol. 12 No. 1 2021
- Hadi, F., & Ristawati, R. (2020). Pemandangan Ibu Kota Indonesia dan Kekuasaan Presiden dalam Perspektif Konstitusi. *Jurnal Konstitusi*, 17(3), 530–557. <https://doi.org/10.31078/jk1734>
- Hutasoit, W. L. (2018). Analisa Pemandangan Ibiukota Negara Oleh : Wesley Liano Hutasoit Dosen Fisipol Universitas 17 Agustus 1945 Samarinda. *Dedikasi*, 19(2), 108–128. <http://ejurnal.untag-smd.ac.id/index.php/dedikasi/article/view/3989>
- Imron, A. (2019). Analisis Sentimen Terhadap Tempat Wisata di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier. *Teknik Informatika*, 10–13. <https://dspace.uin.ac.id/handle/123456789/14268>
- Krisdiyanto, T., Maricha, E., & Nurharyanto, O. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naive Bayes Classifier. *Jurnal CoreIT*, 7(1).
- Kurdi, S., Mubibah, J., & Faizah, U. (2017). Konsep Taat Kepada Pemimpin (Ulil Amri) Di Dalam Surah an-Nisa : 59, Al-Anfal :46 Dan Al-Maidah : 48-49 (Analisis Tafsir Tafsir Al-Qurthubi, Al-Mishbah, Dan Ibnu Katsir). *Jurnal of Islamic Law and Studies*, 1(1), 13. <https://jurnal.uin-antasari.ac.id/index.php/jils/article/view/2552>
- Malik Zuhdi, A., Utami, E., & Raharjo, S. (2019). *Analisa Sentiment Twitter Terhadap Capres Indonesia 2019 Dengan Metode K-NN* (Vol. 5).
- Mardiana, L., & Zi'ni, A. F. (2020). Pengungkapan Diri Pengguna Akun Autobase Twitter @SUBTANYARL. *Jurnal Audience*, 3(1), 34–54. <https://doi.org/10.33633/ja.v3i1.4134>
- Mohammad Al-Aidaros, K., Abu Bakar, A., & Othman, Z. (2010). Naive Bayes Variants in Classification Learning. *Exploring the Relation between*

Organizational Learning and Knowledge Management for Improving Performance, 297–302.

Muhardisyah. Etika Dalam Komunikasi Islam. *Jurnal Peurawi* Vol.1 No. 1 Tahun 2017

Ratniasih, N. L., Sudarma, M., & Gunantara, N. (2017). Penerapan Text Mining Dalam Spam Filtering Untuk Aplikasi Chat. *Majalah Ilmiah Teknologi Elektro*, 16(3), 13. <https://doi.org/10.24843/mite.2017.v16i03p03>

Ridho, A. J. (2020). Kebijakan Pemindahan Ibu Kota Negara Republik Indonesia Dalam Perspektif Fikih Siyasa. In *Molecules*.

Saraswati, M. K., Ariananto, E., Adi, W., Perhubungan, K. S., Sekretariat, U., Ri, K., Hukum, A., Subbidang, S., Iklim, P., Lingkungan, P., & Sekretariat, H. (2022). Pemindahan Ibu Kota Negara Ke Provinsi Kalimantan Timur Berdasarkan Analisis Swot. *Jurnal Ilmu Sosial Dan Pendidikan (JISIP)*, 6(2), 2598–9944. <https://doi.org/10.36312/jisip.v6i1.3086/http>

Setianingrum, A. H., Kalokasari, D. H., & Shofi, I. M. (2018). Implementasi Algoritma Multinomial Naive Bayes Classifier. *JURNAL TEKNIK INFORMATIKA*, 10(2), 109–118. <https://doi.org/10.15408/jti.v10i2.6822>

Sidiq, A. A., & Christanto, F. W. (2020). Algoritma Naive Bayes Untuk Penentuan Pkh (Program Keluarga Harapan) Berbasis Sistem Pendukung Kepu-Tusan (Studi Kasus: Kelurahan Karanganyar Gunung Se-Marang). *Jurnal Riptek*, 14(1), 65–71.

Suhartono, Fachrul Kurniawan, Bahtiar Imran. (2018). Identifikasi tanaman irtual menggunakan jaringan bayesian berdasarkan sistem-L parametrik. *Jurnal Internasional Kemajuan dalam Informatika Cerdas* Vol. 4, No. 1, Maret 2018, hlm. 40-52. ISSN 2442-6571.

Tim Pustaka Phoenix. (2013). *Kamus Besar Bahasa Indonesia, Macet*.

Wibawa, A. P., Guntur, M., Purnama, A., Fathony Akbar, M., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, 3(1).

Yuli Mardi. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . *Jurnal Edik Informatika*. *Jurnal Edik Informatika*, 2(2), 213–219.