PERBANDINGAN PEMBOBOTAN TF-IDF DAN *WORD2VEC* PADA ANALISIS SENTIMEN MASYARAKAT TERHADAP IBU KOTA NUSANTARA (IKN) DI MEDIA SOSIAL X MENGGUNAKAN METODE *NAÏVE BAYES*

SKRIPSI

Diajukan untuk Memenuhi Sebagian Syarat Guna Memperoleh
Gelar Sarjana Program Strata 1 (S.1)
Dalam Ilmu Teknologi Informasi

Diajukan Oleh:

NISFAH LAILI FIKRIA
NIM: 2008096007

PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI WALISONGO SEMARANG

2025

PERNYATAAN KEASLIAN

Yang bertandatangan dibawah ini:

Nama	: Nisfah Laili Fikria
NIM	: 2008096007
Jurusan	: Teknologi Informasi

Menyatakan bahwa skripsi yang berjudul:

PERBANDINGAN PEMBOBOTAN TF-IDF DAN *WORD2VEC* PADA ANALISIS SENTIMEN MASYARAKAT TERHADAP IBU KOTA NUSANTARA (IKN) DI MEDIA SOSIAL X MENGGUNAKAN METODE NAÏVE BAYES

Secara keseluruhan adalah hasil penelitian/karya saya sendiri, kecuali bagian tertentu yang dirujuk sumbernya.

Semarang, 9 Mei 2025

Yang menyatakan,

Fikria

96007

PENGESAHAN

Naskah skripsi berikut ini :

Judul	: Perbandingan pembobotan TF-IDF dan Word2Vec Pada Analisis Sentimen Masyarakat Terhadap Ibu Kota Nusantara (IKN) Di Media Sosial X Menggunakan Metode Naive Bayes
Penulis	: **Nisfah Laili Fikria**
NIM	: 2008096007
Jurusan	: Teknologi Informasi

Telah diajukan dalam sidang tugas akhir oleh Dewan Penguji Fakultas Sains dan Teknologi UIN Walisongo dan dapat diterima sebagai salah satu syarat memperoleh gelar sarjana dalam Teknologi Informasi.

Semarang, 5 Mei 2025

Dewan Penguji

Penguji I,

Hery Mustofa, M.Kom
NIP. 198703172019031007

Penguji II,

Siti Nur'aini, S.Kom., M.Kom
NIP.198401312018012001

Penguji III,

Dr. Wenty Dwi Yuniarti, S.Pd., M.Kom
NIP. 197706222006042005

Penguji IV,

Masy Ari Ulinuha, M.T
NIP.198108122011011007

Pendamping I,

Nur Cahyo Hendro Wibowo, S.T,M.Kom
NIP. 19731222200641001

Pendamping II,

Siti Nur'aini, S.Kom., M.Kom
NIP.198401312018012001

NOTA PEMBIMBING I

Semarang, 19 Februari 2025

Yth. Ketua Program Studi Teknologi Informasi
Fakultas Sains dan Teknologi
UIN Walisongo Semarang

Assalamu'alaikum wr, wb
Dengan ini diberitahukan bahwa saya telah melakukan bimbingan, arahan dan koreksi naskah skripsi dengan :

Judul	:Perbandingan Pembobotan TF-IDF dan *Word2Vec* pada Analisis Sentimen MAsyarakat Terhadap Ibu Kota Nusantara (IKN) di Media Sosial X Menggunakan Metode *Naïve Bayes*
Nama	: **Nisfah Laili Fikria**
NIM	: 2008096007
Jurusan	: Teknologi Informasi

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada Fakultas Sains dan Teknologi UIN Walisongo, untuk diujika dalam sidang munaqosah

Wassalamu'alaikum wr,wb.

Semarang, Februari 2025
Pembimbing I,

Nur Cahyo Hedro Wibowo, S.T., M.Kom
NIP.197312222006041001

NOTA PEMBIMBING II

Semarang, 19 Februari 2025

Yth. Ketua Program Studi Teknologi Informasi
Fakultas Sains dan Teknologi
UIN Walisongo Semarang

Assalamu'alaikum wr, wb

Dengan ini diberitahukan bahwa saya telah melakukan bimbingan, arahan dan koreksi naskah skripsi dengan :

Judul	:Perbandingan Pembobotan TF-IDF dan *Word2Vec* pada Analisis Sentimen MAsyarakat Terhadap Ibu Kota Nusantara (IKN) di Media Sosial X Menggunakan Metode *Naïve Bayes*
Nama	: **Nisfah Laili Fikria**
NIM	: 2008096007
Jurusan	: Teknologi Informasi

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada Fakultas Sains dan Teknologi UIN Walisongo, untuk diujika dalam sidang munaqosah

Wassalamu'alaikum wr,wb.

Semarang, Februari 2025
Pembimbing II,

Siti Nur`aini, S.Kom., M.Kom
NIP.198401312018012001

LEMBAR PERSEMBAHAN

Dengan mengucapkan syukur Alhamdulillah, laporan tugas akhir skripsi ini dapat penulis selesaikan. Karya kecil ini penulis persembahkan untuk:

1. Bapak Makiran dan Ibu Suyati selaku orang tua penulis.
2. Sumaryati, Siti Rohmawatun, Dwik Yulianti selaku kakak penulis.
3. Sherlien Anggie Puspitasari, Nurhidayati Putri dan Usawanti selaku teman dekat penulis.
4. Seluruh dosen Program Studi Teknologi Informasi.
5. Teman-teman Pramuka Walisongo
6. Teman-teman seperjuangan Khusus Jurusan Teknologi Informasi 2020.
7. Almameter Universitas Islam Negeri Walisongo Semarang.

MOTTO

“proses ini seperti jalan yang penuh terjal dan rintangan. Jika kamu dapat bertahan kamu akan mencapai akhir tapi jika kamu menyerah bahkan mimpimu dan orang-orang yang menaruh harapan padamu juga tidak akan pernah tercapai”

“kamu hebat sudah bertahan dan akan selalu bertahan”

ABSTRAK

Perkembangan media sosial saat ini sangat berkembang pesat salah satunya yaitu media sosial X yang sebelumnya dikenal dengan istilah *twitter*. Dengan fitur yang sudah diperbarui sehingga segala informasi sangat mudah didapatkan. Dari informasi yang mudah di dapat memungkinkan terjadinya diskusi publik yang mendasari dari suatu topik salah satunya yaitu topik mengenai Ibu Kota Nusantara, diskusi tersebut kemudian dapat dianalisis dengan menggunakan model klasifikasi *Naïve Bayes*. Untuk memasukkan teks ke model dibutuhkan suatu ektraksi fitur, dengan TF-IDF dan *Word2Vec* yang menjadi dua metode ektraksi umum yang sering digunakan untuk analisis sentimen. TF-IDF terfokus pada pembobotan kata dan frekuensi kata sedangkan *Word2Vec* berfungsi menangkap hubungan makna antar kata. Tujuan dari penelitian ini yaitu untuk membandingkan kinerja pembobotan TF-IDF dan juga *Word2Vec* dalam analisis sentimen menggunakan model klasifikasi *Naïve bayes*. Hasil penelitian menunjukkan bahwa TF-IDF memiliki kinerja lebih baik. Dengan *akurasi* rata-rata sebesar 64%, *presisi* 64%, *recall* 65% dan *F1_Score* 63%. Sedangkan pembobotan *Word2Vec* menunjukkan *akurasi* rata-rata sebesar 43% dengan *presisi* 38%, *recall* 42% dan *F1_Score* 30%. Selain itu, dari data set yang didapatkan terlihat bahwa mayoritas tanggapan masyarakat cenderung positif, dengan 1589 sentimen positif, 1321 sentimen netral dan 1215 sentimen negatif dari data keseluruhan sebesar 4125 data.

Kata Kunci: Analisis sentimen, *Naïve bayes*, TF-IDF, *Word2Vec*

KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena berkat rahmat, karunia, dan kasih-Nya, saya dapat menyelesaikan skripsi ini dengan baik. Skripsi yang berjudul **"Perbandingan Pembobotan Tf-Idf Dan *Word2Vec* Pada Analisis Sentimen Masyarakat Terhadap Ibu Kota Nusantara (IKN) Di Media Sosial X Menggunakan Metode *Naïve Bayes* "** ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana di UIN Walisongo Semarang pada Program Studi Teknologi Informasi.

Penyusunan skripsi ini tidak lepas dari bantuan, dukungan, dan bimbingan dari berbagai pihak. Oleh karena itu, pada kesempatan ini, saya ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. H. Nizar, M.Ag, selaku Rektor Universitas Islam Negeri Walisongo Semarang.
2. Bapak Prof. Dr. H. Musahadi, M.Ag, selaku Dekan Fakultas Teknologi Informasi Universitas Islam Negeri Walisongo Semarang.
3. Bapak Dr. Khotibul Umam, S.T., M.Kom, selaku Ketua Program Studi Teknologi Informasi Universitas Islam Negeri Walisongo Semarang.
4. Bapak Nur Cahyo Hendro Wibowo,S.T, M.Kom, selaku pembimbing utama, yang telah memberikan arahan, bimbingan, dan motivasi yang sangat berharga selama proses penyusunan skripsi ini.

5. Ibu Siti Nur'Aini, S.Kom, M.Kom, yang telah memberikan dukungan dan bimbingan dalam penyelesaian skripsi ini.
6. Kepada kedua orang tua saya, Bapak Makiran dan Ibu Suyati, yang senantiasa memberikan doa, kasih sayang, dan dukungan tanpa batas.
7. Organisasi Pramuka Walisongo yang telah memberikan ruang mengembangkan bakat.
8. Teman-teman, rekan sejawat, dan semua pihak yang telah memberikan bantuan baik secara langsung maupun tidak langsung.

Saya menyadari bahwa skripsi ini masih jauh dari kesempurnaan, baik dalam hal substansi maupun penyajiannya. Oleh karena itu, saran dan kritik yang membangun sangat saya harapkan untuk perbaikan di masa mendatang.

Semoga skripsi ini dapat memberikan manfaat, baik bagi perkembangan ilmu pengetahuan maupun bagi pihak yang berkepentingan. Terima kasih.

Semarang, 10 Maret 2025

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I
PENDAHULUAN

A. Latar Belakang

Ibu Kota Nusantara (IKN) adalah sebuah proyek strategis Indonesia yang bertujuan untuk memindahkan Ibu Kota Nusantara dari Jakarta ke Kalimantan.

IKN dirancang untuk menjadi pusat gravitasi ekonomi baru di Indonesia, termasuk di Kawasan Timur Indonesia, dengan harapan dapat menciptakan pusat-pusat pertumbuhan ekonomi baru dan memaksimalkan potensi sumber daya daerah. Tujuan utama IKN adalah untuk mengatasi tantangan pembangunan yang terkonsentrasi di Jakarta dan pulau Jawa, serta meningkatkan keadilan distribusi sumber daya dan kesejahteraan di seluruh wilayah Indonesia.

Namun, implementasi IKN juga menghadapi beberapa tantangan, dari mulai opini masyarakat pro dan kontra seperti konflik agraria yang melibatkan masyarakat lokal dan perusahaan swasta.

Seiring perkembangan zaman yang terjadi teknologi juga semakin berkembang pesat. Internet telah melahirkan media sosial yang memungkinkan orang berbagi ide, mengekspresikan diri, dan berinteraksi dengan orang lain

secara online. Salah satu media sosial yang cukup terkenal sekarang yaitu *Twitter* atau sekarang diganti nama dengan X. Dengan media ini seseorang dapat dengan bebas mengemukakan ide, pendapat, opini dalam bentuk text ataupun kalimat. Contoh-contoh penggunaan *Twitter* untuk berbagi pengalaman dan pendapat meliputi *tweet* yang berisi cerita pribadi, pendapat tentang berita, dan diskusi dengan orang lain(Mutiara, Antonius, and Leviane 2020).

Indonesia adalah negara dengan pengguna X terbesar di Asia. Berdasarkan data yang diterbitkan oleh Kementerian Komunikasi dan Informatika (Kominfo), Indonesia memiliki jumlah pengguna X sebesar 24 juta orang, yang menempati peringkat kelima di dunia. Jadi tentu saja banyak sekali pengguna aplikasi X yang mengemukakan pendapatnya tentang IKN (Ibu Kota Nusantara yang saat ini marak diperbincangkan. Tanggapan yang diberikan mungkin saja bersifat positif, negatif, ataupun netral. Oleh karena itu kumpulan tanggapan, opini, serta keluh kesah dari respon pengguna aplikasi X ini dapat ditampung dan dapat dimanfaatkan untuk penlitian ini.

عَلَيْكُمْ بِالصِّدْقِ ، فَإِنَّ الصِّدْقَ يَهْدِيْ إِلَى الْبِرِّ ، وَإِنَّ الْبِرَّ يَهْدِيْ إِلَى الْجَنَّةِ

Artinya: '*Hendaklah kalian selalu berlaku jujur, karena*

kejujuran membawa kepada kebaikan, dan kebaikan mengantarkan seseorang ke Surga.'

Dalam hadits ini Nabi Muhamad SAW memerintahkan umatnya untuk berlaku jujur dalam perkataan, perbuatan, ibadah dan segala perkara. Hal ini tidak dipungkiri bahwa walaupun media sosial merupakan tempat untuk segala sumber informasi namun juga sebagai tempat yang banyak menebarkan kebohongan oleh karena itu salah satu tujuan lain diadakannya penelitian ini untuk mengumpulkan berbagai macam opini masyarakat lalu menganalisisnya sehingga mendapatkan hasil yang berguna bagi semua orang.

Banyaknya opini/pendapat yang ditulis pada aplikasi X dapat diklasifikasikan sesuai sentimen yang ada agar mudah untuk mendapatkan kecenderungan tanggapan tentang tanggapan masyarakat mengenai Ibu Kota Nusantara apakah lebih cenderung kearah positif, negatif atau netral. Data X yang didapatkan memiliki karakteristik yang tidak terstruktur dan memuat banyak sekali *noise*. Sehingga diperlukan text mining yang memiliki peranan penting dalam mengolah dan menganalisis data tersebut.

Untuk melakukan analisis sentimen ada beberapa macam metode salah satunya yaitu metode algoritma *Naïve bayes. Naive Bayes* Kelas*ifier* atau NBC merupakan

proses pengklasifikasian probabilitas sederhana yangmengacu pada *Teory Bayes* (Mustofa and Mahfudh 2019). Algoritma ini bekerja berdasarkan prinsip probabilitas bersyarat, *Teorema Bayes* adalah metode untuk menemukan probabilitas ketika kita mengetahui probabilitas tertentu lainnya. *Naïve bayes* Kelas*ifier* telah menunjukkan kinerja yang baik dalam berbagai aplikasi, seperti klasifikasi teks, deteksi spam, dan prediksi cuaca.

Selanjutnya pembobotan dalam konteks pemrosesan teks dan text mining merujuk pada proses memberikan nilai atau bobot pada kata-kata (terms) dalam dokumen untuk menentukan seberapa penting kata-kata tersebut dalam konteks tertentu. Pembobotan ini sangat penting dalam analisis teks karena membantu dalam mengekstraksi informasi yang relevan dan meningkatkan *akurasi* dalam klasifikasi dan pengambilan keputusan. Dalam penelitian ini akan ada dua pembobotan yang akan digunakan dan sebagai bahan perbandiangan yaitu pembobotan TF_IDF dan *Word2Vec*.

Sehingga pada penelitian ini, peneliti melakukan penenlitian mengenai perbandingan pembobotan yang mana akan di simpulkan pembobotan yang lebih efisien untuk klasidikasi sentimen pendapat masyarakat mengenai Ibu Kota Nusantara menggunakan algoritma

Naïve bayes dengan data yang diperoleh dari aplikasi X.

B. Identifikasi Masalah

Dari uraian latar belakang diatas, maka dapat diidentifikasi beberapa masalah, diantaranya:

1. Ibu Kota Nusantara (IKN) merupakan proyek pemerintah yang besar dan sudah berlangsung cukup lama. Seimbang dengan banyaknya pro dan kontra masyarakat yang mengiringinya.
2. Pembobotan yang paling sering digunakan yaitu pembobotan TF_IDF namun dalam penelitian ini akan disandingan dengan pembobotan lain yaitu pembobotan *Word2Vec* yang kemudian akan dinilai efektifitasnya.

C. Rumusan Masalah

Berdasarkan latar belakang diatas, Rumusan masalah yang akan dibahas dalam penelitian ini sebagai berikut :

1. Bagaimana mengimplementasikan algoritma *naive bayes* dalam membantu analisis sentimen masyarakat terhadap Ibu Kota Nusantara (IKN) pada respon pengguna aplikasi X?
2. Bagaimana perbandingan performa pembobotan TF-IDF dan *Word2Vec* pada pada analisis sentimen masyarakat terhadap Ibu Kota Nusantara (IKN) dari respon pengguna aplikasi X?

D. Batasan Masalah

Agar penelitian dapat dilakukan secara objektif dan jelas,maka peneliti menerapkan batasan masalah yang diperlukan

pada penelitian ini. Berikut batasan masalah pada penelitian ini, adalah :

1. Data yang digunakan adalah data dari media sosial X yang berbahasa Indonesia
2. Sentimen analisis dilakukan dengan klasifikasi dari metode *Naïve bayes*.
3. Tanggapan pada aplikasi X akan diklasifikasikan menjadi tiga sentimen yaitu sentimen positif, negatif dan netral.
4. Kata kunci pencarian yang diteliti pada data aplikasi X yaitu "IKN".

E. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk:

1. Menganalisis sentimen positif, negatif, dan netral dengan menggunakan algoritma *Naïve bayes.*
2. Mengetahui performa pembobotan TF-IDF dan *Word2Vec* dalam mengklasifikasi sentimen terhadap study kasus sentimen masyarakat mengenai IKN.

F. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

a. Manfaat teoritis
    1. Membantu untuk mengkasifikasi *tweet*/tanggapan positif, negatif dan netral.
    2. Mengetahui performa algoritma *Naïve bayes* dalam mengklasifikasi sentimen dari media sosial.
    3. Sebagai bahan perbandingan dan referensi pada penelitian-penelitian selanjutnya yang berhubungan dengan perbandingan pembobotan.

b. Manfaat praktis
    1. Bagi masyarakat, dapat mengetahui jumlah respon/tanggapan masyarakat Mengenai Ibu Kota Nusantara (IKN).
    2. Bagi pemerintah, dapat mengetahui pemikiran masyarakat saat ini mengenai pembangunan Ibu Kota Nusantara (IKN).
    3. Bagi mahasiswa, dapat mengetahui keunggulan dari masing-masing pembobotan.

G. Sistematika Penelitian

Laporan penelitian ini secara keseluruhan terdiri dari beberapa bab, agar memudahkan pembaca untuk memahami isi penelitian ini maka peneliti menunjukkan sistematika penelitian. Berikut sistematika penelitian pada penelitian ini yaitu:

BAB I PENDAHULUAN

Dalam bab pendahuluan ini membahas tentang latar belakang permasalahan kemudian membahas identifikasi masalah, rumusan masalah,batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penelitian penyusunan laporan.

BAB II LANDASAN PUSTAKA

Dalam bab landasan pustaka ini membahas kajian pustaka untuk mendukung dilakukannya penelitian ini serta membahas teori-teori dasar yang berkaitan dengan penelitian.

BAB III METODOLOGI PENELITIAN

Dalam bab metodologi penelitian membahas tentang cara peneliti memperoleh data dan kemudian membahas perangkat yang digunakan pada penelitian, kemudian membahas alur pengerjaan penelitian serta gambaran umum yang terdapat pada uraian metodologi.

BAB IV HASIL DAN PEMBAHASAN

Dalam bab hasil dan pembahasan ini membahas hasil penelitian yang dapat menjawab pertanyaan dari analisis permasalahan

pada bab III metodologi penelitian.

BAB V KESIMPULAN DAN SARAN

Dalam bab kesimpulan dan saran, membahas poin kesimpulan dan saran yang didapat dari hasil bab IV yang diuraikan secara singkat dan jelas. Kemudian yang kedua yaitu poin saran yang berisikan saran-saran dari peneliti untuk penelitian selanjutnya.

BAB II

LANDASAN PUSTAKA

A. Text Mining

Text Mining adalah suatu teknik yang digunakan untuk menemukan informasi berharga dari sumber informasi terstruktur dan tidak terstruktur. Tujuan utama dari text mining adalah untuk menemukan pola yang menarik dari sekumpulan data tekstual dan menggunakannya untuk memecahkan berbagai masalah, seperti klasifikasi, *clustering*, ekstraksi informasi, dan penarikan informasi (Firdaus and Firdaus 2021).

Text mining merupakan proses untuk mengekstraksi, memahami, dan mengolah data berupa teks yang tidak terstruktur secara otomatis untuk mendapatkan informasi yang terkandung dalam sebuah kalimat atau opini. Dalam konteks *Twitter*, text mining digunakan untuk menganalisis *tweet* yang dikirim oleh pengguna, memahami makna dari *tweet*, dan mengklasifikasikan *tweet* menjadi bermakna positif, negatif, atau netral.

B. Analisis Sentimen

Analisis sentimen adalah proses menggunakan teknik pemrosesan bahasa alami (NLP), analisis teks, dan linguistik komputasional untuk mengidentifikasi dan mengekstrak informasi subjektif dari teks (Nugraha et al.

2023). Tujuan utama dari analisis sentimen adalah untuk menentukan nada emosional dibalik serangkaian kata, baik untuk memahami sikap, pendapat, atau emosi pembicara atau peneliti mengenai topik tertentu yang kemudian memberikan hasil dalam bentuk teks negatif, positif dan netral (Nugraha et al. 2023).

Berikut merupakan jenis – jenis dari parameter analisis sentimen:

1. Polaritas: inti dari jenis parameter ini yaitu komentar atau tanggapan yang berada di media sosial mengenai topik yang diambil yaitu IKN (positif, netral, dan negatif).
2. Emosi: emosi yang muncul dari tanggapan masyarakat mengenai topik yang dibahas (senang, sedih, kecewa, bersemangat, dan banyak lagi).
3. Urgensi: terfokus pada permasalahan dan solusi permasalahan (mendesak dan dapat menunggu).
4. Niat: fokus untuk mencari tau apakah masyarakat tertarik dengan pembahasan mengenai IKN.

C. Media Sosial

Media sosial dapat dipahami sebagai suatu platform digital yang menyediakan fasilitas untuk melakukan aktivitas sosial bagi setiap penggunanya. Berbagai aktivitas yang dapat dilakukan di media sosial, seperti melakukan

komunikasi atau interaksi hingga memberikan informasi atau konten berupa tulisan, foto, dan video. Media sosial sendiri pada dasarnya adalah bagian dari pengembangan internet. Kehadiran beberapa dekade lalu telah membuat media sosial dapat berkembang dan bertumbuh secara luas dan cepat seperti sekarang.

Media sosial termasuk suatu platform digital yang mengizinkan pemakainya untuk berinteraksi, bertukar informasi, serta membuat konten yang dapat diakses dan dibagikan oleh pengguna lainnya melalui internet. Pengguna media sosial dapat berhubungan dengan orang-orang di seluruh dunia dan membangun jaringan sosial online. Banyak media sosial yang terkenal misalnya Facebook, Instagram, Twitte*r*, LinkedIn, TikTok, dan YouTube. Pengguna Instagram dapat mengkritik foto atau video yang diposting orang-orang di bagian komentar. Komentar yang dibuat dalam bentuk kalimat digunakan sebagai input dan output digunakan dalam bentuk kalimat identifikasi yang mengandung cyberbullying dan non-cyberbullying (Nugraha et al. 2023).

Komentar – komentar inilah yang kemudian akan dijadikan sebagai bahan penelitian kali ini.

D. Aplikasi X

Aplikasi X merupakan aplikasi pengganti *Twitter* dengan pengguna terbanyak di dunia dan menempati 5 besar di asia, oleh karena itu segala informasi entah itu real informasi atau hoax informasi dapat ditemui di aplikasi X. pengguna juga dapat memberikan komentar atau opini secara bebas (Wandani 2021).

Aplikasi X dipilih karena cocok untuk dilakukan teks mining selain itu data dari aplikasi X mudah untuk dianalisis, kebijakan X mengenai data relative liberal dari pada data dari media sosial lainnya. Tidak mudah mengumpulkan data secara terbuka dan otomatis dengan media lain semacam itu. Opini dari sebuah *tweet* dapat ditemukan di dalam bagian teks dari *tweet* tersebut. Penarikan data dalam bentuk teks bebas yang tidak terstruktur dan tidak berstandar (Wandani, 2021). Gambar logo ditunjukkan pada Gambar 2.1.

Gambar 2.1 Aplikasi X

E. *Crawling Data*

Crawling data adalah proses pengumpulan data yang dilakukan oleh sebuah program komputer untuk mengumpulkan data dari berbagai sumber. Proses ini menggunakan *software* atau aplikasi khusus yang disebut "*crawler*" untuk mengakses sumber data dan mengambil informasi yang dibutuhkan. Data yang dikumpulkan melalui *crawling* kemudian dapat diproses dan digunakan untuk berbagai tujuan, seperti analisis data, penelitian, atau pengembangan sistem informasi. *Crawling* data memiliki berbagai tujuan dan fungsi, serta perbedaan dengan data *scraping*. Kinerja dari *crawler* Antara lain untuk mengumpulkan data *tweet*, data produk, dan data statistic (Saputra 2017).

Cara melakukan *crawling* data yaitu dengan membuat program dengan memasukkan kata kunci untuk mencari *tweet* yang sesuai dengan topik yang diambil. Misalnya, "#IKN" program akan mengambil *tweet* yang mention ke hastag/tagar pada upload an mengenai IKN, Kumpulan *tweet* tersebut yang kemudian akan digunakan.

F. *Python*

Pemrograman *Python* adalah suatu bentuk pemrograman komputer yang menggunakan bahasa pemrograman *Python*. *Python* merupakan bahasa

pemrograman tingkat tinggi yang terkenal karena sintaksisnya yang mudah dibaca dan mudah dipahami. Ini digunakan secara luas dalam berbagai bidang seperti pengembangan perangkat lunak, analisis data, kecerdasan buatan, pengembangan *web*, dan banyak lagi.

Python telah menjadi pilihan bahasa pemrograman yang terkenal dan digunakan diseluruh dunia, baik oleh pemula maupun oleh perusahaan teknologi besar. Itu juga menjadi bahasa yang populer dalam pengembangan solusi kecerdasan buatan dan pengolahan data, karena memiliki banyak pustaka dan alat yang mendukung analisis data yang kuat (Mahrozi and Faisal 2023). Berikut gambar logo ditunjukkan pada gambar 2.2.

Gambar 2.2 Python

G. *Text Prepocessing*

Text *prepocessing* adalah suatu proses untuk menyeleksi data text agar menjadi lebih terstruktur lagi

dengan melalui serangkaian tahapan yang meliputi tahapan *Remove duplicate, Case folding, Cleansing, Normalization, Stopword removal, stemming,* dan *Tokenizing*. Tahapan ini dilakukan untuk memudahkan analisis data dan meningkatkan kualitas data. Berikut adalah beberapa contoh tahapan text *prepocessing*:

1. *Cleansing*

   Cleansing mempunyai fungsi menghilangkan informasi yang tidak berhubungan dengan dokumen. Sebagai contoh yaitu code script, link, HTML dan lain sebagainya (Mustofa and Mahfudh 2019).

2. *Case folding*

   Case folding adalah untuk mengubah huruf kapital menjadi huruf kecil agar data menjadi lebih terstruktur. Hal ini dilakukan untuk memastikan bahwa teks tersebut konsisten dan untuk memfasilitasi analisis teks yang tidak peka terhadap huruf besar dan kecil, seperti pencocokan kata kunci atau pencarian teks.

3. *Tokenizing*

   Proses *tokenisasi* adalah suatu proses yang digunakan untuk membagi teks menjadi unit-unit yang lebih kecil dan lebih mudah dipahami, disebut *token*. *Token* ini dapat berupa kata, frasa, atau bahkan karakter. Proses

tokenisasi dilakukan untuk memudahkan analisis dan pemrosesan teks (Mustofa and Mahfudh 2019)Dengan *tokenizing*, kita dapat membedakan mana antara pemisah kata atau bukan. Jika menggunakan bahasa pemrograman *Python*, biasanya *tokenizing* juga mencakup proses removing number, removing punctuation, serta removing whitespace.

4. *Normalization*

   Normalization digunakan untuk mengganti kata yang tidak baku menjadi baku sesuai anjuran KBBI.

5. *Stopword removal*

   Stopword removal adalah tahapan yang digunakan untuk mengambil kata-kata yang penting dari hasil *token* dengan artian *Stopword removal* merupakan langkah untuk membersihkan teks dari elemen-elemen yang tidak diinginkan atau tidak relevan, seperti tanda baca, angka, dan kata-kata umum (*stopwords*. Kata umum yang biasanya muncul dan tidak memiliki makna disebut dengan stopword. Penghilangan *stopword* ini dapat mengurangi ukuran index dan waktu pemrosesan. Selain itu, juga dapat mengurangi level *nois*e.

6. *Stemming*

   Stemming merupakan suatu teknik untuk

mentransformasi kata- kata dalam sebuah dokumen teks menjadi bentuk kata dasar. Proses *stemming* berbeda-beda dalam tiap bahasa. Setiap bahasa memiliki aturan-aturan yang berbeda dalam penggunaan kata berimbuhan dan mempunyai aturan-aturan sendiri. Pada Bahasa Indonesia kopleksitas ada pada variasi imbuhan. Hal tersebut menjadi penting dalam pembentukan kata dasarnya. Contoh: "berjalan", "berjalanmu", "menjalankan" menjadi bentuk dasar "jalan" (Mustofa and Mahfudh 2019).

H. TF-IDF

TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah sebuah metode yang digunakan untuk memilih fitur sebagai hasil ringkasan, dengan penerapannya pada seleksi fitur bobot kata. Metode TF-IDF digunakan untuk memilih fitur yang paling relevan dan penting dalam suatu dokumen, dengan cara menghitung frekuensi kemunculan kata-kata dalam dokumen dan menghitung frekuensi kemunculan kata-kata dalam korpus dokumen.

TF-IDF dihitung dengan mengalikan nilai frekuensi istilah (TF) dengan nilai inversi frekuensi dokumen (IDF). Nilai TF mengukur seberapa sering suatu kata muncul

dalam sebuah dokumen, sedangkan nilai IDF mengukur seberapa umum kata tersebut dalam korpus keseluruhan. Ada beberapa skema pembobotan TF-IDF yang dikembangkan, seperti TF-PDF (frekuensi istilah–frekuensi dokumen seimbang) dan TF-IDF (frekuensi istilah–inversi frekuensi dokumen pengguna). Variasi ini digunakan untuk menyesuaikan TF-IDF dengan kebutuhan spesifik aplikasi.

IDF adalah ukuran yang digunakan untuk mengetahui seberapa umum sebuah kata dalam korpus dokumen. Nilai IDF dihitung dengan membagi total jumlah dokumen dalam korpus dengan jumlah dokumen yang mengandung kata tersebut, dan mengambil logaritma dari hasilnya. Semakin sedikit jumlah dokumen yang mengandung kata tersebut, semakin tinggi nilai IDF-nya.

TF adalah ukuran yang digunakan untuk mengetahui seberapa sering sebuah kata muncul dalam suatu dokumen. Nilai TF dihitung dengan membagi jumlah kemunculan kata tersebut dalam dokumen dengan total jumlah kata dalam dokumen tersebut. Berikut persamaan TFIDF, yang di tunjukkan pada persamaan 2.1.

$$\text{TFIDF}(d, t) = \text{TF}(d, t) \ . \ \text{IDF}(t)$$

Dimana d adalah dokumen dan t adalah kata. Sedangkan TF(d,t) merupakan jumlah kata pada tiap dokumen yang dirumuskan pada persamaan 2.2.

$$\text{TF}(d,t) = \frac{jumlah\ kata\ t\ pada\ dokumen\ d}{total\ kata\ pada\ dokumen\ d}$$

Pada IDF merupakan jumlah dokumen yang mengandung kata tersebut yang dirumuskan pada persamaan 2.3.

$$\text{IDF}(t) = \log \frac{total\ dokumen}{jumlah\ dokumen\ yang\ mengandung\ kata\ t}$$

I. *Word2Vec*

Word2Vec merupakan sekumpulan beberapa model yang saling berkaitan yang digunakan untuk menghasilkan *word embeddings*. *word embeddings* merupakan sebutan dari seperangkat bahasa pemodelan dan teknik pembelajaran fitur atau *feature learning* pada *Natural Language Processing* (NLP) dimana setiap kata dari kosakata (*Vocabulary*) memiliki vektor yang mewakili makna dari kata tersebut dan kata-kata tersebut dipetakan ke dalam bentuk vector bilangan rill(Kencana and Maharani 2017).

Metode *Word2Vec* terdiri dari dua algoritma utama yaitu *vontinuous bag of word* (CBOW) dan skip-gram.

Algoritma CBOW digunakan untuk melihat panjang tertentu dari sebuah kata pada dokumen masukan. Sedangkan algoritma skip-gram digunakan untuk memprediksi konteks kata dengan cara melihat kedekatan sebuah kata dengan kata lain yang posisinya sebelum atau sesudah kata tersebut. Secara arsitektur *Word2Vec* sebenarnya hanya sebuah jaringan syaraf tiruan yang tidak banyak memiliki hidden layer, baik secara node dalam setiap layer maupun banyaknya layer (Prabowo, Marselino, and Suryawiguna 2019).

Model skip-gram dalam analisis sentimen menggunakan rumus untuk memprediksi kata konteks berdasarkan kata target.berikut langkah-langkah dan rumus yang digunakan dalam skip-gram:

Model skip-gram berfungsi untuk mempelajari distribusi probabilitas dari kata konteks berdasarkan kata saat ini (current word) atau *w*(*t*). rumus skip-gram sendiri yaitu:

1. Rumus Dasar Skip-Gram

$$P(w_{t-c}, w_{t+1}, \ldots, w_{t+c} \mid w_t) = \prod_{j=1}^{c} P(wt + j \mid wt)$$

Dimana:

- $P(wt + j \mid wt)$ adalah probabilitas kemunculan kata

konteks $wt + j$ diberikan kata target wt.

- c adalah ukuran jendela (window size), yang menentukan seberapa banyak kata konteks yang akan dipertimbangkan di sekitar kata target.

2. Representasi Vektor

   Setiap kata akan dipresentasikan dalam bentuk vector menggunakan one-hot encoding, dimana setiap kata dalam kosakata memiliki vector unik. Misal kita memiliki kosakata V dengan ukuran m, maka representasi one-hot untuk w adalah

$$v_w = [0,0, \dots \dots, 1, \dots \dots, 0]$$

3. Menghitung Probabilitas

   Untuk menghitung probabilitas $p(w_o \mid w_t)$, kita menggunakan softmax function:

$$p(w_o \mid w_t) = \frac{e^{v_{w_o} \cdot v_{w_t}}}{\sum_{w \in V} e^{v_w \cdot v_{w_t}}}$$

   Dimana:

   - v_{w_o} adalah vector keluaran untuk konteks w_o
   - v_{w_t} adalah vector keluaran untuk kata target w_t

J. *Split Validation Data*

Split Validation adalah teknik validasi yang membagi data menjadi dua bagian secara acak, sebagian sebagai data *training* dan sebagian lainnya sebagai data *testing*. Dengan

menggunakan *Split Validation*, akan dilakukan percobaan *training* berdasarkan *split* ratio yang telah ditentukan sebelumnya, untuk kemudian sisa dari split ratio data *training* akan dianggap sebagai data *testing* (Untari 2018).

Data latih (*training*) adalah data yang akan digunakan dalam melakukan proses pembelajaran (*learning*), sedangkan data uji (*testing*) adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian (Turmudi Zy, Adji Ardiansyah, and Maulana 2021).

Model pengklasifikasi data terbuat dari kumpulan data *Training*, kemudian performa pengklasifikasiannya diukur berdasarkan data *Testing*. Perbandingan antara data *Training* dan data *Testing* pada umumnya adalah 80:20 (80% adalah *Training* dan 20% adalah *Testing*), Hasil yang optimal pada pengklasifikasian bergantung pada data *training*, jika data *training* dapat mencakup sebagian besar data yang dibutuhkan dalam pengujian data *testing* maka hasilnya yang didapatkan akan maksimal (Darmawan and Amini 2022).

K. Klasifikasi

Klasifikasi adalah proses pengelompokan suatu hal berdasarkan persamaan dan perbedaannya. Klasifikasi dapat diartikan sebagai penyusunan bersistem dalam

kelompok atau golongan menurut kaidah atau standar yang ditetapkan. Istilah ini berasal dari bahasa Belanda, 'kelas*ificatie*', yang sendiri berasal dari bahasa Prancis, 'kelas*ification*', yang berarti pengelompokkan atau klasifikasi. Cara kerja Klasifikasi adalah untuk mengelompokkan data menjadi kelas-kelas yang telah ditentukan sebelumnya berdasarkan nilai atribut data. Klasifikasi digunakan untuk memprediksi kelas yang paling cocok untuk suatu objek berdasarkan atribut yang dimiliki oleh objek tersebut.

L. *Naïve bayes*

Algoritma *Naïve bayes* adalah salah satu metode yang paling sederhana dan efektif untuk klasifikasi teks, termasuk analisis sentimen. Algoritma *Naïve bayes* menggunakan *Teorema Bayes* untuk menghitung probabilitas kelas berdasarkan fitur-fitur yang terkait. Dalam NBC, asumsi independensi antara fitur-fitur yang ada memungkinkan perhitungan probabilitas yang lebih sederhana. Dengan demikian, NBC dapat digunakan untuk mengklasifikasikan teks berdasarkan sentimen, seperti analisis sentimen pada media sosial (Zaki 2021).

Algoritma *Naïve bayes* didasarkan pada teorema bayes yang dinyatakan sebagai berikut:

$$P(Y|X)\,\frac{P(X|\mathrm{Y}\,.\,P(Y)}{P(X)}$$

Keterangan:

Y = data X dari kelas spesifik

X = data kelas yang belum diketahui

$P(Y|X)$ = probabilitas kelas Y diberikan fitur X.

$P(X|Y)$ = probabilitas fitur $X\ diberikan\ kelas\ Y$.

$P(Y)$ = probabilitas fitur Y.

$P(X)$ = probabilitas fitur X.

M. Ibu Kota Nusantara (IKN)

Ibu Kota Nusantara (IKN) adalah nama resmi untuk ibu kota baru Indonesia yang terletak di Kalimantan Timur, menggantikan Jakarta. Pemindahan ibu kota ini diumumkan oleh Presiden Joko Widodo pada tahun 2019 dan direncanakan untuk selesai pada tahun 2045. Pemindahan ini bertujuan untuk menciptakan keseimbangan pembangunan di seluruh wilayah Indonesia dan mengurangi ketergantungan pada Pulau Jawa(Sitio, Rumapea, and Lumbanraja 2023).

Gambar 2. 1 Ibu Kota Nusantara

N. Kajian Penelitian yang Relevan

Pada bagian ini, akan membahas penelitian – penelitian terkait yang relevan sebelumnya dengan tujuan untuk memperkuat pelaksanaan penelitian ini. Salah satu penelitian terkait yang mengacu pada perbandingan pembobotan analisis sentimen pernah dilakukan oleh A Dani. Penelitian ini membandingkan kinerja TF-IDF dan *Word2Vec* dalam mendeteksi sentimen cyberbullying di media sosial. Hasil penelitian menunjukkan bahwa TF-IDF lebih baik dibandingkan dengan *Word2Vec*. Dari 3000 data yang menjadi data pengujian menghasilkan TF-IDF dengan akurasi 83%, precision 83%, recall 83%, f1_score 83% sedangkan Word2Vec dengan akurasi 77%, precision76%, recall 80%, f1_score 78% Penelitian ini menggunakan sembilan skenario uji untuk menganalisis performa kedua metode(Dani, Puspaningrum, and Mumpuni 2024).

Penelitian kedua ditulis oleh Aganda Maulana yang membandingkan *Word2Vec* CBOW CNN ternyata lebih baik dari pada Sent2vec TF-IDF LR dalam analisis sentiment youtube dengan selisih hasil *accuracy* 4,09%, precisius 6,75%, *recall* 0,06% FI score 2,81% dan ROC 0,2%. Dengan (TF-IDFF) + Logistic Regresion mendapatkan hasil nilai Accuracy = 84,35%, Precission = 87,63%, Recall = 88,54%, F1 Score = 88%, ROC = 95%, sedangkan untuk model Word2vec (CBOW) dengan metode CNN mendapatkan hasil yang lebih baik yaitu nilai Accuracy = 88,44%, Precission = 94,38%, Recall = 88,6%, F1 Score = 90,81% dan nilai ROC = 95,2%. Keunggulan ini disebabkan kemampuan Word2vec menangkap hubungan semantik antar kata dan arsitektur CNN yang efektif mengekstraksi pola lokal dari teks serta Word2vec merepresentasikan kata dalam vektor yang mempertahankan konteks, sehingga cocok untuk komentar YouTube yang mengandung slang atau ekspresi informal hal ini juga disebabkan oleh data set yang besar (14.605) memungkinkan Word2Vec menghasilkan embedding yang stabil pada data kecil (<10.000) TF-IDF lebih unggul (akurasi 84% vs 77% untuk Word2Vec) (Maulana, Dyantono, and Putra 2023).

Penelitian ketiga yaitu oleh Hendrawan rifky yang

membandingkan metode TF – IDF dan *Word2Vec* pada klasifikasi sentimen masyarakat terhadap produk local Indonesia dengan menghasilkan data 25.581 kemudian hasil perbandingan *Word2Vec* + CNN menghasilkan nilai 0,939% dan TF-IDF +XGBoos 0,935% .Word2Vec lebih unggul karena mampu menangkap hubungan sematik antar kata sedangkan TF-IDF hanya mengandalkan frekuensi kata tanpa mempertimbangkan konteks (Hendrawan Rifky, Utami, and Hartanto Dwi 2022).

Berikut tabel kajian penelitian pada 2.1.

Tabel 2.1 Kajian Penelitian

No	Pustaka	Topik	Metode	Objek	Klasifikasi
1.	(Dani et al. 2024)	Studi Performa TF-IDF dan *Word2Vec* Pada Analisis Sentimen Cyberbullying	*TF-IDF, Word2Vec, dan SVM*	Youtube	Positif, Negatif, Netral
2.	(Maulana et al. 2023)	Perbandingan Sent2vec TF-IDF Logistic Regression dan *Word2Vec* CNN pada hasil	*TF-IDF, Word2Vec, CNN,Word Embadding,SVM*	Youtube	Positif, Negatif

		Sentimen Analysis Youtube Comment			
3.	(Hendrawan Rifky et al. 2022)	Analisis Perbandingan Metode Tf-Idf dan *Word2Vec* pada Klasifikasi Teks Sentimen Masyarakat Terhadap Produk Lokal di Indonesia	*Word2Vec, TF-IDF,CNN, dan XGBoos*	marketplace	Positif, Negatif

O. Evaluasi

Evaluasi dilakukan untuk mengukur keberhasilan sistem dengan cara membandingkan dengan hasil pengujian pada sistem dengan standar yang sudah ada.

Confision matrix merupakan *matrix* yang menampilkan prediksi klasifikasi dan klasifikasi yang actual (Alfiani Mahardhika, Saptono, and Anggrainingsih 2016) Tabel ini menggambarkan hasil prediksi model dibandingkan dengan label yang sebenarnya, memberikan informasi

tentang jumlah prediksi yang benar dan salah untuk setiap kelas.

Dikarenakan *output* sentimen dari penelitian ini ada tiga yaitu positif, negatif, dan netral maka dalam penelitian ini menggunakan *multiclass confusion matrix* 3x3, berikut tabel *multiclass confusion matrix* 2.2

Tabel 2.2 Multiclass confusion matrix 3x3

		True Kelas		
		Negatif	*Netral*	*Positif*
Predicted Kelas	*Negatif*	T Neg	F NegNet	F NegPos
	Netral	F NetNeg	T Net	F NetPos
	Positif	F PosNeg	F PosNet	T Pos

Dalam *multiclass* confusion matriks tersebut, terdapat sembilan nilai yang dijadikan acuan dalam perhitungan, yaitu :

a. T Pos (True Positif), Yaitu jumlah data yang diprediksi positif dan faktanya data itu positif (Sesuai).
b. F PosNeg (False Positve Negatif), Yaitu jumlah data yang diprediksi positif dan faktanya data itu Negatif.
c. F PosNet (False Positif Netral), Yaitu jumlah data

yang diprediksi positif dan faktanya data itu Netral.

d. F NegPos(False Negatif Positif), Yaitu jumlah data yang diprediksi negatif dan faktanya data itu positif.
e. T Neg (True Negatif), Yaitu jumlah data yang diprediksi negatif dan faktanya data itu negatif (Sesuai).
f. F NegNet (False Negatif Netral), Yaitu jumlah data yang diprediksi negatif dan faktanya data itu netral.
g. F NetPos (False Netral Positif), Yaitu jumlah data yang diprediksi netral dan faktanya data itu positif.
h. F NetNeg (False Netral Negatif), Yaitu jumlah data yang diprediksi netral dan faktanya data itu negatif.
i. T Net (True Netral), Yaitu jumlah data yang diprediksi netral dan faktanya data itu netral (Sesuai).

1. *Akurasi*:

Akurasi adalah proporsi prediksi yang benar (baik untuk kelas A, B, maupun C) dibandingkan dengan total seluruh data yang diuji. *Akurasi* memberikan gambaran umum seberapa sering model membuat prediksi yang benar di semua kelas. Rumusnya adalah:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} 100\%$$

2. *Presisi*:

Presisi mengukur tingkat ketepatan model dalam memprediksi suatu kelas tertentu. Untuk setiap kelas, *presisi* adalah rasio antara prediksi benar pada kelas tersebut (True Positive/TP) dengan total prediksi model untuk kelas itu. Rumusnya adalah:

$$Presisi = \frac{TP}{TP+FP} 100\%$$

3. *Recall*:

Recall (atau sensitivitas) mengukur seberapa baik model dalam menemukan semua data aktual pada suatu kelas tertentu. Untuk setiap kelas, *recall* adalah rasio antara prediksi benar pada kelas tersebut (TP) dengan total data aktual pada kelas itu (TP + False Negative/FN). Rumusnya adalah:

$$Recall = \frac{TP}{TP+FN} 100\%$$

4. F1-Score:

F1-Score adalah metrik yang menggabungkan precision dan recall menjadi satu nilai tunggal, yaitu rata-rata harmonis antara precision dan recall. *F1-score* sangat berguna ketika terdapat ketidakseimbangan jumlah data antar kelas. Sehingga *F1-score* memberikan keseimbangan antara *precision*

dan *recall*, sehingga cocok digunakan jika keduanya sama pentingnya. Rumusnya adalah:

$$\text{F1 Score} = 2 \text{ x } \frac{Presisi\ x\ Recall}{Presisi\ x\ Recall} 100\%$$

BAB III
METODOLOGI PENELITIAN

A. Metode Pengumpulan Data

1. Studi Pustaka

Peneliti melakukan studi pustaka dengan memanfaatkan buku, jurnal, skripsi terdahulu dan sejenisnya untuk memepelajari konsep, alur serta permasalahan yang berhubungan dengan judul yang diambil. Peneliti juga melakukan penelitian secara daring dengen mengunjungi website-website yang berhubungan dengan analisis sentimen, perbandingan pembobotan, text mining dan algoritma *naïve bayes*.

2. Studi Lapangan

Pada metode ini peneliti melakukan pengamatan secara langsung dengan melihat komentar-komentar dari aplikasi X mengenai isu Ibu Kota Nusantara.

B. Perangkat Penelitian

Dalam penelitian ini diperlukan perangkat yang mendukung untuk keperluan penelitian, perangkat keras dan perangkat lunak yang digunakan peneliti diantaranya yaitu:

1. Perangkat Keras

Perangkat keras (*hardware*) adalah komponen fisik dari komputer yang dapat dilihat dan disentuh. Berikut tabel 3.1 perangkat keras.

Tabel 3.1 Perangkat Keras

No.	Perangkat Keras	Spesifikasi
1.	Device	ASUS X441BA
2.	Processor	AMD A9-9425 RADEON R5
3.	Memori (RAM)	4 GB
4.	Monitor	14 inch
5.	Keyboard dan Mouse	Standard

2. Perangkat Lunak

Perangkat lunak (*software*) adalah sekumpulan instruksi atau program yang dirancang untuk menjalankan tugas tertentu pada perangkat keras komputer. Berikut tabel 3.2 perangkat lunak.

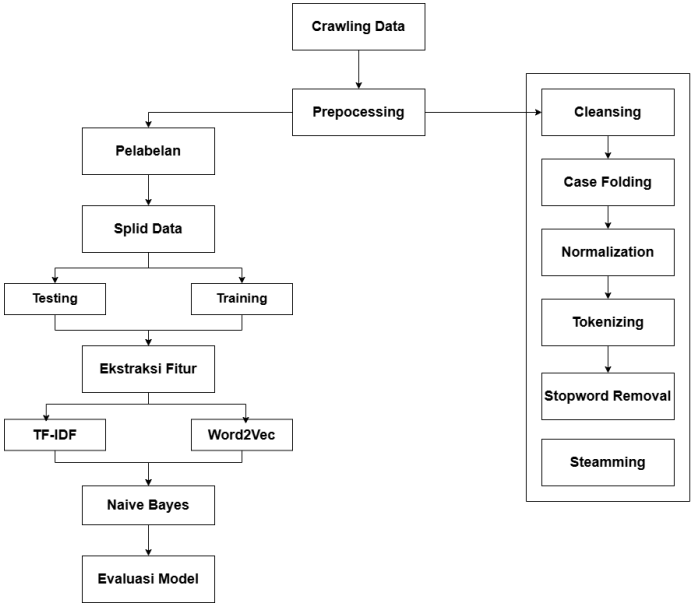
Tabel 3.2 Perangkat Lunak

No.	Perangkat Lunak	Spesifikasi
1.	Sistem Operasi	Wondows 10 64-bit
2.	Bahasa Pemograman	*Python*
3.	Ms.Office	Ms. Word, Ms. Exel

		2013
4.	Google Drive	*Google colab*
5.	Browser	Chrome

C. Alur pengerjaan Penelitian

Dalam sebuah pengerjaan harus ada yang namanya alur sehingga pengerjaan lebih terstruktur. Berikut gambar 3.1 Diagram alur pengerjaan penelitian.



Gambar 3.1 Alur Pengerjaan Penelitian

Langkah pertama dari penelitian ini adalah mengidentifikasi masalah, dengan cara menganalisis masalah yang sekiranya menjadi permasalahan di masyarakat dengan

media pencarian menggunakan aplikasi X sehingga didapatkan permasalahan yaitu mengenai sentimen masyarakat terhadap ibu kota nusantara(IKN).

Langkah kedua yaitu crawling data, dengan mencari dan mengumpulkan sentimen/opini masyarakat terhadap ibu kota nusantara(IKN) di sosial media X menggunakan *google colab* dengan Bahasa pemograman *phyton*, dengan menggunakan keyword "IKN" dapat secara otomatis mengumpulkan berbagai sentimen/opini yang yang berkaitan dengan keyword tersebut. Selanjutnya hasil dari crawling tersebut akan di kumpulkan dalam bentuk csv, setelah semua data didapatkan langkah selanjutnya yaitu proses *prepocessing*, data yang masih mentah dapat dipilah untuk selanjutnya akan dilakukan proses klasifikasi. Dalam proses *prepocessing* ini terdapat enam tahapan yaitu: *Cleansing, Case folding, Normalization, Tokenize, Stopword, dan Stemming*. Tahapan dimulai dari *remove duplicate* dimana pada tahapan ini bertujuan untuk menghapus sentimen ganda dan hanya menyisakan satu sentimen yang serupa. Selanjutnya tahapan *cleansing* dimana berfungsi untuk menghapus simbol, tanda baca dan angka yang tidak memiliki peran penting. Selanjutnya tahapan *case folding*, berfungsi untuk merubah huruf kapital menjadi huruf kecil bertujuan untuk menyamaratakan huruf pada suatu kalimat. Selanjutnya yaitu tahap *Normalization* dimana pada tahap ini bertujuan untuk

merubah kalimat yang tidak baku menjadi kalimat baku yang sesuai dengan KBBI. Selanjutnya yaitu *Tokenize*, berfungsi untuk memecah teks suatu kalimat menjadi beberapa bagian kata dan untuk menghilangkan *whitespace*. Langkah selanjutnya yaitu *stopword* pada tahap ini bertujuan untuk menghapus daftar kata yang tidak memiliki arti penting. Yang terakhir adalah tahap *stemming*, berfungsi untuk merubah kata yang berimbuhan menjadi kata dasar dan menghapus kata yang berulang.

Langkah ketiga yaitu pelabelan data dilakukan oleh satu tenaga ahli yang mampu di bidangnya untuk kemudian akan di bagi menjadi tiga kelas yaitu positif, negatif, dan netral.

Langkah keempat yaitu split data, split validation data merupakan teknik membagi data menjadi dua secara acak. Data dibagi menjadi dua yaitu data latih dan data uji untuk mempermudah perhitungan *akurasi*, data akan di rasiokan menjadi 80:20 dimana 80% dari data latih dan 20% dari data uji.

Langkah kelima yaitu TF-IDF dan *Word2Vec*, berfungsi untuk merubah kata menjadi bilangan angka dan pada tahap ini dilakukannya pembobotan nilai kata untuk mempermudah proses klasifikasi.

Langkah selanjutnya yaitu proses klasifikasi dengan menggunakan metode *naïve bayes* dengan data sentimen yang diperoleh, selanjutnya masuk pada tahapan uji model untuk mengetahui ketepatan klasifikasi dengan membandingkan hasil

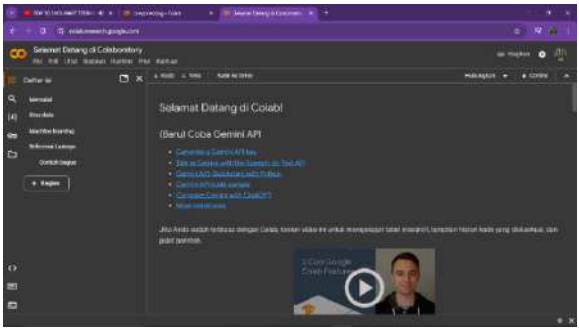
prediksi pada suatu sistem dengan hasil dari data yang sudah dilabeling sebelumnya sehingga menghasilkan *multiclass confusion matrix.* Setelah uji model selanjutnya merupakan tahap evaluasi model untuk menentukan tingkat *akurasi*, *precisson, recall* dan *f1 score,* dengan menggunakan metode *multiclass confusion matrix* untuk mengetahui peforma dari model tersebut.

D. Uraian Metodologi

Pada tahapan ini dilakukan penguraian metode pengerjaan penelitian untuk menjelaskan lebih dalam mengenai penelitian, terdapat beberapa tahapan dalam prosesnya diantaranya yaitu:

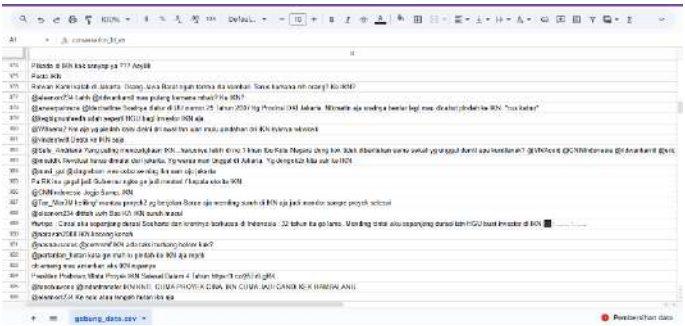
1. Pengambilan Data Aplikasi X

Data diambil dari sentimen/komentar yang ada di aplikasi X dengan keyword Ibu Kota Nusantara(IKN), keseluruhan pengumpulan data melalui proses crawling menggunakan bahasa pemograman *python* yang ditulis pada *google colab*. Berikut gambar *Google colab* ditunjukkan pada gambar 3.2



Gambar 3.2 Google colab

Hasil dari crawling tersebut berupa file csv dari periode tahun 2022-2024 sehingga didapatkan 5000 data mentah yang belum di proses, Dan di dapat 4125 data yang sudah di proses. Hasil crawling ditunjukkan pada gambar 3.3 hasil crawling.



Gambar 3.3 Hasil Crawling

2. *Prepocessing*

a. Cleansing

Cleansing merupakan langkah untuk membersihkan komentar dari hal- hal yang tidak berkaitan dengan penelitian, contohnya yaitu menghilangkan code script, link, HTML, dll.

Contoh penggunaan *Cleansing* pada table 3.3

Tabel 3.3 Penggunaan Cleansing

Input proses	Output proses
IKN menjadi salah satu destinasi wisata favorit Kaltim	IKN menjadi salah satu destinasi wisata favorit Kaltim

https://t.co/8aO81kGlEX	

b. Case folding

Untuk memudahkan penelitian data yang digunakan harus sudah terstruktur dan fungsi dari *Case folding* sendiri yaitu mengubah huruf kapital menjadi huruf kecil sehingga kalimat yang dihasilkan lebih terstruktur. Contoh penggunaan *Case folding* pada table 3.4

Tabel 3.4 Penggunaan Case folding.

Input proses	Output proses
IKN menjadi salah satu destinasi wisata favorit Kaltim	ikn menjadi salah satu destinasi wisata favorit kaltim

c. Normalization

Selanjutnya yaitu proses *Normalization* yang berfungsi untuk merubah kata yang tidak baku menjadi kata baku yang sesuai dengan KBBI. Contoh penggunaan *Normalization* pada table 3.5

Tabel 3.5 Penggunaan Normalization

Input proses	Output proses
ikn menjadi salah satu destinasi wisata favorit kaltim	ikn menjadi salah satu destinasi wisata favorit kaltim

d. Tokenizing

Untuk memudahkan analisis dan proses data *Tokenizing* perperan dalam membagi teks menjadi unit-unit kecil agar mudah dipahami. Contoh penggunaan *Tokenizing* pada table 3.6

Tabel 3.6 Penggunaan Tokenizing

Input proses	Output proses
ikn menjadi salah satu destinasi wisata favorit kaltim	['ikn', 'menjadi', 'salah', 'satu', 'destinasi', 'wisata', 'favorit', 'kaltim']

e. Stopword removal

Stopword merupakan kata yang umum ditemukan namun tidak memiliki makna, menghilangkan *Stopword* dapat mengurangi ukuran index dan waktu pemrosesan. Contoh penggunaan *Sopword Removal* pada table 3.7

Tabel 3.7 penggunaan Sopword Removal

Input proses	Output proses
['ikn', 'menjadi', 'salah', 'satu', 'destinasi', 'wisata', 'favorit', 'kaltim']	['ikn', 'salah', 'destinasi', 'wisata', 'favorit', 'kaltim']

f. Stemming

Fungsi dari *Stemming* yaitu untuk mentransformasikan kata-kata dalam sebuah dokumen menjadi sebuah kata dasar agar mudah untuk dianalisis.

1) Prefiks, imbuhan yang terletak pada awal kata. Prefiks terdiri dari "se-", "ke","me-" dll. Contoh: Me-rasa
2) Suffiks, imbuhan yang terletak pada akhir kata. Contoh dari suffiks adalah "–lah", "-kah", "-pun" dll. Contoh: ajar-kan.
3) Konfiks, imbuhan ini merupakan imbuhan gabungan dari prefiks dan suffiks. Imbuhan terdapat pada awal dan akhir kata. Contoh: me-laku-kan.
4) Infiks, imbuhan yang terletak pada tengah kata. Contoh: k-em-ilau dari kata kilau.
5) Perulangan kata, contoh: anak-anak.

Contoh penggunaan *Stemming* pada table 3.8

Tabel 3.8 Penggunaan Stemming

Input proses	Output proses
['ikn', 'salah', 'destinasi', 'wisata', 'favorit', 'kaltim']	ikn salah destinasi wisata favorit kaltim

3. Pelabelan data

Setelah proses crawling selanjutnya dilakukan pelabelan data, data yang telah disimpan dalam bentuk csv akan dilakukan proses pelabelan. Pelabelan dilakukan dengan bantuan pakar Bahasa yang mampu dibidangnya. Contoh proses pelabelan data ada pada tabel 3.9

Tabel 3.9 Pelabelan Data

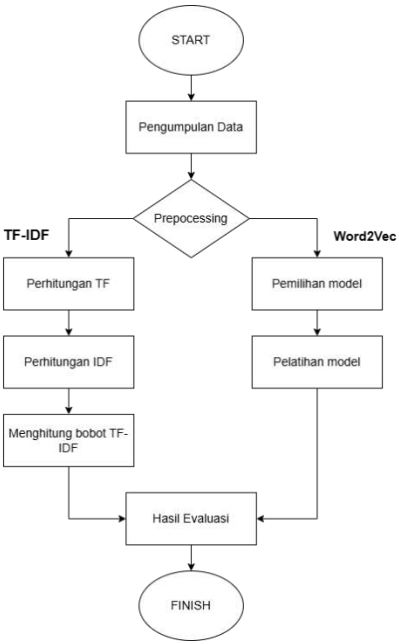
Komentar	Sentimen	kelas
Presiden Prabowo tegas komitmen selesaikan IKN dalam empat tahun https://t.co/Ix3ZJVZAt6	Positif	0
https://t.co/vMTAq061ji Yahya Sinwar guyu ngliat kelakuan kalen. msh byk jembatan blm kebangun buat sekolah. IKN oh IKN. Taek kabeh	Negatif	1
Pembangunan Ibu Kota Nusantara Sudah 58 Persen per Oktober 2024 #IbuKotaNegara #IKNNusantara #IKN https://t.co/ZWcQ6XC7Vm	Netral	2

4. Split Data

Split data, split validation data merupakan teknik membagi data menjadi dua secara acak. Data dibagi menjadi dua yaitu data latih dan data uji untuk mempermudah perhitungan *akurasi*, data akan di rasiokan menjadi 80:20 dimana 80% dari data latih dan 20% dari data uji. Menggunakan 80% dari dataset untuk *Training* memungkinkan model belajar dari sebagian besar data yang tersedia. Hal ini membantu model mengidentifikasi pola dan fitur yang signifikan dalam data (Darmawan and Amini 2022). Pengambilan data dilakukan secara acak dengan bantuan *library* dari *python*.

5. Ekstrasi Fitur

Untuk memahami suatu alur pengerjaan agar mudah untuk dipahami biasanya dibuatlah sebuah flowchart, flowchart adalah representasi diagram yang menggambarkan urutan operasi yang dilakukan untuk mendapatkan solusi masalah(Yuniarti 2019). Berikut gambar 3.4 *Flowcard ekstraksi fitur*.



Gambar 3.4 Flowcart Ektraksi Fitur

Setelah melewati proses *Split data* semua data kemudian akan masuk pada tahap ekstrasi fitur untuk mempermudah tahapan selanjutnya. Pada proses pembuatan fitur membobotan dilakukan menggunakan cara TF-IDF(*Term Frequency-Inverse Document Frequency*) dan *Word2Vec*, merupakan metode yang digunakan untuk memilih fitur sebagai hasil ringkasan, dengan penerapannya pada seleksi fitur bobot kata. Metode TFIDF digunakan untuk memilih fitur yang paling relevan dan penting dalam suatu dokumen, dengan cara menghitung frekuensi

kemunculan kata-kata dalam dokumen dan menghitung frekuensi kemunculan kata-kata dalam korpus dokumen(Septiani and Isabela 2023). Adapun langkah-langkah pembobotan TF-IDF yaitu dimulai dengan menghitung Term Frequency (TF), TF berfungsi untuk mengukur seberapa sering sebuah kata muncul dalam sebuah dokumen, kemudian dilanjutkan dengan menghitung Inverse Document Frequency (IDF) IDF sendiri berfungsi untuk mengukur seberapa penting sebuah kata dalam keseluruhan koleksi dokumen. Yang terakhir yaitu menghitung TF-IDF. Nilai TF-IDF didapat dengan cara mengalikan nilai TF dengan IDF untuk setiap kata dalam dokumen. Sedangkan *Word2Vec*, yang merupakan singkatan dari "word to vector," mengubah kata menjadi vektor numerik dalam ruang berdimensi tinggi. Vektor ini memungkinkan mesin untuk memahami hubungan semantik antara kata-kata, di mana kata-kata dengan konteks serupa akan memiliki representasi vektor yang berdekatan satu sama lain. Pembobotan dalam *Word2Vec* adalah teknik yang digunakan untuk merepresentasikan kata-kata dalam bentuk vektor numerik, yang memungkinkan pemodelan

hubungan semantik dan sintaktik antar kata.

Word2Vec mampu memproses data besar dan cepat, vector yang dihasilkan dapat menangkap hubungan sematik dan sintaksis antar kata sehingga memungkinkan untuk analisis lebih mendalam. Dalam analisis sentimen *Word2Vec* mengidentifikasi opini positif dan negatif dalam teks berdasarkan hubungan antar kata. Ada dua arsitektur utama untuk menghasilkan representasi vector yaitu:

Skip-gram: Model ini memprediksi konteks dari sebuah kata berdasarkan kata target. Ini efektif untuk menangkap hubungan antara kata-kata dalam konteks yang lebih luas.

Continuous Bag of Words (CBOW): Model ini memprediksi kata target berdasarkan kata-kata di sekitarnya, lebih cepat dalam pelatihan dan biasanya lebih baik untuk kata-kata yang sering muncul(Dani et al. 2024).

Untuk mendapatkan hasil vector tidak harus menggunakan dua duanya bisa salah satu sesuai dengan data yang dipilih dan pada data ini akan digunakan pemodelan algoritma skip-gram untuk menghitung perhitungan pembobotan *Word2Vec*.

6. Klasifikasi *Naïve bayes*

Selanjutnya yaitu tahap klasifikasi, dengan menggunakan metode algoritma *Naïve bayes* yaitu salah satu metode yang paling sederhana dan efektif untuk klasifikasi teks, termasuk analisis sentimen. Dalam NBC, asumsi independensi antara fitur-fitur yang ada memungkinkan perhitungan probabilitas yang lebih sederhana. Dengan demikian, NBC dapat digunakan untuk mengklasifikasikan teks berdasarkan sentimen, seperti analisis sentimen pada media sosial (Zaki 2021).

Pengklasifikasian dibagi menjadi tiga kelas yaitu positif, negatif, dan netral. Data yang digunakan adalah data hasil setelah dilakukannya tahapan *Prepocessing*, selanjutnya data yang telah berhasil di*training* akan diuji menggunakan data test untuk menguji hasil ketepatan klasifikasi yang dilakukan. Kelas yang memiliki jumlah terbanyak pada setiap *tweet* akan dianggap menjadi pemenangnya.

7. Evaluasi Model

Evaluasi model dilakukan dengan menguji tingkat kinerja metode melalui *Multiclass confusion matrix* 3x3. *Matrix* konfusi berisi informasi yang

membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang diperlukan(Turmudi Zy et al. 2021). Data uji yang diuji dengan data latih akan menghasilkan daftar kelas dari data tersebut yang disebut prediksi kelas, kemudian prediksi kelas tersebut akan dibandingkan dengan kelas sebenarnya dari data yang disembunyikan. Sehingga dapat dilihat pada peforma model *Naïve bayes* yang berupa tingkat *akurasi*, *pressicion, recall, dan fi score.*

BAB IV
HASIL DAN PEMBAHASAN

A. Pengambilan Data Aplikasi X

Dalam pengambilan data hingga selesainya keseluruhan proses penelitian ini menggunakan bahasa pemograman *python* yang ditulis pada *google colab*, platform berbasis cloud yang memungkinkan pengguna menulis dan menjalankan code *python* secara gratis melalui browser, akses gratis GPU,serta memudahkan kolaborasi. *Google colab* sering digunakan untuk analisis data dan pelatihan model machine learning lainnya(Yanuar 2024). Selain itu *google colab* ini sangat mudah digunakan terutama untuk orang yang baru belajar pemograman.

Berikut gambar *google colab* di tunjukkan pada gambar 4.1

Gambar 4.1 Google colab

Dalam pengambilan data dilakukan menggunakan *library* yang dapat digunakan pada *python* yaitu *library pandas*. *Library pandas* sering digunakan penelitian analisis data untuk mengolah, menganalisis dan menyajikan data dengan efisien(Isa Albanna and R. Tri hadi laksono 2022).

Untuk menggunakan *library pandas* harus dilakukan instalasi terlebih dahulu menggunakan perintah pip. Berikut cara instalasi *library pandas* seperti pada gambar 4.2 dibawah.

```
!pip install pandas
```

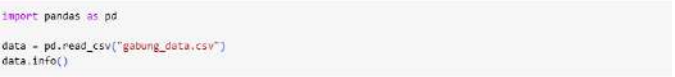
Gambar 4.2 Instal Library Pandas

Kemudian untuk mengumpulkan data digunakan *token* auntentikasi yang diperlukan untuk mengakses API *Twitter* juga sebagai penghubung antara aplikasi X dengan *code*, *token* auntentikasi diambil dari akun aplikasi X yang tentunya setiap akun memiliki kode yang berbeda agar dapat membedakan kepemilikan akun dan agar tidak melanggar kebijakan dari aplikasi X itu sendiri. Kunci utama dalam pengambilan data yaitu keyword atau kata kunci pencarian dalam penelitian ini keyword yang digunakan yaitu #IKN, ibukotanegara, ibukotanusantara, dan nusantara.

Berikut pada gambar 4.3 *source code* penerapan proses crawling data.

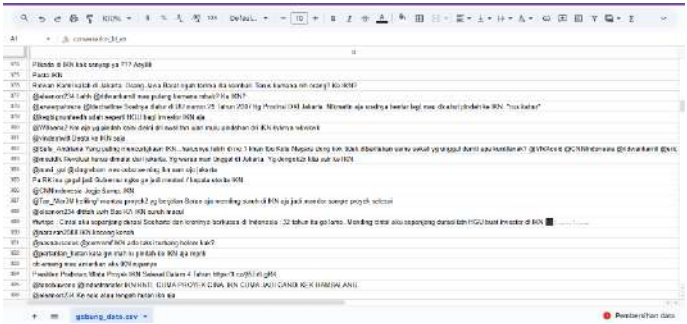
Gambar 4.3 Source Code Crawling

Data yang berhasil diambil akan di simpan dalam bentuk CSV(Comma Separated Values) yang berfungsi untuk menyimpan data dalam bentuk tabel. Format ini dapat dengan mudah di baca oleh manusia maupun mesin karena hanya menggunakan teks biasa. Dikarenakan keyword yang dicari lebih dari satu maka diperlukan file tersendiri untuk menggabungkannya. Berikut pada gambar 4.4 *source code* gabung data.



Gambar 4.4 Source Code Gabung Data

Berikut hasil dari crawling data yang berisi kumpulan keyword yang dicari pada gambar 4.5 hasil crawling dibawah ini.



Gambar 4.5 Hasil Crawling

B. *Prepocessing*

Tahap ini terdiri dari beberapa proses karena data yang diperoleh masih merupakan data mentah dan masih memiliki banyak *nois*. Maka dari itu pada tahapan ini bertujuan untuk merubah data yang mentah menjadi data bersih untuk selanjutnya dapat dilakukan proses pengklasifikasian.

a. Cleansing

Cleansing merupakan proses awal dari *prepocessing* yang bertujuan untuk memperbaiki dan membersihkan hal yang tidak diperlukan seperti tanda baca, *hastag, username* dan lainnya. Agar analisis dapat dilakukan dengan lebih efektif. Berikut beberapa tahapan dalam proses *cleansing:*

1. Menghapus URL (http:// atau https://)
2. Menghapus tanda "@nama" *username*

3. Menghapus HTML
4. Menghapus emoji
5. Menghapus symbol
6. Menghapus angka serta tanda baca

Dengan menerapkan sembilan tahapan *cleansing* ini, data komentar akan menjadi lebih bersih dan terstruktur, sehingga memudahkan analisis selanjutnya. Proses *cleansing* sangat penting untuk memastikan bahwa hasil analisis akurat dan dapat diandalkan, serta memberikan wawasan yang lebih baik dari data teks yang dianalisis. Dalam prosesnya diperlukan beberapa *library python* diantaranya yaitu *library* re, string, dan nltk. Berikut gambar proses import *library* re,string, dan nltk 4.6 dibawah ini.

Gambar 4.6 Import Library re, string,nltk

Masing–masing *library* pempunyai perannya sendiri yaitu: *"re"* berfungsi untuk mengidentifikasi elemen yang perlu di hapus dan mengganti teks, *"string"* berfungsi untuk menghapus tanda baca yang tidak diperlukan dan terakhir ada *library "nltk"* fungsi dari *library* ini yaitu menyediakan daftar stopword

yang dapat digunakan menghapus kata-kata tersebut dari teks membantu menyederhanakan analisis dengan focus pada kata kata yang lebih bermakna. Berikut gambar 4.7 *source code* proses *Cleansing.*

```
# Fungsi untuk menghapus URL
def remove_URL(tweet):
    if tweet is not None and isinstance(tweet, str):
        url = re.compile(r'https?://\S+|www\.\S+')
        return url.sub(r'', tweet)
    else:
        return tweet

# Fungsi untuk menghapus HTML
def remove_html(tweet):
    if tweet is not None and isinstance(tweet, str):
        html = re.compile(r'<.*?>')
        return html.sub(r'', tweet)
    else:
        return tweet

# Fungsi untuk menghapus emoji
def remove_emoji(tweet):
    if tweet is not None and isinstance(tweet, str):
        emoji_pattern = re.compile("["
            u"\U0001F600-\U0001F64F"  # emoticons
            u"\U0001F300-\U0001F5FF"  # symbols & pictographs
            u"\U0001F680-\U0001F6FF"  # transport & map symbols
            u"\U0001F700-\U0001F77F"  # alchemical symbols
            u"\U0001F780-\U0001F7FF"  # Geometric Shapes Extended
            u"\U0001F800-\U0001F8FF"  # Supplemental Arrows-C
            u"\U0001F900-\U0001F9FF"  # Supplemental Symbols and Pictographs
            u"\U0001FA00-\U0001FA6F"  # Chess Symbols
            u"\U0001FA70-\U0001FAFF"  # Symbols and Pictographs Extended-A
            u"\U0001F004-\U0001F0CF"  # Additional emoticons
```

```
            u"\U0001F1E0-\U0001F1FF"  # flags
                                      "]+", flags=re.UNICODE)
        return emoji_pattern.sub(r'', tweet)
    else:
        return tweet

# Fungsi untuk menghapus simbol
def remove_symbols(tweet):
    if tweet is not None and isinstance(tweet, str):
        tweet = re.sub(r'[^a-zA-Z0-9\s]', '', tweet)  # Menghapus semua simbol
    return tweet

# Fungsi untuk menghapus angka
def remove_numbers(tweet):
    if tweet is not None and isinstance(tweet, str):
        tweet = re.sub(r'\d', '', tweet)  # Menghapus semua angka
    return tweet

def remove_username(text):
    import re
    return re.sub(r'@[^\s]+', '', text)

df['cleaning'] = df['full_text'].apply(lambda x: remove_URL(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_username(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_html(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_emoji(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_symbols(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_numbers(x))

df.head(5)
```

Gambar 4.7 Source Code Cleansing

Berikut hasil dari penerapan *cleansing* yang ditunjukkan pada gambar 4.8 dibawah ini.

index	username	full_text	cleaning
0	berkehematbot	Presiden Prabowo berkomitmen untuk membangunan IKN yg tidak sekadar memindahkan pusat pemerintahan tetapi juga menciptakan kota dengan fasilitas dengan standar dunia. IKN sebagai kota modern ramah lingkungan dan nyaman bagi ... /Agus Wahid- Gunbot Bot#3 Pertemuan Pejuambo Kuat lengan di cairo? #GMGY#3	Presiden Prabowo berkomitmen untuk membangunan IKN yg tidak sekadar memindahkan pusat pemerintahan tetapi juga menciptakan kota dengan fasilitas dengan standar dunia. IKN sebagai kota modern ramah lingkungan dan nyaman bagi ... Agus Wahid- Gunbot Bot#3 Pertemuan pejuambo Kuat
1	IRGunBot#gc95	@BosPanus Pile #IKN Mangkrak dan Jokodok Bajilol Lapaar tangan Ga Mau Tanggung Jawab#JokoWi#Gibran Wakah Hokcerit yg marah Tanggung jawab Wakah Presiden PAI 20 Periode bagi Kacau Idiot 100% yg Dimenangkan Secara Curang#gacor#Perwir Tipuldaya. Bbe Anaknya SIPUAPala	Pile #IKN Mangkrak dan Jokodok Bajilol Lapaar tangan Ga Mau Tanggung Jawab#JokoWi#Gibran Wakah Hokcerit yg marah Tanggung jawab Wakah Presiden PAI 20 Periode bagi Kacau Idiot 100% yg Dimenangkan Secara Curang#gacor#Perwir Tipuldaya. Bbe Anaknya SIPUAPala
2	jaehuni	Hary visi ragam IKN ini baru	Hary visi ragam IKN ini baru
3	draganrobo	@Rajaa2AKABot Jaahya ini gimana bro? Lalu?	Jaahya ini gimana bro? Lalu?
4	EdwardSirait2010	@OppositeAll008 Siapa bilang 1 VS 1??? Memang Anies itu Siapa??Pemimpin itu Siapa? Selain KPK mana lagi yang Berwenang dalam tangani Ibu Kota Ini? Kemudian bukan DKI Jakarta itu IKN, artinya siapa yg menjadi Gubernur disana yg akan menjadi Calon Presiden	Siapa bilang 1 VS 1??? Memang Anies itu Siapa??Pemimpin itu Siapa? Selain KPK mana lagi yang Berwenang dalam tangani Ibu Kota Ini? Kemudian bukan DKI Jakarta itu IKN, artinya siapa yg menjadi Gubernur disana yg akan menjadi Calon Presiden

Gambar 4.8 Hasil Cleansing

Tabel 4.1 Hasil Manual Cleansing

Proses Sistem	Proses Manual
Presiden Prabowo berkomitmen untuk membangunan IKN yg tidak sekadar	Presiden Prabowo berkomitmen untuk membangunan IKN yg tidak sekadar

memindahkan pusat pemerintahan tetapi juga merancang masa depan Indonesia dengan menjadikan IKN sebagai kota modern ramah lingkungan dan nyaman huni Agus Wulan Guritno Bahlil Pertamax jakartabo Nasi	memindahkan pusat pemerintahan tetapi juga merancang masa depan Indonesia dengan menjadikan IKN sebagai kota modern ramah lingkungan dan nyaman huni

b. Case folding

Case folding adalah proses mengubah semua karakter dalam teks menjadi huruf kecil *(lowercase)* untuk memastikan konsistensi. Teknik ini penting dalam analisis teks karena menghindari pengenalan kata yang sama dalam bentuk berbeda (misalnya, "Data" dan "data" akan dianggap sama). Ini membantu dalam mengurangi kompleksitas data dan meningkatkan *akurasi* saat melakukan pencarian atau analisis. Berikut *cource code Case folding* pada gambar 4.9 dibawah ini.

```
def case_folding(text):
    if isinstance(text, str):
        lowercase_text = text.lower()
        return lowercase_text
    else:
        return text

df['case_folding'] = df['cleaning'].apply(case_folding)
df.head(5)
```

Gambar 4.9 Source Code Case Folding

Berikut hasil dari penerapan proses *case folding* pada gambar 4.10 dibawah ini.

Index	username	full_text	cleaning	case_folding
0	hantamantabat	Presiden Prabowo berkomitmen untuk membangunan IKN yg tidak sekadar memindahkan pusat pemerintahan tetapi juga merancang masa depan indonesia dengan menjadikan IKN sebagai kota modern ramah lingkungan dan nyaman huni ___ Agus Wulan Guritno Bahlil Pertamax #jakartabo Nasi https://t.co/orTuOeOYNO	Presiden Prabowo berkomitmen untuk membangunan IKN yg tidak sekadar memindahkan pusat pemerintahan tetapi juga merancang masa depan indonesia dengan menjadikan IKN sebagai kota modern ramah lingkungan dan nyaman huni Agus Wulan Guritno Bahlil Pertamax jakartabo Nasi	presiden prabowo berkomitmen untuk membangunan ikn yg tidak sekadar memindahkan pusat pemerintahan tetapi juga merancang masa depan indonesia dengan menjadikan ikn sebagai kota modern ramah lingkungan dan nyaman huni agus wulan guritno bahlil pertamax jakartabo nasi
1	RKiarEsterlng99	@thotPrince Fix IKN Mangkrak dan Jokowi di Impiti Lopez dengan Ga Mau Tanggung jawabkarna ditutotin Mulek Kekuyel yg harus Tanggungjawab Nulah Presiden PKI 2Periode bagi Kaum Idiot 54% yg Dimenangkan Secara Curangkarna Pemuk TipuDaya. Bln Anehnya SIFuluf ala	Fix IKN Mangkrak dan Jokowi Impiti Lopez dengan Ga Mau Tanggung jawabkarna ditutotin Mulek Kekuyel yg harus Tanggungjawab Nulah Presiden PKI 2Periode bagi Kaum Idiot yg Dimenangkan Secara CuangkarnaPemuk TipuDaya Bln Anehnya SIFuluf ala	fix ikn mangkrak dan jokowi impiti lopez dengan ga mau tanggung jawabkarnaditutotin mulek kekuyel yg harus tanggungjawab nulah presiden pki periode bagi kaum idiot yg dimenangkan secara curangkarnapemuk tipudaya bln anehnya sifuluf ala
2	jaehenn_	Holy shit rajah IKN is here	Holy shit rajah IKN is here	holy shit rajah ikn is here
3	dagaxandalo	@Raja Addriansah Jualnya ikn gnna bro? Laku?	Jualnya ikn gnna bro Laku	jualnya ikn gnna bro laku
4	EdwardSimk3493	@OposisiMK666 Siapa bilang 1 VS 1977 Mainung Antes itu Siapa? Hmmmm Itu Siapa? kalau KIM mau bisa saja Semuanya diambil tanpa tersisa Ibu Kota Ind nantinya bukan DKI Jakarta tp IKN artinya siapa yg menjadi Gubernur disana yg akan menjadi Calon Presiden	Siapa bilang VS mainung Antes itu SiapaPrumono itu Siapa kalau KIM mau bisa saja Semuanya diambil tanpa tersisa Ibu Kota Ind nantinya bukan DKI Jakarta tp IKN artinya siapa yg menjadi Gubernur disana yg akan menjadi Calon Presiden	siapa bilang vs mainung antes itu siapaprumono itu siapa kalau kim mau bisa saja semuanya diambil tanpa tersisa ibu kota ind nantinya bukan dki jakarta tp ikn artinya siapa yg menjadi gubernur disana yg akan menjadi calon presiden

Gambar 4.10 Hasil Case Folding

Tabel 4.2 Hasil Manual Case Folding

Proses Sistem	Proses Manual
presiden prabowo berkomitmen untuk membangunan ikn yg tidak sekadar memindahkan pusat pemerintahan tetapi juga merancang masa depan indonesia dengan menjadikan ikn sebagai kota modern ramah lingkungan dan nyaman huni agus wulan guritno bahlil pertamax jakartabo nasi	presiden prabowo berkomitmen untuk membangunan ikn yg tidak sekadar memindahkan pusat pemerintahan tetapi juga merancang masa depan indonesia dengan menjadikan ikn sebagai kota modern ramah lingkungan dan nyaman huni

c. *Normalization*

Normalization digunakan untuk mengganti kata yang tidak baku menjadi baku sesuai anjuran KBBI.

Dengan artian mengubah kata yang tidak sesuai/typo menjadi kata yang dapat dimengerti oleh computer. Berikut *cource code Normalization* pada gambar 4.11 dibawah ini.

```
import pandas as pd

# fungsi penggantian kata tidak baku
def replace_taboo_words(text, kamus_tidak_baku):
    if isinstance(text, str):
        words = text.split()
        replaced_words = []
        kalimat_baku = []
        kata_diganti = []
        kata_tidak_baku_hash = []

        for word in words:
            if word in kamus_tidak_baku:
                baku_word = kamus_tidak_baku[word]
                if isinstance(baku_word, str) and all(char.isalpha() or char.isspace() for char in baku_word):
                    replaced_words.append(baku_word)
                    kalimat_baku.append(baku_word)
                    kata_diganti.append(word)
                    kata_tidak_baku_hash.append(hash(word))
            else:
                replaced_words.append(word)
        replaced_text = ' '.join(replaced_words)
    else:
        replaced_text = ''
        kalimat_baku = []
        kata_diganti = []
        kata_tidak_baku_hash = []

    return replaced_text, kalimat_baku, kata_diganti, kata_tidak_baku_hash
```

Gambar 4.11 Source Code Normalization

Dalam proses *normalisasi* di *library python* tentunya tidak menyediakan pengubahan kata tidak baku menjadi baku sehingga dilakukan dengan manual yaitu dengan cara mengumpulkan kosa kata dalam satu file yang diberi nama kamuskatabaku. Berikut tabel 4.3 kamus kata baku.

Tabel 4.3 Kata Baku

No.	Tidak baku	Kata baku
1.	woww	wow
2.	amiin	amin

3.	met	selamat
4.	keberpa	keberapa
5.	ehhhhhhh	eh
6.	Kata2nyaaa	Kata-katanya
….	…..	…..

Setelah dilakukan proses pengimport an file kemudian masuk pada proses pengubahan kata menjadi kata baku. Dimana kata yang tidak baku akan secara otomatis berubah menjadi kata baku sesuai dengan kumplan kata yang berada di kamus kata baku yang telah di input. Berikut 4.12 source code pemanggilan file dan penerapan file.

Gambar 4.12 Pemanggilan Dan Penerapan File

Berikut hasil dari penerapan proses *Normalization* pada gambar 4.13 dibawah ini.

Gambar 4.13 Hasil Normalization

Tabel 4.4 Hasil Manual Normalization

Proses Sistem	Proses Manual
presiden prabowo berkomitmen untuk membangunan ikn yang tidak sekadar memindahkan pusat pemerintahan tetapi juga merancang masa depan indonesia dengan menjadikan ikn sebagai kota modern ramah lingkungan dan nyaman huni agus wulan guritno bahlil pertamax jakartabo nasi	presiden prabowo berkomitmen untuk membangunan ikn yang tidak sekadar memindahkan pusat pemerintahan tetapi juga merancang masa depan indonesia dengan menjadikan ikn sebagai kota modern ramah lingkungan dan nyaman huni

d. *Tokenizing*

Tokenizing adalah proses memecah teks menjadi unit-unit kecil yang disebut *token*, biasanya berupa kata atau frasa. Proses ini penting dalam pemrosesan bahasa alami (NLP) karena memungkinkan analisis lebih lanjut pada setiap kata atau frasa secara individual. Di proses ini setiap kata akan dipisah dengan koma untuk memudahkan proses filtering. Berikut *cource code tokenizing* pada gambar 4.14 dibawah ini.

```python
def tokenize(text):
    tokens = text.split()
    return tokens

df['tokenize'] = df['hasil_normalisasi'].apply(tokenize)

df.head(5)
```

Gambar 4.14 Source Code Tokenizing

Berikut hasil dari penerapan proses *tokenizing* pada gambar 4. 15 dibawah ini.

	username	full_text	cleaning	case_folding	hasil_normalisasi	tokenize
0	[illegible]	Presiden Prabowo berkomitmen untuk membangunan...	Presiden Prabowo berkomitmen untuk membangunan...	presiden prabowo berkomitmen untuk membangunan...	presiden prabowo berkomitmen untuk membangunan...	[presiden, prabowo, berkomitmen, untuk, memban...
1	[illegible]	[illegible] Fix IKN Mangkrak dan Jokodok [illegible]...	Fix IKN Mangkrak dan Jokodok [illegible]...	fix ikn mangkrak dan jokodok [illegible]...	fix ikn mangkrak dan jokodok [illegible]...	[fix, ikn, mangkrak, dan, jokodok, [illegible]...
2	[illegible]	Holy shit rajan IKN is here	Holy shit rajan IKN is here	holy shit rajan ikn is here	holy shit rajan ikn is here	[holy, shit, rajan, ikn, is, here]
3	[illegible]	[illegible]	Jualnya ikn gmna bro Laku	jualnya ikn gmna bro laku	jualnya ikn bagaimana bro laku	[jualnya, ikn, bagaimana, bro, laku]
4	[illegible]	[illegible] Siapa bilang 1 VS 1517 Memang An...	Siapa bilang VS Memang Anies itu SiapaPramo...	siapa bilang vs memang anies itu siapapramo...	siapa bilang vs memang anies itu siapapramono...	[siapa, bilang, vs, memang, anies, itu, siapap...

Gambar 4.15 Hasil Tokenizing

Tabel 4.5 Hasil Manual Tokenizing

Proses Sistem	Proses Manual
['presiden', 'prabowo', 'berkomitmen', 'untuk', 'membangunan', 'ikn', 'yang', 'tidak', 'sekadar', 'memindahkan', 'pusat', 'pemerintahan', 'tetapi', 'juga', 'merancang', 'masa', 'depan', 'indonesia', 'dengan', 'menjadikan', 'ikn', 'sebagai', 'kota', 'modern', 'ramah', 'lingkungan', 'dan', 'nyaman', 'huni', 'agus', 'wulan', 'guritno',	['presiden', 'prabowo', 'berkomitmen', 'untuk', 'membangunan', 'ikn', 'yang', 'tidak', 'sekadar', 'memindahkan', 'pusat', 'pemerintahan', 'tetapi', 'juga', 'merancang', 'masa', 'depan', 'indonesia', 'dengan', 'menjadikan', 'ikn', 'sebagai', 'kota', 'modern', 'ramah', 'lingkungan', 'dan', 'nyaman', 'huni']

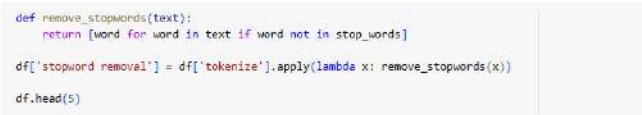
'bahlil', 'pertamax', 'jakartabo', 'nasi']	

e. *Stopword removal*

Stopword removal adalah langkah untuk menghapus kata-kata umum yang tidak memberikan informasi signifikan atau kata yang tidak penting dalam analisis teks, seperti "dan", "atau", "adalah". Menghilangka *stopwords* membantu menyederhanakan dataset dan meningkatkan fokus pada kata-kata yang lebih bermakna dalam konteks analisis. Ini juga membantu mengurangi dimensi data dan meningkatkan efisiensi model. Berikut 4.16 Penginstalan *Stopword*

Gambar 4.16 Install Stopword

Berikut *cource code Stopword* pada gambar 4.17 dibawah ini.



Gambar 4.17 Source Code Stopword Removal

Berikut hasil dari penerapan proses *stopword removal* pada gambar 4. 18 dibawah ini.

	username	full_text	cleaning	case_folding	hasil_normalisasi	tokenize	stopword removal
0	tantotmantabs	Presiden Prabowo berkomitmen untuk membangunan...	Presiden Prabowo berkomitmen untuk membangunan...	presiden prabowo berkomitmen untuk membangunan...	presiden prabowo berkomitmen untuk membangunan...	['presiden', 'prabowo', 'berkomitmen', 'untuk', 'membangunan...	['presiden', 'prabowo', 'berkomitmen', 'membangunan...
1	RikanSantang99	@BosPurwo Fix MOI Mangkrak dan jokodok Bajilol Lepas la...	Fix IKN mangkrak dan jokodok Bajilol Lepas la...	fix ikn mangkrak dan jokodok bajilol lepas la...	fix ikn mangkrak dan jokodok bajilol lepas lan...	['fix', 'ikn', 'mangkrak', 'dan', 'jokodok', 'bajilol', 'le...	['fix', 'ikn', 'mangkrak', 'jokodok', 'bajilol', 'lepas', ...
2	jaunarn_	Holy shit rajah IKN is here	Holy shit rajah IKN is here	holy shit rajah ikn is here	holy shit rajah ikn is here	['holy', 'shit', 'rajah', 'ikn', 'is', 'here']	['holy', 'shit', 'rajah', 'ikn', 'is', 'here']
3	diagnerado	@RajaJukVictim Jualnya ikn gimna bro? Laku?	Jualnya ikn gimna bro Laku	jualnya ikn gimna bro laku	jualnya ikn bagaimana bro laku	['jualnya', 'ikn', 'bagaimana', 'bro', 'laku']	['jualnya', 'ikn', 'bro', 'laku']
4	DewaBiru03800	@Opposite888 Bapa biang ! VS !!?? Memang Ar...	Bapa biang VS memang Aries itu SapaPramo...	bapa biang vs memang aries itu sapapramo...	bapa biang vs memang aries itu sapapramono ...	['bapa', 'biang', 'vs', 'memang', 'aries', 'itu', 'sapap...	['biang', 'vs', 'aries', 'sapapramono', 'kim', 'diambil...

Gambar 4.18 Hasil Stopword Removal

Tabel 4.6 Hasil Manual Stopword Removal

Proses Sistem	Proses Manual
['presiden', 'prabowo', 'berkomitmen', 'membangunan', 'ikn', 'memindahkan', 'pusat', 'pemerintahan', 'merancang', 'indonesia', 'menjadikan', 'ikn', 'kota', 'modern', 'ramah', 'lingkungan', 'nyaman', 'huni', 'agus', 'wulan', 'guritno', 'bahlil', 'pertamax', 'jakartabo', 'nasi']	['presiden', 'prabowo', 'berkomitmen', 'membangunan', 'ikn', 'sekadar', 'memindahkan', 'pusat', 'pemerintahan', 'merancang', 'masa', 'depan', 'indonesia', 'menjadikan', 'ikn', 'kota', 'modern', 'ramah', 'lingkungan', 'nyaman', 'huni']

f. *Stemming*

Stemming adalah proses mengubah kata-kata ke bentuk dasarnya (stem) atau standar dengan cara menghapus akhiran atau prefiks. Proses ini membantu

menyatukan variasi kata yang memiliki makna serupa sehingga memudahkan analisis dan pengelompokan data. Pada proses *stemming* diperlukan penginstalan salah satu *library* yaitu *library sastrawi* selanjunya mengimport *library* dengan memanggil kelas *StemmerFactory.* Berikut *cource code* pemanggilan *library sastrawi* pada gambar 4.19 dibawah ini.

```
!pip install Sastrawi

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from nltk.stem import PorterStemmer
from nltk.stem.snowball import SnowballStemmer
```

Gambar 4.19 Install Sastrawi

Kemudian memasukkan *code* proses *stemming.* Pada gambar 4.20

```
factory = StemmerFactory()
stemmer = factory.create_stemmer()

def stem_text(text):
    return [stemmer.stem(word) for word in text]

df['steming_data'] = df['stopword removal'].apply(lambda x: ' '.join(stem_text(x)))
df.head(5)
```

Gambar 4.20 Source Code Stemming

Berikut hasil dari penerapan proses *stemming* pada gambar 4. 21 dibawah ini.

	username	full_text	cleaning	case_folding	hasil_normalisasi	tokenize	stopword removal	stemming_data
0	herkatmanabak	Presiden Prabowo berkomitmen untuk membangunan...	Presiden Prabowo berkomitmen untuk membangunan...	presiden prabowo berkomitmen untuk membangunan...	presiden prabowo berkomitmen untuk membangunan...	[presiden, prabowo, berkomitmen, untuk, memban...	[presiden, prabowo, berkomitmen, membangunan, ...	presiden prabowo komitmen bangun ikn pindah pu...
1	RKiantSantang09	@BosPanes Fix IKN Mangkrak dan Jokodok Bajilol Lepas ta...	fix ikn mangkrak dan jokodok bajilol lepas ta...	fix ikn mangkrak dan jokodok bajilol lepas tan...	[fix, ikn, mangkrak, dan, jokodok, bajilol, le...	[fix, ikn, mangkrak, jokodok, bajilol, lepas, ...	fix ikn mangkrak jokodok bajilol lepas tangan ...	fix ikn mangkrak jokodok bajilol lepas tangan ...
2	jauharo_	Holy shit rajah IKN is here	Holy shit rajah IKN is here	holy shit rajah ikn is here	holy shit rajah ikn is here	[holy, shit, rajah, ikn, is, here]	[holy, shit, rajah, ikn, is, here]	holy shit rajah ikn is here
3	dragoenisole	@RajaJukiNotoni Jualnya ikn gmna bro? Laku?	Jualnya ikn gmna bro? Laku	jualnya ikn gmna bro laku	jualnya ikn bagaimana bro laku	[jualnya, ikn, bagaimana, bro, laku]	[jualnya, ikn, bro, laku]	jual ikn bro laku
4	Edwardtien036933	@Opposite6888 Siapa bilang VS Memang Anies itu SiapaPramo...	siapa bilang vs memang anies itu siapapramo...	siapa bilang vs memang anies itu siapapramono ...	siapa, bilang, vs, memang, anies, itu, siapap...	[siapa, bilang, vs, memang, anies, siapapramon...	[bilang, vs, anies, siapapramono, kim, dambil...	bilang vs anies siapapramono kim ambil sisa ko...

Gambar 4.21 Hasil Stemming

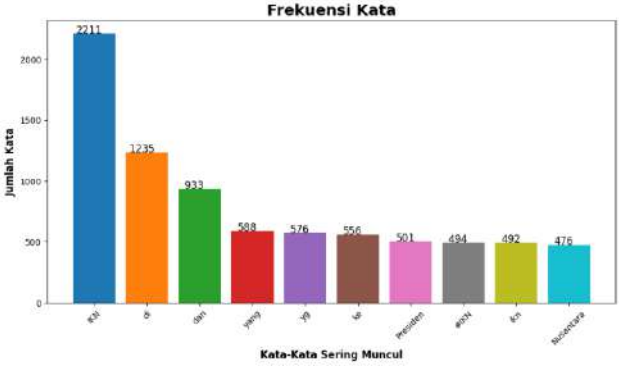
Tabel 4.7 Hasil Manual Stemming

Proses Sistem	Proses Manual
presiden prabowo komitmen bangun ikn pindah pusat perintah rancang indonesia jadi ikn kota modern ramah lingkung nyaman huni agus wulan guritno bahlil pertamax jakartabo nasi	presiden prabowo komitmen bangun ikn pindah pusat pemerintah rancang masa depan indonesia jadi ikn kota modern ramah lingkungan nyaman huni

Sebelum dilakukannya proses *prepocessing* data terkesan lebih berantakan dan masih banyak *nois* yang tersebar sehingga hasil wordcloud masih belum sesuai dengan apa yang dibahas. Berikut gambar 4.22 Wordcloud sebelum *prepocessing*.

Gambar 4.22 Wordcloud Sebelum Prepocessing

Frekuensi kata sebelum dilakukan proses *prepocessing* juga cenderung banyak yang tidak penting dan keyword kurang menonjol berikut pada gambar 4.23 Frekuensi kata sebelum *prepocessing*.

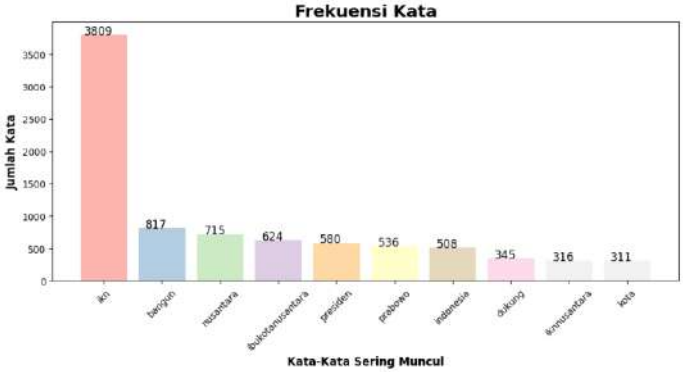


Gambar 4.23 Frekuensi Kata Sebelum Prepocessing

Setelah dilakukannya proses *prepocessing* dapat dilihat pada gambar 4.24 Bahwa wordcloud lebih terstruktur dan menonjolkan poin utama yang dicari juga lebih terlihat.

Gambar 4.24 Wordcloud Setelah Prepocessing

Frekuensi kata setelah *prepocessing* juga sudah berubah perbandingan frekuensi munculnya kata sebelum dan sesudah proses *prepocessing* sangat jauh berbeda. Berikut pada gambar 4.25 Frekuensi kata setelah *prepocessing*



Gambar 4.25 Frekuensi Kata Setelah Prepocessing

C. Pelabelan Data

Agar mempermudah penentuan kelas, tahapan ini dilakukan setelah proses *prepocessing* data selesai dimana data mentah sudah diolah dan disortir sehingga data sudah fix dan sudah tidak memiliki *nois*. Dalam proses pelabelan data ini peneliti menentukan nilai dari data jika opini positif artinya setuju atau pro dengan apa yang dibahas, negatif berarti user tidak setuju atau menentang pembahasan, serta netral artinya user tidak memihak salah satunya atau bisa jadi berisi informasi dari apa yang di bahas dalam hal ini mengenai IKN. Dalam proses pelabelan ini diperlukan seorang pakar Bahasa atau seseorang yang paham dibidangnya untuk menentukan kelas dari data sentimen yang diperoleh hanya saja dengan menganalisis banyak data berarti juga membutuhkan waktu yang cukup lama oleh karena itu peneliti menggunakan *lexicon* yaitu pelabelan otomatis yang menggunakan mesin untuk mempermudah pakar dalam melabeling sentimen. Kekurangan dari lexicon sendiri yaitu kurang akurat karena *lexicon* mendeteksi kalimat sesuai dengan rumus kata yang sudah di input oleh karena itu dalam penelitian ini tetap dibutuhkan pakar untuk melabeling manual. Berikut gambar 4.26 code *lexicon*

```
import pandas as pd

# Unduh kamus leksikon positif dan negatif dari URL yang diberikan
positive_lexicon = set(pd.read_csv("positive.tsv", sep="\t", header=None)[0])
negative_lexicon = set(pd.read_csv("negative.tsv", sep="\t", header=None)[0])

# Fungsi untuk menentukan sentimen dan menghitung skornya
def determine_sentiment(text):
    if isinstance(text, str):
        positive_count = sum(1 for word in text.split() if word in positive_lexicon)
        negative_count = sum(1 for word in text.split() if word in negative_lexicon)
        sentiment_score = positive_count - negative_count
        if sentiment_score > 0:
            sentiment = "Positif"
        elif sentiment_score < 0:
            sentiment = "Negatif"
        else:
            sentiment = "Netral"
        return sentiment_score, sentiment
    return 0, "Netral"

# Tentukan sentimen dan skor untuk setiap ulasan
df[['Score', 'Sentiment']] = df['steming_data'].apply(lambda x: pd.Series(determine_sentiment(x)))

df.head(10)
```

Gambar 4.26 Code Lexicon

Hasil yang diperoleh dalam pelabelan lexicon dan manual jelas berbeda. Berikut gambar 4.27 Hasil pelabelan *lexicon*

	tanggal	waktu	steming_data	Score	Sentiment
0	Fri Nov 29	23:51:02	presiden prabowo komitmen bangun ikn pindah pu...	3	Positif
1	Fri Nov 29	23:36:57	fix ikn mangkrak jokodok bajilol lepas tangan ...	-1	Negatif
2	Fri Nov 29	23:36:29	holy shit rajjah ikn is here	0	Netral
3	Fri Nov 29	23:30:44	jual ikn bro laku	1	Positif
4	Fri Nov 29	23:29:28	bilang vs anies siapapramono kim ambil sisa ko...	-3	Negatif
5	Fri Nov 29	23:29:26	ikn	0	Netral
6	Fri Nov 29	23:05:46	uang asih upeti lancar urus lihat naga kemarin...	-1	Negatif
7	Fri Nov 29	23:01:19	konon info wakil ketua ikn kang	-1	Negatif
8	Fri Nov 29	22:43:15	pindah pindah mangkrak	0	Netral
9	Fri Nov 29	22:40:03	lapor telepon cakeda tagih janji pindah ikn	0	Netral

Gambar 4.27 Hasil Lexicon

Pelabelan manual dilakukan oleh salah satu mahasiswa jurusan Bahasa dan sastra dari universitas

negeri semarang bernama Raihan Adib Ghifari. Berikut tabel 4.8 hasil data labeling manual.

Tabel 4.8 Hasil Labeling Manual

No.	Sentimen	kelass
1	presiden prabowo komitmen bangun ikn pindah pusat perintah rancang indonesia jadi ikn kota modern ramah lingkung nyaman huni agus wulan guritno bahlil pertamax jakartabo nasi	Positif
2	fix ikn mangkrak jokodok bajitol lepas tangan tanggung jawabampdisalahin rakyat tanggungjawab presiden pki periode kaum idiot menang curangamppenuh tipudaya anak sifufufafa	Negatif
3	bilang vs anies siapapramono kim ambil sisa kota ind dki jakarta ikn gubernur sana calon presiden	Positif
4	ikn pindah pusat perintah rancang indonesia integrasi teknologi canggih budaya lokal	Positif
5	uang asih upeti lancar urus lihat naga kemarin suruh kumpul ikn tagih bangun nyetor upeti gede pilkada kemarin menang	Negatif
6	ikn kota baru	Netral
7	presiden prabowo proyek ikn selesai	Positif
….	…..	…..
4124	lapor telepon cakeda tagih janji pindah ikn	Positif
4125	lapor bos kantor ikn	Netral

Dikarenakan dalam proses sebelumnya yaitu proses *prepocessing* data mentah sudah di proses sehingga hanya menyisakan 4125 data jadi dari data mentah sebanyak 5000. Dan setelah memasuki proses labeling yang menentukan kelas sentimen didapatkan bahwa sentimen positif berjumlah 1589 data *tweet*, sentimen netral sejumlah 1321 data *tweet* dan sentimen negatif sebanyak 1215 data *tweet*. Dapat dilihat pada gambar 4.28 Hasil sentimen.

```
df['Sentiment'].unique()
```

```
array(['Positif', 'Negatif', 'Netral'], dtype=object)
```

```
df.Sentiment.value_counts()
```

Sentiment	count
Positif	1589
Netral	1321
Negatif	1215

dtype: int64

Gambar 4.28 Jumlah Sentimen

D. Ekstrasi Fitur

Selanjutnya yaitu proses ektraksi fitur, pada tahap ini diperlukan *library sklearn*, *Library* ini juga mendukung model regresi untuk memprediksi nilai kontinu. Algoritma seperti regresi linier dan regresi polinomial dapat digunakan untuk analisis hubungan antar variabe(Riadi Silitonga 2019). Berikut gambar 4.29 Pengimport an *library sklearn*.

```
import re
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
```

Gambar 4.29 Import Sklearn

Selanjutnya yaitu proses split validation data untuk mempermudah proses klasifikasi yaitu proses membagi data menjadi data uji dan data latih. Pada proses ini persentase yang digunakan yaitu 80:20. Proses pembagian dapat dilihat pada 4.30 Dibawah ini.

Gambar 4.30 Source Code Split Data

Setelah dilakukannya proses split validation data yang membagi data 80:20 didapatkan hasil sebanyak 3278 sebagai data latih dan 820 sebagai data uji dari data keseluruhan.

Tahap selanjutnya yaitu proses ekstraksi fitur, ektraksi fitur sendiri merupakan teknik yang digunakan untuk mengidentifikasi dan memilih informasi penting dari teks. Dalam analisis sentimen fitur yang di ekstrak biasanya mencakup kata-kata atau frasa yang dapat menunjukkan emosi atau opini yang kemudian akan dirubah menjadi bilangan angka agar dapat di proses oleh sistem dalam proses klasifikasi. Untuk melaksanakan tahapan ini dibutuhkan metode pembobotan kata, dalam penelitian ini metode pembobotan yang digunakan ada dua yaitu metode TF-IDF dan *Word2Vec*.

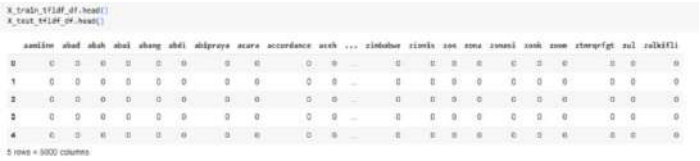
Masing-masing metode memiliki kebaikan dan kekurangan, jika TF-IDF berfungsi untuk mengukur seberapa sering dan seberapa penting sebuah kata muncul dan lebih efisien jika data pendek dan sedikit maka *Word2Vec* berfungsi untuk merubah kata kata

menjadi vector numerik dalam ruang vector berdimensi tinggi dan lebih efisien jika data panjang dan banyak. Berikut source code pembobotan TF-IDF dapat dilihat pada gambar 4.31

```
import pandas as pd
import numpy as np

# Membuat DataFrame untuk data TF-IDF
X_train_tfidf_df = pd.DataFrame.sparse.from_spmatrix(X_train_tfidf, columns=tfidf_vectorizer.get_feature_names_out())
X_test_tfidf_df = pd.DataFrame.sparse.from_spmatrix(X_test_tfidf, columns=tfidf_vectorizer.get_feature_names_out())

# Menambahkan label pada data latih dan uji
X_train_tfidf_df['label'] = y_train.tolist()
X_test_tfidf_df['label'] = y_test.tolist()

# Menyimpan dalam format CSV
train_csv_path = 'tfidf_train_data.csv'
test_csv_path = 'tfidf_test_data.csv'

X_train_tfidf_df.to_csv(train_csv_path, index=False)
X_test_tfidf_df.to_csv(test_csv_path, index=False)

train_csv_path, test_csv_path
```

Gambar 4.31 Source Code Pembobotan TF-IDF

Setelah tahap pembobotan, hasil akan di input menjadi sekumpulan array yang menjadi suatu matrix dimana setiap baris mewakili setiap dokumen dan pada setiap kolom mewakili setiap kata pada keseluruhan teks. Berikut gambar 4.32 hasil dari pembobotan TF-IDF



```
X_train_tfidf_df.head()
X_test_tfidf_df.head()
```

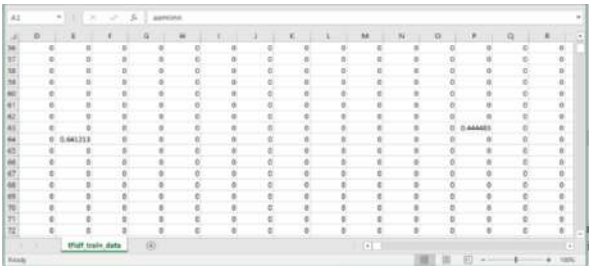
	aamiine	abad	abah	abai	abang	abdi	abipraya	acara	accordance	aceh	...	zimbabwe	zinnia	zoo	zona	zonasi	zonk	zoom	zteregrfgt	zul	zulkifli
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

5 rows × 5000 columns

Gambar 4.32 Hasil Pembobotan TF-IDF

Dikarenakan kata yang di proses sangat banyak sekali maka hasil tidak hanya ditunjukkapan pada

pemanggilan array hasil juga diinput pada file CSV. Pada file CSV hasil tidak hanya berupa nominal kosong namun dapat dilihat bahwa terdapat nilai tertentu yang dihasilkan. Berikut hasil dalam file CSV pada gambar 4.33



Gambar 4.33 CSV Hasil Pembobotan TF-IDF

Sebagai contoh, penelitian menggunakan tiga komentar untuk perhitungan manualisasi untuk pembobotan TF-IDF sebagai berikut:

Doc = dokumen/komentar

(Doc1)=“IKN milik bangsa indonesia ”

(Doc2)=“Kayak IKN megah tak ada penghuninya”

(Doc3)=“ikn memajukan bangsa indonesia”

Setelah sistem melakukan *prepocessing* maka menjadi seperti berikut:

(Doc1)=['ikn’,’milik’,’bangsa’,’indonesia']

(Doc2)=['kayak', 'ikn', 'megah',tidak’, 'huni']

(Doc3)=['ikn', 'maju', 'bangsa', 'indonesia']

Pada tahap selanjutnya dilakukan perhitungan dengan metode TF-IDF dengan tujuan membentuk word vector yang telah diberi bobot nilai. Pada TF-IDF sendiri mempunyai dua kata yaitu TF(Term Frequency) dan IDF(Invers Document Frequency) Dimana TF merupakan jumlah sebuah kata dari tiap dokumen, sedangkan IDF berfungsi mengurangi bobot suatu kata apabila kemunculannya tersebar di banyak dokumen. Langkah pertama untuk pembobotan TF-IDF yaitu mencari nilai TF terlebih dahulu. Berikut perhitungan rumus TF secara manualisasi.

$$\mathrm{TF}(t,d) = \frac{jumlah\ kemunculan\ term\ t\ dalam\ dokumen\ d}{jumlah\ total\ kata\ dalam\ dokumen\ d}$$

Dokumen 1

$\mathrm{TF}(ikn,d) = \frac{1}{4} = 0{,}25$

$\mathrm{TF}(milik,d) = \frac{1}{4} = 0{,}25$

$\mathrm{TF}(bangsa,d) = \frac{1}{4} = 0{,}25$

$\mathrm{TF}(indonesia,d) = \frac{1}{4} = 0{,}25$

Dokumen 2

$\mathrm{TF}(kayak,d) = \frac{1}{5} = 0{,}2$

$\mathrm{TF}(ikn,d) = \frac{1}{5} = 0{,}2$

$\text{TF}(megah, d) = \frac{1}{5} = 0{,}2$

$\text{TF}(tidak, d) = \frac{1}{5} = 0{,}2$

$\text{TF}(huni, d) = \frac{1}{5} = 0{,}2$

Dokumen 3

$\text{TF}(ikn, d) = \frac{1}{4} = 0{,}25$

$\text{TF}(maju, d) = \frac{1}{4} = 0{,}25$

$\text{TF}(bangsa, d) = \frac{1}{4} = 0{,}25$

$\text{TF}(indonesia, d) = \frac{1}{4} = 0{,}25$

Setelah didapatkan nilai TF selanjutnya yaitu masuk pada tahapan perhitungan IDF namun sebelum masuk pada tahapan IDF harus mengetahui DF terlebih dahulu, DF(Document Frequency) merupakan jumlah dokumen dalam korpus yang mengandung istilah tertentu. Dalam konteks perhitungan TF-IDF, DF digunakan untuk menghitung nilai Inverse Document Frequency (IDF), yang merupakan komponen penting dalam menentukan seberapa signifikan suatu istilah dalam konteks koleksi dokumen yang lebih besar. Fungsi DF sendiri yaitu mengukur seberapa sering istilah muncul di dalam dokumen. Jika sebuah istilah muncul dalam banyak dokumen, maka DF-nya tinggi.

Sebaliknya, jika istilah tersebut hanya muncul di sedikit dokumen, DF-nya rendah.

Tabel 4.9 Mencari DF

TOKEN	TF			DF
	D1	D2	D3	
Ikn	1	1	1	3
Milik	1	0	0	1
Bangsa	1	0	1	2
Indonesia	1	0	1	2
Kayak	0	1	0	1
Megah	0	1	0	1
Tidak	0	1	0	1
Huni	0	1	0	1
maju	0	0	1	1

Setelah didapatkan nilai DF, selanjutnya yaitu masuk pada perhitungan IDF (Inverse Document Frequency) untuk menilai seberapa penting suatu kata dalam konteks koleksi dokumen yang lebih besar. IDF membantu mengurangi bobot kata-kata yang umum dan memberikan bobot lebih pada kata-kata yang jarang muncul di seluruh dokumen. Semakin jarang suatu istilah muncul di antara dokumen, semakin tinggi nilai IDF-nya. Ini berarti bahwa kata-kata yang lebih spesifik dan kurang umum akan mendapatkan bobot yang lebih tinggi, sehingga lebih relevan dalam konteks analisis. Diketahui jumlah dokumen(D) yang digunakan sebagai contoh disini yaitu sebanyak tiga

komentar, sehingga diketahui D=3. Berikut perhitungan IDF.

$$\text{IDF}(t, D) = \log\left(\frac{N}{DF(t)}\right)$$

Note: N adalah total jumlah dokumen

$$\text{IDF}(ikn) = \log\left(\frac{3}{3}\right) = 0$$

$$\text{IDF}(milik) = \log\left(\frac{3}{1}\right) = 0{,}4771$$

$$\text{IDF}(bangsa) = \log\left(\frac{3}{2}\right) = 0{,}1760$$

$$\text{IDF}(indonesia) = \log\left(\frac{3}{2}\right) = 0{,}1760$$

$$\text{IDF}(kayak) = \log\left(\frac{3}{1)}\right) = 0{,}4771$$

$$\text{IDF}(megah) = \log\left(\frac{3}{1}\right) = 0{,}4771$$

$$\text{IDF}(tidak) = \log\left(\frac{3}{1}\right) = 0{,}4771$$

$$\text{IDF}(huni) = \log\left(\frac{3}{1}\right) = 0{,}4771$$

$$\text{IDF}(maju) = \log\left(\frac{3}{1}\right) = 0{,}4771$$

Tabel 4.10 Mencari IDF

TOKEN	DF	D/DF	IDF (Log(D/DF)	HASIL
Ikn	3	1	Log 1	0
Milik	1	3	Log 3	0,4771
Bangsa	2	1,5	Log 1,5	0,1760
Indonesia	2	1,5	Log 1,5	0,1760
Kayak	1	3	Log 3	0,4771

Megah	1	3	Log 3	0,4771
Tidak	1	3	Log 3	0,4771
Huni	1	3	Log 3	0,4771
maju	1	3	Log 3	0,4771

Setelah nilai TF dan IDF di dapatkan selanjutnya yaitu masuk pada perhitungan TF-IDF. Dengan mengalikan nilai TF dengan IDF.

TF-IDF$(t, d, D) = TF(t, d) \times IDF(t, D)$

Dokumen 1

TF-IDF(ikn, $doc1) = 0{,}25 \times 0 = 0$

TF-IDF(milik, $doc1) = 0{,}25 \times 0{,}4771 = 0{,}1192$

TF-IDF(bangsa, $doc1) = 0{,}25 \times 0{,}1760 = 0{,}044$

TF-IDF(indonesia, $doc1) = 0{,}25 \times 0{,}1760 = 0{,}044$

Dokumen 2

TF-IDF(kayak, $doc2) = 0{,}2 \times 0{,}4771 = 0{,}352$

TF-IDF(ikn, $doc2) = 0{,}2 \times 0 = 0$

TF-IDF(tidak, $doc2) = 0{,}2 \times 0{,}4771 = 0{,}352$

TF-IDF(megah, $doc2) = 0{,}2 \times 0{,}4771 = 0{,}352$

TF-IDF(huni, $doc2) = 0{,}2 \times 0{,}4771 = 0{,}352$

Dokumen 3

TF-IDF(ikn, $doc3) = 0{,}25 \times 0 = 0$

TF-IDF(maju, $doc3) = 0{,}25 \times 0{,}4771 = 0{,}1192$

TF-IDF(bangsa, $doc3) = 0{,}25 \times 0{,}1760 = 0{,}044$

TF-IDF(indonesia, $doc3) = 0{,}25 \times 0{,}1760 = 0{,}044$

Tabel 4.11 Mencari TF-IDF

TOKEN	TF			IDF	TF * IDF		
	D1	D2	D3		D1	D2	D3
Ikn	0,25	0,2	0,25	0	0	0	0
Milik	0,25	0	0	0,477	0,119	0	0
Bangsa	0,25	0	0,25	0,176	0,044	0	0,044
Indonesia	0,25	0	0,25	0,176	0,044	0	0,44
Kayak	0	0,2	0	0,477	0	0,352	0
Megah	0	0,2	0	0,477	0	0,352	0
Tidak	0	0,2	0	0,477	0	0,352	0
Huni	0	0,2	0	0,477	0	0,352	0
maju	0	0	0,25	0,477	0	0	0,119

Sehingga apabila hasil TF-IDF ditulis dalam array akan menjadi seperti di bawah ini:

Array([
[0, 0.119, 0.044, 0.044, 0, 0, 0, 0, 0, 0, 0,]
[0, 0, 0.352, 0.352, 0.352, 0.352, 0, 0, 0,]
[0.044, 0.044, 0, 0, 0, 0, 0.119])

Dimana setiap baris menginterpretasikan tiap dokumen, dan setiap kolom menginterpretasikan tiap kata pada seluruh teks.

Selanjutnya yaitu pembobotan *Word2Vec* adalah teknik yang digunakan untuk merepresentasikan kata-kata dalam bentuk vektor numerik, yang

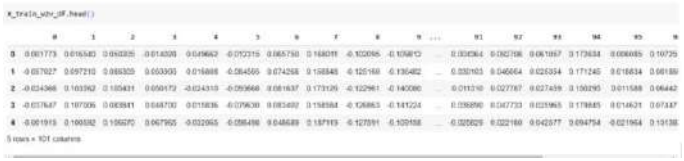
memungkinkan pemodelan hubungan semantik dan sintaktik antar kata.

Word2Vec mampu memproses data besar dan cepat, vector yang dihasilkan dapat menangkap hubungan sematik dan sintaksis antar kata sehingga memungkinkan untuk analisis lebih mendalam. Dalam analisis sentimen *Word2Vec* mengidentifikasi opini positif dan negatif dalam teks berdasarkan hubungan antar kata. Berikut gambar 4.34 *source code* pembobotan *Word2Vec*

```
import pandas as pd

# Membuat DataFrame untuk data Word2Vec
X_train_w2v_df = pd.DataFrame(X_train_w2v)
X_test_w2v_df = pd.DataFrame(X_test_w2v)

# Menambahkan label pada data latih dan uji
X_train_w2v_df['label'] = y_train.values
X_test_w2v_df['label'] = y_test.values

# Simpan ke file CSV
X_train_w2v_df.to_csv("X_train_w2v.csv", index=False)
X_test_w2v_df.to_csv("X_test_w2v.csv", index=False)

print("Data latih Word2Vec disimpan sebagai 'X_train_w2v.csv'")
print("Data uji Word2Vec disimpan sebagai 'X_test_w2v.csv'")

# Membaca dan menampilkan file CSV
print("\nTampilkan 5 data pertama (X_train_w2v):")
print(pd.read_csv("X_train_w2v.csv").head())

print("\nTampilkan 5 data pertama (X_test_w2v):")
print(pd.read_csv("X_test_w2v.csv").head())
```

Gambar 4.34 Source Code Pembobotan Word2Vec

Adapun hasil dari pembobotan *Word2Vec* seperti pada gambar 4.35



Gambar 4.35 Hasil Pembobotan Word2Vec

Dalam Perhitungan *Word2Vec* ada dua algoritma utama yaitu vontinuous bag of word(CBOW) dan skip-gram. Algoritma CBOW digunakan untuk melihat panjang tertentu dari sebuah kata pada dokumen masukan. Sedangkan algoritma skip-gram digunakan untuk memprediksi konteks kata dengan cara melihat kedekatan sebuah kata dengan kata lain yang posisinya sebelum atau sesudah kata tersebut.

Sebagai contoh, penelitian menggunakan salah satu algoritma *Word2Vec* yaitu algoritma skip-gram, Model Skip-gram digunakan untuk menghasilkan representasi vektor dari kata-kata dalam dataset. Representasi ini dibuat dengan mempelajari hubungan semantik antar kata berdasarkan konteksnya. Berikut contoh perhitungan manualisasi pembobotan *Word2Vec*.

1. Pemilihan kata

   ['kayak', 'ikn', 'megah',tidak', 'huni']

2. Pilih kata target

   Semisal kata target “megah”

3. Tentukan ukuran jendela

   Semisal kita menggunakan ukuran jendela *c=2*. Artinya kita akan mempertimbangkan dua kata sebelah kiri dan dua kata sebelah kanan dari target.

4. Identifikasi kata konteks

   Dengan window size *c*=2, konteks untuk kata "megah" adalah:

   Sebelum : "kayak","ikn"

   Sesudah :"tidak","huni"

   Sehingga konteksnya adalah:

   '["kayak", "ikn", "tidak", "huni"]'

5. One-Hot Encoding

   Buat representasi one-hot untuk setiap kata

   '["kayak", "ikn", "megah", "tidak", "huni"]'

   Representasi one-hot untuk setiap kata adalah sebagai berikut:

   Kayak : '[1, 0, 0, 0, 0]'

   Ikn : '[0, 1, 0, 0, 0]'

   Megah : '[0, 0, 1, 0, 0]'

   Tidak : '[0, 0, 0, 1, 0]'

   Huni : '[0, 0, 0, 0, 1]'

6. Menghitung Probabilitas

Probabilitas ini dihitung dengan menggunakan fungsi softmax

$$p(w_o \mid w_t) = \frac{e^{v_{w_o} \cdot v_{w_t}}}{\sum_{w \in V} e^{v_w \cdot v_{w_t}}}$$

- kayak: 0.9*0.2 + 0.6*0.5 + (-0.2)*(-0.1) = 0.18 + 0.3 + 0.02 = 0.5
- ikn: 0.9*0.7 + 0.6*(-0.3) + (-0.2)*0.4 = 0.63 - 0.18 - 0.08 = 0.37
- megah: 0.9*0.9 + 0.6*0.6 + (-0.2)*(-0.2) = 0.81 + 0.36 + 0.04 = 1.21
- tidak: 0.9*0.4 + 0.6*0.1 + (-0.2)*(-0.5) = 0.36 + 0.06 + 0.1 = 0.52
- huni: 0.9*0.3 + 0.6*(-0.7) + (-0.2)*0.8 = 0.27 - 0.42 - 0.16 = -0.31

sofmax

Total = e^0.5 + e^0.37 + e^1.21 + e^0.52 + e^-0.31 ≈ 1.65 + 1.45 + 3.35 + 1.68 + 0.73 ≈ 8.86

Probabilitas:

kayak: e^0.5 /8.86 ≈ 1.65/8.86 ≈ 0.186

ikn: e^0.37 /8.86 ≈ 1.45/8.86 ≈ 0.164

tidak: e^0.52 /8.86 ≈ 1.68/8.86 ≈ 0.190

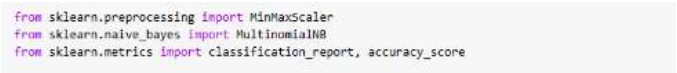
huni: e^-0.31 /8.86 ≈ 0.73/8.86 ≈ 0.082

Fungsi softmax digunakan pada layer output untuk menghitung distribusi probabilitas, fungsi ini mengubah input menjadi probabilitas antara 0 dan 1 yang jumlahnya sama dengan 1. Kata-kata dengan probabilitas tertinggi diidentifikasi sebagai kata-kata kontekstual yang di prediksi.

E. Klasifikasi *Naïve bayes*

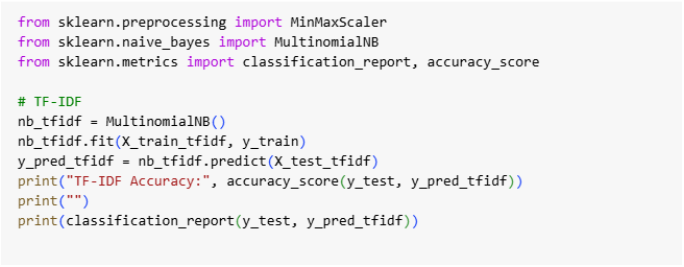
Setelah data melewati proses *prepocessing* kemudian pelabelan dan ekstraksi fitur, selanjutnya data masuk pada tahapan klasifikasi *naïve bayes*, dimana pada proses ini data akan dikelompokkan atau diklasifikasi dengan cara *naïve bayes*. Data yang digunakan merupakan data uji dan data latih dari masing masing pembobotan yaitu pembobotan TF-IDF dan pembobotan *Word2Vec*. Data akan diuji untuk menilai ketepatan suatu sistem dalam mengklasifikasikan data.

Pada proses ini menggunakan *library sklearn* yang kemudian akan digunakan untuk klasifikasi multinominalNB, *accuracy*_score, *precision*_score, *recall*_score,*F1_Score*, kelas*ification_report*, dan *confusion matrix*. Berikut import library pada gambar 4.36



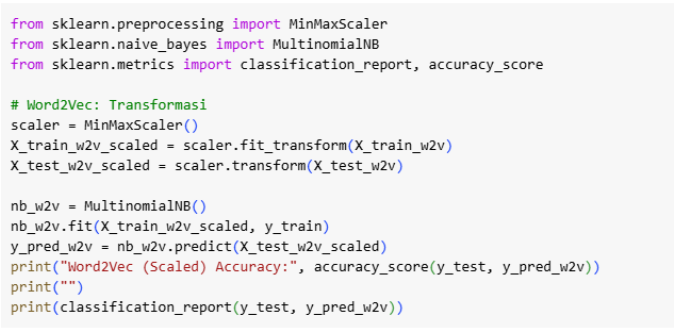
Gambar 4.36 Import Library Untuk Klasifikasi

Selanjutnya proses pengklasifikasian menggunakan *naïve bayes*. Adapun source code klasifikasi TF-IDF dapat dilihat pada gambar 4.37



Gambar 4.37 Source Code Klasifikasi TF-IDF

Adapun untuk klasifikasi *Word2Vec* dapat dilihat pada gambar 4.38



Gambar 4.38 Source Code Klasifikasi Word2Vec

Berikut contoh perhitungan manual Naïve bayes

2. Data uji

Tabel 4.12 Contoh Data Uji

No.	Teks	Kata Kunci	Sentimen
1.	“ IKN maju cepat “	[maju, cepat]	positif
2.	“ Proyek boros “	[boros]	Negatif
3.	“ Pembangunan ramah lingkungan “	[ramah, lingkungan]	Positif
4.	“ Macet parah di IKN “	[macet, parah]	Negatif
5.	“ Infrastruktur biasa saja “	[biasa]	Netral

3. Hitung Prior Probabilitas

- Total data: 5
- Positif: 2/5 = 0.4
- Negatif: 2/5 = 0.4
- Netral: 1/5 = 0.2

4. Hitung Likelihood

Misalkan data uji: "Pembangunan IKN maju dan ramah lingkungan" (Kata kunci: [maju, ramah, lingkungan).

a. Likelihood Positif:

- $P(maju \mid Positif) = \frac{1}{2} = 0{,}5$
- $P(ramah \mid Positif) = \frac{1}{2} = 0{,}5$
- $P(lingkungan \mid Positif) = \frac{1}{2} = 0{,}5$
- Total: 0,5 X 0,5 X 0,5 = 0.125

b. Likelihood Positif:

- $P(maju \mid Negatif) = \frac{0}{2} = 0$ → Gunakan Laplace/smoothing (missal a = 1) : (0+1)/ (2+3) = 0.2
- $P(ramah \mid Negatif) = 0.2$
- $P(maju \mid Negatif) = 0.2$
- Total : 0.2 X 0.2 X 0.2 = 0.008

c. Likelihood Netral :

- $P(ramah \mid Netral) = \frac{0}{1} = 0 \rightarrow \frac{0+1}{1+3} = 0.25$
- $P(ramah \mid Netral) = 0.25$
- $P(ramah \mid Netral) = 0.25$
- Total : 0.25 X 0.25 X 0.25 = 0.0156

5. Hitung Posterior Probabilitas

- Positif : 0.125 X 0.4 = 0.05
- Negatif : 0.008 X 0.4 = 0.0032
- Netral : 0.0156 X 0.2 = 0.0031

6. Normalisasi

Total : 0.05 + 0.0032 + 0.0031 = 0.0563

Probabilitas akhir :

- Positif: 0.05/0.0563 ≈ 88.8%
- Negatif : 0.0032/0.0563 ≈ 5.7%
- Netral : 0.0031/0.0563 ≈ 5.5%

Dari contoh diatas menunjukkan bahwa kelas positif memiliki probabilitas tinggi (88.8%), sehingga data uji diklasifikasikan sebagai positif.

a. Uji Model

Setelah melalui proses klasifikasi *naïve bayes*, selanjutnya data akan masuk pada proses uji model. Tujuan dilakukannya uji model sendiri yaitu untuk mendapatkan ketepatan model dalam pengklasifikasian dan mendapatkan hasil klasifikasi, Data yang digunakan yaitu data uji sebanyak 20% atau 820 data yang nantinya akan ditunjukkan pada tabel *multiclass confusion matrix.* Pada tabel *multiclass confusion matrix* terdapat dua kelas yaitu kelas prediksi (*predicted class*) dan kelas sebenarnya (*True class*). Pada penelitian ini digunakan *multiclass confusion matrix* karena terdapat tiga kelas yaitu positif, negatif, dan netral. Berikut tabel 4.13 Tabel *multiclass confusion matrix.*

Tabel 4.13 Multiclass Confusion Matrix 3x3

		True Kelas		
		Negatif	*Netral*	*Positif*
Predicted Kelas	*Negatif*	T Neg	F NegNet	F NegPos
	Netral	F NetNeg	T Net	F NetPos
	Positif	F PosNeg	F PosNet	T Pos

Untuk mengetahui ketepatan model dalam pengklasifikasian maka harus diketahui nilai *akurasi*nya. Cara untuk mengetahui nilai *akurasi* yaitu dengan cara, menjumlahkan total data yang diprediksi sesuai dengan data sebenarnya kemudian dibagi dengan jumlah keseluruhan data yang diuji. Dikarenakan dalam penelitian ini menggunakan dua metode pembobotan yang akan dibandingkan maka perhitungan *akurasi* akan dilakukan dua kali. Berikut rumus perhitungan nilai *akurasi* dari pembobotan TF-IDF.

$$Akurasi = \frac{T\ Neg + T\ Net + T\ Pos}{Total\ data\ yang\ diuji} \times 100\%$$

$$Akurasi = \frac{166 + 101 + 263}{166 + 40 + 40 + 50 + 101 + 106 + 25 + 29 + 263} \times 100\%$$

$$Akurasi = \frac{530}{820} \times 100\%$$

$$Akurasi = 0{,}6463 \times 100\%$$

$$Akurasi = 65\%$$

Berikut rumus perhitungan nilai *akurasi* dari pembobotan *Word2Vec*.

$$Akurasi = \frac{T\ Neg + T\ Net + T\ Pos}{Total\ data\ yang\ diuji} \times 100\%$$

$$Akurasi = \frac{0 + 47 + 303}{0 + 0 + 0 + 10 + 47 + 14 + 236 + 210 + 303} \times 100\%$$

$$Akurasi = \frac{350}{820} \times 100\%$$

$$Akurasi = \ 0{,}4268 \times 100\%$$

$$Akurasi = \ 43\%$$

Setelah dilakukan tahapan uji model, didapatkan hasil dari nilai *akurasi* TF-IDF dan *Word2Vec*. Berikut gambar hasil nilai *akurasi* pembobotan TF-IDF pada gambar 4.39

```
TF-IDF Accuracy: 0.6463414634146342
```

Gambar 4.39 Akurasi TF-IDF

Berikut gambar hasil nilai *akurasi* pembobotan *Word2Vec* pada gambar 4.40

```
Word2Vec (Scaled) Accuracy: 0.4268292682926829
```

Gambar 4.40 Akurasi Word2Vec

Dari dua gambar diatas dapat disimpulkan bahwa hasil uji model dari TF-IDF diperoleh nilai 0.65 atau 65% Dan hasil uji model dari *Word2Vec* diperoleh nilai 0.43 atau 43%.

b. Evaluasi Model

Setelah dilakukan proses uji model, selanjutnya yaitu msuk pada tahap evaluasi model yang digunakan untuk menilai kesesuaian model dengan hasil klasifikasi. Pada tahap ini akan ditampilkan perbandingan hasil peforma model dari data yang didapat oleh pembobotan TF-IDF dan *Word2Vec*. Hasil peforma tersebut didapat dari dilakukannya perhitungan *confusion matrix* 3x3. Nilai peforma diukur dari hasil perhitungan nilai *accuracy*, *precision*, *recall* dan *F1_Score*.

1. *Accuracy*

Akurasi mengukur seberapa banyak prediksi yang benar dari seluruh prediksi yang dibuat oleh model. Berikut rumus *accuracy* seperti dibawah ini

$$Akurasi = \frac{T\ Neg + T\ Net + T\ Pos}{Total\ data\ yang\ diuji} \times 100\%$$

2. *Precision*

Presisi mengukur *akurasi* dari prediksi positif untuk setiap kelas. Berikut rumus *precision* seperti dibawah ini

$$Presisi_i = \frac{TP_i}{TP_i + FP_i}$$

3. *Recall*

Recall mengukur seberapa baik model mendeteksi semua label positif yang sebenarnya untuk setiap kelas. Berikut rumus *recall* seperti dibawah ini

$$Recall_i = \frac{TP_i}{TP_i + FN_i}$$

4. *F1_Score*

F1_Score adalah metrix yang menggabungkan *presisi* dan *recall*, dan memberikan keseimbangan antara keduanya. Berikut rumus *F1_Score* seperti di bawah ini

$$F1\_Score_i = \frac{2 \times Presisi_i \times Recall_i}{Presisi_i + Recall_i}$$

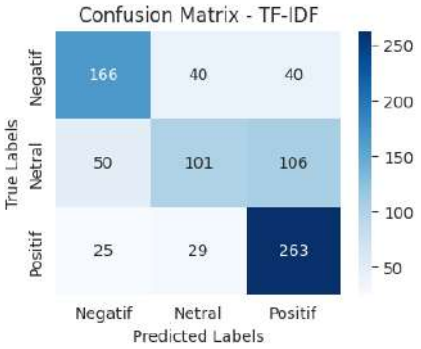
5. Weighted Average

Weighted Average yaitu menghitung rata-rata dengan cara mengalikan setiap nilai dengan bobot, menjumlahkan hasilnya kemudian membaginya dengan total bobot. Berikut rumus *Weighted Average sebagai berikut.*

$$Weighted\ Average = \frac{\sum(nilai \times support)}{\sum support}$$

Multiclass confusion matrix 3x3 merupakan alat evaluasi yang menunjukkan jumlah prediksi yang benar dan salah dalam setiap kelas dalam model klasifikasi dalam konteks *matrix* 3x3 terdapat tiga

kelas berbeda, setiap elemen dalam *matrix* mewakili jumlah prediksi model. Berikut *multiclass confusion matrix* 3x3 TF-IDF pada gambar 4.41



Gambar 4.41 Multiclass Confusion Matrix 3x3 TF-IDF

Multiclass confusion matrix 3x3 diatas digunakan untuk melakukan perhitungan nilai *accuracy*, precission, *recall*, dan *F1_Score* berikut perhitungan manualnya:

a. Kelas positif:

TPpositif = 263

FPpositif= 106+40 =146

FNpositif=25+29=54

b. Kelas negatif

TPnegatif=166

FPnegatif=50+25=75

FNnegatif=40+40=80

c. Kelas netral

TPnetral=101

FPnetral=40+29=69

FNnetral=50+106=156

d. Penilaian kelas positif

Presisi

$$Presisi = \frac{TP}{TP + FP} = \frac{263}{263 + 146} = \frac{263}{409} = 0{,}64$$

Recall

$$Recall = \frac{TP}{TP + FN} = \frac{263}{263 + 54} = \frac{263}{317} = 0{,}83$$

F1_Score

$$F1\_Score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} = \frac{2 \times 0{,}64 \times 0{,}83}{0{,}64 + 0{,}83}$$

$$= \frac{1{,}0624}{1{,}47} = 0{,}72$$

e. Penilaian kelas negatif

Presisi

$$Presisi = \frac{TP}{TP + FP} = \frac{166}{166 + 75} = \frac{166}{241} = 0{,}69$$

Recall

$$Recall = \frac{TP}{TP + FN} = \frac{166}{166 + 80} = \frac{166}{246} = 0{,}67$$

F1_Score

$$F1\_Score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} = \frac{2 \times 0{,}69 \times 0{,}67}{0{,}69 + 0{,}67}$$

$$= \frac{0{,}9246}{1{,}36} = 0{,}68$$

f. Penilaian kelas netral

Presisi

$$Presisi = \frac{TP}{TP + FP} = \frac{101}{101 + 69} = \frac{101}{170} = 0{,}59$$

Recall

$$Recall = \frac{TP}{TP + FN} = \frac{101}{101 + 156} = \frac{101}{257} = 0{,}39$$

F1_Score

$$F1\_Score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} = \frac{2 \times 0{,}59 \times 0{,}39}{0{,}59 + 0{,}39}$$

$$= \frac{0{,}4602}{0{,}98} = 0{,}47$$

g. $Weighted\ Presisi = \frac{\sum(nilai \times support)}{\sum support}$

$$= \frac{(0.69 \times 246) + (0.59 \times 257) + (0.64 \times 317)}{246 + 257 + 317} =$$

$$\frac{169.74 + 151.63 + 202.88}{820} = \frac{524.25}{820} = 0.64$$

h. $Weighted\ Recall = \frac{\sum(nilai \times support)}{\sum support}$

$$= \frac{(0.67\times246)+(0.39\times257)+(0.83\times317)}{246+257+317} =$$

$$\frac{164.82+100.23+263.11}{820} = \frac{528.16}{820} = 0.65$$

i. $Weighted\ F1\_Score = \frac{\sum(nilai \times support)}{\sum support}$

$$= \frac{(0.68\times246)+(0.47\times257)+(0.72\times317)}{246+257+317} =$$

$$\frac{167+120.79+228.24}{820} = \frac{516.31}{820} = 0.63$$

Berdasarkan perhitungan diatas diketahui peforma dari *naïve bayes* dengan metode pembobotan TF-IDF memperoleh *accuracy*, *presisi*, *recall*, dan *F1_Score*. Adapun perhitungan yang dilakukan sistem dapat dilihat pada gambar 4.42

TF-IDF Accuracy: 0.6463414634146342

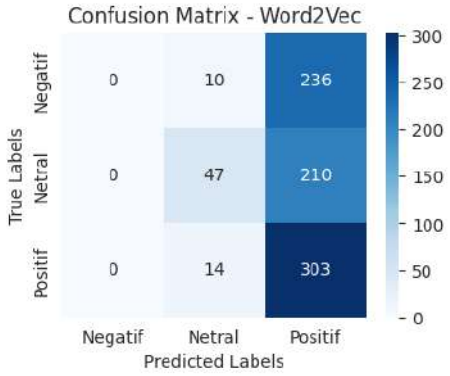
	precision	recall	f1-score	support
Negatif	0.69	0.67	0.68	246
Netral	0.59	0.39	0.47	257
Positif	0.64	0.83	0.72	317
accuracy			0.65	820
macro avg	0.64	0.63	0.63	820
weighted avg	0.64	0.65	0.63	820

Gambar 4.42 Perhitungan Sistem TF-IDF

Dari gambar 4.42 Diatas dapat disimpulkan bahwa hasil perhitungan *accuracy*, *precision*, *recall*, *F1_Score* metode pembobotan TF-IDF secara manual dengan hasil sistem terbukti sama dan didapatkan nilai accurasi dari metode pembobotan TF-IDF sebesar 0,65 atau 65%, *precision* dari kelas negatif sebesar 0,69 atau 69%, netral sebesar 0,59 atau 59% dan positif sebesar 0,64 atau 64%. Kemudian *recall* dari kelas negatif sebesar 0,67 atau 67%, netral sebesar 0,39 atau 39% dan positif sebesar 0,83 atau 83%. Kemudian *F1_Score* dari kelas negatif sebesar 0,68 atau 68%, netral sebesar 0,47 atau 47% dan positif sebesar 0,72 atau 72%, sehingga diperoleh total keseluruhan dari nilai *precision*, *recall*, dan *F1_Score* pada gambar 4.2 Didapatkan nilai *precision* sebesar 0,64 atau 64%, *recall* sebesar 0,65 atau 65%, dan *F1_Score* sebesar 0,63 atau 63%. Didapatkan hasil bahwa pada pengujian menggunakan metode TF-IDF sudah mendapatkan hasil yang baik terbukti dengan rata-rata hasil diatas dari 50%.

Selanjutnya perhitungan peforma algoritma *naïve bayes* dengan metode pembobotan *Word2Vec* dilakukan untuk menghitung nilai *accuracy*, *recall*, *presisi*, dan *F1_Score.* Berdasarkan perhitungan dari

multiclass confusion matrix 3x3. Berikut gambar 4.43 Mulrikelas *Confusion matrix* 3x3 *Word2Vec*



Gambar 4.43 Multiclass Confusion Matrix 3x3 Word2Vec

Multiclass confusion matrix 3x3 diatas digunakan untuk melakukan perhitungan nilai *accuracy*, precission, *recall*, dan *F1_Score* berikut perhitungan manualnya:

Kelas positif:

a. TPpositif = 303

   FPpositif= 236+210=446

   FNpositif=0+14=14

b. Kelas negatif

   TPnegatif=0

   FPnegatif=0+0=0

   FNnegatif=10+236=246

c. Kelas netral

TPnetral=47

FPnetral=10+14=24

FNnetral=0+210=210

d. Penilaian kelas positif

Presisi

$$Presisi = \frac{TP}{TP + FP} = \frac{303}{303 + 446} = \frac{303}{749} = 0{,}40$$

Recall

$$Recall = \frac{TP}{TP + FN} = \frac{303}{303 + 14} = \frac{303}{317} = 0{,}96$$

F1_Score

$$F1\_Score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} = \frac{2 \times 0{,}40 \times 0{,}96}{0{,}40 + 0{,}96}$$

$$= \frac{0{,}768}{1{,}36} = 0{,}57$$

e. Penilaian kelas negatif

Presisi

$$Presisi = \frac{TP}{TP + FP} = \frac{0}{0 + 0} = \frac{0}{0} = 0$$

Recall

$$Recall = \frac{TP}{TP + FN} = \frac{0}{0 + 246} = \frac{0}{246} = 0$$

F1_Score

$$F1\_Score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} = \frac{2 \times 0 \times 0}{0 + 0} = \frac{0}{0} = 0$$

f. Penilaian kelas netral

Presisi

$$Presisi = \frac{TP}{TP + FP} = \frac{47}{47 + 24} = \frac{47}{71} = 0{,}66$$

Recall

$$Recall = \frac{TP}{TP + FN} = \frac{47}{47 + 210} = \frac{47}{257} = 0{,}18$$

F1_Score

$$F1\_Score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} = \frac{2 \times 0{,}66 \times 0{,}18}{0{,}66 + 0{,}18}$$

$$= \frac{0{,}2376}{0{,}84} = 0{,}29$$

g. $Weighted\ Presisi = \frac{\sum(nilai \times support)}{\sum support}$

$$= \frac{(0.00 \times 246) + (0.66 \times 257) + (0.40 \times 317)}{246 + 257 + 317} =$$

$$\frac{(0) + (169.62) + (126.8)}{820} = \frac{296.42}{820} = 0.36$$

h. $Weighted\ Recall = \frac{\sum(nilai \times support)}{\sum support}$

$$= \frac{(0.00 \times 246) + (0.18 \times 257) + (0.96 \times 317)}{246 + 257 + 317} =$$

$$\frac{(0) + (46.26) + (304.32)}{820} = \frac{350.58}{820} = 0.43$$

i. $Weighted\ F1\_Score = \frac{\sum(nilai \times support)}{\sum support}$

$$= \frac{(0.00 \times 246)+(0.29 \times 257)+(0.57 \times 317)}{246+257+317} =$$

$$\frac{(0)+(74.53)+(180.69)}{820} = \frac{255.22}{820} = 0.31$$

Berdasarkan perhitungan diatas diketahui peforma dari *naïve bayes* dengan metode pembobotan *Word2Vec* memperoleh *accuracy*, *presisi*, *recall*, dan *F1_Score*. Adapun perhitungan yang dilakukan sistem dapat dilihat pada gambar 4.44

```
Word2Vec (Scaled) Accuracy: 0.4268292682926829

              precision    recall  f1-score   support

      Negatif       0.00      0.00      0.00       246
       Netral       0.66      0.18      0.29       257
      Positif       0.40      0.96      0.57       317

     accuracy                           0.43       820
    macro avg       0.36      0.38      0.29       820
 weighted avg       0.36      0.43      0.31       820
```

Gambar 4.44 Perhitungan Sistem Word2Vec

Dari gambar 4.44 Diatas dapat disimpulkan bahwa hasil perhitungan *accuracy*, *precision*, *recall*, *F1_Score* metode pembobotan *Word2Vec* secara manual dengan hasil sistem terbukti sama dan didapatkan nilai accurasi dari metode pembobotan *Word2Vec* sebesar 0,43 atau 43%, *precision* dari kelas negatif sebesar 0,00 atau 0%, netral sebesar 0,66 atau 66% dan positif

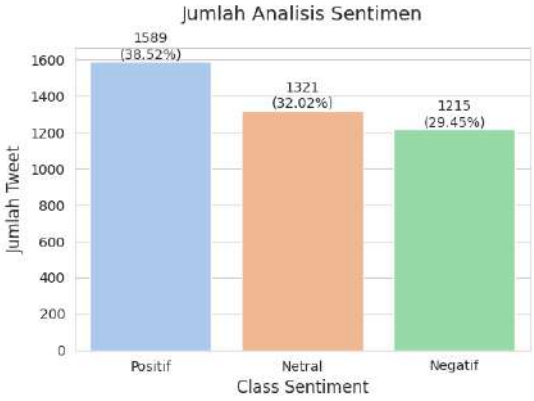
sebesar 0,40 atau 40%. Kemudian *recall* dari kelas negatif sebesar 0,00 atau 0%, netral sebesar 0,18 atau 18% dan positif sebesar 0,96 atau 96%. Kemudian *F1_Score* dari kelas negatif sebesar 0,00 atau 0%, netral sebesar 0,29 atau 29% dan positif sebesar 0,57 atau 57%, sehingga diperoleh total keseluruhan dari nilai *precision*, *recall*, dan *F1_Score* dari *Word2Vec* pada gambar 4.44 Didapatkan nilai *precision* sebesar 0,36 atau 36%, *recall* sebesar 0,43 atau 43%, dan *F1_Score* sebesar 0,31 atau 31%. Didapatkan bahwa pada pengujian menggunakan metode *Word2Vec* belum mendapatkan hasil yang baik terbukti dengan rata-rata hasil kurang dari 50%.

c. Visualisasi

Tahap terakhir yaitu memvisualisasikan hasil dari analisis sentimen menggunakan diagram dan *wordcloud*. *Wordcloud* akan digunakan untuk memvisualisasikan hasil yang diperoleh setelah dilakukannya proses analisis klasifikasi. Tujuan dari visualisasi ini yaitu untuk mengetahui hasil persentase yang diperoleh dari pembahasan analisis sentimen masyarakat terhadap ibu kota nusantara.

Diketahui sebelumnya hasil dari proses *prepocessing* menghasilkan data sebanyak 4125 dengan data 1589

merupakan sentimen positif, 1321 merupakan sentimen netral, dan 1215 merupakan sentimen negatif. Persetase data dapat dilihat pada gambar 4.45

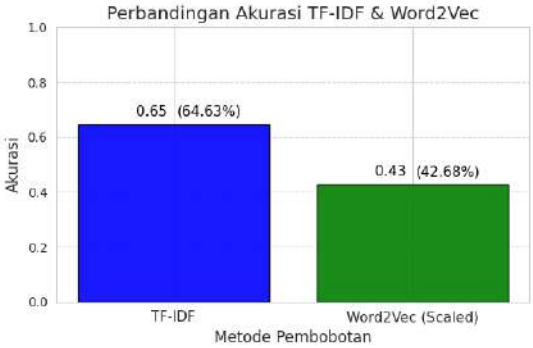


Gambar 4.45 Persentase Kelas Sentimen

Pada gambar 4.45 Dapat disimpulkan bahwa sentimen positif memiliki persentase tertinggi yaitu 38.52%, disusul dengan sentimen netral dengan persentase 32.02% dan terakhir di posisi ter rendah yaitu sentimen negatif dengan persentase 29.45%. walaupun selisih ketiga persentase sentimen terpaut sedikit namun dapat disimpulkan bahwa mayoritas masyarakat Indonesia masih mendukung kebijakan adanya ibu kota nusantara baru dibuktikan dengan baiknya sentimen positif. Sisanya sebanyak 32.02% berisikan tanggapan netral dan29.45% berisi kontra

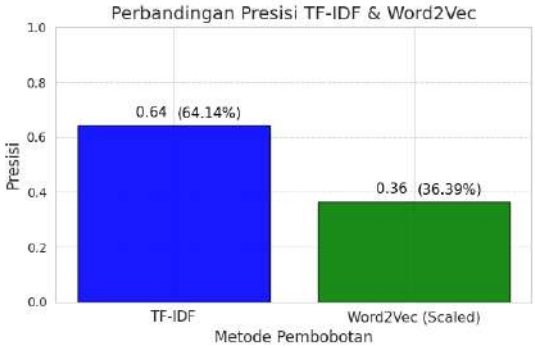
atau ketidaksetujuan adanya kebijakan ibu kota nusantara yang baru.

Selain visualisasi hasil sentimen peneliti juga ingin memperlihatkan visualisasi persentase perbandingan efisiensi metode pembobotan yaitu metode pembobotan TF-IDF dan *Word2Vec*. Walaupun kedua metode memiliki kebaikan dan kekurangan masing-masing namun dalam pengaplikasian dengan data sentimen masyarakat terhadap ibu kota nusantara didapatkan bahwa metode pembobotan TF-IDF lebih efisien digunakan dari pada metode pembobotan *Word2Vec* dengan hasil persentase *akurasi* metode pembobotan TF-IDF sebesar 64.63% dan persentase *akurasi* metode pembobotan *Word2Vec* sebesar 42.44%. visualisasi *akurasi* perbandingan metode pembobotan dapat dilihat pada gambar 4.46



Gambar 4.46 Persentase Akurasi

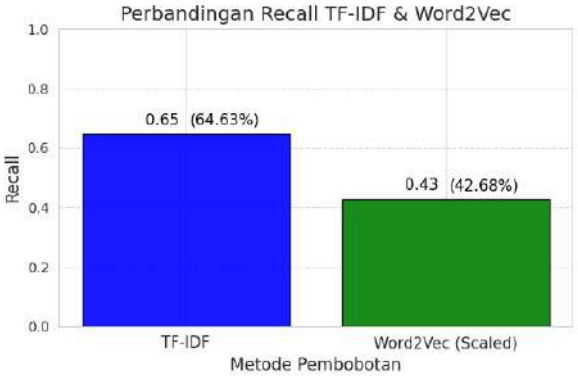
Pada gambar 4.46 Dapat dilihat bahwa *akurasi* metode TF-IDF lebih tinggi yaitu 64.63% dari pada metode *Word2Vec* dengan *akurasi* 42.68% dalam pengaplikasian studi kasus penilaian sentimen masyarakat terhadap ibu kota nusantara. TF-IDF telah menunjukkan performa lebih baik untuk data seperti analisis sentimen dibandingkan dengan metode lain dikarenakan TF-IDF efektif jika digunakan untuk analisis teks yang bersifat pendek dan langsung, serta fungsi TF-IDF sendiri yaitu penekanan pada kata kunci yang dapat menjadi indikator penting dalam sentimen sehingga meningkatkan *akurasi* klasifikasi.



Gambar 4.47 Persentase Presisi

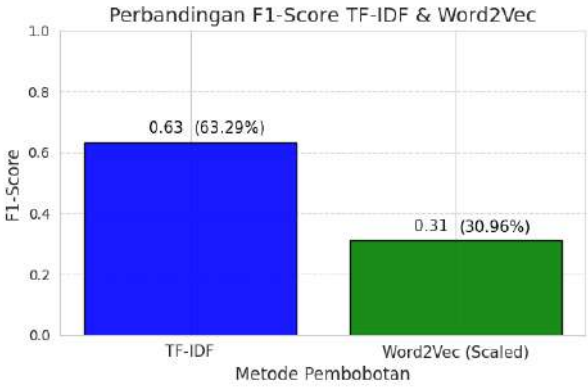
Pada gambar 4.47 dapat dilihat bahwa persentase *presisi* pembobotan TF-IDF lebih tinggi dari pada pembobotan *Word2Vec* yakni pembobotan TF-IDF

memiliki persentase *presisi* sebesar 64% dan pembobotan *Word2Vec* sebesar sebesar 36%. Dapat diartikan bahwa TF-IDF lebih baik dalam mendeteksi nilai *presisi*.



Gambar 4.48 Persentase Recall

Pada gambar 4.48 dapat dilihat bahwa persentase *recall* pembobotan TF-IDF lebih tinggi dari pada pembobotan *Word2Vec* yakni pembobotan TF-IDF memiliki persentase *recall* sebesar 65% dan pembobotan *Word2Vec* sebesar sebesar 43%. Dapat diartikan bahwa pembobotan TF-IDF lebih baik dalam mengukur nilai *recall* dari pada pembobotan *Word2Vec*.



Gambar 4.49 Persentase F1_Score

Pada gambar 4.49 dapat dilihat bahwa persentase *F1_Score* pembobotan TF-IDF lebih tinggi dari pada pembobotan *Word2Vec* yakni pembobotan TF-IDF memiliki persentase *F1_Score* sebesar 63% dan pembobotan *Word2Vec* sebesar sebesar 31%. Dapat diartikan bahwa nilai *F1_Score* TF-IDF lebih tinggi dari pada nilai *F1_Score Word2Vec*.

Selain itu peneliti juga ingin memperlihatkan kata-kata yang banyak dibicarakan oleh pengguna media sosial X di Indonesia mengenai ibu kota nusantara, berikut visualisasi *wordcloud* pada gambar 4.50

Gambar 4.50 Wordcloud IKN

Berdasarkan pada gambar 4.50 Dapat dilihat bahwa ada beberapa kata yang tercetak lebih besar dari pada yang lain, dengan artian kata yang tercetak lebih besar merupakan kata yang sangat sering di bincangkan dan menjadi kunci utama seperti ”IKN”, “bangun ikn” di susul dengan kata kunci “presiden prabowo”, “ikn ibukotanusantara” sebagai kata yang lumayan sering muncul. Selain itu ada beberapa kata lagi yang masuk dalam wordcloud seperti ‘jakarta’, ’rakyat’, ’jokowi’ dan masih banyak lainnya.

Selanjutnya peneliti ingin menunjukkan hasil dari visualisasi *wordcloud* sentimen positif yang tentunya berisi kata-kata yang pro atau setuju dengan pembahasan ibu kota nusantara. Berikut visualisasi *wordcloud* sentimen positif ditunjukkan pada gambar 4.51

Gambar 4.51 Wordcloud Positif

Dapat dilihat pada gambar 4.51 Bahwa kata-kata yang tertera pada *wordcloud* sentimen positif berisi kat-kata yang setuju pada pembahasan ibu kota nusantara seperti “presiden prabowo”, “bangun ikn, “ikn ibukotanusantara”, “ya” dan lain sebagainya.

Selanjutnya peneliti ingin memperlihatkan visualisasi *wordcloud* sentimen negatif tentunya berisi kata-kata kontra atau ketidak setujuan pada pembahasan ibu kota nusantara. Berikut visualisasi

wordcloud sentimen negatif pada gambar 4.52

Gambar 4.52 Wordcloud Negatif

Dapat dilihat pada gambar 4.52 Bahwa *wordcloud* berisi kata kata kontra atau ketidak setujuan masyarakat terhadap ibu kota nusantara sehingga kata-kata lebih ber variatif dan mungkin beberapa kurang pantas. Walaupun terkesan kurang baik namun itu semua adalah pendapat masyarakat mengenai permasalahan kenegaraan.

Yang terakhir, peneliti ingin memperlihatkan *wordcloud* sentimen netral dimana sentimen bersifat netral atau tidak memihak. Berikut visualisasi *wordcloud* sentimen netral ditunjukkan pada gambar 4. 53

Gambar 4.53 Word Cloud Netral

Dapat dilihat pada gambar 4.53 Kata yang paling tebal pada *wordcloud* sentimen netral hanya satu yaitu "ikn" dan selebihnya merata hanya kecil namun berfariatif, hal itu menunjukkan bahwa sebagian masyarakat hanya terfokus pada kata "ikn" dan mungkin hanya berisi informasi atau tidak memihak siapapun.

BAB V

SIMPULAN DAN SARAN

d. Kesimpulan

Berdasarkan penelitian yang telah dilaksanakan, dapat disimpulkan bahwa:

1. Dari penelitian ini didapatkan data sejumlah 5000 data sentimen mentah dari aplikasi X kemudian setelah melalui proses preprocessing menjadi 4125 data yang kemudian diproses dan menghasilkan sejumlah 1589 sentimen positif, 1321 sentimen netral dan 1215 sentimen negatif. Sehingga diketahui kelas sentimen positif memiliki nilai persentase tertinggi dengan 38.52% di peringkat pertama, selanjutnya yaitu kelas sentimen netral dengan 32.02% di peringkat kedua dan terakhir yaitu kelas sentimen negatif dengan 29.45% sebagai peringkat terakhir.
2. Berdasarkan hasil perbandingan performa dua metode pembobotan yaitu pembobotan TF-IDF dan *Word2Vec* dapat diartikan bahwa metode TF-IDF lebih efektif jika digunakan untuk analisis sentimen seperti analisis sentimen masyarakat terhadap ibu kota nusantara(IKN). Dengan hasil metode TF-IDF menghasilkan nilai *accuracy*

sebesar 65%, precision 64%, *recall* 65% dan *F1_Score* sebesar 63%. sedangkan pada pembobotan *Word2Vec* menghasilkan nilai *accuracy* sebesar 43%, precision 36%, *recall* 43% dan *F1_Score* 31%. hal ini membuktikan bahwa peforma metode TF-IDF lebih baik dari pada peforma metode *Word2Vec* untuk analisis sentimen.

e. Saran

Berdasarkan penelitian yang telah dilaksanakan, peneliti berharap kepada peneliti selanjutnya untuk dapat dikembangkan kembali dan terdapat beberapa saran dari peneliti yaitu:

1. Dapat menggunakan metode klasifikasi yang berbeda selain Naïve Bayes atau melakukan perbandingan dengan metode yang lain. Dengan begitu dapat membandingkan hasil untuk mencari peforma metode klasifikasi yang terbaik.
2. Dapat membandingkan metode pembobotan lain selain TF-IDF dan *Word2Vec* seperti Unigram dan Bigram, TF-CHI, TF-RF dan lainnya agar dapat mengetahui metode pembobotan yang lebih akurat jika digunakan untuk analisis sentimen.

DAFTAR PUSTAKA

Alfiani Mahardhika, Aisha, Ristu Saptono, and Rini Anggrainingsih. 2016. "Sistem Klasifikasi Feedback Pelanggan Dan Rekomendasi Solusi Atas Keluhan Di UPT Puskom UNS Dengan Algoritma *Naive bayes* Kelasifier Dan Cosine Similiarity." *Jurnal Teknologi & Informasi ITSmart* 4(1):36. doi: 10.20961/its.v4i1.1806.

Dani, Ahmad Hilman, Eva Yulia Puspaningrum, and Retno Mumpuni. 2024. "Studi Performa TF-IDF Dan *Word2Vec* Pada Analisis Sentimen Cyberbullying." *Router : Jurnal Teknik Informatika Dan Terapan* 2(2):94–106.

Darmawan, Rizky, and Safrina Amini. 2022. "Perbandingan Hasil Sentimen Analysis Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor Pada *Twitter* Comparison of Sentimen Analysis Results Using Naïve Bayes and K-Nearest Neighbor Algorithm on *Twitter*." *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Jakarta-Indonesia* (September):495–501.

Firdaus, Ali, and Wahyu Istalama Firdaus. 2021. "Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan)." *Jurnal JUPITER* 13(1):66.

Hendrawan Rifky, Ivan, Ema Utami, and Anggit Hartanto Dwi. 2022. "Analisis Perbandingan Metode Tf-Idf Dan *Word2Vec* Pada Klasifikasi Teks Sentimen Masyarakat Terhadap Produk Lokal Di Indonesia." *Smart Comp: Jurnalnya Orang Pintar Komputer* 11(3):497–503. doi: 10.30591/smartcomp.v11i3.3902.

Isa Albanna, and R. Tri hadi laksono. 2022. "Implementasi Pandas Data Frame Sebagai Agregasi Dan Tabulasi Penyajian Data Luaran Survei Kepuasan Pengguna Proses

Pembelajaran Dalam Pendidikan Tinggi." *Seminar Nasional Sains Dan Teknologi Terapan X 2022* 1–6.

Kencana, I. Kadek Bayu Arys Wisnu, and Warih Maharani. 2017. "Klasifikasi Opini Pada Fitur Produk Berbasis Graph." *E-Proceeding of Engineering* 4(2):3148.

Mahrozi, Nanang, and Muhammad Faisal. 2023. "Analisis Perbandingan Kecepatan Algoritma Selection Sort Dan Bubble Sort." *Jurnal Ilmiah Sain Dan Teknologi* 1(2):89–98.

Maulana, Aganda, Dan Dyantono, and Ricky Eka Putra. 2023. "Perbandingan Sent2vec TF-IDF Logistic Regression Dan *Word2Vec* CNN Pada Hasil Sentimen Analysis Youtube Comment." *Journal of Informatics and Computer Science* 05:63–72.

Mustofa, Hery, and Adzhal Arwani Mahfudh. 2019. "Klasifikasi Berita Hoax Dengan Menggunakan Metode *Naive bayes*." *Walisongo Journal of Information Technology* 1(1):1. doi: 10.21580/wjit.2019.1.1.3915.

Mutiara, Zaskya, Boham Antonius, and Jackelin Hera Lotulung Leviane. 2020. "*Twitter* Sebagai Media Mengungkapkan Diri Pada Kalangan Milenial." *Fakultas Ilmu Sosial Dan Politik, Universitas Sam Ratulangi* 1–8.

Nugraha, Damar, Puji Astuti, Program Studi Informatika, Universitas Nusa Mandiri, Program Studi Informatika, and Universitas Nusa Mandiri. 2023. "ANALISIS SENTIMEN CYBERBULLYING PADA SOSIAL MEDIA INSTAGRAM MENGGUNAKAN METODE SUPPORT VECTOR MACHINE." 8(2):153–64.

Prabowo, Yulius Denny, Tedi Lesmana Marselino, and Meylisa Suryawiguna. 2019. "Pembentukan Vector Space Model Bahasa Indonesia Menggunakan Metode Word to Vector."

Jurnal Buana Informatika 10(1):29–40. doi: 10.24002/jbi.v10i1.2053.

Riadi Silitonga, Yosua. 2019. "Sistem Pendeteksi Berita Hoax Di Media Sosial Dengan Teknik Data Mining Scikit Learn." *Jurnal Ilmu Komputer* 4:173.

Saputra, Pramana Yoga. 2017. "Implementasi Teknik Crawling Untuk Pengumpulan Data Dari Media Sosial *Twitter*." *Jurnal Dinamika* 8(2):160–68.

Septiani, Dwi, and Ica Isabela. 2023. "Analisis Term Frequency Inverse Document Frequency (TF-IDF) Dalam Temu Kembali Informasi Pada Dokumen Teks." *SINTESIA: Jurnal Sistem Dan Teknologi Informasi Indonesia* 1(2):81–88.

Sitio, Ginni Yema, Agustina Rumapea, and Drs Posma Lumbanraja. 2023. "Analisis Sentimen Pemindahan Ibu Kota Negara Di Media Sosial *Twitter* Menggunakan Metode Convolutional Neural Network (CNN)." *Methotika : Jurnal Ilmiah Teknik Informatika* 3(2):97–104.

Turmudi Zy, Ahmad, Lutfi Adji Ardiansyah, and Donny Maulana. 2021. "Implementasi Algoritma Naïve Bayes Dalam Mendiagnosa Penyakit Angin Duduk." *Jurnal Pelita Teknologi* 16(1):52–65.

Untari, Dwi. 2018. "Data Mining Untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree C4.5." *Fakultas Ilmu Komputer Universitas Dian Nuswantoro* 2(November):31–48.

Wandani, Aprilia. 2021. "Sentimen Analisis Pengguna *Twitter* Pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, Dan *Naive bayes*." *Jurnal Sains Komputer & Informatika (J-SAKTI* 5(2):651–65.

Yanuar, Roni Andarsyah; Amri. 2024. “Jurnal Teknik Informatika, Vol. 16, No. 2, April 2024.” 16(2):1–7.

Yuniarti, Wenty Dwi. 2019. *DASAR- DASAR PEMOGRAMAN DENGAN PYTHON Python Merupakan Salah Satu Bahasa Pemograman Yang Bersifat Open Source, Menyediakan Dukungan Untuk Pengelolaan*.

Zaki, Ahmad. 2021. “Penerapan Metode Bayes Dalam Prediksi Segementasi Pasar Penjualan Smartphone.” *Journal Computer Science and Informatic Sistems : J-Cosys* 1(1):40–45. doi: 10.53514/jc.v1i1.15.

Alfiani Mahardhika, Aisha, Ristu Saptono, and Rini Anggrainingsih. 2016. “Sistem Klasifikasi Feedback Pelanggan Dan Rekomendasi Solusi Atas Keluhan Di UPT Puskom UNS Dengan Algoritma *Naive bayes* Kelasifier Dan Cosine Similiarity.” *Jurnal Teknologi & Informasi ITSmart* 4(1):36. doi: 10.20961/its.v4i1.1806.

Dani, Ahmad Hilman, Eva Yulia Puspaningrum, and Retno Mumpuni. 2024. “Studi Performa TF-IDF Dan *Word2Vec* Pada Analisis Sentimen Cyberbullying.” *Router : Jurnal Teknik Informatika Dan Terapan* 2(2):94–106.

Darmawan, Rizky, and Safrina Amini. 2022. “Perbandingan Hasil Sentimen Analysis Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor Pada *Twitter* Comparison of Sentimen Analysis Results Using Naïve Bayes and K-Nearest Neighbor Algorithm on *Twitter*.” *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Jakarta-Indonesia* (September):495–501.

Firdaus, Ali, and Wahyu Istalama Firdaus. 2021. “Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan).” *Jurnal JUPITER* 13(1):66.

Hendrawan Rifky, Ivan, Ema Utami, and Anggit Hartanto Dwi.

2022. “Analisis Perbandingan Metode Tf-Idf Dan *Word2Vec* Pada Klasifikasi Teks Sentimen Masyarakat Terhadap Produk Lokal Di Indonesia.” *Smart Comp: Jurnalnya Orang Pintar Komputer* 11(3):497–503. doi: 10.30591/smartcomp.v11i3.3902.

Isa Albanna, and R. Tri hadi laksono. 2022. “Implementasi Pandas Data Frame Sebagai Agregasi Dan Tabulasi Penyajian Data Luaran Survei Kepuasan Pengguna Proses Pembelajaran Dalam Pendidikan Tinggi.” *Seminar Nasional Sains Dan Teknologi Terapan X 2022* 1–6.

Kencana, I. Kadek Bayu Arys Wisnu, and Warih Maharani. 2017. “Klasifikasi Opini Pada Fitur Produk Berbasis Graph.” *E-Proceeding of Engineering* 4(2):3148.

Mahrozi, Nanang, and Muhammad Faisal. 2023. “Analisis Perbandingan Kecepatan Algoritma Selection Sort Dan Bubble Sort.” *Jurnal Ilmiah Sain Dan Teknologi* 1(2):89–98.

Maulana, Aganda, Dan Dyantono, and Ricky Eka Putra. 2023. “Perbandingan Sent2vec TF-IDF Logistic Regression Dan *Word2Vec* CNN Pada Hasil Sentimen Analysis Youtube Comment.” *Journal of Informatics and Computer Science* 05:63–72.

Mustofa, Hery, and Adzhal Arwani Mahfudh. 2019. “Klasifikasi Berita Hoax Dengan Menggunakan Metode *Naive bayes*.” *Walisongo Journal of Information Technology* 1(1):1. doi: 10.21580/wjit.2019.1.1.3915.

Mutiara, Zaskya, Boham Antonius, and Jackelin Hera Lotulung Leviane. 2020. “*Twitter* Sebagai Media Mengungkapkan Diri Pada Kalangan Milenial.” *Fakultas Ilmu Sosial Dan Politik, Universitas Sam Ratulangi* 1–8.

Nugraha, Damar, Puji Astuti, Program Studi Informatika,

Universitas Nusa Mandiri, Program Studi Informatika, and Universitas Nusa Mandiri. 2023. "ANALISIS SENTIMEN CYBERBULLYING PADA SOSIAL MEDIA INSTAGRAM MENGGUNAKAN METODE SUPPORT VECTOR MACHINE." 8(2):153–64.

Prabowo, Yulius Denny, Tedi Lesmana Marselino, and Meylisa Suryawiguna. 2019. "Pembentukan Vector Space Model Bahasa Indonesia Menggunakan Metode Word to Vector." *Jurnal Buana Informatika* 10(1):29–40. doi: 10.24002/jbi.v10i1.2053.

Riadi Silitonga, Yosua. 2019. "Sistem Pendeteksi Berita Hoax Di Media Sosial Dengan Teknik Data Mining Scikit Learn." *Jurnal Ilmu Komputer* 4:173.

Saputra, Pramana Yoga. 2017. "Implementasi Teknik Crawling Untuk Pengumpulan Data Dari Media Sosial *Twitter*." *Jurnal Dinamika* 8(2):160–68.

Septiani, Dwi, and Ica Isabela. 2023. "Analisis Term Frequency Inverse Document Frequency (TF-IDF) Dalam Temu Kembali Informasi Pada Dokumen Teks." *SINTESIA: Jurnal Sistem Dan Teknologi Informasi Indonesia* 1(2):81–88.

Sitio, Ginni Yema, Agustina Rumapea, and Drs Posma Lumbanraja. 2023. "Analisis Sentimen Pemindahan Ibu Kota Negara Di Media Sosial *Twitter* Menggunakan Metode Convolutional Neural Network (CNN)." *Methotika : Jurnal Ilmiah Teknik Informatika* 3(2):97–104.

Turmudi Zy, Ahmad, Lutfi Adji Ardiansyah, and Donny Maulana. 2021. "Implementasi Algoritma Naïve Bayes Dalam Mendiagnosa Penyakit Angin Duduk." *Jurnal Pelita Teknologi* 16(1):52–65.

Untari, Dwi. 2018. "Data Mining Untuk Menganalisa Prediksi

Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree C4.5." *Fakultas Ilmu Komputer Universitas Dian Nuswantoro* 2(November):31–48.

Wandani, Aprilia. 2021. "Sentimen Analisis Pengguna *Twitter* Pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, Dan *Naive bayes*." *Jurnal Sains Komputer & Informatika (J-SAKTI* 5(2):651–65.

Yanuar, Roni Andarsyah; Amri. 2024. "Jurnal Teknik Informatika, Vol. 16, No. 2, April 2024." 16(2):1–7.

Yuniarti, Wenty Dwi. 2019. *DASAR- DASAR PEMOGRAMAN DENGAN PYTHON Python Merupakan Salah Satu Bahasa Pemograman Yang Bersifat Open Source, Menyediakan Dukungan Untuk Pengelolaan*.

Zaki, Ahmad. 2021. "Penerapan Metode Bayes Dalam Prediksi Segementasi Pasar Penjualan Smartphone." *Journal Computer Science and Informatic Sistems : J-Cosys* 1(1):40–45. doi: 10.53514/jc.v1i1.15.

DAFTAR LAMPIRAN

Lampiran 1 : Hasil Pengumpulan Data

No.	username	Sentimen
1	harkatmartabak	Presiden Prabowo berkomitmen untuk membangunan IKN yg tidak sekadar memindahkan pusat pemerintahan tetapi juga merancang masa depan Indonesia dengan menjadikan IKN sebagai kota modern ramah lingkungan dan nyaman huni. ___ Agus Wulan Guritno Bahlil Pertamax #jakartabo Nasi https://t.co/orTzOsGYNO
2	RKianSantang99	@BosPurwa Fix IKN Mangkrak dan Jokodok Bajitol Lepas tangan Ga Mau Tanggung jawab&diSalahin. Malah Rakyat yg harus TanggungJawab. Itulah Presiden PKI 2Periode bagi Kaum Idiot 58% yg Dimenangkan Secara Curang&Penuh TipuDaya. Blm Anaknya SiFufuFafa
3	jauhann_	Holy shit rajjah IKN is here
4	dragoeinside	@RajaJuliAntoni Jualnya ikn gmna bro? Laku?
5	EdwardSimb38933	@Opposite6888 Siapa bilang 1 VS 15?? Memang Anies itu

		Siapa??Pramono itu Siapa? kalau KIM mau bisa saja Semuanya diambil tanpa tersisa Ibu Kota Ind nantinya bukan Dki Jakarta ttp IKN. artinya siapa yg menjadi Gubernur disana yg akan menjadi Calon Presiden
6	edyponidei	@OposisiCerdas IKN gmn?
7	AliBej0	IKN tak sekadar memindahkan pusat pemerintahan tetapi juga merancang masa depan Indonesia dengan mengintegrasikan teknologi canggih dan budaya lokal.
…	…..	….
4999	pamungkasbudhi	@adrieutama @footballinanews Ya kan yg mau bangun maunya stadion di IKN bang. Bukan di daerah lain
5000	sniper7787	@niwseir ini makanya banyak crazy rich di Kalimantan ga suka ama IKN. Yang gini2 bakal kena radar pemerintah kalau pemerintahan ud di Kalimantan. Dan mereka bakal bersaing dengan orang2 baru dan yang lebih pintar.

Lampiran 2 : User Samaran

No.	User	Sentimen
1	Dokumen 1	presiden prabowo komitmen bangun ikn pindah pusat perintah rancang indonesia jadi ikn kota modern ramah lingkung nyaman huni agus wulan guritno bahlil pertamax jakartabo nasi
2	Dokumen 2	fix ikn mangkrak jokodok bajitol lepas tangan tanggung jawabampdisalahin rakyat tanggungjawab presiden pki periode kaum idiot menang curangamppenuh tipudaya anak sifufufafa
3	Dokumen 3	bilang vs anies siapapramono kim ambil sisa kota ind dki jakarta ikn gubernur sana calon presiden
4	Dokumen 4	ikn pindah pusat perintah rancang indonesia integrasi teknologi canggih budaya lokal
5	Dokumen 5	uang asih upeti lancar urus lihat naga kemarin suruh kumpul ikn tagih bangun nyetor upeti gede pilkada kemarin menang
6	Dokumen 6	ikn kota baru
7	Dokumen 7	presiden prabowo proyek ikn selesai
….		…..
4124	Dokumen 4124	lapor telepon cakeda tagih janji pindah ikn
4125	Dokumen 4125	lapor bos kantor ikn

Lampiran 3 : Pelabelan Lexicon

No			
1.	presiden prabowo komitmen bangun ikn pindah pusat perintah rancang indonesia jadi ikn kota modern ramah lingkung nyaman huni agus wulan guritno bahlil pertamax jakartabo nasi	3	Positif
2.	fix ikn mangkrak jokodok bajitol lepas tangan tanggung jawabampdisalahin rakyat tanggungjawab presiden pki periode kaum idiot menang curangamppenuh tipudaya anak sifufufafa	-2	Negative
3.	holy shit rajjah ikn is here	0	Netral
4.	jual ikn bro laku	1	Positif
5.	bilang vs anies siapapramono kim ambil sisa kota ind dki jakarta ikn gubernur sana calon presiden	-2	Negative
6.	ikn	0	Netral
7.	ikn pindah pusat perintah rancang indonesia integrasi teknologi canggih budaya lokal	2	Positif
…	….	…	…
4124	pt ppi barusan pos pt ppi anak usaha bumn idfood pt rajawali nusantara indonesia tau deh anggap swasta	-2	Negative
4125	pt industri gula nusantara ign impor thailand india total impor capai ton	0	Netral

Lampiran 4: Pelabelan Manual

No.	Sentimen	kelass
1	presiden prabowo komitmen bangun ikn pindah pusat perintah rancang indonesia jadi ikn kota modern ramah lingkung nyaman huni agus wulan guritno bahlil pertamax jakartabo nasi	Positif
2	fix ikn mangkrak jokodok bajitol lepas tangan tanggung jawabampdisalahin rakyat tanggungjawab presiden pki periode kaum idiot menang curangamppenuh tipudaya anak sifufufafa	Negatif
3	bilang vs anies siapapramono kim ambil sisa kota ind dki jakarta ikn gubernur sana calon presiden	Positif
4	ikn pindah pusat perintah rancang indonesia integrasi teknologi canggih budaya lokal	Positif
5	uang asih upeti lancar urus lihat naga kemarin suruh kumpul ikn tagih bangun nyetor upeti gede pilkada kemarin menang	Negatif
6	ikn kota baru	Netral
7	presiden prabowo proyek ikn selesai	Positif
….	…..	…..
4124	lapor telepon cakeda tagih janji pindah ikn	Positif
4125	lapor bos kantor ikn	Netral

Lampiran 5 : kamus Kata Baku

No.	tidak_baku	kata_baku
1	woww	wow
2	aminn	amin
3	met	selamat
4	netaas	menetas
5	keberpa	keberapa
6	eeeehhhh	eh
7	kata2nyaaa	kata-katanya
8	hallo	halo
9	kaka	kakak
10	ka	kak
11	daah	dah
12	aaaaahhhh	ah
13	yaa	ya
14	smga	semoga
15	slalu	selalu
16	amiin	amin
17	kk	kakak
18	trus	terus
19	kk	kakak
20	sii	sih
21	nyenengin	menyenangkan
22	bgt	banget
23	gemess	gemas
24	akuuu	aku
25	jgn	jangan
…	…	…
15183	udah	sudah
15184	gitu	begitu
15185	aja	saja

Lampiran 6 : Kata Positif

word weight
hai
merekam
ekstensif
paripurna
detail
pernik
belas
welas
kabung
rahayu
maaf
hello
promo
terimakasih
cover
mohon
mengawal
statistik
keluangan
jalan terbuka
banyaknya
lebar
bentang
hendaknya
silahkan
semboyan
ditunggu
akses
penerangan
....
dibantu
makasih

Lampiran 7 : Kata Negatif

word weight
putus tali gantung
gelebah
gobar hati
tersentuh (perasaan)
isak
larat hati
nelangsa
remuk redam
tidak segan
gemar
tak segan
sesal
pengen
penghayatan
absorpsi
linu
salah benang
sakit
lara
zuhud
mencederai
mengingkari
maaf
mengkhianat
mencelakai
mulu
ngga
borong
lever
....
gamau
doang

RIWAYAT HIDUP

A. Identitas Diri

1. Nama Lengkap	: Nisfa Laili Fikria
2. Tempat & Tanggal Lahir	: Demak, 17 September 2002
3. Alamat Rumah	: Desa Sari RT.08/RW.02, Kec. Gajah, Kab. Demak
4. Hp	: 085695489100
5. Email	: nisfalalili04@gmail.com

B. Riwayat Pendidikan

1. SDN Sari 1
2. SMP N 1 Gajah
3. SMA N 2 Demak

Semarang, 30 Februari 2025

Nisfah Laili Fikria
NIM. 2008096007