**Tugas Akhir**

**Analisis Sentimen Pandang Audiens Youtube terhadap video tentang berita pembangunan IKN**



**Disusun oleh:**

|  |  |
| --- | --- |
| Zaenal Abidin Syah | 210411100186 |
|  |  |
|  |  |
|  |  |

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS TRUNOJOYO MADURA**

**2024**

# BAB I Pendahuluan

## **Latar Belakang**

Pembangunan Ibu Kota Negara (IKN) semakin populer di platform media sosial seperti Youtube. Youtube, salah satu platform berbagi video terbesar di dunia, memberikan kesempatan kepada pemirsa untuk mengekspresikan pandangan, komentar, dan pendapat mereka tentang topik tertentu. Analisis sentimen terhadap komentar-komentar tersebut memungkinkan kita mengetahui bagaimana pandangan masyarakat umum terhadap perkembangan IKN.

Penelitian ini menggunakan teknik web mining untuk meng-crawl data komentar melalui API klien Google. Data yang diterima meliputi penulis, publikasi\_at, komentar, dan jumlah suka. Kami kemudian melakukan analisis sentimen menggunakan teknik Gated Recurrent Unit (GRU), yang merupakan jenis arsitektur jaringan saraf berulang (RNN). Pendekatan ini diharapkan memungkinkan kami mengidentifikasi sentimen positif, negatif, atau netral dari komentar audiens berdasarkan hasil analisis.

Sebuah studi baru-baru ini mengkaji sentimen masyarakat terhadap rencana pemindahan ibu kota Indonesia ke Kalimantan Timur. Studi lain yang menggunakan Sebuah penelitian menggunakan SMOTE Naive Bayes mencapai akurasi seimbang sebesar 76,01% dalam mengklasifikasikan emosi menjadi positif, netral, atau negatif (Huwaida et al., 2024). Namun kekhawatiran mengenai dampak lingkungan dan ketidakpastian juga telah diidentifikasi (Munawaroh dkk., 2024). Menganalisis data media sosial seperti YouTube, Twitter, Instagram, dan Facebook memberikan wawasan mengenai opini masyarakat bahwa kegembiraan adalah emosi yang dominan (Munawaroh et al., 2024). Hasil-hasil ini memberikan data berharga bagi para pengambil keputusan dan menyoroti pentingnya mempertimbangkan opini publik ketika mengambil keputusan kebijakan.

## Rumusan Masalah

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah:

* Bagaimana cara melakukan crawling data komentar YouTube menggunakan Google Client API?
* Bagaimana penerapan metode GRU untuk analisis sentimen komentar audiens?
* Seberapa akurat model GRU dalam mengklasifikasikan sentimen komentar YouTube?

## Tujuan Penelitian

Penelitian ini bertujuan untuk:

* Mengambil data komentar audiens dari video YouTube menggunakan Google Client API.
* Menerapkan metode GRU untuk melakukan analisis sentimen terhadap data komentar.
* Mengevaluasi kinerja model GRU dalam melakukan klasifikasi sentimen.

## **Manfaat Penelitian**

## Bagi Peneliti: Menambah pemahaman dan keterampilan dalam implementasi web mining serta analisis sentimen menggunakan GRU.

## Bagi Masyarakat: Memberikan gambaran tentang pandangan audiens YouTube terhadap pembangunan IKN.

## Bagi Akademisi: Sebagai referensi dalam penelitian terkait analisis sentimen dan metode GRU.

# BAB II TINJAUAN PUSTAKA

## Web Mining

Web mining adalah proses yang kompleks dan beragam yang bertujuan untuk mengekstraksi informasi atau pola dari data yang tersedia di internet. Proses ini melibatkan penggunaan teknik dan algoritma untuk menganalisis data yang tidak terstruktur, seperti teks, gambar, dan video, yang dapat ditemukan di berbagai situs web. Dalam konteks penelitian ini, kami memanfaatkan Google Client API untuk melakukan crawling data komentar di platform YouTube. Dengan menggunakan API ini, kami dapat mengumpulkan data komentar dari berbagai video, yang kemudian dapat dianalisis lebih lanjut untuk mendapatkan wawasan yang berharga. Proses crawling ini memungkinkan kami untuk mengakses dan mengumpulkan data dalam jumlah besar, yang sangat penting untuk analisis yang mendalam dan komprehensif.

## Analisis Sentimen

Analisis sentimen adalah proses yang bertujuan untuk mengidentifikasi, mengevaluasi, dan mengklasifikasikan opini yang terkandung dalam teks menjadi kategori tertentu, seperti positif, negatif, atau netral. Proses ini sangat penting dalam memahami bagaimana orang merespons suatu produk, layanan, atau isu tertentu. Dalam konteks penelitian ini, analisis sentimen digunakan untuk mengevaluasi komentar-komentar yang dikumpulkan dari YouTube, sehingga kami dapat memahami persepsi publik terhadap konten yang ditampilkan. Dengan menggunakan teknik pemrosesan bahasa alami (Natural Language Processing/NLP), kami dapat menganalisis kata-kata dan frasa yang digunakan dalam komentar untuk menentukan sentimen keseluruhan. Hasil dari analisis ini dapat memberikan wawasan yang berharga bagi pembuat konten, pemasar, dan peneliti dalam memahami dinamika opini publik.

## Gated Recurrent Unit (GRU)

Gated Recurrent Unit (GRU) adalah salah satu varian dari Recurrent Neural Network (RNN) yang dirancang khusus untuk menangani masalah yang sering muncul dalam data sekuensial, seperti teks dan urutan waktu. GRU memiliki struktur yang lebih sederhana dibandingkan dengan Long Short-Term Memory (LSTM), namun tetap efektif dalam mengatasi masalah vanishing gradient yang sering terjadi pada RNN tradisional. Dengan menggunakan GRU, model dapat lebih efisien dalam memproses informasi dari urutan data yang panjang, sehingga meningkatkan akurasi dalam prediksi dan klasifikasi. Dalam penelitian ini, GRU digunakan untuk menganalisis data komentar yang telah dikumpulkan, memungkinkan kami untuk menangkap pola dan hubungan dalam data sekuensial yang kompleks. Keunggulan GRU dalam efisiensi komputasi dan kemampuannya untuk menangani data sekuensial menjadikannya pilihan yang ideal untuk analisis sentimen dalam konteks komentar YouTube.

2.4 Penelitian Terkait

* Analisis Sentimen Komentar YouTube terhadap Pemindahan Ibu Kota Negara Menggunakan Metode Naïve Bayes (Huwaida et al., 2024).
* SENTIMENT ANALYSIS DENGAN NAÏVE BAYES BERBASIS ORANGE TERHADAP RESIKO PEMBANGUNAN IKN (Munawaroh et al., 2024).

# BAB III Metodologi Penelitian

## Alur Penelitian

Pengumpulan Data: Pada tahap ini, data komentar dari YouTube dikumpulkan menggunakan Google Client API. Proses crawling ini memungkinkan peneliti untuk mengakses dan mengumpulkan data dalam jumlah besar dari berbagai video, yang menjadi dasar untuk analisis lebih lanjut.

Preprocessing Data: Setelah data dikumpulkan, langkah selanjutnya adalah preprocessing untuk mempersiapkan data agar siap digunakan dalam model. Proses ini meliputi:

* Tokenisasi: Memecah teks komentar menjadi kata-kata atau token yang lebih kecil.
* Padding: Menyesuaikan panjang input agar seragam, sehingga model dapat memproses data dengan lebih efisien.
* Encoding data menggunakan LabelEncoder: Mengubah label kategori menjadi format numerik yang dapat dipahami oleh model.

Pemodelan Sentimen: Pada tahap ini, arsitektur GRU digunakan untuk membangun model analisis sentimen. Model dilatih menggunakan dataset yang telah dibagi menjadi dua bagian: data pelatihan (train) dan data pengujian (test). Proses pelatihan ini bertujuan untuk mengajarkan model bagaimana mengklasifikasikan sentimen berdasarkan pola yang ada dalam data.

Evaluasi Model: Setelah model dilatih, langkah selanjutnya adalah evaluasi kinerja model. Pada tahap ini, metrik seperti accuracy (akurasi) dan loss (kerugian) digunakan untuk menilai seberapa baik model dalam mengklasifikasikan sentimen dari komentar yang diuji.

Interpretasi Hasil: Tahap terakhir adalah interpretasi hasil dari evaluasi kinerja model GRU. Di sini, peneliti menganalisis hasil yang diperoleh untuk memahami efektivitas model dalam mengidentifikasi sentimen dari komentar YouTube. Hasil ini dapat memberikan wawasan yang berharga bagi pengembangan lebih lanjut dan aplikasi praktis dari model analisis sentimen.

## Pengumpulan Data

Pengumpulan data merupakan langkah awal yang krusial dalam penelitian ini, di mana kami fokus pada pengumpulan komentar dari platform YouTube. Komentar-komentar ini diambil dari video yang berkaitan dengan berita pembangunan Ibu Kota Negara (IKN) yang baru. Proses ini tidak hanya melibatkan pengumpulan data, tetapi juga pemilihan sumber yang relevan dan representatif untuk memastikan bahwa analisis yang dilakukan dapat memberikan wawasan yang akurat dan bermanfaat.

Sumber data utama dalam penelitian ini adalah komentar yang terdapat pada video-video YouTube yang membahas topik pembangunan IKN. YouTube, sebagai salah satu platform media sosial terbesar di dunia, menyediakan ruang bagi pengguna untuk berbagi pendapat, pandangan, dan reaksi mereka terhadap berbagai isu, termasuk pembangunan infrastruktur dan kebijakan pemerintah. Dengan memilih video yang relevan, kami dapat mengumpulkan komentar yang mencerminkan beragam perspektif masyarakat mengenai proyek pembangunan IKN.

Pentingnya memilih video yang tepat tidak dapat diabaikan, karena komentar yang diambil harus mencakup berbagai sudut pandang, baik yang mendukung maupun yang menentang, untuk memberikan gambaran yang lebih komprehensif tentang sentimen publik. Oleh karena itu, kami melakukan analisis awal untuk mengidentifikasi video-video yang memiliki jumlah komentar yang signifikan dan relevan dengan topik yang diteliti.

## Teknologi yang Digunakan

Untuk melakukan pengumpulan data, kami memanfaatkan beberapa teknologi yang mendukung proses crawling dan pengolahan data. Berikut adalah rincian teknologi yang digunakan:

Google Client API adalah alat yang sangat berguna untuk mengakses data dari berbagai layanan Google, termasuk YouTube. Dengan menggunakan API ini, kami dapat melakukan crawling untuk mengumpulkan komentar dari video yang telah dipilih. Proses ini memungkinkan kami untuk mengakses data dalam jumlah besar secara efisien dan terstruktur. API ini juga menyediakan berbagai parameter yang dapat disesuaikan, seperti batasan jumlah komentar yang diambil, sehingga kami dapat mengoptimalkan pengumpulan data sesuai dengan kebutuhan penelitian.

Python adalah bahasa pemrograman yang sangat populer dalam bidang analisis data dan pengembangan model machine learning. Dalam penelitian ini, Python digunakan untuk berbagai tujuan, termasuk pengolahan data yang telah dikumpulkan, serta pembangunan model analisis sentimen. Dengan menggunakan pustaka seperti Pandas untuk manipulasi data, NumPy untuk perhitungan numerik, dan TensorFlow atau Keras untuk membangun model GRU, kami dapat melakukan analisis yang mendalam dan efisien. Python juga menyediakan berbagai pustaka untuk pemrosesan bahasa alami (Natural Language Processing/NLP), yang sangat penting dalam menganalisis komentar teks.

## Arsitektur GRU

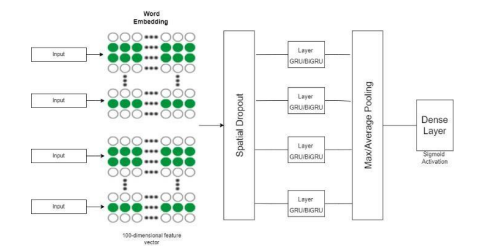
Arsitektur Gated Recurrent Unit (GRU) yang digunakan dalam penelitian ini dirancang untuk menganalisis data sekuensial, khususnya dalam konteks analisis sentimen dari komentar YouTube. GRU merupakan salah satu jenis Recurrent Neural Network (RNN) yang memiliki kemampuan untuk menangani masalah vanishing gradient, yang sering terjadi pada model RNN tradisional. Dengan struktur yang lebih sederhana dibandingkan dengan Long Short-Term Memory (LSTM), GRU tetap mampu menangkap pola dalam data sekuensial dengan efisiensi yang tinggi. Berikut adalah penjelasan lebih mendalam mengenai setiap komponen dari arsitektur GRU yang digunakan dalam penelitian ini:

1. Input Layer Pada tahap awal, data komentar yang telah diproses akan dimasukkan ke dalam Input Layer. Di sini, kami menggunakan Embedding Layer untuk memetakan kata-kata ke dalam representasi vektor berdimensi rendah. Embedding Layer berfungsi untuk mengubah kata-kata yang bersifat diskrit menjadi representasi numerik yang lebih kompak, sehingga model dapat lebih mudah memproses informasi tersebut.

* Pemetaan Kata: Setiap kata dalam korpus data akan dipetakan ke dalam vektor berdimensi tetap, yang memungkinkan model untuk menangkap hubungan semantik antara kata-kata. Misalnya, kata-kata yang memiliki makna serupa akan memiliki representasi vektor yang berdekatan dalam ruang vektor. Proses ini sangat penting dalam analisis sentimen, karena konteks dan makna kata dapat mempengaruhi interpretasi sentimen secara keseluruhan. 2. Hidden Layer Setelah data melewati Input Layer, informasi akan diteruskan ke Hidden Layer, yang merupakan inti dari arsitektur GRU. Pada tahap ini, kami menggunakan GRU Layer dengan ukuran hidden\_dim = 32.
* Fungsi GRU: GRU memiliki dua gerbang utama, yaitu update gate dan reset gate, yang memungkinkan model untuk memutuskan informasi mana yang harus disimpan dan mana yang harus dilupakan. Dengan menggunakan ukuran hidden\_dim yang ditetapkan, model dapat menangkap pola temporal dalam data sekuensial, sehingga dapat memahami konteks dari komentar yang diberikan. Ukuran hidden\_dim yang lebih kecil, seperti 32, memungkinkan model untuk beroperasi dengan efisiensi komputasi yang lebih baik, sambil tetap mempertahankan kemampuan untuk menangkap informasi penting dari data.
* Keunggulan GRU: Salah satu keunggulan utama dari GRU adalah kemampuannya untuk mengatasi masalah vanishing gradient yang sering dihadapi oleh RNN tradisional. Dengan struktur yang lebih sederhana, GRU dapat lebih cepat dilatih dan lebih efisien dalam memproses data sekuensial yang panjang, sehingga sangat cocok untuk analisis sentimen yang melibatkan teks yang bervariasi dalam panjang dan kompleksitas.

1. Output Layer Setelah informasi diproses melalui Hidden Layer, hasilnya akan diteruskan ke Output Layer. Di sini, kami menggunakan Linear Layer untuk klasifikasi sentimen.

* Klasifikasi Sentimen: Output Layer bertugas untuk mengklasifikasikan hasil analisis menjadi tiga kategori sentimen: positif, negatif, dan netral. Dengan menggunakan fungsi aktivasi seperti softmax, model dapat menghasilkan probabilitas untuk setiap kategori, yang memungkinkan kami untuk menentukan sentimen dominan dari komentar yang dianalisis.
* Proses Klasifikasi: Setelah model dilatih, Output Layer akan memberikan prediksi berdasarkan input yang diberikan. Misalnya, jika model mengidentifikasi bahwa komentar tersebut memiliki lebih banyak kata positif, maka model akan mengklasifikasikan komentar tersebut sebagai sentimen positif. Sebaliknya, jika terdapat banyak kata negatif, maka komentar tersebut akan diklasifikasikan sebagai sentimen negatif.



## Evaluasi Model

Evaluasi model adalah langkah penting dalam proses pengembangan model machine learning, termasuk dalam analisis sentimen menggunakan arsitektur Gated Recurrent Unit (GRU). Pada tahap ini, kami akan menilai kinerja model berdasarkan beberapa metrik yang relevan, yang akan memberikan gambaran yang jelas tentang seberapa baik model dapat mengklasifikasikan sentimen dari komentar YouTube. Evaluasi ini dilakukan dengan membandingkan hasil yang diperoleh dari data pelatihan (train) dan data pengujian (test). Berikut adalah penjelasan lebih mendalam mengenai metrik yang digunakan dalam evaluasi model:

1. Train Score dan Test Score

Train Score dan Test Score adalah dua metrik utama yang digunakan untuk mengukur akurasi model.

* Train Score: Metrik ini menunjukkan seberapa baik model dapat memprediksi sentimen pada data pelatihan. Data pelatihan adalah subset dari dataset yang digunakan untuk melatih model, dan biasanya terdiri dari sebagian besar data yang tersedia. Train Score dihitung dengan membandingkan prediksi model terhadap label sebenarnya dari data pelatihan. Nilai akurasi yang tinggi pada train score menunjukkan bahwa model telah belajar dengan baik dari data yang diberikan.

* Test Score: Metrik ini mengukur akurasi model pada data pengujian, yang merupakan subset terpisah dari dataset yang tidak digunakan selama proses pelatihan. Test Score sangat penting karena memberikan indikasi tentang kemampuan model untuk generalisasi, yaitu seberapa baik model dapat melakukan prediksi pada data baru yang belum pernah dilihat sebelumnya. Nilai akurasi yang tinggi pada test score menunjukkan bahwa model tidak hanya menghafal data pelatihan, tetapi juga dapat menerapkan pengetahuan yang diperoleh untuk data yang berbeda.

Perbandingan antara Train Score dan Test Score dapat memberikan wawasan tambahan mengenai kinerja model. Jika Train Score jauh lebih tinggi daripada Test Score, ini bisa menjadi indikasi bahwa model mengalami overfitting, di mana model terlalu kompleks dan hanya belajar pola dari data pelatihan tanpa mampu generalisasi ke data baru. Sebaliknya, jika kedua skor tersebut seimbang, ini menunjukkan bahwa model memiliki kemampuan generalisasi yang baik.

1. Train Cost dan Test Cost

Train Cost dan Test Cost adalah metrik tambahan yang digunakan untuk mengevaluasi kinerja model, yang berfokus pada nilai kerugian (loss) yang dihasilkan selama pelatihan dan pengujian.

* Train Cost: Metrik ini mengukur nilai kerugian pada data pelatihan. Kerugian adalah ukuran seberapa baik model memprediksi label yang benar; semakin rendah nilai kerugian, semakin baik model dalam memprediksi. Train Cost dihitung menggunakan fungsi kerugian yang sesuai, seperti binary cross-entropy untuk klasifikasi biner atau categorical cross-entropy untuk klasifikasi multi-kelas. Nilai kerugian yang rendah pada data pelatihan menunjukkan bahwa model telah belajar dengan baik dari data yang diberikan.
* Test Cost: Metrik ini mengukur nilai kerugian pada data pengujian. Seperti halnya Train Cost, Test Cost memberikan gambaran tentang seberapa baik model dapat memprediksi label yang benar pada data yang tidak terlihat sebelumnya. Nilai kerugian yang rendah pada data pengujian menunjukkan bahwa model dapat melakukan prediksi yang akurat dan efektif pada data baru.

Analisis perbandingan antara Train Cost dan Test Cost juga memberikan wawasan penting mengenai kinerja model. Jika Train Cost jauh lebih rendah daripada Test Cost, ini bisa menjadi indikasi bahwa model mengalami overfitting, di mana model sangat baik dalam memprediksi data pelatihan tetapi tidak mampu melakukan hal yang sama pada data pengujian. Sebaliknya, jika kedua nilai kerugian tersebut seimbang atau Test Cost lebih rendah, ini menunjukkan bahwa model memiliki kemampuan generalisasi yang baik.

# BAB IV HASIL DAN PEMBAHASAN

## Hasil Crawling Data

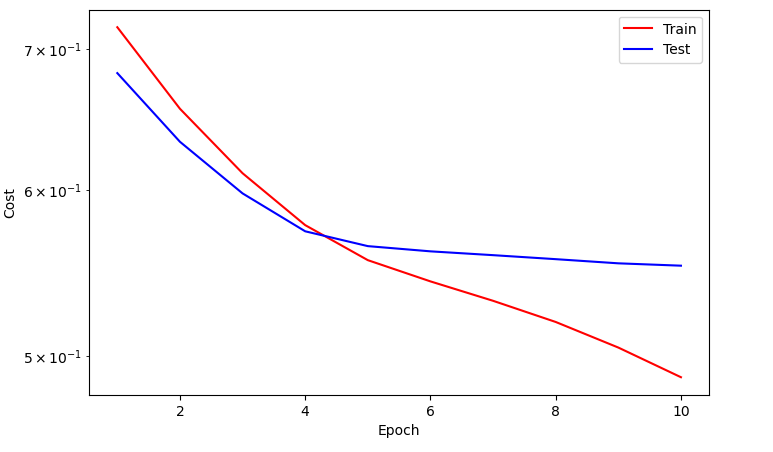
Setelah proses crawling data komentar dari video YouTube yang berkaitan dengan pembangunan Ibu Kota Negara (IKN) selesai dilakukan, kami berhasil mengumpulkan sejumlah data yang relevan. Data yang berhasil diambil mencakup beberapa kolom penting yang memberikan informasi mendetail tentang setiap komentar. Berikut adalah kolom-kolom yang terdapat dalam dataset:

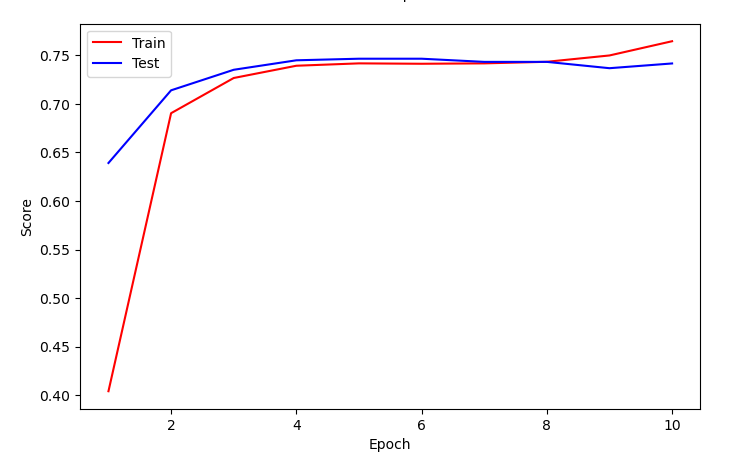


Hasil Analisis Sentimen

Setelah data komentar berhasil dikumpulkan dan diproses, langkah selanjutnya adalah melakukan analisis sentimen menggunakan model GRU yang telah dibangun. Model ini dirancang untuk mengklasifikasikan komentar ke dalam tiga kategori sentimen: positif, negatif, dan netral. Hasil dari analisis sentimen ini memberikan gambaran tentang kinerja model dalam mengidentifikasi dan mengklasifikasikan sentimen dari komentar yang dianalisis. Berikut adalah hasil kinerja model GRU yang ditampilkan dalam bentuk table dan grafik cost dan score:

|  |  |
| --- | --- |
| **Metrik** | **Nilai** |
| Train\_score | 76.43% |
| Test\_score | 74.63% |
| Train\_cost | 0.4887 |
| Test\_cost | 0.5522 |





# BAB V KESIMPULAN DAN SARAN

## Kesimpulan

Berdasarkan penelitian yang telah dilakukan, kami berhasil mencapai beberapa temuan penting yang memberikan wawasan mengenai analisis sentimen terhadap komentar YouTube terkait pembangunan Ibu Kota Negara (IKN). Berikut adalah kesimpulan yang dapat diambil dari penelitian ini:

* Pengumpulan Data yang Efektif: Proses pengumpulan data komentar dari video YouTube berhasil dilakukan dengan menggunakan Google Client API. Metode ini memungkinkan kami untuk mengakses dan mengumpulkan data dalam jumlah besar secara efisien dan terstruktur. Data yang berhasil diambil mencakup informasi penting seperti nama penulis komentar, waktu publikasi, isi komentar, dan jumlah "like" yang diterima, yang semuanya sangat berharga untuk analisis lebih lanjut.
* Kinerja Model GRU: Model Gated Recurrent Unit (GRU) yang dibangun dalam penelitian ini menunjukkan kinerja yang baik dalam mengklasifikasikan sentimen. Dengan akurasi sebesar 74.63% pada data uji, model ini mampu memberikan hasil yang cukup memuaskan. Meskipun ada sedikit penurunan akurasi dibandingkan dengan data pelatihan, nilai ini masih menunjukkan bahwa model dapat menggeneralisasi pengetahuan yang diperoleh dari data pelatihan ke data baru yang tidak terlihat sebelumnya.
* Kemampuan Klasifikasi Sentimen: Metode GRU terbukti efektif dalam mengklasifikasikan sentimen komentar menjadi tiga kategori: positif, negatif, dan netral. Hasil analisis menunjukkan bahwa model dapat menangkap nuansa dalam komentar yang diberikan oleh pengguna, sehingga memberikan gambaran yang lebih jelas tentang pandangan publik terhadap pembangunan IKN. Kemampuan ini sangat penting, mengingat kompleksitas bahasa alami dan variasi dalam ekspresi sentimen yang dapat muncul dalam komentar.
* Pentingnya Evaluasi Model: Proses evaluasi model yang dilakukan dengan menggunakan metrik seperti Train Score, Test Score, Train Cost, dan Test Cost memberikan wawasan yang berharga mengenai kinerja model. Hasil evaluasi menunjukkan bahwa meskipun model memiliki akurasi yang baik, masih ada ruang untuk perbaikan, terutama dalam hal mengurangi nilai kerugian pada data pengujian. Hal ini menunjukkan bahwa model dapat terus ditingkatkan untuk mencapai hasil yang lebih optimal.

## Saran

Berdasarkan temuan dan kesimpulan yang diperoleh dari penelitian ini, kami memberikan beberapa saran untuk penelitian selanjutnya yang dapat membantu meningkatkan akurasi dan efektivitas analisis sentimen:

* Eksplorasi Metode Lain: Penelitian selanjutnya dapat mengeksplorasi penggunaan metode lain seperti Long Short-Term Memory (LSTM) atau Transformer. LSTM, yang merupakan pengembangan dari RNN, memiliki kemampuan yang lebih baik dalam menangkap dependensi jangka panjang dalam data sekuensial. Sementara itu, arsitektur Transformer, yang telah terbukti sangat efektif dalam berbagai tugas NLP, dapat memberikan hasil yang lebih baik dalam analisis sentimen dengan memanfaatkan mekanisme perhatian (attention mechanism) yang canggih.
* Peningkatan Dataset: Untuk meningkatkan akurasi model, disarankan untuk memperluas dataset dengan mengumpulkan lebih banyak komentar dari berbagai video dan sumber. Dengan memiliki dataset yang lebih besar dan beragam, model dapat belajar dari lebih banyak variasi dalam bahasa dan ekspresi sentimen, yang pada gilirannya dapat meningkatkan kemampuan generalisasi model.
* Penggunaan Teknik Preprocessing yang Lebih Lanjut: Penelitian selanjutnya dapat mempertimbangkan untuk menerapkan teknik preprocessing yang lebih canggih, seperti penghapusan stop words, stemming, atau lemmatization, untuk meningkatkan kualitas data yang digunakan dalam pelatihan model. Teknik-teknik ini dapat membantu mengurangi noise dalam data dan meningkatkan akurasi klasifikasi.
* Analisis Sentimen Multimodal: Mengingat bahwa komentar di YouTube sering kali terkait dengan video, penelitian selanjutnya dapat mengeksplorasi analisis sentimen multimodal yang menggabungkan analisis teks dengan analisis visual dari video. Dengan mempertimbangkan konteks visual, model dapat memberikan analisis yang lebih komprehensif dan akurat terhadap sentimen publik.

# DAFTAR PUSTAKA

Huwaida, S.F., Kusumawati, R., & Isnaini, B. (2024). Analisis Sentimen Komentar YouTube terhadap Pemindahan Ibu Kota Negara Menggunakan Metode Naïve Bayes. *Jambura Journal of Informatics*.

Munawaroh, A., Ridhoi, R., & Rudiman, R. (2024). SENTIMENT ANALYSIS DENGAN NAÏVE BAYES BERBASIS ORANGE TERHADAP RESIKO PEMBANGUNAN IKN. *JATI (Jurnal Mahasiswa Teknik Informatika)*.