



Analisis Sentimen dalam Opini Publik di Chanel Youtube Indonesia *Lawyers Club* Tentang Isu Populer dengan Menggunakan Metode LSTM dan Bi-LSTM

Bagus Wicaksono^{1*}, Vinna Rahmayanti Setyaning Nastiti²

^{1,2}Universitas Muhammadiyah Malang, Indonesia

**email:* vinastiti@umm.ac.id

Info Artikel

Dikirim: 24 Juli 2024
Diterima: 12 Desember 2024
Diterbitkan: 12 Desember 2024

Kata kunci:

Berital
Dataset;
Komentar;
Opini Publik;
YouTube.

ABSTRAK

Salah satu media sosial yang sering membagikan berita dan populer digunakan di Indonesia adalah *YouTube*. Melalui platform *YouTube* pembaca dapat memberikan komentar untuk berbagi pendapat mereka dibawah vidio terkait. Komentar ini telah menjadi sumber informasi dan penelitian yang bagus. Artikel ini menyajikan kumpulan data yang berisi 368.299 komentar publik dan balasan dari 73 berita video yang diterbitkan dari 01 Januari 2024 hingga 18 Mei 2024 dari saluran “*YouTube Indonesia Lawyers Club*” yang terkenal. Sebagai upaya memastikan privasi komentator maka nama komentator dikodekan dalam dataset. Dataset ini terbuka untuk digunakan oleh para peneliti dengan link akses <https://data.mendeley.com/datasets/h9335fgsgr/1>. Data ini dapat membantu para peneliti untuk mengidentifikasi pola-pola dalam opini publik dan menganalisis apa isu yang populer dibicarakan di chanel youtube Indonesia lawyers club tahun 2024. Hasilnya isu populer yang sering dibicarakan adalah terkait dengan pemerintahan Jokowi dan pemilu, sedangkan sentimen yang paling banyak dari isu tersebut adalah negatif yang menghasilkan nilai akurasi metode LSTM sebesar 99.61 % sedangkan dengan metode Bi-LSTM akurasi sebesar 98.11 %.

1. PENDAHULUAN

Platform media sosial telah menjadi sumber utama bagi masyarakat untuk mengakses informasi dan menyampaikan opini yang menimbulkan berbagai sinisme dikalangan masyarakat terhadap isu-isu tertentu yang berkembang [1]. Salah satu media sosial yang sering digunakan oleh masyarakat Indonesia adalah *YouTube* yang menjadi salah satu platform media sosial terkemuka, memegang peranan penting dalam penyebaran berita dan pembentukan opini publik di Indonesia. Menurut survei *databoks*, indonesia masuk kedalam *list* daftar negara nomor 4 pengguna *YouTube* terbanyak di dunia dengan jumlah 139 juta pengguna [2]. Jumlah pengguna yang terus meningkat, menjadikan *YouTube* sebagai wadah bagi masyarakat untuk mengonsumsi konten berita dan mengekspresikan pandangan mereka terhadap isu-isu tersebut.

Penyebaran informasi yang telah mengalami degradasi dari metode konvensional menuju era teknologi informasi dengan menggunakan internet seperti saat ini dimana penyebaran informasi sangat mudah dikuasai oleh pihak tertentu untuk kepentingan propaganda. Bahkan sebaliknya digunakan untuk tujuan positif seperti upaya mengatasi pelanggaran norma-norma sosial, pelanggaran hukum, pencegahan dan penanganan kesehatan, pencegahan kejahatan serta kekerasan [3][4], edukasi kesehatan [5][6], perkembangan ilmu pengetahuan [7][8], keamanan ekologis [9]. Opini publik pada era globalisasi ini perlu dilakukan penanganan

serta tindakan yang efektif untuk mencegah propaganda dan berita yang tidak benar terhadap suatu topik serta melakukan penilaian berbagai topik pada individu atau sekelompok orang [10]. Dataset dikumpulkan dan dipergunakan untuk berbagai kebutuhan salah satunya terkait dengan analisis opini publik [11][12]. Dataset ini dapat memberikan wawasan berharga bagi pemangku kepentingan, seperti pemerintah, media, dan peneliti, untuk memahami isu-isu yang menjadi perhatian publik, serta mengidentifikasi dan analisis tren dan pola sentimen masyarakat.

Penelitian ini memiliki kesamaan dengan penelitian terdahulu, seperti penelitian dari Rozi, Pramono dan Dahlan yang mencoba mengimplementasikan opinion mining pada data opini publik diperguruan tinggi [13] penelitian itu secara jelas membedakan mana yang termasuk opini dan mana yang bukan. Sejalan dengan ini penelitian yang akan diajukan juga membedakan opini dengan kalimat lainnya pada dokumen teks. Peneliti mengklasifikasikan kedalam 2 jenis yaitu opini dan berita. Menurut Gerung (2017) Opini merupakan persepsi individu yang terikat secara subjektif [14]. Sedangkan berita menurut M. Lyle Spencer, News Writing dalam Effendy (2023) [15] merupakan kenyataan atau ide yang benar dan dapat menarik perhatian sebagian besar pembaca. Klasifikasi ini juga penting dilihat secara tajam dengan logika berfikir yang tepat bahwa memungkinkan dalam menganalisis suatu dokumen teks komentar terdapat opini dan berita. Pada setiap opini misalnya “saya tidak suka dan benci kepada anda” dapat ditentukan apakah opini tersebut termasuk kedalam sentimen yang bermuatan positif atau negarif atau sebaliknya dalam sentimen negatif seperti “saya tidak suka anda” bisa diklasifikasikan kedalam kelompok opini suka/tidak suka atau benci/tidak benci.

Sedangkan secara metode penelitian ini berbeda dengan penelitian sebelumnya seperti penelitian Fitri, Fikrillah dan Karimah yang melakukan analisis dengan menggunakan metode *Naïve Bayes Classifier (NBC)*, *Random Forest (RF)*, dan *K-Nearest Neighbor (KNN)* [16][17][18][19], penelitian dari Mualfah yang menggunakan metode *Support Vector Machines (SVM)* [20], penelitian dari Mujahidin yang menggunakan metode *Gaussian Naïve Bayes (GNN)* [21], penelitian dari Nurmawiyi yang menggunakan metode *Latent Dirichlet Allocation (LDA)* [22], penelitian dari Srivastava yang menggunakan metode *Logistic Regression (LR)*, *Artificial Neural Network (ANN)* [23], dan penelitian dari Wahid yang menggunakan metode *Decision Tree (DT)* [24], sedangkan penelitian ini merupakan penelitian eksperimen yang melihat opini dari berbagai isu aktual, penelitian ini juga menyumbang lebih banyak jumlah dataset yang digunakan dari penelitian sebelumnya untuk mewakili beberapa isu pembahasan yang relevan dan disukai oleh masyarakat di Indonesia selama periode waktu tertentu dari tanggal 01 Januari 2024 hingga 18 Mei 2024 di chanel youtube Indonesia lawyers club, serta melakukan model analisis dengan menggunakan metode *Long Short-Term Memory (LSTM)*.

Selanjutnya penelitian dari Islam, Kaur dan Lestandy yang menggunakan metode yang sama yaitu *LSTM* [25][26][27], metode ini dapat menyimpan informasi penting dalam waktu yang lama atau jauh sebelumnya yang sangat berguna dalam melakukan analisis sentimen dibandingkan metode *RNN* lainnya. Perbedaan dengan penelitian sebelumnya terletak pada fokus dan tujuan yang berbeda dimana fokus penelitian ini adalah melakukan eksplorasi opini publik terhadap topik yang paling disukai di Indonesia di chanel youtube Indonesia lawyers club dari tanggal 01 Januari 2024 hingga 18 Mei 2024.

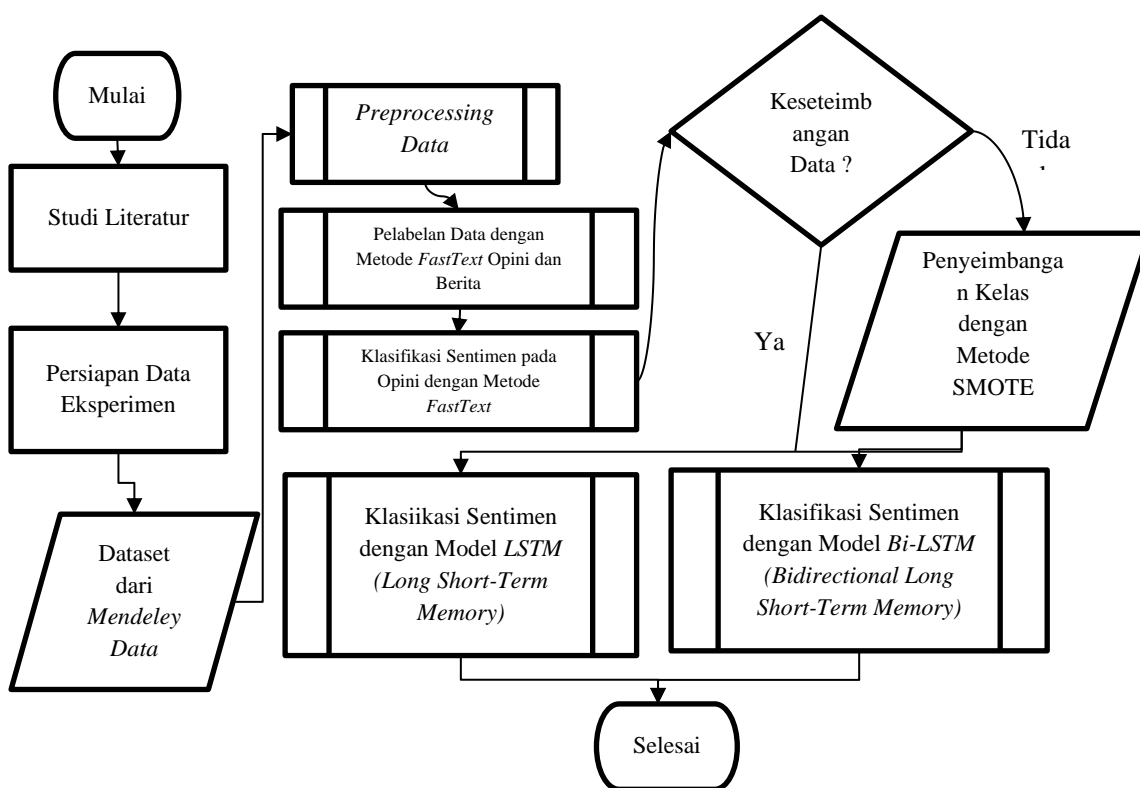
Dataset ini akan membantu para peneliti yang tertarik dengan *Natural Language Processing (NLP)* untuk mengidentifikasi sentimen publik terhadap artikel berita serta perbedaan budaya Indonesia dengan Negara lain. Dataset ini berisi total 368.299 komentar publik dari 73 vidio berita. Para pencari data dapat memilihnya berdasarkan topik tertentu untuk penelitian selanjutnya yang akan datang. Data tersebut dapat digunakan untuk mendeteksi pemalsuan dalam berita dan peristiwa yang dapat membantu menghentikan penyebaran informasi yang salah. Dataset ini dapat digunakan untuk menggambarkan bagaimana media sosial dan berita mempromosikan suatu topik yang banyak disukai oleh publik sehingga berguna untuk para pihak media untuk membuat berita yang disukai oleh masyarakat Indonesia. Dataset ini dapat digunakan oleh pemerintah untuk membuka diskusi terhadap topik yang viral sehingga mendukung kepentingan publik.

Penelitian ini bertujuan untuk mengetahui bagaimana opini publik Indonesia terhadap berbagai isu aktual yang tercermin dalam komentar di chanel *YouTube "Indonesia Lawyers Club"* pada tahun 2024. Sehingga dapat diidentifikasi topik-topik yang menjadi perhatian utama masyarakat Indonesia berdasarkan komentar di chanel

YouTube "Indonesia Lawyers Club". Setelahnnya dapat dilakukan modeling analisis opini publik pada komentar tersebut. Hasil dari penelitian ini dapat dimanfaatkan untuk mengevaluasi kebijakan yang telah di buat, menentukan isu populer yang disukai oleh masyarakat Indonesia yang berguna untuk promosi bagi pemilik media berita tanah air untuk menentukan topik yang populer di bicarakan.

2. METODE PENELITIAN

Penelitian ini berjenis kuantitatif dengan metode penelitian eksperimen, pada metode ini ada beberapa tahapan perancangan eksperimen (*experiment design*) yaitu pengumpulan data, persiapan data dan analisis data. Jenis penelitian kuantitatif adalah investigasi yang dilakukan secara sistematis mengenai sebuah fenomena dengan cara mengumpulkan data yang dapat diukur dengan teknik statistik, matematika, atau komputasi [28]. Metode penelitian ekperimen adalah suatu penelitian yang berusaha melihat hubungan sebab akibat dari satu atau lebih data, baik dengan cara memanipulasi beberapa variable untuk tujuan tertentu, mengelompokan data, membandingkan beberapa kelompok data untuk tujuan tertentu [29].



Gambar 1. Flowchart Penelitian

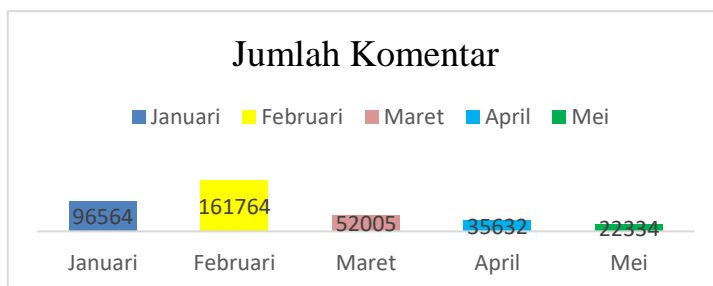
Penelitian ini dilakukan dengan melakukan studi literatur seperti terlihat pada Gambar 1 yang menghasilkan rumusan masalah tertentu, tujuan, urgensi, *research gap* serta metode yang digunakan untuk melakukan penelitian. Metode penelitian eksperimen dilakukan dengan cara mengumpulkan data, persiapan data dan *preprocessing* serta analisis data. Pengumpulan data eksperimen didapatkan dari mendeley data. Persiapan data dilakukan dengan semi manual menggunakan *Microsoft Excell* dan *google collab*. Tahap terakhir dalam melakukan penelitian eksperimen ini adalah analisis data dengan menggunakan program *google collab* untuk melakukan wordcloud dan analisis korelasi antar variabel kemudian melakukan visualisasi sebelum dan sesudah dilakukannya proses *preprocessing*. Visualisasi tidak menghentikan eksplorasi dataset yang dilakukan [30]. Pada penelitian ini juga melakukan klasifikasi opini publik terhadap isu populer yang disukai oleh masyarakat Indonesia dalam dataset yang diperoleh kedalam kategori opini dan berita kemudian pada opini yang didapat akan di klasifikasikan sentimen apa yang terkandung dalam setiap opini individu.

3. HASIL DAN PEMBAHASAN

Dataset ini dibagikan untuk penelitian lebih lanjut dalam file Excel. Perancangan penelitian eksperimen (*experiment design*) yang dilakukan melibatkan beberapa langkah yaitu pengumpulan data, *preprocessing* data dan analisis data.

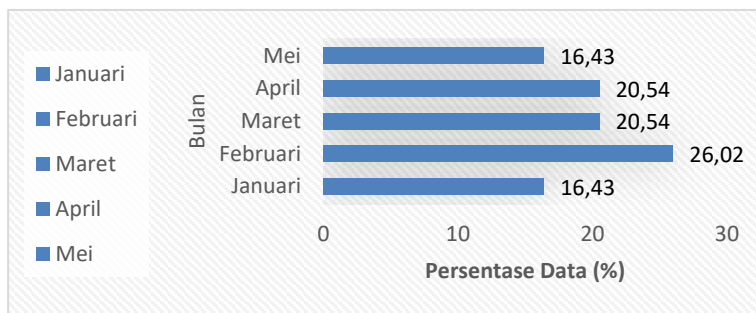
3.1 Pengumpulan Data

Penting untuk memiliki kumpulan data yang memadai yang mencakup berbagai isu populer untuk memahami bagaimana perasaan individu-individu yang berbahasa Indonesia terhadap isu populer yang disukai. Dengan tujuan tersebut, proses pengumpulan data dimulai. Data dihimpun dari sumber data mentahan yang telah tersedia secara publik di *mendeley data* dengan link <https://data.mendeley.com/datasets/h9335fgsgr/1>. Data tersebut adalah data komentar pada video yang di upload dari tanggal 01 Januari 2024 hingga 18 Mei 2024. Data mentah disimpan dalam file Excel dan terdapat 368.299 komentar berita berhasil dikumpulkan dari 73 berita video *YouTube*. Dataset diambil dari chanel *YouTube “Indonesia Lawyers Club”* yang dikenal sebagai saluran berita *YouTube* Indonesia yang Independent serta memiliki 5,41 Juta pelanggan. Dataset ini mencakup beragam isu populer, seperti politik, berita lokal, nasional, dan internasional, ekonomi, hiburan, olahraga, dan pendidikan dan masih banyak lagi. Dataset ini berisi vidio yang diupload dari 01 Januari 2024 hingga 18 Mei 2024. Saat mengumpulkan data, total 15 fitur untuk setiap sampel yang dikumpulkan yaitu *video url, title, totallikes, views, date, hashtags, description, comment author, comment time, comment, comment likes, reply author, reply time, reply, dan reply likes*. Untuk menjamin terpenuhinya hak cipta moral maka penulis memuatnya rincian pada setiap item dengan fitur deskripsi. Pada persiapan data nama penulis dalam data diprivasi dengan cara di ganti dengan notasi tertentu menggunakan program *Microsoft Excell*. Selanjutnya data diurutkan berdasarkan tanggal penerbitannya dalam urutan menaik. Terdapat 15 atribut dengan berbagai jenis informasi yang terpisah satu atribut dengan atribut lainnya. Data kemudian disimpan dalam file *Microsoft Excell* untuk tindakan lebih lanjut.



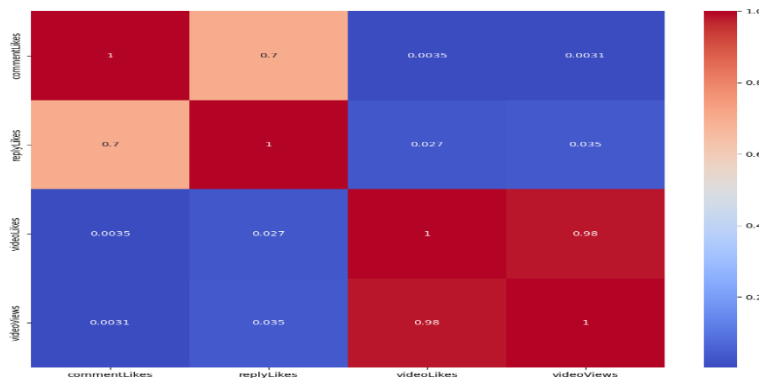
Gambar 2. Jumlah Komentar Bulanan

Berdasarkan Gambar 2 menampilkan vidio yang di upload pada 01 Januari 2024 hingga 18 Mei 2024 yang memiliki jumlah total komentar 368.299 paling banyak terdapat pada bulan februari sebanyak 161.764 komentar atau sekitar 43 % dari total jumlah komentar. Komentar paling sedikit terdapat pada rentang video yang di upload pada bulan Mei yaitu sebesar 22.334 komentar.



Gambar 3. Jumlah Persentase Vidio yang di Upload pada setiap Bulan

Berdasarkan Gambar 3 menampilkan video yang paling banyak di upload oleh chanel *YouTube* tersebut terdapat pada bulan Februari senilai 26.02 % dari total 73 vidio *YouTube*, sedangkan paling sedikit video yang diupload pada bulan Januari dan Mei masing-masing senilai 16.43 % dari total 73 vidio yang diupload. Terhadap data tersebut sebelum melakukan *preprocessing* data juga di visualisasikan dalam bentuk *wordcloude* dan frekuensi kata yang paling sering muncul.

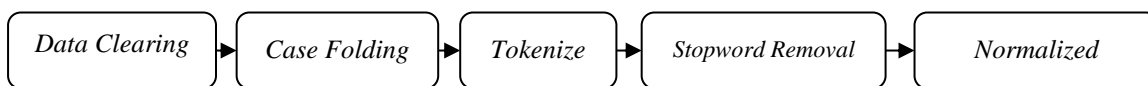


Gambar 4. Korelasi Antar Variabel

Terlihat pada Gambar 4 hasil analisis korelasi yang dilakukan untuk menentukan pola fitur numerik data. Nilai korelasi menggambarkan bahwa semua atribut memiliki korelasi positif antar variabel. Korelasi antar variabel *videoLikes* dan *videoViews* memiliki korelasi positif yang paling kuat sebesar 0,98, yang diharapkan karena video dengan lebih banyak suka pada umumnya cenderung memiliki lebih banyak penayangan. Korelasi variabel *commentLikes* dan *replyLikes* memiliki korelasi positif yang kuat sebesar 0,7 menunjukkan bahwa video dengan lebih banyak komentar yang disukai cenderung memiliki lebih banyak balasan suka. Korelasi variabel *commentLikes* dan *videoLikes* memiliki korelasi positif yang sangat lemah sekitar 0,0035 menunjukkan bahwa jumlah komentar yang disukai tidak terlalu terkait dengan jumlah video yang disukai. Korelasi variabel *commentLikes* dan *videoViews* juga memiliki korelasi positif yang sangat lemah yaitu sekitar 0.0031 menunjukkan bahwa jumlah komentar yang disukai tidak memiliki hubungan yang signifikan dengan jumlah penayangan video. Korelasi variabel *replyLikes* dan *videoLikes* memiliki korelasi positif yang lemah sekitar 0,027 yang menunjukkan bahwa video dengan lebih banyak balasan yang disukai cenderung memiliki lebih banyak video yang disukai. Korelasi antar variabel *ReplyLikes* dan *videoViews* memiliki korelasi positif yang sedikit lebih kuat sekitar 0,035 yang menunjukkan bahwa video dengan lebih banyak balasan suka cenderung memiliki lebih banyak penayangan video.

3.2 Preprocessing Data

Selanjutnya melakukan *preprocessing* data dengan menggunakan *Google Collab* dilakukan pembersihan data.



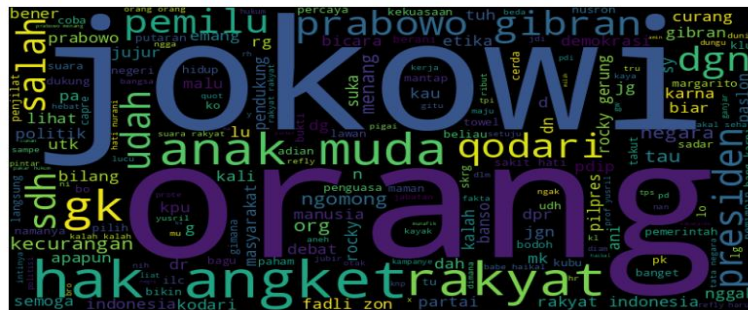
Gambar 5. Tahapan *Preprocessing*

Data dibersihkan dari emoji, url, html, mention twitter, hashtag, titik, koma, symbol, baris kosong, kolom kosong. Pada tahapan data *clearing* pertama kali dilakukan penghapusan data duplikat yang ada di komentar sehingga diperoleh data komentar sejumlah 323.682 data dari sebelumnya 368.299 data komentar.

Tabel 1. Hasil Tahapan *Preprocessing*

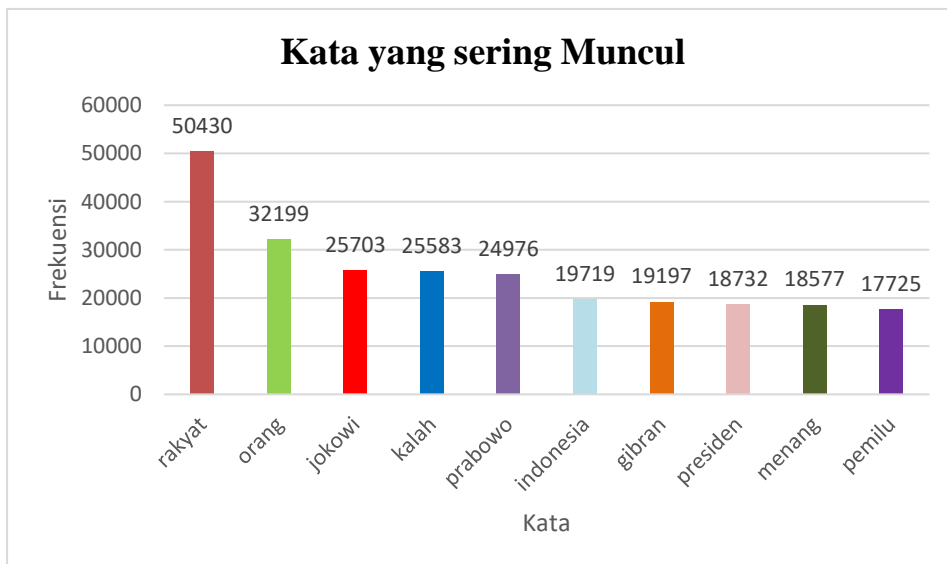
| Komentar | Clening | Case_Folding | Tokenize | Filtering/Stopword Removal | Normalized |
|--|--|--|--|--------------------------------------|--------------------------------------|
| Tempo siapa yang punya,, jangan 2 USA punya,,dia tidak dapat duit kalau tidak hilirisasi | Tempo siapa yang punya jangan USA punyadia tidak dapat duit kalau tidak hilirisasi | tempo siapa yang punya jangan usa punyadia tidak dapat duit kalau tidak hilirisasi | tempo,siapa,yang, punya,jangan,usa, punyadia,tidak,da pat,duit,kalau,tida k,hilirisasi | tempo,usa,p unyadia,duit, hilirisasi | tempo,usa,p unyadia,duit, hilirisasi |

Tahapan selanjutnya melakukan *case folding* atau melakukan konversi semua huruf dalam teks menjadi huruf kecil, dilanjutkan dengan melakukan tokenisasi yang bertujuan untuk memisahkan kata per kata untuk selanjutnya dilakukan penghilangan kata yang tidak penting seperti kata penghubung yang tidak memiliki makna tertentu dan tahapan terakhir yang dilakukan adalah normalisasi data untuk menghapus karakter lain selain kata seperti tanda baca, simbol, spasi. Tahapan *preprocessing steaming* data atau menghilangkan kata yang berhamburan tidak dilakukan karena jumlah data yang digunakan sangat banyak.



Gambar 6. 100 Kata yang Paling sering Muncul setelah *Preprocessing*

Pada Gambar 6 setelah dilakukan *preprocessing* data, kata yang sering muncul sudah mengalami perubahan dari sebelumnya dimana data yang ditampilkan tidak lagi memuat simbol, tanda baca, karakter atau emoji tertentu dan kata hubung sehingga dapat dianalisa lebih lanjut.



Gambar 7. Frekuensi 10 Kata yang Paling sering Muncul setelah *Preprocessing*

Pada Gambar 7 terlihat bahwa 10 kata yang paling sering muncul diantaranya rakyat, orang dan Jokowi, kalah, Prabowo, Indonesia, Gibran, presiden, menang dan pemilu. Secara keseluruhan, data ini dapat memberikan

gambaran tentang topik atau konteks yang dibahas terkait dengan politik, pemerintahan, atau pemilihan umum di Indonesia.

3.3 Analisis Data

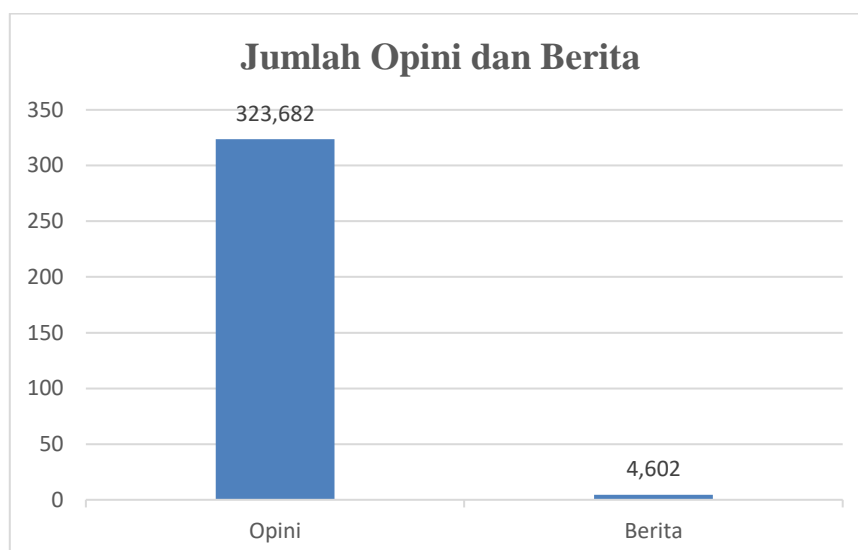
Peneliti mengklasifikasikan dokumen teks yang telah didapatkan kedalam 2 jenis yaitu opini dan berita. Menurut Gerung (2017) Opini merupakan persepsi individu yang terikat secara subjektif [14]. Sedangkan berita menurut M. Lyle Spencer, News Writing dalam Effendy (2023) [15] merupakan kenyataan atau ide yang benar yang dapat diverevisi dan dapat menarik perhatian sebagian besar pembaca. Klasifikasi ini juga penting dilihat secara tajam dengan logika berfikir yang tepat bahwa memungkinkan dalam menganalisis suatu dokumen teks komentar terdapat opini dan berita. Pada setiap opini misalnya “saya tidak suka dan benci kepada anda” dapat ditentukan apakah opini tersebut termasuk kedalam sentimen yang bermuatan positif atau negatif atau sebaliknya dalam sentimen negatif seperti “saya tidak suka anda” bisa diklasifikasikan kedalam kelompok opini suka/tidak suka atau benci/tidak benci.

Sebelum melakukan analisis maka dilakukan pengklasifikasian atau pelabelan data yang telah dinormalisasi tersebut. Setelah diketahui bahwa opini publik terhadap isu populer di video *YouTube* tersebut yaitu terkait dengan pemerintahan Jokowi maka dapat diklasifikasikan komentar yang berupa opini dan berita menggunakan *labeling* dengan *library FastText* [31].

Tabel 2. Hasil Labeling Komentar dengan *FastText*

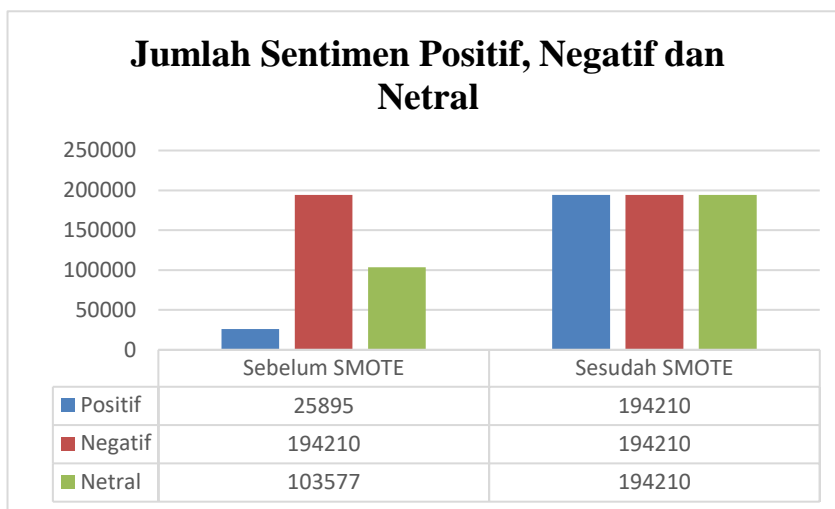
| | <i>Normalized</i> | <i>Label</i> |
|---|--|--------------|
| 0 | tempo usa punyadia duit hilirisasi | opini |
| 1 | ok gas orde gaya | opini |
| 2 | bhaya pakar politik indonesiabagaimana demokras... | opini |
| 3 | sahkan uud jokowi biar orangquot membenci muk... | opini |
| 4 | betulllll | opini |

Komentar di setiap video youtube yang di unggah dapat dipisahkan antara opini dan berita. Labeling berita dan opini dengan *library FastText* didapat bahwa dalam komentar tersebut banyak terdapat opini dengan jumlah 323.682 termasuk label opini dan 4.602 termasuk label berita.



Gambar 8. Jumlah Opini dan Berita pada Komentar dengan *FastText*

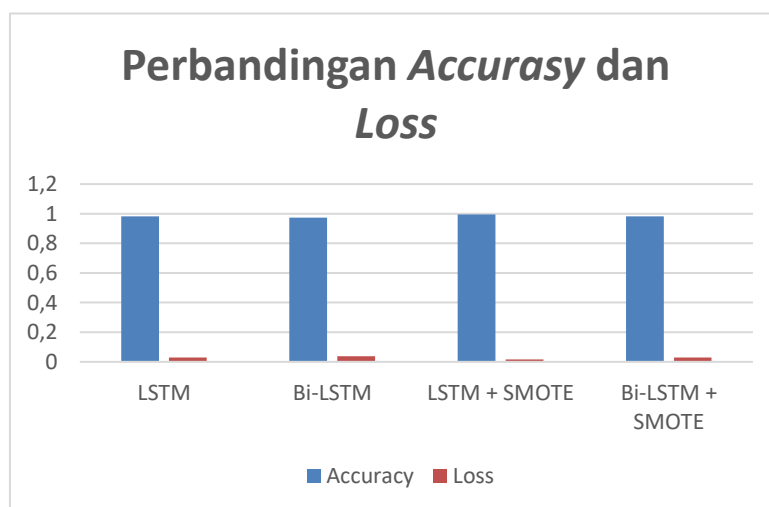
Selanjutnya dari opini yang didapat akan di generalisasikan kedalam sentimen opini publik terhadap isu populer pada tahun 2024 di acara Indonesia lawyers club. Pelabelan dilakukan dengan menggunakan *FastText* untuk sentimen didapati hasil bahwa kelas *sentimen* tidak seimbang sehingga dilakukan proses lebih lanjut dengan metode *SMOTE* dengan *library imblearn over sampling* hal ini dilakukan agar data seimbang.



Gambar 9. Hasil Analisis Sentimen setelah di SMOTE

Sentimen analisis dilakukan dengan metode LSTM (*Long Short Term Memory*). LSTM merupakan varian dari *Recurrent Neural Network* (RNN) yang dirancang untuk menangani masalah yang terkait dengan data sekuensial, seperti teks atau urutan kata. LSTM mampu menangkap konteks atau urutan kata dalam sebuah kalimat atau teks, yang sangat penting dalam menentukan sentimen. Misalnya, kalimat "Saya kurang setuju pemilu diulang" memiliki sentimen negatif, sedangkan "Saya setuju pemilu diulang" memiliki sentimen positif. LSTM dapat membedakan nuansa seperti ini dengan mempertimbangkan urutan kata. LSTM dapat membantu dalam menganalisis teks yang panjang atau kompleks. Pada penelitian ini teknik representasi vektor dari kata atau *word embeddings* yang digunakan dengan *library FastText* karena *FastText* lebih cepat dan lebih efisien dari *Word2Vec*, terutama untuk kosakata yang besar.

Hasil data analisis *sentimen* dengan model LSTM nilai *loss* pada data di akhir *epoch* 0.0162 dengan *accuracy* model pada data di akhir *epoch* sebesar 0.9961. Hasil data analisis sentimen dengan model Bi-LSTM menghasilkan nilai *loss* di akhir *epoch* 0.0284 dengan *accuracy* model sebesar 0.9811. Perbandingan nilai antara kedua metode tersebut dapat dilihat pada Gambar 10.



Gambar 10. Hasil Analisis Sentimen Metode LSTM dan Bi-LSTM

Selanjutnya dilakukan perhitungan hasil kualifikasi dan percobaan kombinasi algoritma untuk mendapatkan hasil kualifikasi yang paling baik. Hasil *evaluation* dapat dilihat pada Tabel 3.

Tabel 3. Hasil Evaluasi

| Algoritma | Precision | Recall | F1- Score | Accuracy |
|------------------------|-----------|--------|-----------|----------|
| <i>LSTM</i> | 96.23% | 96.23% | 96.23% | 98.24% |
| <i>Bi-LSTM</i> | 95.39% | 96.39% | 95.39% | 97.39% |
| <i>LSTM + SMOTE</i> | 99.60% | 99.60% | 99.60% | 99.61% |
| <i>Bi-LSTM + SMOTE</i> | 96.48% | 97.48% | 97.48% | 98.11% |

4. KESIMPULAN

Dari penelitian yang telah dilakukan dapat disimpulkan bahwa isu populer yang sering diperbincangkan di chanel youtube Indonesia lawyers club dari tanggal 01 Januari 2024 hingga 18 Mei 2024 di saluran berita *YouTube* adalah rakyat, orang dan Jokowi, kalah, Prabowo, Indonesia, Gibran, presiden, menang dan pemilu. Secara keseluruhan, data ini dapat memberikan gambaran tentang topik atau konteks yang dibahas terkait dengan pemerintahan jokowi dan pemilu maka dapat diklasifikasikan apakah komentar tersebut termasuk kedalam opini atau berita dengan menggunakan *library fasttext* menghasilkan opini berjumlah 323.682 komentar dan komentar berupa berita berjumlah 4.602. Sejalan dengan itu komentar juga dilakukan pelabelan sentimen menggunakan *library fasttext* menghasilkan komentar positif berjumlah 25.895 dan negatif 194.210 dengan sisanya memilih netral sebanyak 103.577. Hasil dari analisis sentimen dengan model *LSTM* (*Long Short Term Memory*) nilai *loss* pada data di akhir *epoch* 0.0162 dengan *accuracy* model pada data di akhir *epoch* sebesar 0.9961. Hasil data analisis sentimen dengan model *Bi-LSTM* (*Bidirectional Long Short Term Memory*) menghasilkan nilai *loss* di akhir *epoch* 0.0284 dengan *accuracy* model sebesar 0.9811.

REFERENSI

- [1] K. Shanthi, "The Evolution of Authoritarian Digital Influence Grappling with the New Normal," *Prism*, vol. 9, no. 1, pp. 32–51, 2020.
- [2] M. Yasir, M. Grace Haque, R. Suraji, and C. Author, "Analisis Sentimen Terhadap Kontroversi Fatwa MUI Nomor 83 Tahun 2023 Tentang Pemboikotan Produk yang Terafiliasi Israel," *J. Ekon. Manaj. Sist. Inf.*, vol. 5, no. 4, pp. 409–422, Mar. 2024, doi: 10.31933/JEMSI.V5I4.1845.
- [3] T. Daglis and K. P. Tsagarakis, "A LinkedIn-Based Analysis Of The U.S. Dynamic Adaptations In Healthcare During The COVID-19 Pandemic," *Healthc. Anal.*, vol. 5, p. 100291, Jun. 2024, doi: 10.1016/J.HEALTH.2023.100291.
- [4] S. S. Zaidi, A. Perveen, M. A. Alam, J. Kishore, and U. D. Bhardwaj, "Study to Assess the Effectiveness of Behavioural Change Communication Aid for the Anger Management in Adolescents: A Quasi Experimental Study in the Selected Juvenile Aid Center, New Delhi," *Brain Behav. Immun. Integr.*, vol. 6, p. 100057, Apr. 2024, doi: 10.1016/J.BBII.2024.100057.
- [5] S. M. Alhashmi, I. A. T. Hashem, and I. Al-Qudah, "Artificial Intelligence applications in healthcare: A bibliometric and topic model-based analysis," *Intell. Syst. with Appl.*, vol. 21, p. 200299, Mar. 2024, doi: 10.1016/J.ISWA.2023.200299.
- [6] N. Novianty, S. Syarif, and M. Ahmad, "Influence of Breast Milk Education Media on Increasing Knowledge About Breast Milk: Literature Review," *Gac. Sanit.*, vol. 35, pp. S268–S270, Jan. 2021, doi: 10.1016/J.GACETA.2021.10.031.
- [7] A. Năstasă, T. C. Dumitra, and A. Grigorescu, "Artificial intelligence and sustainable development during the pandemic: An overview of the scientific debates," *Heliyon*, vol. 10, no. 9, May 2024, doi: 10.1016/J.HELIVON.2024.E30412.
- [8] Z. M. Yusoff, N. Ismail, and S. A. Nordin, "Dataset for Five Recent Years (2019 – 2023) Agarwood Essential Oil Research Trends: A Bibliometric Analysis," *Data Br.*, vol. 54, p. 110310, Jun. 2024, doi: 10.1016/J.DIB.2024.110310.
- [9] M. Yang, J. He, L. Shi, Y. Lv, and J. Li, "Integrating policy quantification analysis into ecological

- security pattern construction: A case study of Guangdong–Hong Kong–Macao Greater Bay Area,” *Ecol. Indic.*, vol. 162, p. 112049, May 2024, doi: 10.1016/J.ECOLIND.2024.112049.
- [10] R. Pils and P. Schoenegger, “Scientific Realism, Scientific Practice, and Science Communication: An Empirical Investigation of Academics and Science Communicators,” *Stud. Hist. Philos. Sci.*, vol. 105, pp. 85–98, Jun. 2024, doi: 10.1016/J.SHPSA.2024.05.005.
- [11] L. H. Chowdhury, S. Islam, and S. Shatabda, “A Bengali News And Public Opinion Dataset From Youtube,” *Data Br.*, vol. 52, p. 109938, Feb. 2024, doi: 10.1016/J.DIB.2023.109938.
- [12] Y. Dokuz, “Discovering Popular And Persistent Tags From Youtube Trending Video Big Dataset,” *Multimed. Tools Appl.*, vol. 83, no. 4, pp. 10779–10797, Jan. 2024, doi: 10.1007/S11042-023-16019-Z/METRICS.
- [13] I. Rozi, S. Pramono, and E. Dahlan, “Implementasi Opinion Mining (Analisis Sentimen) Untuk Ekstraksi Data Opini Publik Pada Perguruan Tinggi,” *J. EECCIS*, vol. 6, no. 1, pp. 37–43, 2012.
- [14] R. Gerung, “Pengetahuan Membebaskan Kita dari Opini yang Keliru,” *Jurnal Perempuan*, p. 1, 2017. Accessed: Oct. 24, 2024. [Online]. Available: <https://www.jurnalperempuan.org/warta-feminis/rocky-gerung-pengetahuan-membebas-kita-dari-opini-yang-keliru>
- [15] E. Effendy, Zakaria, Azlisa, and Anggarana, “Dasar Dasar Penulisan Berita,” *J. Pendidik. dan Konseling*, vol. 5, no. 2, pp. 4042–4044, 2023, [Online]. Available: <https://journal.universitaspahlawan.ac.id/index.php/jpdk/article/view/13888>
- [16] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, “Sentiment Analysis of Social Media Twitter with Case of Anti-LGBT Campaign in Indonesia using Naïve Bayes, Decision Tree, and Random Forest Algorithm,” *Procedia Comput. Sci.*, vol. 161, pp. 765–772, Jan. 2019, doi: 10.1016/J.PROCS.2019.11.181.
- [17] H. Nurriqfi, F. Fikrillah, and D. Kurniadi, “Rekomendasi Pemilihan Program Studi Menggunakan Algoritma Naïve Bayes,” *J. Algoritma.*, vol. 20, no. 1, pp. 42–49, May 2023, doi: 10.33364/ALGORITMA/V.20-1.1236.
- [18] A. Karimah, G. Dwilestari, and M. Mulyawan, “Analisis Sentimen Komentar Video Mobil Listrik Di Platform Youtube Dengan Metode Naive Bayes,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 767–737, Mar. 2024, doi: 10.36040/JATI.V8I1.8373.
- [19] P. Y. Saputra, D. H. Subhi, and Fahmi Zain Afif Winatama, “Implementasi Sentimen Analisis Komentar Channel Video Pelayanan Pemerintah Di Youtube Menggunakan Algoritma Naïve Bayes,” *J. Inform. Polinema*, vol. 5, no. 4, pp. 209–213, Aug. 2019, doi: 10.33795/jip.v5i4.259.
- [20] D. Mualfah, R. Gunawan, D. Mulyadipa Suratno, and A. Sentimen Komentar YouTube TvOne Tentang Ustadz Abdul Somad Dideportasi Dari Singapura Menggunakan, “Analisis Sentimen Komentar YouTube TvOne Tentang Ustadz Abdul Somad Dideportasi Dari Singapura Menggunakan Algoritma SVM,” *J. FASILKOM*, vol. 13, no. 01, pp. 72–80, Jul. 2023, doi: 10.37859/JF.V13I01.4920.
- [21] S. Mujahidin, B. Prasetyo, and M. C. C. Utomo, “Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode Gaussian naïve bayes,” *Voteteknika (Vocational Tek. Elektron. dan Inform.)*, vol. 10, no. 3, pp. 17–24, Sep. 2022, doi: 10.24036/VOTETEKNIKA.V10I3.118299.
- [22] Nurmawiyana and K. A. Harvian, “Public Sentiment Towards Face-to-Face Activities During the COVID-19 Pandemic in Indonesia,” *Procedia Comput. Sci.*, vol. 197, pp. 529–537, Jan. 2022, doi: 10.1016/J.PROCS.2021.12.170.
- [23] S. Srivastava, C. Chakraborty, and M. K. Sarkar, “Leveraging Machine Learning and Dimensionality Reduction for Sports And Exercise Sentiment Analysis,” *Meas. Sensors*, vol. 33, p. 101182, Jun. 2024, doi: 10.1016/J.MEASEN.2024.101182.
- [24] A. Wahid and G. Saputri, “Analisis Sentimen Komentar Youtube Tentang Relawan Patwal Ambulance Menggunakan Algoritma Naïve Bayes dan Decision Tree,” *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, pp. 319–326, Dec. 2022, doi: 10.30865/json.v4i2.4941.
- [25] M. S. Islam, M. A. T. Rony, M. Ahammad, S. M. N. Alam, and M. S. Rahman, “An Innovative Novel Transformer Model and Datasets for Safeguarding Religious Sensitivities in Online Social Platforms,” *Procedia Comput. Sci.*, vol. 233, pp. 988–997, Jan. 2024, doi: 10.1016/J.PROCS.2024.03.288.
- [26] S. Kaur, S. Singh, and S. Kaushal, “Deep Learning-Based Approaches For Abusive Content Detection And Classification For Multi-Class Online User-Generated Data,” *Int. J. Cogn. Comput. Eng.*, vol. 5,

- pp. 104–122, Jan. 2024, doi: 10.1016/J.IJCCE.2024.02.002.
- [27] M. Lestandy and Abdurrahim, “Exploring the Impact of Word Embedding Dimensions on Depression Data Classification Using BiLSTM Model,” *Procedia Comput. Sci.*, vol. 227, pp. 298–306, Jan. 2023, doi: 10.1016/J.PROCS.2023.10.528.
- [28] B. P. Nayoga, R. Adipradana, R. Suryadi, and D. Suhartono, “Hoax Analyzer for Indonesian News Using Deep Learning Models,” *Procedia Comput. Sci.*, vol. 179, pp. 704–712, Jan. 2021, doi: 10.1016/J.PROCS.2021.01.059.
- [29] D. Priadana, Sidik; dan Sunarsi, “Metode Penelitian Kuantitatif,” in *Education*, 1st ed., Tangerang Selatan: Pascal Books, 2021, ch. 1, pp. 1–215.
- [30] R. Akbar, R. A. Siroj, M. Win Afgani, and U. Islam Negeri Raden Fatah Palembang Abstract, “Experimental Research Dalam Metodologi Pendidikan,” *J. Ilm. Wahana Pendidik.*, vol. 9, no. 2, pp. 465–474, Jan. 2023, doi: 10.5281/ZENODO.7579001.
- [31] C. Li *et al.*, “Pyvisvue3d3: Python Visualization From Hierarchy Tree To Call Graph,” *SoftwareX*, vol. 26, p. 101689, May 2024, doi: 10.1016/J.SOFTX.2024.101689.