TESIS – ES5401

PEMETAAN SENTIMEN MEDIA DAN HUBUNGAN ENTITAS DALAM BERITA TEKNOLOGI PERTANIAN

ELISA NUR SYAFIATUL

NRP 6025232010

Dosen Pembimbing

Amalia Utamima, S.Komn, MBA., Ph.D

NIP 197103120091001

Program Studi Magister

Departemen Sistem Informasi

Fakultas Teknologi Elektro dan Informastika Cerdas

Institut Teknologi Sepuluh Nopember

Surabaya

2025

TESIS – ES5401

PEMETAAN SENTIMEN MEDIA DAN HUBUNGAN ENTIAS DALAM BERITA PERTANIAN

ELISA NUR SYAFIATUL MUFIDA

NRP 6026232010

Dosen Pembimbing

Amalia Utamima, S.Kom., MBA., Ph.D

NIP 197103120091001

Program Studi Magister

Departemen Sistem Informasi

Fakultas Teknologi Elektro dan Informatika Cerdas

Institut Teknologi Sepuluh Nopember

Surabaya

2025

THESIS – ES5401

MAPPING MEDIA SENTIMENT AND ENTITY RELATIONSHIP IN AGRICULTURAL TECHNOLOGY NEWS

ELISA NUR SYAFIATUL MUFIDA

NRP 6026232010

Advisor

Amalia Utamima, S.Kom., MBA., Ph.D

NIP 197103120091001

Study Program Magister

Department of Information System

Faculty of Intelligent Electrical and Informatics Technology

Institut Teknologi Sepuluh Nopember

Surabaya

2025

LEMBAR PENGESAHAN TESIS

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Sistem Informasi (M.Kom.)
di
Institut Teknologi Sepuluh Nopember

Oleh:
Elisa Nur Syafiatul Mufida
NRP: 6026232010

Tanggal Ujian: 24 November 2025
Periode Wisuda ITS: 132

Disetujui oleh:
Pembimbing:

Amalia Utamima, S.Kom, MBA, Ph.D
NIP: 198612132015042001

Penguji:

Faizal Mahananto, S.Kom, M.Eng, Ph.D
NIP: 5200.201301010

Retno Aulia Vinarti, S.Kom., M.Kom., Ph.D
NIP: 1988201812010

Surabaya, 06 Agustus 2025
Kepala Departemen Sistem Informasi
Fakultas Teknologi Elektro dan Informatika Cerdas

LP/P/25/438

Prof. Dr. Wiwik Anggraeni, S.Si, M.Kom
NIP: 197601232001122002

PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini:

Nama mahasiswa / NRP	:	Elisa Nur Syafiatul Mufida / 6026232010
Program studi	:	S2 Sistem Informasi
Dosen Pembimbing / NIP	:	Amalia Utamima, S.Kom, MBA, Ph.D / 198612132015042001

dengan ini menyatakan bahwa Tesis dengan judul "PEMETAAN SENTIMEN MEDIA DAN HUBUNGAN ENTITAS DALAM TEKNOLOGI PERTANIAN" adalah hasil karya sendiri, bersifat orisinal, dan ditulis dengan mengikuti kaidah penulisan ilmiah.

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya
bersedia menerima sanksi sesuai dengan ketentuan yang berlaku di Institut Teknologi Sepuluh Nopember.

Surabaya, 05 Agustus 2025

Mengetahui

Dosen Pembimbing

Amalia Utamima, S.Kom, MBA, Ph.D
NIP. 198612132015042001

Mahasiswa

Elisa Nur Syafiatul Mufida
NRP. 6026232010

ABSTRAK

PEMETAAN SENTIMEN MEDIA DAN HUBUNGAN ENTITAS DALAM BERITA TEKNOLOGI PERTANIAN

Nama Mahasiswa /NRP	**:**	**Elisa Nur Syafiatul Mufida / 6026232010**
Departemen	**:**	**Sistem Informasi - ITS**
Dosen Pembimbing	**:**	**Amalia Utamima, S.Kom., MBA., Ph.D.**

Abstrak

Digitalisasi pertanian memiliki perkembangan yang begitu pesat dengan adanya dukungan teknologi seperti Big Data, Internet of Things, dan Artificial Intelligence yang didorong oleh peran media sebagai pembentuk opini publik. Penelitian ini mengintegrasikan Machine Learning, dan Knowledge Graph untuk memetakan sentimen dan relasi antar entitas dalam 1.295 artikel oleh BBC, The Guardian, The Independent, dan Reuters, pada publikasi 2013-2025. Hasil klasifikasi menunjukan model BERT dengan Random Forest unggul dengan hasil akurasi 92% dan F1-macro rata-rata 87.2% menjadi baseline terbaik dibanding dengan BERT + SVM (91%), dan BERT+XGBoost (91%). Pada tahap pemetaan relasi, triples hasil NER dan Relation Extraction dibangun pada Knowledge Graph, kemudian dieksplorasi menggunakan Graph Convolutional Network. Klaster K-Means menunjukkan lebih dari 70% relasi masih berupa open triad, menandakan celah koneksi. Enrichment diikuti pemisahan komunitas Louvain meningkatkan modularitas menjadi 0,45 dan membentuk empat komunitas yang sejalan dengan enam topik utama: *Food Security & Risk Management*, *Economic & Family Farming*, *Supply Chain & Trade*, *Technology Adoption*, *Climate & Environment*, serta *Investment & Market*. Sentimen Pro mendominasi 70% pada topik ketahanan pangan, sedangkan komunitas perdagangan memuat porsi Contra tertinggi, yakni 40 %. Secara keseluruhan, pendekatan gabungan KG–GCN–ML efektif memperkaya analisis sentimen berita teknologi pertanian, mengungkap pola topik sentimen sekaligus peluang perluasan relasi antar entitas untuk riset kebijakan dan strategi adopsi teknologi pertanian.

Kata kunci: ***Bidirectional Encorder Representations from Transformers, Random Forest, Konowledge Graph, XGBoost, Support Vector Machine***.

ABSTRACT

MAPPING MEDIA SENTIMENT AND ENTITY RELATIONSHIP IN AGRICULTURAL TECHNOLOGY NEWS

Student Name / NRP	**:**	**Elisa Nur Syafiatul Mufida / 6026232010**
Department	**:**	**Sistem Informasi - ITS**
Advisor	**:**	**Amalia Utamima, S.Kom., MBA., Ph.D.**

Abstract

Agricultural digitalization propelled by IoT, AI, and Big Data, and amplified by media influence is rapidly evolving. This study combines Knowledge Graphs (KG) and Machine Learning to map sentiment across 1,295 articles news from BBC, *The Guardian*, *The Independent*, and Reuters published between 2013 and 2025. The best classifier, a BERT model paired with a Random Forest, attains 92% accuracy and an 88 %F1-macro, outperforming BERT-SVM and BERT-XGBoost. For relation mapping, NER- and RE-derived triples are built into a KG and explored with a Graph Convolutional Network. Initial k-means clustering shows that over 70 % of the triads are open, revealing connectivity gaps; community enrichment with Louvain raises modularity to 0.45 and yields four communities aligned with six key topics *Food Security and Risk Management*; *Economic and Family Farming*; *Supply Chain and Trade; Technology Adoption*; *Climate and Environment; and Investment and Market*. Pro sentiment dominates food-security coverage at roughly 70 %, while trade-focused articles carry the highest Contra share, at 40 %. Overall, the integrated KG–GCN–ML framework enriches agritech sentiment analysis, unveiling topic–sentiment patterns and opportunities for expanding entity relations in policy and technology-adoption research.

Keywords: ***Bidirectional Encoder Representations from Transformers, Graph Convolutional Network, Knowledge Graph, Random Forest, Support Vector Machine***.

KATA PENGANTAR

Segala puji dan rasa syukur penulis sampaikan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, tesis yang berjudul "Pemetaan Sentimen Media dan Hubungan Entitas dalam Berita Teknologi Pertanian"dapat diselesaikan dengan baik dan tepat waktu. Adapun tesis berikut sebagai salah satu syarat untuk memperoleh gelas Magister pada Program Studi Sistem Informasi, Fakultas Teknologi Elektro dan Informatika Cerdas (FTEIC), Institut Teknologi Sepuluh Nopember (ITS). Tesis ini disusun sebagai bentuk kontribusi akademik dalam menjawab berbagai tantangan dan dinamika perkembangan teknologi pada sektor pertanian, khususnya dalam konteks peran media dalam membentuk opini publik melalui pendekatan analisis sentimen dan pemetaan hubungan entitas dalam berita menggunakan pendekatan grafik.

Penulis menyadari bahwa tanpa bimbingan, dukungan, dan do'a dari berbagai pihak, penyusunan tesis ini tidak akan dapat terselesaikan dengan baik. Oleh karena itu, pada bagian selanjutnya penulis ingin menyampaikan rasa terima kasih yang tulus kepada semua pihak yang telah membantu secara langsung maupun tidak dalam penyusunan dalam proses penyusunan tesis ini.

1. Kedua orang tua penulis, Bapak dan Ibu atas kasih sayang, maupun segala bentuk dukungan berupa do'a, harapan, motivasi, dan kepercayaan yang selalu diberikan kepada penulis selama proses pengerjaan tesis ini hingga akhir tahapan.
2. Kepada nenek penulis atas segala macam bentuk do'a, dukungan, motivasi, serta harapan dan kepercayaan yang diberikan kepada penulis hingga penulis dapat menyelesaikan sampai tahapan terakhir dalam kepenulisan, maupun serangkaian proses yang dilalui.
3. Ibu Amalia Utamima, S.Kom., MBA., Ph.D., selaku dosen pembimbing penulis atas segala bentuk dukungan, memberikan motivasi, arahan, dan saran yang membangun pada disetiap kesempatan, maupun bimbingan yang tiada henti kepada penulis, sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik, sesuai dengan kaidah akademik, dan tepat waktu.
4. Bapak Faizal Mahananto, S.Kom., M.Eng., Ph.D, selaku dosen penguji I, atas segala saran dan arahan yang membangun dalam proses akhir tahapan tesis berikut, sehingga tesis berikut dapat memberikan hasil yang baik dan sesuai dengan kaidah akademik sebagaimana mestinya.

5. Ibu Retno Aulia Vinarti, S.Kom., M.Kom., Ph.D selaku dosen penguji II atas saran, dan arahan yang membangun pada proses akhir pengerjaan tesis berikut, sehingga penulis dapat memahami disetiap celah pada penelitian untuk memberikan hasil yang baik sesuai dengan kaidah akademik, maupun segala bentuk dukungan, serta bimbingan sehingga penulis dapat menyelesaikan tesis berikut dengan tepat waktu.
6. Kepada rekan-rekan terdekat penulis, Soraya Mitha, Nadya, Yulanda, Diah, Ratna dan Desinta yang telah menjadi sandaran, dan menemani perjalanan penulis dalam segala proses. Terima kasih atas do'a, harapan, dukungan, kepercayaan, motivasi yang tiada henti diberikan kepada penulis, serta masukan yang membangun disetiap kesempatan.
7. Kepada rekan-rekan seperjuangan penulis, Citra, Mitari, Millatina, Rifdah, dan teman-teman seperjuangan lainnya di S2-Sistem Informasi Semester Genap 2024. Terima kasih telah menjadi teman diskusi, menemani perjalanan penulis, dan selalu memberikan insight, maupun motivasi, harapan, serta do'a kepada penulis dalam segala macam bentuk kesempatan.
8. Kepada pihak-pihak terkait lainnya yang memberikan do'a, motivasi, dukungan, informasi, serta diberikannya kepercayaan kepada penulis hingga penulis dapat menyelesaikan kepenulisan tesis berikut dengan baik, dan tepat waktu.

Dalam tahapan kepenulisan tesis ini, penulis menyadari masih memiliki keterbatasan dan kekurangan. Sejalan dengan hal tersebut, penulis sangat terbuka terhadap kritik serta saran yang membangun pada kepenulisan maupun riset yang dilakukan penulis.

Sebagai penutup, penulis berharap tesis ini dapat memberikan kontribusi ilmiah bagi pengembangan pengetahuan, khususnya dalam bidang analisis sentimen media, digitalisasi sektor pertanian, dan pemanfaatan teknologi machine learning dalam riset sosial. Semoga kepenulisan berikut bermanfaat bagi akasemisi, praktisi, dan semua pihak yang memiliki perhatian terhadap isu inovasi teknologi dalam pembangunan pertanian berkelanjutan.

Surabaya, 24 Juli 2025

Elisa Nur Syafiatul Mufida

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Ketahanan pangan global dan pembangunan berkelanjutan tidak dapat dilepaskan dari peran vital pada sektor pertanian. Dalam beberapa dekade terakhir, sektor pertanian berada di garis depan dalam menghadapi tuntutan pada perubahan iklim yang sering berubah secara signifikan tanpa adanya tanda-tanda yang nyata sebelumnya hingga peningkatan kebutuhan pangan akibat pertumbuhan populasi global. *Food and Agriculture Organization of the United Nations* (FAO), memperkirakan kebutuhan pangan global akan meningkat seiring jumlah populasi dunia yang mencapai 9,7 miliar pada tahun 2050. Hal tersebut perlu adanya analisis lebih lanjut sebagai upaya peningkatan ketahanan pangan secara global di era modernisasi seperti yang terjadi saat ini. Untuk mengatasi masalah tersebut, inovasi digitalpun berkembang secara berkelanjutan, seperti *Internet of Things* (*IoT*), *Artifical Intelligence* (*AI*), dan *Big Data* yang telah diadopsi untuk menunjang produktivitas, efisiensi, dan keberlanjutan pada sektor pertanian (Mohr & Hohler, 2023). Adanya teknologi yang berkembang memberikan kemudahan bagi petani untuk memantau kondisi lahan secara *real-time*, memprediksi hasil panen secara akurat, serta analisis data lingkungan guna menunjang hasil panen yang berkualitas (Xu et al., 2022).

Digitalisasi memainkan peran penting dalam transformasi sektor pertanian, peran tersebut dapat menawarkan solusi sebagai peningkatan produktivitas dan efisiensi sumber daya. Namun, penerapan teknologi yang ada memunculkan tantangan baru pada masyarakat, terutama terkait adopsi teknologi oleh petani yang sering kali dipengaruhi opini yang dibentuk media massa (Z. Chen et al., 2022). Media memiliki peran dalam membentuk opini publik dan mempengaruhi regulasi melalui *framing* berita. Studi terbaru menunjukkan bahwa *framing* oleh media dapat membentuk pemahaman dan sikap publik terhadap isu-isu kompleks (Lindgren et al., 2022). Pembentukan opini publik oleh pemberitaan media mendorong penggunaan teknologi terus berkembang maupun sebaliknya, dimana adopsi teknologi dapat terhambat oleh beberapa pemberitaan yang dirasa kurang menguntungkan. Pemahaman terhadap narasi

media dalam membingkai suatu berita perlu ditekankan, dimana hal tersebut akan mempengaruhi opini publik guna menunjang keberhasilan adopsi teknologi (Chen et al., 2022).

Beberapa studi terkini telah mengeksplorasi peran media dalam mendorong adopsi teknologi di sektor pertanian melalui pendekatan analisis yang beragam. Analisis wacana dan analisis sentimen menjadi pendekatan umum yang digunakan sebagai bahan evaluasi pandangan masyarakat terhadap isu-isu tertentu termasuk digitaliasai sektor pertanian. Studi-studi tersebut memanfaatkan berbagai jenis media, mulai dari media sosial *Twitter* (Wester et al., 2023) hingga media cetak (Z. Chen et al., 2022) dan portal berita (Mohr & Höhler, 2023);(Wester et al., 2023). Selain itu penilitian terkait pengolahan data menggunakan *machine learning* sebagai analisis sentimen pengaruh media yang ada pada perdesaan daerah terpencil di Fuzhou, China (Z. Chen et al., 2024). Metode ini bertujuan untuk memahami tren digitalisasi dan membangun strategi komunikasi yang lebih efektif guna mendorong transformasi di sektor pertanian (Ancín et al., 2022).

Kendati demikian, penelitian sebelumnya seringkali berfokus pada media sosial seperti *Twitter*, *Facebook*, dan artikel berita pada media cetak. Penggunaan penelitian serupa pada portal berita berbasis website masih terbatas. Portal berita seperti *BBC*, *Reuters*, *The Guardian* dan *The Independent* memiliki pemberitaan yang berbeda dibandingkan media sosial, dimana pemberitaan memiliki pendekatan jurnalistik yang lebih mendalam, dengan editorial yang ketat dan berbasis riset (Hermida et al., 2011). Selain itu, gaya peliputan dan *framing* isu pada media portal juga lebih kompleks dibandingkan dengan media sosial (Fletcher et al., 2020). Pemilihan portal berita berikut dalam penelitian ini di dasari adanya jangkauan audiens global dan kualitas jurnalististik yang tinggi, serta banyaknya media-media tersebut dijadikan rujukan utama dalam perumusan kebijakan publik dan regulasi internasional. Pemberitaan melalui portal yang dipilih juga menyajikan pemberitaan yang cenderung bebas dari bias lokal sehingga memungkinkan analisis yang lebih objektif dan representatif terhadap narasi global pada isu digitalisasi pertanian. Sejalan dengan hal tersebut, media pada negara-negara maju memiliki kebijakan digitalisasi

pertanian yang baik dan banyak disorot oleh media global sebagai media yang memiliki pengaruh signifikan terhadap opini publik dan regulasi di bidang pertanian (Rust et al., 2021). Dengan demikian, pemahaman bagaimana media tersebut membingkai narasi tentang digitaliasasi pertanian menjadi esensial sebagai pemahaman pola opini publik serta faktor-faktor apa saja yang dapat mempengaruhi adopsi teknologi di skala global (Pérez-Mesa et al., 2023) .

Penelitian ini bertujuan untuk mengeksplorasi lebih dalam bagaimana portal berita membingkai isu digitalisasi pertanian menggunakan pendekatan *Natural Laguange Processing* (*NLP*) dan *Machine Learning* sebagai media yang lebih efektif untuk menganalias sentimen dalam artikel berita. Penggunaan algoritma seperti *Bidirectional Encorder Representations from Transformers* (BERT), kombinasi *Bidirectional Encorder Representations from Transformers* (BERT) + *Support Vector Machine* (SVM), kombinasi *Bidirectional Encorder Representations from Transformers* (BERT) + *Random Forest*, serta kombinasi *Bidirectional Encorder Representations from Transformers* (BERT) + *XGBoost* memiliki kemampuan dalam mengolah teks dan analisa lebih lanjut terkait klasifikasi sentimen dalam artikel berita menjadi kategori positif, negatif, atau netral (*"Pro", Contra","Neutral").*

Selain pendekatan algoritma sebelumnya, terdapat pendekatan *Knowledge Graph* sebagai media untuk memahami keterkaitan antara entitas dalam teks berita, seperti hubungan antara teknologi pertanian, kebijakan pemerintah, maupun persepsi atau opini publik.

Berdasarkan studi sebelumnya, terdapat beberapa kesenjangan dalam literatur terkait peran media dalam digitalisasi pertanian, yaitu minimnya penelitian yang menggabungkan pendekatan multimodel NLP dan *machine learning* dalam klasifikasi sentimen berita berbasis website. Keterbatasan selanjutnya pemanfaatan *knowledge graph* sebagai pemetaan hubungan antar entitas dan analisis *framing media* dalam konteks digitalisasi pertanian, serta minimnya eksplorasi hubungan antara komunitas entitas motif, dan distribusi sentimen dalam visualisasi GCN. Hal-hal tersebut memberikan celah bagi peneliti untuk melakukan penelitian untuk menjawab kesenjangan tersebut dengan mengusulkan pendekatan terkait analisis sentimen dan *knowledge graph*.

Dengan pendekatan berbasis *machine learning* dan integrasi *knowledge graph* dapat mewujudkan tujuan dari penelitian ini sebagai analisis sentimen yang mendalam secara pemetaan keterkaitan antar entitas yang muncul dalam berita terkait digitaliasasi pertanian. Hasil penelitian yang diharapkan dapat memberikan informasi yang lebih dalam bagi *stakeholder*, akademisi, dan industri teknologi pertanian tentang bagaimana opini publik serta strategi komunikasi yang dapat dioptimalkan untuk mendorong transformasi digital di sektor pertanian.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini berfokus pada bagaimana media berita terkenal membingkai isu digitaliasasi pertanian serta bagaimana analisa sentimen menggunakan *machine learning* dan *knowledge graph* dapat digunakan untuk memahami pola pemberitaan. Sejalan dengan hal tersebut, media sendiri memiliki peran penting dalam membentuk opini publik, terutama dalam mempertajam persepsi terhadap adopsi teknologi pada pertanian.

Berdasarkan hipotesis yang terbentuk sebelumnya, rumusan masalah yang ingin dicapai dalam riset ini adalah sebagai berikut:

1. Bagaimana analisis sentimen berbasis *Machine Learning* dapat mengidentifikasi pola pemberitaan teknologi pertanian?
2. Bagaimana *Knowledge Graph* dapat memetakan relasi antar entitas dalam pemberitaan digitalisasi pertanian?

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk mengkaji hubungan antara *framing* media, sentimen berita, serta keterkaitan antar-entitas dalam pemberitaan digitalisasi pertanian. Selaras dengan tujuan tersebut, penelitian memadukan pendekatan *Knowledge Graph* yang digunakan sebagai pemetaan hubungan antar-entitas dalam teks berita serta *Machine Learning* berbasis *Bidirectional Encorder Representations from Transformers* (BERT), kombinasi *Bidirectional Encorder*

Representations from Transformers (BERT) + *Support Vector Machine* (SVM), *Bidirectional Encorder Representations from Transformers* (BERT) + *Random Forest*, serta *Bidirectional Encorder Representations from Transformers* (BERT) + *XGBoost* dalam analisis sentimen. Melalui penelitian yang dilakukan, peneliti mengharapkan dapat diperoleh informasi yang lebih mendalam mengenai pemberitaan terhadap opini publik pada adopsi teknologi pertanian. Hal tersebut diharapkan juga dapat menjadi dasar dalam perumusan strategi komunikasi yang lebih efektif bagi *stakeholder* dalam sektor pertanian.

1.4 Kontribusi Penelitian

Penulis melakukan kontribusi baru dalam penelitian terkait analisis sentimen pada berita digitaliasi pertanian yang belum terdapat pada penelitian sebelumnya. Kontribusi ini ditujukan untuk menjawab beberapa kesenjangan gap, antara lain:

1. Mengisi kekosongan integrasi pada pendekatan multimodel *Natural Language Processing* dan *machine learning*, seperti *Bidirectional Encorder Representations from Transformers* (BERT), *Bidirectional Encorder Representations from Transformers* (BERT) + *Support Vector Machine* (SVM)**,** kombinasi *Bidirectional Encorder Representations from Transformers* (BERT) + *Random Forest*, dan *Bidirectional Encorder Representations from Transformers* (BERT) + *XGBoost* untuk klasifikasi sentimen "Pro", "Contra", dan "Neutral" pada korpus berita teknologi pertanian.
2. Menjawab keterbatasan pemanfaatan *Knowledge Graph* dalam memetakan relasi entitas dan sentimen dalam wacana media berita digitalisasi pertanian, terlebih portal berita berbasis website seperti *BBC*, *Reuters*, *The Guardian*, dan *The Independent* sebagai portal media yang lebih luas dan mendalam terhadap pemberitaan.
3. Mengisi celah eksplorasi framing media berbasis topik dan komunitas entitas, dengan menerapkan analisis *community detection* (Louvain), distribusi motif relasi sentimen, serta visualisasi entitas dan topik pada

hasil *Knowledge Graph* dengan menggunakan *Graph Convotuinal Network* (GCN).

Secara keseluruhan, penelitian berikut berkontribusi dalam mengisi kesenjangan literatur terkait penerapan *machine learning* dan *knowledge graph* pada berita digitalisasi pertanian, serta menyajikan pendekatan yang lebih mendalam dan terstruktur dalam pemahaman terkait opini publik yang dipengaruhi oleh *framing media* terhadap adopsi teknologi pertanian.

1.5 Batasan Masalah

Dalam penelitian yang diajukan, terdapat beberapa batasan permasalahan yang dihadapi, seperti:

1. Penelitian menggunakan algoritma *Machine Learning* seperti *Natural Language Processing* (NLP) dan algoritma *hybrid machine learning* seperti *Bidirectional Encorder Representations from Transformers* (BERT), dan *Bidirectional Encorder Representations from Transformers* (BERT) + *Support Vector Machine* (SVM)**,** *Bidirectional Encorder Representations from Transformers* (BERT) + *Random Forest* dan *Bidirectional Encorder Representations from Transformers* (BERT) + *XGBoost***.**
2. Sumber data yang digunakan hanya mencangkup sumber data berita media berbasis web (*BBC, Reuter, The Guardian, dan The Independent*) dalam kurun waktu 2013 – 2025.
3. Menggunakan analisis *Knowledge Graph* dengan fokus pada hubungan antar entitas yang terkait dengan digitalisasi pertanian, seperti teknologi, kebijakan, dan *stakeholder* tanpa melakukan eksplorasi mendalam terhadap tema lain dalam berita.

BAB 2 TINJAUAN PUSTAKA

2.1 Hasil Penelitian Terdahulu

Penelitian yang dilakukan terkait analisis sentimen media dalam digitalisasi teknologi pertanian mengisi beberapa kesenjangan pada literatur. Beberapa aspek dari peran media dalam mempengaruhi opini publik terhadap digitaliasasi pertanian sebelumnya telah diteliti, namun beberapa penelitian perlu adanya pengembangan lebih lanjut. Pada penelitian sebelumnya menunjukkan bahwa media memiliki pengaruh besar dalam membentuk opini publik dan memengaruhi penerimaan teknologi di sektor pertanian selama beberapa dekade terakhir. Misalnya, dalam penelitian oleh **Ancín, Pindado, dan Sánchez (2022)** menggunakan data dari *Twitter* untuk mengeksplorasi transformasi digital di sektor *agrifood* secara global. Studi tersebut berfokus pada adopsi teknologi digital seperti *Artificial Intellegience*(AI) dan *Internet of Things* (IoT) pada sektor agrifood ataupun sektor pertanian lainnya dapat membangun persepsi publik terhadap inovasi yang berkembang. Peneliti mengemukakan bahwa pandemi COVID-19, memberikan dampak terhadap perkembangan teknologi pertanian, didorong oleh banyaknya diskusi di media sosial terkait pentingnya digitalisasi dalam *agrifood*. Studi tersebut menggunakan analisis dengan pendekatan data dan *sentiment analysis* untuk mengklasifikasikan sentimen publik menjadi positif, negatif, atau netral. Salah satu hasil utama yang ditemukan dalam penelitian ini adalah sekitar 80% dari *tweet* yang dianalisis memiliki sentimen positif terhadap transformasi digital dalam sektor *agrifood* yang menunjukkan antusias yang tinggi masyarakat terhadap implementasi teknologi digital dalam agrifood. Dalam penelitian juga disebutkan bahwa sentimen positif di dominasi oleh negara-negara maju yang memiliki infrastruktur teknologi yang lebih berkembang, sementara negara-negara berkembang menunjukkan perbedaan adopsi akibat tantangan dalam akses dan literasi digital. Selain hal tersebut, ditemukan bahwa isu keberlanjutan dan perubahan iklim berperan dalam membentuk persepsi publik terhadap digitaliasi dalam sektor pertanian dan makanan. Wacana yang muncul dalam media sosial menyoroti bahwa teknologi digital dapat membantu menciptakan sistem pertanian yang lebih berkelanjutan dengan efisiensi sumber data yang lebih baik. Namun,

pada penelitian ini masih memiliki keterbatasan dalam penerapan *machine learning* pada penggunaan *sentiment analysis*, seperti penggunaan *deep learning* atau *knowledge graph* untuk menghubungkan berbagai entitas dalam diskusi tentang digitalisasi pertanian. (Ancín et al., 2022). Pada studi yang dilakukan oleh **Mohr & Hohler (2023)**, mengeksplorasi bagaimana media masa membingkai digitalisasi dalam sektor pertanian melalui analisis konten berbasis *Natural Language Processing* (NLP). Di mana penelitian ini memiliki fokus pada analisis pemberitaan digitalisasi pertanian di Jerman dengan metode koding tematik untuk mengklasifikasikan argumen bersifat pro, Contra, dan netral dalam berita media cetak. Dataset pada penelitian ini menggunakan 88 artikel berita dari berbagai media besar yang dianalisis menggunakan *inductive content analysis* dengan perangkat lunak *MaxQDA*. Hasil penelitian menunjukkan bahwa 59% pemberitaan bersifat positif (pro), 23, 4 % negative (Contra), dan 17,6% netral. Topik atau argument beberapa pemberitaan di analisis berdasarkan kalimat yang sering muncul pada beberapa kategori pro, Contra, dan netral. Pada argumen yang sering muncul dalam pemberitaan positif seperti kemudahan kerja, pengurangan penggunaan pupuk dan peptisida, dan peningkatan keberlanjutan. Sedangkan argumen Contra yang sering muncul pada topik jaringan internet, dominasi pasar oleh penyedia teknologi, dan isu privasi data. Dalam konteks penelitian yang dilakukan oleh Mohr & Hohler (2023), pendekatan analisis berbasis *Natural Language Processing* (NLP), kemudian pengkategorian dari beberapa topik bahasan pada berita diadaptasi dalam penelitian ini untuk memahamai bagaimana media membentuk opini publik terhdap digitalisasi pertanian (Mohr & Höhler, 2023).

Penelitian oleh **Mulyani, Saifurrahman, Arinia, dan Rizqiawan (2024)**, menganalisis wacana publik terhadap adopsi *photovoltaic* (PV) di Indonesia menggunakan *topic-based sentiment analysis* berdasarkan *machine learning* dan *deep learning*. Studi ini berfokus pada identifikasi topik utama yang sering di diskusikan serta mengevaluasi metode terbasik dalam analisis sentimen yang digunakan. Pendekatan yang digunakan mencangkup *Bidirectional Encorder Representations from Transformers* (BERT) dan *Latent Dirichlet Allocation* (LDA), dengan analisis data dari media berbasis *website* dan media sosial. Hasil penelitian menunjukkan bahwa isu utama dalam diskusi publik mencangkup pengetahuan,

skeptisisme, regulasi, serta biaya investasi PV. Portal media cenderung lebih positif dibandingkan media sosial, meskipun sentimen negatif terhadap kebijakan pemerintah tetap mendominasi. Penerapan model *BERT* menunjukkan performa yang baik, meskipun demikian penelitian yang dilakukan masih memiliki keterbatasan dalam pemahaman semantik dan pengaruh faktor sosial. Sehingga, penelitian yang akan diajukan akan mengisi kesenjangan penerapan model BERT dalam analisis sentimen dengan pemetaan pada *knowledge graph*, serta perbandingan algoritma *machine learning* yang lain sebagai perbandingan algoritma terbaik untuk analisis sentimen (Mulyani et al., 2024).

Penelitian oleh **Daza, Rueda, Sanchez, Espirith, dan Quinones (2024)**, mengeksplorasi penerapan *sentiment analysis* pada ulasan produk *e-commerce* dengan pendekatan *machine learning* dan *deep learning*. Studi ini befokus pada analisis metode terbaik sebagai klasifikasi sentimen, serta mengevaluasi teknik dan metrik yang paling umum dalam penelitian terkait. Dengan menggunakan pendekatan *bibliometric analysis* dan *systematic literature review* dengan *PRISMA methodology*, di mana pendekatan dilakukan menggunakan pencarian pada empat database utama, yaitu *ScienceDirect, Scopus, ProQuest*, dan *Web of Science*. Dari hasil analisis terhadap 20 artikel yang diterbitkan antara tahun 2018 hingga 2024, ditemukan bahwa *Support Vector Machine* (SVM) merupakan algoritma yang paling banyak digunakan dalam analisis sentimen, dengan tingkat penggunaan mencapai 40% dari total studi yang dianalisis. Metode validasi yang digunakan dalam penelitian tersebut menggunakan metode *cross validation* untuk meningkatkan akurasi model, sedangkan *F1-Score* menjadi metrik utama dalam mengevaluasi performa klasifikasi. Hasil penelitian menunjukkan tingkat akurasi pada pendekatan SVM mencapai tingkat akurasi hingga 89,10% dalam tugas klasifikasi sentimen pada ulasan *e-commerce*, di mana hal tersebut lebih tinggi dibandingkan *Logistic Regression* (84,25%), dan *VADER* (88,50%). Namun, meskipun *Support Vector Machine* (SVM) memiliki performa yang baik, studi ini juga menyoroti beberapa keterbatasan seperti penggunaan konteks linguistik yang kompleks, dan pemahaman semantik antar kata tidak dapat dilakukan secara sempurna. Dalam konteks penelitian oleh (Daza et al., 2024) dapat dijadikan rujukan dalam penggunaan NLP dan SVM untuk analisis sentimen berbasis teks.

Namun, penelitian ini masih memiliki keterbatasan dalam penerapan model yang lebih kompleks dalam penggunaan *machine learning* dengan *knowledge graph* sehingga penelitian yang akan diajukan akan memperluas pendekatan yang lebih luas guna mengintegrasikan konsep *machine learning* dengan *knowledge graph*. Di mana hal tersebut digunakan sebagai pemahaman bagaimana opini publik terhadap digitaliasasi dalam sektor pertanian berkembang di media digital (Daza et al., 2024).

Studi yang dilakukan oleh **Smairi, Abadilla, Brahim, dan Chaari (2024)**, mengeksplorasi penerapan *sentiment analysis* menggunakan *BERT* dan *Support Vector Machine* (SVM) dalam mengklasifikasikan emosi pengguna media sosial. Penelitian berfokus pada klasifikasi sentimen berbasis *machine learning* dengan mengoptimalkan *SVM* menggunakan *Genetic Algorithm* (GA). Dalam studi yang dilakukan, dataset yang digunakan adalah IMDB mengandung banyak elemen ironi, dan sarkasme, di mana hal tersebut digunakan untuk menguji performa berbagai metode. *BERT*, dan *Word2Vec* diterapkan untuk mengekstrak fitur teks sebelum dilakukan klasifikasi sentimen menggunakan *SVM*. Selain itu, *Genetic Algorithm* (GA) digunakan sebagai media untuk mengoptimalkan *hyperparameter SVM*, sehingga meningkatkan akurasi klasifikasi. Hasil penelitian menunjukan bahwa kombinasi *BERT-GSVM* mencapai akurasi 91%, *precision* 89%, dan *F1-Score* 88%, di mana hal tersebut lebih unggul disbandingkan *Word2Vec-SVM* yang memiliki akurasi 90%, *precision* 88%, dan *F1-Score* 86%. Pada model *SVM* tanpa adanya *hybrid model*, memiliki tingkat akurasi sekitar 88%, *precision* 85%, dan *F-Score* 80%. Sementara itu teknik pendekatan menggunakan *Random Forest* (RF) menunjukkan performa terendah dengan akurasi 86%, *precision* 80%, dan *F1-Score* 82%. Penelitian menunjukkan integrasi *BERT* dan *SVM* serta optimasi menggunkan *GA* mampu meningkatkan aklurasi dibandingkan metode lainnya. Dalam konteks penelitian yang dilakukan oleh (Smairi et al., 2024), penelitian tersebut dapat kami adapatasi pada penelitian yang akan kami ajukan dengan pendekatan *BERT* dan *SVM* dalam analisis sentimen berita digitaliasi pertanian, serta mengekplorasi bagaimana pengguna *Knowledge Graph* dapat meningkatkan pemahaman terhadap hubungan natar entitas dalam berita digital (Smairi et al., 2024).

Penelitian oleh **Wan, Chen, Lin, Jiayuan, dan Chen (2024)** mengembangkan *Knowledge Augmented Heterogeneous Graph Convolutional Network* (KAHBGN)

sebagai analisis sentimen. Model yang digunakan menggabungkan *Natutal Languange Processing* (NLP), *Graph Convolutional Networks* (GCN), dan *Knowledge Graph* untuk meningkatkan pemahaman sentimen dalam teks dan gambar. Model ini diuji pada dataset Twitter-15 dan Twitter-17, dengan hasil yang menunjukkan bahwa KAHGCN mencapai F1-Score 73,35% pada Twitter-15, dan 70,35 % pada Twitter-17. Di mana metode tersebut mengunguli model lainnya seperti *ModalNet-BERT*. Integrasi *Knowldege Graph* dalam model ini terbukti mampu memperkaya pemahaman kontekstual, dan menungkinkan pengenalan model sentimen yang lebih akurat dibandingkan pendekatan berbasis teks. Hasil penelitian menunjukkan bahwa model ini secara signifikan meningkatkan akurasi analisis sentimen dibandingkan dengan metode sebelumnya, terutama dalam menangkap ekspresi sentimen implisit yang sering diabaikan. Relevansi penelitian ini dengan studi yang diajukan terletak pada penerapan *graph-based machine learning* yang dapat diadaptasi dalam analisis sentimen pada media berita terkait digitalisasi teknologi pertanian (Wan et al., 2024).

Berdasarkan analisis penelitian yang telah dilakukan sebelumnya, penelitian ini menjawab kesenjangan penting dalam literatur terkait dampak media berbasis website pada persepsi dan penerimaan publik terhadap digitalisasi teknologi dalam sektor pertanian khususnya dengan pendekatan *Natural Language Processing* (NLP) dan *Machine Learning*. Penelitian sebelumnya telah banyak mengkaji peran media dalam membentuk opini publik terhadap teknologi pertanian, namun sebagaian besar berfokus pada media sosial atau sumber lain yang terbatas pada satu aspek. Oleh karena itu, penelitian ini menggunakan pendekatan multimodel analisis sentimen dengan kombinasi *BERT* dan *SVM* untuk mengevaluasi sentimen dalam berita pada berbagai media seperti *BBC*, *Reuters, The Guardian*, dan *The Independent*. Selain itu, penelitian ini mengintegrasikan dengan *Knowledge Graph* sebagai pemahaman keterkaitan media, topik pemberitaan, dan sentimen publik. Langkah berikut belum banyak dilakukan dalam penelitian sebelumnya, sehingga berpotensi dalam pemahaman infomasi yang lebih mendalam dan terstruktur dalam menganalisis bagaimana pemberitaan digital mempengaruhi adopsi teknologi pertanian. Dengan demikian, penelitian ini memberikan kontribusi yang signifikan

dalam mengisi kesenjangan literatur terkait peran media dalam membentuk persepsi publik terhadap inovasi teknologi pertanian.

Berdasarkan tinjauan literatur sebelumnya, dapat diberikan kesimpulan bahwa masih terdapat kesenjangan dalam pemanfaatan pendekatan multimodel dan *graph-based machine learning* dalam analisis sentimen media terhadap digitalisasi sektor pertanian. Selain hal tersebut, studi yang mengintegrasikan analisis sentimen*, topic modeling*, dan *knowledge graph* dalam konteks pemberitaan berbasis portal web. Oleh karena itu, penelitian berikut memiliki tujuan untuk mengisi kekosongan tersebut melalui integrasi *BERT*, *SVM*, *Random Forest*, *XGBoost*, *dan Graph Convolutional Networks*, sehingga memberikan kontribusi terhadap pemahaman peran media dalam membentuk opini publik terhadap inovasi teknologi pertanian.

Penelitian	Metode									Data	
	BERT	**SVM**	**BERT + SVM**	**GA**	**RF, XGBoost**	**BERT + SVM + GA**	**NLP**	**GCN**	**Knowledge Graph**	**Actual Data**	**Social Media**
Ancín et al., 2022)							✓				✓
(Daza et al., 2024)		✓					✓			✓	
(Hao et al., 2024)	✓				✓			✓	✓	✓	✓
(Luo & Mu, 2022)							✓			✓	
(Mohr & Höhler, 2023)							✓			✓	
(Ittefaq et al., 2025)							✓			✓	
(Mulyani et al., 2024)	✓						✓			✓	✓
(Lin et al., 2022)	✓						✓				✓
(Smairi et al., 2024)	✓	✓	✓	✓	✓	✓	✓				✓
(Wan et al., 2024)	✓						✓	✓	✓		✓
Penelitian yang diajukan	✓	✓	✓		✓		✓	✓		✓	

Tabel 2.1. Hasil Perbandingan Penelitian Terdahulu

2.2 Dasar Teori

Penelitian ini membutuhkan beberapa teori pendukung guna menunjang adanya landasasan teori pada masalah yang diangkat, Penjelasan terkait teori-teori tersebut berdasarkan beberapa sumber yang berasal dari artikel, jurnal, buku, maupun pernyataan oleh para ahli yang terkutip untuk meningkatkan pemahaman dalam uraian permasalahan pada penelitian yang diajukan. Adapun landasan yang yang digunakan pada penelitian diuraikan sebagai berikut:

2.2.1 Digitalisasi Pertanian

Digitalisasi pertanian dalam perkembangan era ini secara signifikan mengacu pada penerapan teknologi digital seperti *Internet of Things* (IoT), *Artificial Intelligence* (AI), dan *Big Data* sebagai alat untuk mengoptimalkan berbagai aspek dalam sektor pertanian. Dengan teknologi tersebut, petani dapat memantau dan mengelola lahan pertanian secara lebih efisien, dapat meningkatkan produktivitas, serta menjaga keberlanjutan lingkungan. Digitaliasi tidak hanya mencangkup oyomasi proses dalam pengolahan pertanian, namun juga pemanfaatan data sebagai pengambilan keputusan yang lebih baik dan dapat dikenali dengan sebutan teknologi cerdas.

Perkembangan dan penerapan teknologi digital dalam pertanian seperti teknologi *Internet of Things* (IoT) misalnya, di mana pada lahan pertanian dipasang sensor untuk mengumpulkan data secara *real-time* mengenai kondisi tanah, kelembapan, cuaca, dan hama. Data ini digunakan untuk memberikan rekomendasi spesifik kepada petani dalam menentukan waktu yang tepat untuk menyiram, memupuk, atau memanen tanaman (Xu et al., 2022). Selain itu, *Artificial Intelligence* (AI) dan analitik *Big Data* memungkinkan prediksi hasil panen dan perencanaan produksi dengan lebih akurat berdasarkan data historis dan kondisi saat ini (Mohr & Höhler, 2023). Dengan demikian, teknologi ini memberikan dampak signifikan dalam meningkatkan efisiensi dan produktivitas pertanian.

Selain itu, manfaat digitalisasi pertanian sebagai peningkatan produktivitas dengan pemantauan dan analisis data secara *real-time*, petani dapat mengambil tindakan yang tepat waktu untuk mengoptimalkan pertumbuhan tanaman, mengurangi kehilangan hasil panen, dan meningkatkan kualitas produk. Pada pemanfaatan efisisensi sumber daya, teknologi digital membantu mengurangi penggunaan air, pupuk, dan pestisida secara berlebihan, sehingga mengurangi dampak negatif terhadap lingkungan. Penggunaan sumber daya yang

lebih efisien ini juga berkontribusi pada keberlanjutan pertanian jangka panjang (Xu et al., 2022). Sedangkan pemanfaatan keberlanjutan lingkungan sendiri, IoT dan AI membantu meminimalkan dampak negatif aktivitas pertanian terhadap lingkungan, seperti mengurangi emisi karbon dan penggunaan pestisida berlebihan. Hal ini mendukung praktik pertanian yang lebih ramah lingkungan dan berkelanjutan (Siegrist & Hartmann, 2020).

Meskipun digitalisasi teknologi dalam pertanian memiliki manfaat, terdapat tantangan dalam penerapan teknologi tersebut. Adapun tantangan-tantangan tersebut biasanya sering terjadi pada wilayah ataupun negara-negara berkembang, di mana tantangan-tantangan tersebut adalah biaya implementasi yang tinggi (Z. Chen et al., 2022), keterbatasan teknologi dan infrastruktur, serta hambatan pada pengetahuan dan keahlian, di mana implementasi teknologi digital memerlukan pengetahuan dan keahlian teknis. Banyak petani, terutama generasi yang lebih tua, mungkin kesulitan untuk memanfaatkan teknologi ini tanpa pelatihan yang memadai (Mohr & Höhler, 2023).

Peran media dalam membentuk opini publik juga sangat penting dalam mendorong adopsi digitalisasi pertanian. Media dapat mempengaruhi pandangan petani tentang teknologi baru, baik secara positif maupun negatif, tergantung pada framing berita yang diberikan. Media yang fokus pada keuntungan teknologi, seperti peningkatan produktivitas dan efisiensi, akan lebih mendorong adopsi teknologi tersebut. Sebaliknya, pemberitaan yang menyoroti tantangan seperti biaya tinggi atau risiko teknis dapat menurunkan minat petani untuk mengadopsi teknologi digital (Holton et al., 2023).

2.2.2 Peran Media dalam Pembentukan Opini Publik

Media massa memiliki pengaruh besar dalam membentuk pandangan publik, khususnya terkait inovasi dan teknologi baru seperti digitalisasi di sektor pertanian. Melalui pemberitaannya, media tidak hanya menyampaikan informasi, tetapi juga membentuk pandangan masyarakat melalui proses yang dikenal sebagai *agenda setting* dan *framing*. Dalam *agenda setting*, media berperan dalam menetapkan isu-isu yang dianggap penting oleh publik dengan memberikan perhatian lebih pada topik tertentu. Misalnya, ketika media besar seperti *BBC*, *Reuters*, *The Guardian*, dan *The Independent* terus menerus memberitakan teknologi seperti *Internet of Things* (IoT) dan *Artificial Intelligence* (AI) dalam sektor pertanian, masyarakat menjadi lebih menyadari pentingnya topik tersebut, yang pada gilirannya mempengaruhi pandangan mereka terhadap teknologi tersebut (Holton et al., 2023).

Framing merupakan istilah yang menjelaskan bagaimana media menyusun dan menekankan aspek tertentu dari berita untuk mempengaruhi persepsi publik. Media dapat memunculkan persepsi positif atau negatif tergantung pada bagaimana mereka menyajikan berita. Ketika media fokus pada manfaat teknologi pertanian digital, seperti efisiensi sumber daya dan peningkatan hasil panen, publik cenderung merespon dengan sikap positif dan terbuka terhadap teknologi tersebut. Namun, jika pemberitaan lebih banyak menyoroti tantangan, seperti biaya tinggi atau hambatan teknis dalam penerapannya, hal tersebut dapat memicu keraguan dan sikap negatif, yang akhirnya menghambat penerimaan teknologi tersebut (Siegrist & Hartmann, 2020).

Sebagai alat komunikasi yang kuat, media berfungsi sebagai institusi sosial yang membentuk wacana dan pandangan masyarakat. Media sering kali menjadi perantara antara inovasi teknologi dan penerimaan publik, mempercepat atau memperlambat adopsi teknologi berdasarkan cara pemberitaan mereka. Dalam konteks digitalisasi pertanian di berbagai negara, media bisa mempromosikan ataupun menghambat penerimaan teknologi baru, tergantung pada narasi yang disajikan. Sebuah studi oleh Mohr dan Höhler (2023) menunjukkan bahwa pemberitaan positif mengenai teknologi digital dalam pertanian dapat mempercepat adopsi teknologi tersebut oleh para petani. Namun, jika risiko dan tantangan dari teknologi lebih banyak ditonjolkan, hal ini bisa memunculkan kekhawatiran atau ketidakpastian di kalangan masyarakat, yang pada akhirnya mempengaruhi tingkat penerimaan teknologi tersebut.

Dalam teori difusi inovasi, media massa memiliki peran penting sebagai saluran utama dalam menyebarluaskan informasi mengenai inovasi kepada masyarakat. Media tidak hanya bertugas menyampaikan informasi, tetapi juga mempengaruhi keputusan individu atau kelompok masyarakat dalam menerima teknologi baru. Media dapat berperan strategis dalam mengarahkan pandangan publik terhadap keuntungan teknologi digital dan menyoroti pentingnya inovasi ini untuk menciptakan pertanian yang lebih berkelanjutan dan efisien (Wester et al., 2023).

Dari beberapa uraian yang telah diberikan, perlu adanya pemahaman bagaimana peran media mempengaruhi pandangan publik, khususnya dalam strategi komunikasi dan kebijakan publik yang bertujuan untuk meningkatkan penerimaan teknologi digital di sektor pertanian. Melalui proses *agenda setting* dan *framing*, media massa dapat menjadi instrumen yang efektif dalam membentuk sikap publik terhadap inovasi, selama narasi yang dibangun mendukung penerimaan teknologi tersebut.

2.2.3 Analisis Sentimen dalam Berita

2.2.3.1. Analisis Sentimen

Membahas terkait analisis sentimen, berikut dapat didefinisikan sebagai proses yang melibatkan pengumpulan dan analisis opini serta kesan individu dalam berbabagai topik, produk, maupun subjek tertentu. Analisis sentimen telah menjadi salah satu teknik utama yang digunakan untuk memahami sikap dan opini publik terhadap berbagai topik yang diungkapkan dalam bentuk teks (Liu, 2022). Dalam definisi tertentu, analisis sentimen merupakan teknik yang digunakan untuk mengidentifikasi dan mengekstrak opini atau emosi dalam teks, guna menentukan apakah suatu pernyataan bersifat positif, negatif, maupun netral (Costola et al., 2023).

Dalam penggunaan teknik tersebut, teknik ini umumnya digunakan dalam berbagai bidang seperti pada bidang pemasaran, layanan pelanggan, politik, ekonomi, maupun dalam dunia pertanian, guna mengidentifikasi pola sentimen yang dapat mempengaruhi pengambilan keputusan (Medhat et al., 2014). Secara teknis, analisis sentimen merupakan cabang dari *Natural Langage Processing* (NLP) (Liu, 2022). Di mana tahapan tersebut terdiri dari berbagai tahapan seperti, tahap pertama merupakan *data collection* yang melibatkan ekstraksi teks dari berbagai sumber seperti media sosial, berita online, atau dari ulasan pelanggan. Tahap kedua, *text preprocessing*, di mana data teks difilter melalui proses seperti *tokenisasi*, *stemming*, dan, *stopword removal* untuk menghilangkan kata-kata yang tidak relevan. Pada tahap selanjutnya, tahap *sentiment classification* yang dilakukan menggunakan pendekatan berbasis *lexicon* atau metode *machine learning* (Pang, 2008). Untuk tahap akhir dari pengolahan tersebut, analisis digunakan sebagai pemahaman pola sentimen dan impilkasinya terhadap pengambilan keputusan di berbagai domain.

Seiring perkembangannya, metode *sentiment analysis* juga berkembang dengan menggunakan beberapa pendekatan dari *Lexicon* menjadi lebih banyak menggunakan *machine learning* dan *deep learning*. Pada *Lexicon* untuk menentukan polaritas suatu kaya dalam teks, pendekatan ini memiliki keterbatasan dalam menangkap konteks bahasa yang lebih kompleks. Sebaliknya, pendekatan menggunakan *machine learning* seperti *Support Vector Machine*, *Naïve*

Bayes, memungkin sistem untuk mempelajari dataset uang lebih luas, sehingga mampu mengenali pola sentimen yang lebih baik (Socher et al., n.d.). Model *Deep Learning* juga terbukti dapat memahami konteks dan hubungan antar kata yang lebih

akurat, sehingga sering dijadikan salah satu metode unggulan dalam analisis sentimen, salah satu teknik yang paling popular adalah *Bidirectional Encorder Representations from Transformers* (BERT) ((Costola et al., 2023).

Keakuratan analisis sentimen sangat bergantung pada kualitas data yang digunakan dalam proses *text processing*. Tantangan utama dalam implementasi teknik tersebut meliputi deteksi sarcasm, irony, serta analisis sentimen dalam bahasa yang memiliki banyak makna ganda (Cambria et al., n.d., 2017). Dengan demikian, banyak penilitian terbaru menggabungkan metode berbasis *lexicon* dengan *machine learning* untuk meningkatkan akurasi prediksi (Loughran,T., & McDonald, B., 2011) Perkembangan teknologi seperti *Artifical Intelligence* (AI), analisis sentimen diharapkan dapat terus beradaptasi dan mampu memberikan wawasan yang lebih mendalam terkait opini maupun emosi publik terhdadp berbagai isu global.

2.2.3.2. Pendekatan Analisis Sentimen dalam Berita

Dalam perkembangan teknologoi yang melibatkan adanya *Natural Language Processing* (NLP) dan *Artificial Intelligence* (AI) telah mendorong adopsi *sentiment analysis* dalam berbagai bidang, termasuk analisis berita (Liu, 2022). Berita sendiri memerankan peran penting dalam membentuk opini publik dan sentimen pasar, sehingga dalam analisis sentimen sendiri peran tersebut terhadap berita menjadi salah satu metode yang banyak digunakan dalam berbagai bidang (Medhat et al., 2014). Sejalan dengan hal tersebut, volume berita berbasis portal daring dan publikasi digital menggunakan metode konvensional dalam analisis teks tidak lagi cukup untuk mengangani kompleksitas bahasa yang digunakan dalam berita (Pang, 2008). Dengan demikian, berbagai pendekatan telah dikembangkan sebagau upaya meningkatkan akurasi sentiment analysis dalam berita.

Dalam konteks berita, analisis sentimen menghadapi berbagai tantangan unik, seperti ambiguitas bahasa, maupun anacaman lainnya yang dapat terhambat untuk menginterpretasikan suatu model (Cambria et al., 2017). Dengan demikian, banyak penelitian terbaru yang menggabungkan pendekatan berbasis *lexicon* dengan *machine learning* untuk meningkatkan akurasi dalam mengklasifikasikan sentimen dalam teks berita (Socher et al., 2017). Pada bagian berikut, akan dibahas secara lebih dalam terkait pendekatan utama untuk sentiment analysis dalam berita, yaitu *Lexicon based Sentiment Analysis* dan *Machine Learning based Sentiment Analysis*.

a. Lexicon-based Sentiment Analysis

Pendefinisian *Lexicon based Sentiment Analysis* dapat diuraikan sebagai metode untuk mengumpulkan kata yang telah diberikan nilai sentimen tertentu, seperti positif, negatif, dan netral. Metode tersebut mengandalkan kamus kata sebagai media untuk mengidentifikasi polaritas sentimen dalam teks tanpa perlu menggunakan *training dataset* (Taboada et al., 2011). Beberapa *sentiment lexicons* yang sering digunakan dalam analisis berita dalah *SentiWordNet*, *VADER* (*Valnce Aware Dictionary and Sentiment Reasoner*), dan *AFINN* (Liu, 2022).

Dalam tahapan proses *Lexicon based Sentiment Analysis* adalah tahap *tokenization. Lexicon matching*, dan *sentiment scoring*. Pendekatan tersebut memiliki keunggulan dalam interpretabilitas dan kemudahaan implementasi, karena tidak memerlukan data yang besar (Medhat et al., 2014). Namun, metode ini memiliki keterbatasan dalam menangani konteks yang lebih kompleks, seperti sarkarsme maupun kata-kata yang memiliki makna ganda dalam berbagai situasi (Taboada et al., 2011).

b. Machine Learning-based Sentiment Analysis

Dalam penerapannya pada *sentiment analysis* untuk berita, *machine learnig* menggunakan *supervised learning algorithms* sebagai klasifikasi sentimen dalam teks berdasakran *patterns* yang ditemukan dalam *dataset* (Pang, 2008). Berbeda dengan *lexicon-based methods* yang mengandalkan *predefined word list*, pendekatan berikut memungkinkan model untuk *learn from data*, sehingga lebih efektif dalam menangani *contextual sentiment*, *sarcsm*, dan *complex language structures*. Beberapa algoritma yang sering digunakan dalam sentiment analysis sendiri seprti *Naive Bayes*, *Support Vector Machine*, *LSTM*, maupun *BERT* (Devlin et al., 2018).

Keunggulan pendekatan menggunakan *machine learning* adalah memliki kemampuan dalam *adapting to new data*, menangkap *nuanced sentiment*, dan akurasi yang lebih baik dibandingkan *lexicon-based methods*. Namun, secara teknis *machine learning* membutuhkan dataset yang besar, dan *high computational power*, sehingga berisiko kategori yang bersifat bias jika *training* tidak seimbang (Sun et al., n.d., 2019).

2.2.4 Natural Language Processing

Natural Language Processing (NLP) merupakan cabang ilmu pengetahuan dari *artificial intelligence* (AI) yang berfokus pada interaksi antara komputer dan bahasa manusia. NLP memungkinkan komputer untuk menganalisis, memahami, dan memproses

bahasa alami dengan cara yang mirip dengan bagaimana manusia berkomunikasi. Tujuan utama *Natural Language Processing* (NLP), yaitu sebagai alat yang menjembatani kesenjangan antara pemahaman komputer dan bahasa manusia sehingga mesin dapat menginterpretasikan dan merespons teks atau ucapan secara efektif (Jurafsky, 2019).

Dalam penerapannya, *Natural Languange Processing* (NLP) sendiri melbibatkan beberapa komponen dan Teknik utama yang membantu mesin untuk memahami dan memproses bahasa manusia. Berikut merupakan komponen utama dari *Natural Languange Processing* (NLP):

2.2.4.1. Preprocessing Teks

a. Tokenisasi

Dalam teknik tokenisasi melibatkan proses pemecahan teks menjadi unit yang lebih kecil atau sederhana dengan pendefinisian yang sering disebut sebagai token. Penggunaan token dalam proses tokenisasi memiliki struktur berupa kata, frasa, atau karakter untuk pengolahan lebih lanjut. Dengan adanya tahapan maupun struktur dari tokenisasi yang telah diuraikan sebelumnya, Manning (2008) mendefinisikan tokenisasi sebagai suatu tahapan pertama dalam banyak tugas *Natural Language Processing* (NLP), sebagai unit dasar dari teks yang akan dipahami oleh computer sebelum di analisis lebih lanjut oleh mesin (Christopher D. Manning, 2008). Proses penyederhanaan berikut dikarenakan komputer memiliki kemampuan bahasa yang sederhana untuk dapat memahami data yang dimasukkan, sehingga dapat diolah menjadi bentuk data yang sesuai dengan permintaan *user* (pengguna).

b. Stemming dan Lemmatization

Kedua teknik (*Stemming* dan *Lemmatizato*n) berikut umumnya digunakan sebagai teknik pengubaan kata menjadi bentuk kata asal atau dapat dikatakan sebagai bentuk dasar. Teknik *Stemming* umumnya digunakan sebagai teknik memotong akhiran kata untuk mendapatkan bentuk dasar, dapat diberikan contoh pada kata "*running*" yang akan menjadi bentuk dasar dari kata, yaitu "*run*".

Sementara teknik *lemmatization* umumnya menggunakan aturan linguistik yang digunakan dalam menemukan bentuk akar kata, dapat diberikan contoh misalnya dari kata "*better*" menjadi "*good*" dalam bentuk lain dari better yang merupakan bentuk dasar dari kata sebelumnya. Dalam proses processing data sendiri, kedua teknik tersebut

digunakan untuk mengurangi kompleksitas teks dan memastikan bahwa berbagai bentuk kata yang sama diperlakukan sebagai entitas yang sama (Jurafsky, 2019).

c. Stopword Removal

Pendefinisian *stopword*s dapat diuraikan sebagai istilah kata-kata yang sangat umum dalam suatu bahasa seperti "*the*", "*is*", "*which*", dan "*on*". Kata-kata umum yang termasuk dalam bagian teks tersebut dalam proses pendekatan processing perlu adanya penyederhanaan, hal tersebut dikarenakan kata-kata tersebut sering dibaikan dalam pemrosesan bahasa dan sistem pencarian karena dianggap tidak memberikan nilai signifikan dalam pemilihan dokumen yang relevan. Dengan demikian, perlu adanya penyederhanaan kata dengan penghapusan beberapa kata yang umum dengan pendefinisian *stopwords removal* dengan tujuan membantu mengurangi ukuran yang diekstraksi (Christopher D. Manning, 2008) .

2.2.4.2. Word Embedding

Salah satu teknik dalam *Natural Language Processing* (NLP) adalah *word embedding*, di mana *word embedding* sendiri dapat diuraikan sebagai teknik yang mempresentasikan kata-kata dalam bentuk *vector numerical representations* dengan menangkap makna semantik dan hubungan antar kata dalam ruang berdimensi tinggi (Mikolov et al., 2013) Dalam penggunaan teknik berikut Vaswani (2017), memberikan penyataan hal tersebut memungkinkan model *Natural Language Processing* (NLP) dapat memaham *word similarity*, *context*, dan *meaning* dengan lebih efektif dibandingkan menggunakan pendekatan berbasis *bag-of-words* yang hanya mempertimbangkan frekuensi kata tanpa memperhatikan konteks (Vaswani et al., 2017).

Dalam perkembangan model *Natural Language Processing* sendiri memberikan pengaruh pada pengembangan teknik *word embedding*, di mana terdapat beberapa pendekatan statistik seperti *TF-IDF*. Hingga teknik berbasis *networks* seperti *Word2Vec*, dan *BERT Embeddings*, yang mampu menangkap hubungan kontekstial yang lebih dalam (Devlin et al., 2018)

a. TF-IDF

TF-IDF didefinisikan oleh Havrlant dan Kreinovich (2017), sebagai metode numerik dari *Word Embedding* yang digunakan dalam *text mining* dan *information retrieval* sebagai penentu seberapa penting suatu kata dalam sebuah dokumen dibandingkan dengan kumpulan dokumen lainnya (Havrlant & Kreinovich, 2017).

Pada *TF-IDF* sendiri memiliki dua komponen utama dalam pemrosesan dokumen yang akan diolah oleh teknik tersebut, yaitu *Term Frequency* (TF), dan Inverce Document Frequency (IDF). Komponen Term Frequency (TF) dijelaskan sebagai alat yang mengukur seberapa sering suatu kata muncul dalam sebuah dokumen tertentu, di mana kata yang muncul akan memiliki bobot lebih tinggi daripada kata lainnya yang memiliki presentase muncul lebih kecil dalam suatu kalimat. Sedangkan dalam komponen *Inverce Document Frequency* (IDF) berfokus pada pengurangan bobot kata-kata umum yang muncul di banyak dokumen sehingga dapat diketahui seberapa jarang kata tersebut ditemukan dalam koleksi dokumen sebelumnya (Sammut & Webb, 2017) .

Sejalan dengan definisi kedua komponen TF-IDF pada penelitian lainnya, *TF-IDF* diuraikan dengan mengalikan kedua nilai pada komponen dengan memberikan bobot lebih tinggi pada kata-kata yang memiliki frekuensi tinggi dalam dokumen tetapi jarang ditemukan di seluruh koleksi dokumen, sehingga membantu dalam *keyword extraction*, *text classification*, dan *document ranking* (Ramos, n.d., 2003).

b. Word2Vec

Word2Vec merupakan model yang dikembangkan oleh Mikolov et al (2013), yang digunakan untuk mempresentasikan kata-kata dalam bentuk *dense verctors* dalam ruang vektor berdimensi tinggi (Mikolov et al., 2013). Berbeda dengan pendekatan berbasis *one-hot encoding* atau *TF-IDF*, teknik *Word2Vec* mampu menangkap hubungan semantik antar kata berdasarkan konteks penggunaannya dalam teks. Model yang digunakan pada *Word2Vec* menggunakan dua pendekatan utama, yaitu *Continours Bag of Words* (CBOW) sebagai prediksi kata berdasarkan kata-kata yang berada disekitarnya. Kemudian pendekatan menggunakan Skip-gram, di mana pendekatan berikut dapat memprediksi kata-kata disekitar berdasarkan kata yang diberikan (Mikolov et al., 2013).

Dalam penerapannya *Word2Vec* memiliki kemampuan menangkap *word similarity and analogies*, misalnya "*king -man + woman = queen*". Namun, *Word2Vec* sendiri juga memiliki keterbatasan dalan memahami *polysemy* (*multiple meaning of words*) *dan long-range dependencies*, karena hanya mempertimbangkan kata-kata terdekat dalam suatu jendela (Goldberg & Levy, 2014).

Word2Vec sendiri telah digunakan dalam berbagai aplikai NLP, termasuk *sentiment analysis*, *document classification*, dan *recommendation systems* (Zhang et al., n.d, 2020.). Namun, dengan perkembangan model berbasis Transformers seperti *BERT*, penggunaan *Word2Vec* telah berkurang dalam tugas-tugas NLP yang membutuhkan pemahaman lebih kompleks (Rogers et al., 2020).

c. BERT Embeddings

Bidirectional Encorder Representations from Transformers (BERT) merupakan model *deep learning* berbais t*ransformers* yang dikembangkan oleh Google (Devlin et al., 2018). Berbeda dengan *Word2Vec* yang hanya mempertimbangkan hubungan kata dalam satu arah (*left to right* atau *right to left*). Dalam teknik BERT sendiri dapat memahami kata dalam dua arah (*bidirectional*), sehingga teknik ini dapat menangkap makna kontekstual yang lebih kompleks.

Pada penerapannya, BERT sendiri mampu menangkap *contextual embeddings*, di mana satu kata dapat memiliki makna yang berbeda tergantung pada konteksnya. Rogers (2017), memberikan uraian terkait keunggulan utama BERT, yaitu mampu memahami *polysem*i dan konteks yang lebih kompleks dibandingkan *Word2Ve*c, mampu digunakan dalam banyak *state of the art NLP tasks* seperti *text classification*, *sentiment analysis*, dan *machine translation*, dan keunggulan yang terakhir adalah memiliki *high computational* dan *large scale pretraining datasets* untuk menghasilkan hasil optimal (Rogers et al., 2020).

2.2.5 Machine Learning

Dunia IT yang semakin menunjukkan perkembangan secara terus-menerus memberikan ruang pada inovasi algoritma yang memiliki kerterkaitan luas pada adopsi digital. Salah satu inovasi algoritma yang sering digunakan adalah *machine learning*. Pendefinisian *machine learning* dapat didefinisikan sebagai cabang dari *Artificial Intelligience* (AI) dengan tujuan untuk memungkinkan sistem mempelajari data dan membuat prediksi atau keputusan tanpa adanya intruksi yang diberikan secara eksplisit. *Machine Learning* sendiri menggunakan model matematis untuk mengenali pola dan tren dari data yang diberikan dan menerapkannya pada data baru untuk melakukan prediksi ataupun pengambilan keputusan. Konsep tersebut menjadi dasar bagi banyak aplikasi modern, seperti pengenalan suara, analisis teks, maupun pengenalan gambar.

Terdapat tiga jenis utama dalam *machine learning*, di mana tiga jenis utama tersebut dikategorikan sebagai *Supervised Learning*. Algoritma berikut menggunakan data yang telah diberi label sebelumnya. Dalam artian lain, model pembelajaran terkait *machine learning* yang mempelajari proses *input* dan *output* yang berasal dari data yang belum pernah dilihat sebelumnya (Christopher M. Bishop, 2006). Pada kategori kedua terdapat *Unsupervised Learning*, di mana algoritma ini bertujuan untuk menemukan pola atau struktur tersembunyi dalam data tanpa terdapat output tertentu yang menjadi acuan, di mana metode tersebut sering digunakan sebagai pengelompokan data dan segmentasi pasar (Russell, 2020)

Dalam kategori berikutnya *Reinforcement Learning*, yang digunakan untuk membuat keputusan dalam lingkungan tertentu dengan cara memaksimalkan "*reward*" atau hasil terbaik dari setiap tindakan yang diambil. Algoritma ini belajar melalui eksperimen dan memperbaiki strategi berdasarkan hasil yang diperoleh

Algoitma dalam *Machine Lear*ning yang digunakan dalam penelitian dapat disajikan dalam uraian sebagai berikut:

2.2.5.1. Bidirectional Encoder Representations from Transformers

Dalam *Machine Learning*, algoritma *Bidirectional Encorder Representations from Transformers* (BERT) sendiri terdapat dalam kategori *Deep Learning*, di mana *Deep Learning sendiri* merupakan sub-bidang dari *Machine Learning* yang menggunakan jaringan syaraf tiruan (*Neural Networks*) dengan kepemilikan beberapa lapisan (*Deep Networks*). Pada penerapannya, *Deep Learning* memiliki kegunaan sebagai media dalam pengolahan data yang besar dan kompleks, seperti gambar, suara, dan teks (Duryea et al., 2016). Dalam penelitian yang menganut sistem analisis teks yang memiliki kecenderungan kalimat kompleks, *Bidirectional Encorder Representations from Transformers* (BERT) merupakan salah satu model *deep learning* yang popular dan sering digunakan, khususnya dalam *Natural Languange Processing* (NLP).

Bidirectional Encorder Representations from Transformers (BERT) merupakan model berbasis *transformer* yang dikembangkan oleh *Google* dan digunakan untuk tugas-tugas *Natural Language Processing* (NLP), seperti klasifikasi teks, analisis sentimen, dan penerjemahan otomatis. Sejalan dengan hal tersebut, pendekatan berikut memiliki keunggulan utama, yaitu terkait kemampuan dalam memahami konteks kata pada dua arah di dalam sebuah kalimat, yaitu dari kiri ke kanan maupun sebaliknya (Devlin et al., 2018).

Berbeda dengan model *Natural Language Processing* (NLP) tradisional yang hanya membaca teks dari satu arah, *Bidirectional Encorder Representations from Transformers* (BERT) membaca dari kedua arah secara bersamaan, sehingga dapat memahami makna kata lebih mendalam.

Dalam penerapannya, *Bidirectional Encorder Representations from Transformers* (BERT) sendiri menggunakan arsitektur *transformer* yang terdiri dari beberapa lapisan *self-attention* dan lapisan *feed-forward*. *Self-attention* memungkinkan model memperhatikan kata-kata di sekitar kata target, baik sebelum maupun sesudahnya, untuk menghasilkan representasi yang lebih kaya dan kontekstual(Devlin et al., 2018);(Vaswani et al., 2017).

Model berikut menggunakan dua metode utama: *masked language modeling* dan *next sentence prediction*. Pada *masked language modeling*, beberapa kata dalam kalimat dihapus, dan model dilatih untuk memprediksi kata yang hilang, sementara pada *next sentence prediction*, *Bidirectional Encorder Representations from Transformers* (BERT) memprediksi apakah sebuah kalimat mengikuti kalimat sebelumnya, memungkinkan pemahaman hubungan antar kalimat (Devlin et al., 2018);(Vaswani et al., 2017). *Bidirectional Encorder Representations from Transformers* (BERT) sendiri memiliki keefektifan dalam tugas tugas *Natural Language Processing* (NLP), dikarenakan kemampuan yang digunakan mampu menangkap konteks komprehensif. Namun, pada pelatihan dan penggunaanya menurut (Vaswani et al., 2017), *Bidirectional Encorder Representations from Transformers* (BERT) memerlukan sumber daya komputasi yang tinggi serta waktu yang banyak sehingga model ini lebih cocok untuk aplikasi yang memiliki infrastruktur komputasi yang memadai.

2.2.5.2. Support Vector Machine

Suport Vector Machine (SVM) merupakan algoritma yang digunakan sebagai klasifikasi dan regresi. Algoritma berikut bekerja dengan menemukan *hyperplane* terbaik yang memisahkan data menjadi dua kelas dengan margin terbesar(C. Cortes et al., 1995) . Konsep ini memungkinkan *Suport Vector Machine* (SVM) bekerja dengan baik pada data yang bersifat non-linear dengan menggunakan kernel untuk memetakan data ke ruang dimensi yang lebih tinggi.

Pada penerepannya *Suport Vector Machine* (SVM) menggunakan konsep *margin* untuk memisahkan data menjadi dua kelas. Algoritma ini mencari garis atau *hyperplane*

yang memiliki jarak maksimum dari titik data terdekat di setiap kelas. Untuk data yang tidak dapat dipisahkan secara linear, *Suport Vector Machine* (SVM) menggunakan kernel, seperti *polynomial kernel* atau *radial basis function (RBF)*, yang memetakan data ke ruang dimensi yang lebih tinggi sehingga dapat dipisahkan lebih mudah (C. , & V. V. Cortes, 1995) *Suport Vector Machine* (SVM) sangat efektif untuk menangani data berdimensi tinggi dan data dengan hubungan antar fitur yang kompleks. Algoritma ini juga cukup fleksibel karena penggunaan kernel, namun menjadi kurang efisien saat menangani dataset yang sangat besar, karena membutuhkan waktu komputasi yang cukup besar, terutama ketika data tidak seimbang (C., & V. V. Cortes, 1995).

2.2.5.3. Random Forest

Metode *Random Forest* merupakan algoritma ensemble berbasis *decision tree* yang diperkenalkan oleh Breiman (2001), di mana metode ini algoritma berbasis "*bagging*" yang membangun ratusan hingga ribuan *decision tree* terlatih pada *bootstrap sample* yang berbeda kemudian menggabungkan prediksinya melalui mayoritas suara (classification) atau rata-rata (*regression*). Mekanisme dari *Random Forest* sendiri menggunakan dua mekanisme inti dari mulai pengambilan sampel baris secara acak (*bootstrap*) dan pemilihan subset fitur secara acak di setiap node sehingga menciptakan korelasi rendah antar pohon sekaligus mempertahankan bias rendah, sehingga *Random Forest* sendiri mampu menekan overfiting dibandingkan single-tree dan beberapa metode ansambel lain (Breiman, 2001).

Kajian terkini menegaskan keunggulan *Random Forest* dengan beberapa implementasi seperti oleh Galiano et al (2015), yang mampu menunjukkan bahwa *Random Forest* yang dioptimasi dengan *random-search* tetap konsisten mencapai akurasi >90% dalam klasifikasi citra hipersketral berskla besar meskipun jumlah fitur mencapai ribuan, berkat sifat "*feature bagging*" yang meminmalkan dominasi atribut tertentu(Rodriguez-Galiano et al., 2012). Selain studi kasus yang memberikan pembuktian terkait implementasi metode ini, penelitian lainnya oleh Jallal, et al (2022) melaporkan bahwa dalam ranah klasifikasi opini *Random Forest* setelah melalui proses pre-processing teks dan ekstraksi-gram mencapai akurasi >90% pada tugas multi kelas sentimen Twitter melampau metode klasifikasi opini lainnya. Kinerja ini dikaitkan dengan kemampuan *Random Forest* mengeksplor variasi sintaksis leksikal melalui ratusan *decision tree* yang dibangun secara acak(Jalal et al., 2022).

Secara pariktis *Random Forest* bersifat *Robust* dengan bekerja baik pada data berdimensi tinggi, berkskala non-linear, dan mengandung noise. Hal lain mengenai *Random Forest* juga bersifat Non-parametik dengan tidak mengasumsikan distribusi kelas maupun linearitas relasi antar fitur. Sifat *Explainable* pada metode ini menyediakan skor Gini atau *Permutation importance* yang memudahkan analisis variabel kunci, serta metrik *out-of-bag* sebagai estimasi generalisasi tanpa *hold-out* terpisah. Terakhir, metode ini berikut termasuk relatif efisien.

2.2.5.4. XGBoost

XGBoost (*eXtreme Gradient Boosting*) merupakan implementasi *gradient boosting decision trees* yang dioptimalkan untuk performa tinggi pada dataset berskala besar dan berdimensi tinggi. Dengan menformulasikan fungsi objektif yang teregularisasi secara eksplisit, setiap iterasi menambahkan pohon regresi yang meminimalkan *second-order Taylor series expansion* dari *loss*, yaitu:

$$L^{(t)} = \sum_{i=1}^{n} [g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2] + \Omega(f_t)$$

Dimana $\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2$ mempitasi penalti atas jumlah leaf T dan *leaf weight* ω_j untuk mencegah *overfitting* dan menjaga kesederhanaan model. Inovasi ini memungkinkan *XGBoost* memilih pohon yang memaksimalkan *trade-off* antara fit data dan kompleksitas model (T. Chen & Guestrin, 2016). Sedangkan untuk mencegah *overfitting* pada data berdimensi tinggi, *XGBoost* menerapkan berbagai teknik regulasi termasuk *shrinkage* (*learning rate*), *column subsampling*, dan *sparsity aware split findin*g, serta mendukung konstruksi pohon secara paralel. Selain itu, *XGBoost* mempercepat proses pencarian *split* melalui *weighted quantile sketch*, yang mengelola distribusi nilai fitur berbobot tanpa perlu menyortir ulang seluruh kolom. Oleh karena itu, untuk data yang bersifat *sparse* atau mengandung nilai yang hilang, algoritma berikut juga merapkan *sparsity aware split finding* dengan menetapkan *default direction* pada setiap node sehingga hanya entri *non-missing* yang diproses dan kemudian *overhead* komputasi dapat lebih ditekankan (Mesut et al., 2023).

Dalam implementasi model, *XGBoost* terbukti *robust* di berbagai domain aplikasi. Misalnya, pada studi prediksi skor *SOFA* pasien COVID-19 *XGBoost* menunjukkan akurasi melebihi 80% pada dataset besar, berkat kemampuan *out-of-core computing* dan optimasi paralel lainnya (Montomoli et al., 2021). Untuk interperetabilitas, *XGBoost*

menyediakan metrik *feature importance* sepergi *gain, cover,* dan *frequency* dan dapat dipadukan dengan SHAP (*Shapley Additive exPlanations*) untuk menjelaskan kontribusi tiap fitur secara lokal. Dengan kombinasi efisiensi, skalabilitas, dan mekanisme regularisasi yang kuar, *XGBoost* tetap menjadi pilihan utama dalam kompetisi data science dan pebelitian ilmiah terkini.

2.2.6 Knowledge Graph

Definisi *Knowledge Graph* dipresentasikan sebagai data berbasis graph yang menghubungkan entitas dengan relasi yang dapat dipahami untuk menangkap hubungan semantik antar entitas. Model *knowledge graph* sendiri dalam penerapannya sering digunakan dalam berbagai aplikasi seperti *search engines*, *recommendation systems*, dan *Natural Language Processing* (NLP) (Hogan et al., 2021). Struktur utama knowledge graph sendiri terdiri dari entitas, relasii, dan atribut, yang membentuk jaringan informasi yang dapat dianalisis secara semantik (Ji et al., 2022).

Paulheim (2017), menjelaskan terkait *knowledge graph* sendiri dapat membantu strukturisasi informasi dari sumber yang tidak terstruktur, sehingga memungkinkan analisis pada hubungan tiap entitas. Google memperkenalkan *Google Knowledge Graph* pada tahun 2012 untuk meningkatkan hasil pencarian dengan memahami hubungan antara konsep yang dicari oleh pengguna (Cimiano & Paulheim, 2016).

Dalam penerapannya, knowledge graph memiliki tantangan utama seperti scalability, data incompleteness, dan entity aligment, yang mmerlukan pendekatan seperti Knowledge Graph Completion (KGC) untuk mengatasi informasi yang tidak lengkap (Ji et al., 2022). Pada perkembangannya saat ini knowledge graph mengintegrasikan Graph Neural Networks (GNN), dan Graph Convolutional Networks (GCN) sebagai upaya dalam meningkatkan kemampuan reasoning dan prediksi berbasis graph (Wu, Pan, et al., 2019).

2.2.6.1. Named Entity Recognition

Dalam *Natural Language Processing* (NLP), teknik *Named Entity Recognition* (NER) merupakan teknik yang digunakan untuk mengidentifikasi dan mengklasifikasikan entitas dalam teks ke dalam kategori tertentu seperti nama orang, lokasi, organisasi, tanggal, dan produk (Ji et al., 2020). *Named Entity Recognition* berperan penting dalam ekstraksi informasi, sehingga memungkonkan *knowledge graph* dapat menghubungkan teks tidak terstruktur dengan entitas yang telah berada didalamnya.

Named Entity Recognition (NER) umumnya menggunakan metode berbasis machine learning, seperti *Conditional Random Fields* (CRF) dan model berbasis *deep learning*, seperti *Bidirectional LSTM-CRF* atau *Transformer-based models* (BERT-NER) (Yadav & Bethard, 2019).

2.2.6.2. Relation Extraction

Relation Extraction (RE) merupakan proses dalam *information extraction* yang bertujuan untuk mengidentifikasi hubungan antar entitas dalam suatu teks. Teknik ini digunakan dalam *automatic knowledge graph contruction* untuk membangun hubungan antar entitas yang telah dikenali sebelumnya oleh NER (Zhu et al., 2023).

Relation Extraction sendiri dalam pendekatannya menggunakan pendekatan berbasis rule based, supervised learning, maupun deep learning, terasuk model seperti *Graph Neural Networks* (GNN) dan *BERT based Relation Extraction*. Dalam penerapannya, teknik ini disgunakan dalam berbagai aplikasi seperti *question answering*, *biomedical knowledge discovery*, maupun *financial analytics* 22(Zhong et al., 20.).

2.2.6.3. Graph Convolutional Networks

Graph Convolutional Networks (GCN) merupakan model *deep learning* berbasis graph yang memperluas *Convolutional Neural Networks* (CNN) ke dalam struktur berbasis node dan edge, sehingga memungkinkan pembelajaran berbasis *structured data* dalam *Knowledge Graph* (Kipf & Welling, 2016).

Graph Convolutional Networks (GCN) bekerja dengan menggunakan *message passing mechanism*, di mana setiap node memperbarui representasinya berdasarkan informasi dari tetangganya. Pendekatan ini memungkinkan model memahami konteks hubungan antar-entitas, sehingga model ini dapat digunakan dalam tugas *node classification*, *link prediction*, dan *knowledge graph completion***.**

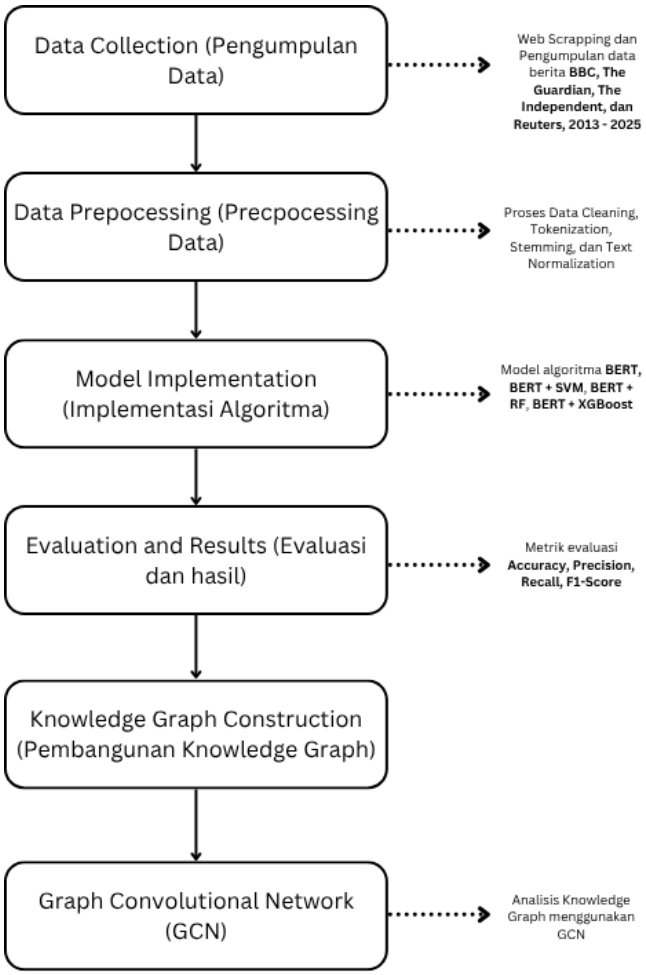
Penerapan *Graph Convolutional Networks* (GCN) mencakup *fraud detection*, di mana sistem dapat mengidentifikasi pola transaksi mencurigakan dalam jaringan keuangan, serta dalam analisis media sosial, di mana ia digunakan untuk memprediksi hubungan antar-pengguna berdasarkan interaksi mereka (Wu, Chen, et al., 2019).

BAB 3 METODOLOGI

Bab ini menjelaskan secara rinci mengenai metode yang digunakan dalam penelitian. Metode yang digunakan meliputi desain penelitian, prosedur pengumpulan dan pengolahan data, alur penelitian, algoritma dan teknik yang ditetapkan, serta evaluasi model. Setiap bagian-bagian metode penelitian tersebut akan dijelaskan secara mendalam guna memberikan gambaran yang lebih jelas mengenai tahapan penelitian yang dilakukan.

3.1 Uraian Metodologi

Penelitian ini menggunakan metode pendekatan kuantitatif melalui *Natural Language Processing* (NLP) berfokus pada analisis sentimen dalam portal berita besar, seperti *BBC*, *Reuters, The Guardian*, dan *The Independent*. Data dikumpulkan melalui *web scraping* maupun pengumpulan secara manual dengan rentang waktu 2013 hingga 2025. Proses *Natural Language Processing* (NLP) diterapkan dalam tahap *prepocessin*g, mencangkup *tokenisas*i, *stemming*, dan *stopword remov*al guna meningkatkan data sebelum analisis lebih lanjut. Studi ini dilakukan untuk membangun *Knowledge Graph* dengan menerapkan *Named Entity Recognotion* (NER) dan *Relation Extraction* (RE) guna mengidentifikasi entitas serta hubungan antar entitas dalam berita. *Graph Convolutional Network* (GCN) sebagai analisis *knowledge graph* lebih lanjut untuk memahami hubungan antar entitas secara lebih mendalam, serta menemukan pola keterkaitan yang tidak terlihat secara eksplisit dalam teks berita. Hasil yang didapat dari proses tersebut dilakukan untuk mengidentifikasi pola distribusi sentimen yang terdapat dalam berita digitalisasi teknologi pertanian serta korelsi dengan topik yang sebelumnya telah digunakan dalam proses klasifikasi artikel berita.



Gambar 3.1. Uraian Metodologi Penelitian

3.1.1 Persiapan Data

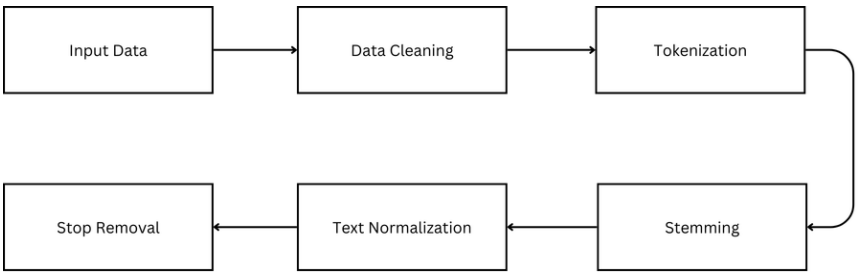
3.1.1.1. Informasi Dataset

Dalam proses persiapan data, dataset yang digunakan dalam penelitian ini terdiri dari artikel berita yang diumpulkan dari *BBC*, *Reuters*, *The Guardian*, dan *The Independent*, di mana artikel membahas terkait topik digitalisasi teknologi pertanian dalam rentang waktu 2013-2025. Pengumpulan data dilakukan melalui metode *web scraping* dan pengumpulan data manual yang diperoleh berdasarkan relevansi terhadap topik penelitian.

3.1.1.2. Prepocessing Data

Tahap *prepocessing* data merupakan langkah yang digunakan dalam proses analisis untuk mengkategorikan artikel maupun proses pembersihan data pada artikel. Di mana data yang telah dikumpulkan melalui proses *web scraping* sering kali mengandung elemen-elemen yang tidak relevan, seperti karakter khusus, tag HTML, maupun beberapa simbol sehingga perlu adanya proses permbersihan untuk meningkatkan kualitas data. Adapun

langkah – langkah *prepocessing* yang dilakukan dalam penelitian ini adalah sebagai berikut:



Gambar 3.2. Tahapan Prepocessing Data

Proses *preprocessing* data merupakan tahap penting dalam memastikan kualitas data sebelum digunakan dalam proses analisis sentiment pada algoritma *machine learning*. Tahapan awal yang digunakan adalah *data cleaning* atau pembersihan data, yang bertujuan untuk menghilangkan elemen-elemen yang tidak relevan, seperti karakter khusus, seperti tanda baca berlebihan, simbol @ atau #, *tag HTML*, kemudian dalam kategori spasi berlebihan yang sering muncul dalam suatu data yang dihasilkan dari pengumpulan artikel berita pada *web scraping* maupun proses manual lainnya. Proses ini memastikan teks yang dihasilkan hanya berisi informasi penting yang akan dianalisis oleh mesin.

Tahapan prepocessing data berikutnya adalah *tokenization***,** yaitu proses memecah teks menjadi unit-unit kecil yang disebut *token*. Tokenisasi mempermudah analisis lebih lanjut dengan mengubah teks menjadi kata-kata atau frasa pendek. Sebagai contoh dalam sebuah kalimat pada artikel berita, seperti "*Digitalisasi pertanian memberikan dampak positif*." Pada tokenisasi akan dipecah menjadi token-token seperti "*Digitalisasi*", "*pertanian*", "*memberikan*", "*dampak*", dan "*positi*f". Langkah ini dirancang untuk membantu algoritma *machine learning* memproses teks dalam unit terkecil yang bermakna.

Setelah proses tokenisasi berlangsung, dilakukan proses *stemming***,** yaitu langkah mengembalikan kata-kata ke bentuk dasarnya atau *root word*. Proses ini bertujuan menyederhanakan variasi morfologis kata yang memiliki makna serupa. Sebagai contoh dalam artikel berita, kata-kata seperti "*pupuk organik*", "*pemupuka*n", dan "*pupuk*" pada proses stemming akan disederhanakan menjadi "*pupuk*". Dalam penelitian ini, algoritma *stemming* seperti *Porter Stemmer* atau *Sastrawi Stemmer* digunakan, dimana keduanya telah terbukti efektif untuk teks dalam artikel. Dengan menyederhanakan kata-kata ke bentuk dasarnya, kompleksitas data dapat berkurang, sehingga memudahkan algoritma dalam mengenali pola linguistik.

Tahapan selanjutnya adalah ***text normalization***, di mana tahapan berikut merupakan tahapan yang memastikan format teks menjadi lebih konsisten dan standar. Normalisasi melibatkan langkah-langkah seperti mengubah semua huruf menjadi huruf kecil (*lowercase*), hal tersebut dilakukan untuk menghindari masalah duplikasi akibat perbedaan kapitalisasi, serta mengganti kata-kata informal atau slang pada artikel berita, seperti "*tdk*" atau "*tidaklah*" menjadi "*tidak*". Proses ini dilakukan dengan bantuan kamus normalisasi yang dirancang khusus untuk bahasa Indonesia, yang menyelaraskan kata-kata tidak baku ke dalam bentuk yang lebih formal dan baku.

Proses berikutnya setelah dilakukan *text normalization* adalah proses *stopword removal*, dimana terdapat penghapusan kata-kata yang sering muncul tetapi tidak memiliki makna signifikan dalam analisis. *Stopwords* pada artikel berita sering ditemukan pada kata "*yang*", "*di*", "*adalah*", dan "*dan*" dihapus untuk memastikan analisis lebih fokus pada kata-kata yang relevan dengan konteks.

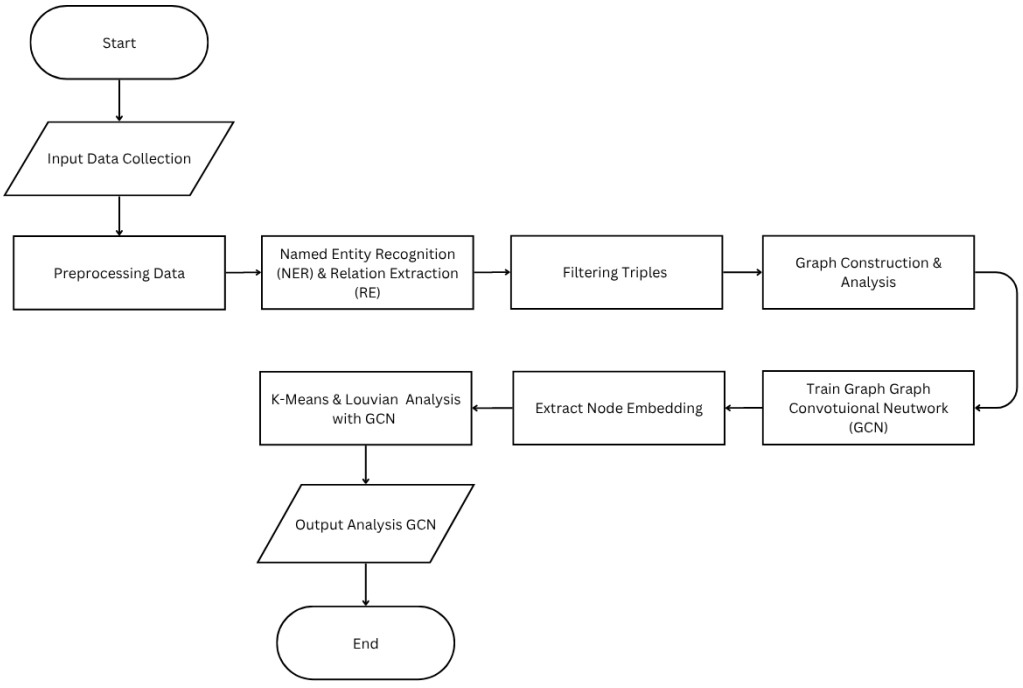
Tahapan preprocessing telah divisualisasikan dalam diagram proses Gambar 3.2, dilakukan secara sistematis dan menyeluruh untuk memastikan data memiliki kualitas tinggi sebelum digunakan dalam proses pembangunan *Knowledge Graph* dan pelatihan model seperti *Support Vector Machine* (SVM)**,** dan *Bidirectional Encoder Representations from Transformers* (BERT).

3.1.2 Pembangunan Knowledge Graph

Pembangunan *knowledge graph* dalam penelitian ini memiliki tujuan mempresentasikan hubungan semantik antar entitas dalam teks berita terkait digitalisasi pertanian khususnya pada persepsi publik untuk adopsi terknologi di era perkembangannya. Sejalan dengan hal tersebut, *knowledge graph* digunakan untuk memungkinkan struktur informasi dari teks berita menjadi bentuk jaringan-jaringan yang membentuk hubungan antar entitas. Dengan pendekatan ini, informasi yang awalnya tidak terstruktur pada artikel berita sebelumnya dapat dianalisis lebih lanjut guna memahami pola hubungan yang terbentuk dari berbagai macam artikel dan topik yang telah dikerucutkan. Proses pembangunan *knowledge graph* sendiri memiliki beberapa pendekatan sebagai proses ekstraksi data, proses dimulai dengan *Named Entity Recognition* (NER) yang digunakan sebagai identifikasi dan kategorisasi entitas penting dalam teks ke dalam beberapa kategori utama. Contoh nyata dari proses ini adalah terdapat kategori Organisasi meliputi *The Guardian, FAO*, kategori Teknologi meliputi *Artificial Intelligence, Smart Farming*. Dan beberapa kategori lainnya yang digunakan sebagai ektraksi pengkategorian kata dalam berita yang relevan dengan digitalisasi pertanian.

Pendekatan kedua menggunakan Ekstraksi hubungan dengan *Relation Extraction* (RE), di mana setelah entitas berhasil diidentifikasi, tahap selanjutnya adalah mengekstraksi hubungan antar entitas yang bertujuan untuk menentukan bagaimana entitas-entitas dalam teks saling berhubungan. Contoh nyata pada proses ini adalah *"BBC melaporkan bahwa digitalisasi pertanian mempengaruhi penggunaan Artificial Intelligence (AI) dalam pertanian di Inggris*". Dengan ekstraksi *Relation Extraction* (RE), maka didapatkan kata (*melaporkan, AI*) yang kemudian diperoleh informasi hubungan (*AI, digunakan dalam, pertanian Inggris*). Metode ini digunakan dikarenakan memiliki kemampuan dalam identifikasi relasi yang relevan dari teks, terutama dalam artikel berita yang memiliki struktur bahasa yang kompleks.

Setelah entitas dan hubungan berhasil diekstraksi, hasil dipresentasikan dalam bentuk graph dengan node sebagai dipresentasikan sebagai entitas dan edge sebagai penghubung node yang mempresentasikan hubungan antar entitas. Dalam penelitian ini, *knowledge graph* dibangun menggunakan teknologi graph database *gephi* dan visualisai pada python dengan library `networkx` yang memungkinkan visualisasi dan eksplorasi hubungan antar entitas secara interaktif. Model *knowledge graph* berikut memungkinkan analisis hubungan antar konsep yang lebih mendalam serta mendukung pemahaman pola informasi dalam liputan media. Tahapan ini telah divisualisasikan dalam Gambar 3.3, yang menggambarkan alur proses pembangunan *knowledge graph* dimulai dari *prepocessin*g dan *cleaning*, *Named Entity Recognition* (NER), *Relation Extraction*, dan *Graph Construction*.



Gambar 3.3. Alur Proses Pembentukan Knowledge Graph dan Analisis GCN

3.1.3 Analisis Knowledge Graph dalam Graph Convolutional Networks (GCN)

Knowledge Graph yang telah dibangun dengan mempresentasikan hubungan antar entitas dalam berita terkait digitalisasi pertanian perlu adanya analisis secara mendalam guna memberikan informasi yang menimbulkan keterkaitan antar entitas serta mengkkasifikasikan hubungan dalam knowledge graph menggunakan pendekatan *Graph Convolutional Networks* (GCN). Pendekatan *Graph Convolutional Networks* (GCN) merupakan salah satu metode *deep learning* yang diperluas ke data berbasis graph, di mana model ini mampu menangkap informasi dari *node* (entitas) dan *edge* (hubungan antar entitas). Dalam konteks penelitian ini, *Graph Convolutional Networks* (GCN) digunakan untuk menganalisis hubungan antar entitas yang terbentuk dalam *knowledge graph*, seperti mengklasifikasikan entitas ke dalam kategori tertentu, memprediksi hubungan baru antar entitas berdasarkan pola yang dtemukan dalam graph, dan memahami bagaimana media menghubungkan berbagai aspek digitalisasi pertanian.

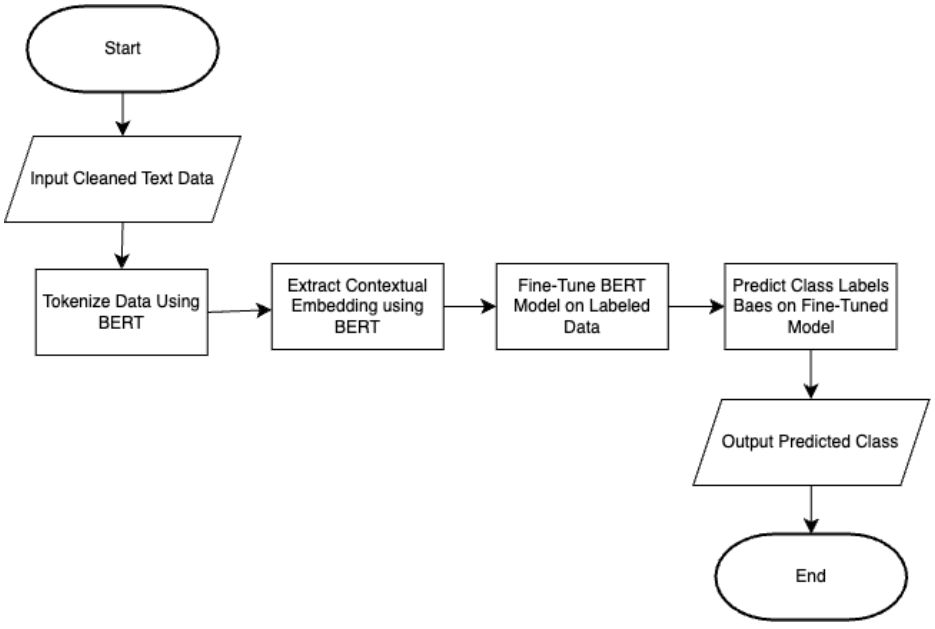
Proses *Graph Convolutional Networks* (GCN) diawali dengan representasi graph dalam bentuk *adjaency matrix*, kemudian digunakan sebagai input ke dalam model *Graph Convolutional Networks* (GCN). Sejalan dengan proses tersebut, *Graph Convolutional Networks* (GCN) mampu mengolah informasi dari *node* dan *edge* yang berdekatan untuk memperoleh representasi fitur yang lebih kaya. Hasil dari tahapan berikut dapat memungkinkan identifikasi oleh hubungan antara berbagai aspek dalam berita digitalisasi pertanian serta membantu prediksi keterkaitan antar konsep. Dengan demikian, pendekatan ini mampu menggabungkan kemampuan *knowledge graph* dalam mempresentasikan struktur relasi antar entitas dengan kekuatan fitur GCN dalam mengekstraksi representasi fitur yang kaya konteks. Node embedding yang dihasilkan oleh GCN selanjutnya dapat dimanfaatkan untuk berbagai tugas analisis mulai dari deteksi komunitas dan *clustering* entitas, prediksi kemunculan hubungan baru, hingga analisis sentimen. Sebagai analisis lebih luas pada sentimen, *embedding* diumpankan pada klasifikasi *machine learning* sebelumnya yang telah dilatih pada lebel sentimen *Pro, Contra*, dan *Neutral* sehingga memugkinkan penilaian polaritas sentimen secara granular pada tiap entitas dalam konteks digitalisasi pertanian. Hasilnya adalah wawasan komprehensif mengenai dinamika pemberitaan tentang bagaimana aspek-aspek teknologi dihubungkan pada media, serta sentimen kolektif yang terbentuk pada masing-masing komunitas entitas.

3.1.4 Analisis Sentimen Berita

Dalam tahapan analisis sentimen pada studi yang akan dilakukan bertujuan untuk mengidentifikasi pola sentimen dalam artikel berita terkait digitalisasi pertanian dengan memanfaatkan pendekatan algoritma *machine learning*. Penelitian ini mengkategorikan sentimen berita menjadi tiga kelas utama, yaitu *pro, contra*, dan *neutral* untuk memahami bagaimana media membingkai digitalisasi pertanian. Adapun algoritma *machine learning* yang digunakan dalam mengklasifikasikan sentimen diekstraksi dari teks berita melalui pendekatan *Biddirectional Encorder Representations from Transformers* (BERT), *Support Vector Machine* (SVM), *Random Forest*, yang akan diuraikan sebagai berikut:

3.1.2.1. Bidirectional Encorder Representations from Transformers

Biddirectional Encorder Representations from Transformers (BERT) merupakan salah satu model transformer yang dirancang untuk memahami konteks kata dalam teks dengan lebih dalam. Dalam penelitian yang dilakukan, *Bidirectional Encorder Representations from Transformers* (BERT) memiliki kemampuan untuk menangkap hubungan semantik antar kata dalam kalimat yang lebih unggul daripada proses *Natural Language Processing* (NLP).



Gambar 3.4. Alur Proses Algoritma BERT

Proses dimulai dengan *Input Cleaned Text Data***,** di mana data teks berita telah melalui tahap *prepocessing*. Pada tahapan ini, teks tidak hanya dibersihkan dari karakter yang tidak relevan, namun juga telah diperkaya dengan informasi hubungan antar entitas yang

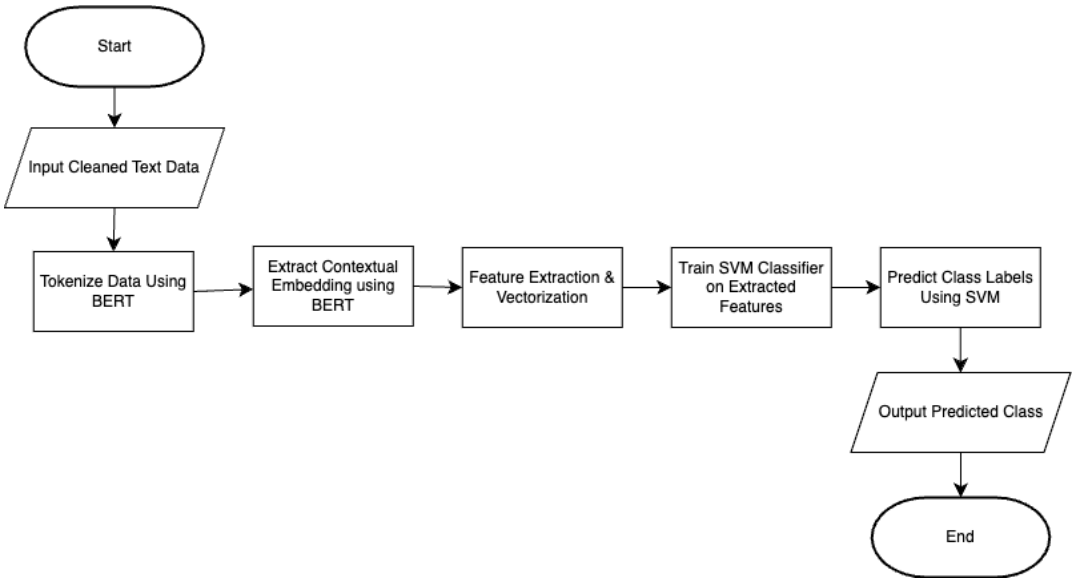
diperoleh dari *knowledge graph*. Setelah data dibersihkan, tokenisasi menggunakan *BERT Tokenizer*, di mana proses tersebut bertujuan untuk mengkonversi teks menjadi token yang dapat dimengerti oleh model. Dengan demikian, embedding kontekstual diekstraksi menggunakan model *Biddirectional Encorder Representations from Transformers* (BERT) yang belum digunakan untuk memahami representasi kata dalam konteks yang lebih luas.

Metode *Biddirectional Encorder Representations from Transformers* (BERT) kemudian dilakukan *fine tuning* pada dataset yang telah diberi sentimen positif, negatif, dan netral. Melalui pendekatan berikut, model dapat memahami pola sentimen dalam berita dan meningkatkan akurasi klasifikasi yang selanjutnya model digunakan untuk memprediksi kelas sentimen dari berita yang diuji.

Hasil klasifikasi sentimen kemudian dievaluasi menggunakan model metriks evaluasi seperti *accuracy*, *precission*, *recall,* dan *F1-Score* guna menilai performa model. Evaluasi ini memastikan bahwa pendekatan berbasis *Biddirectional Encorder Representations from Transformers* (BERT) memberikan hasil yang optimal dalam menganalisis sentimen berita terkait digitaliasi pertanian.

3.1.2.2. Biddirectional Encorder Representations from Transformers dan Support Vector Machine

Pendekatan *Bidirectional Encorder Representations from Transformers dan Support Vector Machine* menggabungkan keunggulan *Bidirectional Encorder Representations from Transformers* (BERT) dalam memahami konteks teks dengan *Support Vector Machine* (SVM) sebagai metode klasifikasi berbasis *hyperlane*. Proses ini dilakukan dalam beberapa tahap utama, yaitu ekstraksi fitur menggunakan *Biddirectional Encorder Representations from Transformers* (BERT), pelatihan model *Support Vector Machine* (SVM), *Bidirectional Encorder Representations from Transformers* (BERT) + *XGBoost,* prediksi sentimen, dan evaluasi performa model.



Gambar 3.5. Alur Proses Algoritma BERT + SVM

Tahapan dimulai dengan pengambilan data yang telah melalui *preprocessing* diberikan sebagai input ke dalam *BERT tokenizer* model. *Biddirectional Encorder Representations from Transformers* (BERT) bekerja dengan menghasilkan *contextual vector representations* dari teks menggunakan model *pretrained BERT* di mana setiap kata dalam teks dipetakan menjadi *embeddin*g yang menangkap hubungan semantik dengan kata lain dalam konteksnya. *Embeddings* ini dihasilkan melalui mekanisme yang memungkinkan model memahami kenteks kata secara lebih mendalam dibandingkan dengan metode *word embedding* tradisional seperti *TF-IDF* atau *Word2Ve*c.

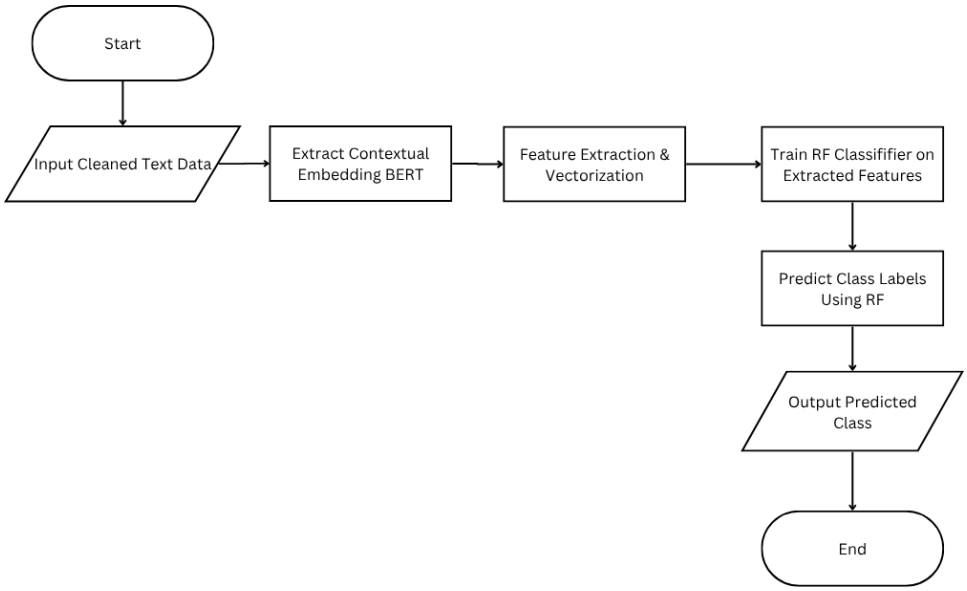
Setelah *feature extraction proce*ss selesai, representasi vektor yang dihasilkan oleh *Biddirectional Encorder Representations from Transformers* (BERT) digunakan sebagai **fitur masukan** untuk *Support Vector Machine* (SVM) model. Berbeda dengan pendekatan konvensional yang biasanya menggunakan *softmax classifier***,** metode klasifikasi default dalam *Biddirectional Encorder Representations from Transformers* (BERT) model pendekatan ini justru menggunakan *Support Vector Machine* (SVM) sebagai *primary classifier***.** *Support Vector Machine* (SVM) model kemudian dilatih menggunakan vector fitur yang diperoleh dari BERT, dengan tujuan membangun *optimal hyperplane* yang mampu memisahkan sentimen berita menjadi tiga kelas utama dalam "*Pro", "Contra",* dan *"Neutral*".

Tahap berikutnya adalah prediksi sentimen, di mana model yang telah dilatih digunakan untuk mengklasifikasikan artikel baru berdasarkan fitur yang dipelajari. Proses ini dilakukan dengan mengukur *hyperlane* yang telah dipelajari oleh *Support Vector Machine*

(SVM) selama proses *training*. Semakin jauh jarak suatu titik data dari hyperplane**,** semakin tinggi level model dalam menentukan sentimen berita. Terakhir, performa model BERT + SVM dilakukan menggunakan klasifikasi metriks, seperti *accuracy*, *precision*, *recall,* dan *F1-score***.** Evaluasi ini bertujuan untuk menilai efektivitas pendekatan melalui *Bidirectional Encorder Representations from Transformers* (BERT) + *Support Vector Machine* (SVM)**,** kombinasi *Bidirectional Encorder Representations from Transformers* (BERT) + *Random Forest* dan *Bidirectional Encorder Representations from Transformers* (BERT) + *XGBoost* yang dibandingkan**.**

3.1.2.3. Biddirectional Encorder Representations from Transformers dan Random Forest

Tahapan *Bidirectional Encorder Representations from Transformers* (BERT) dan *Random Forest* dimulai dengan ekstraksi fitur menggunakan model Bidirectional Encoder Representations from Transformers (BERT). Setelah korpus teks melalui proses *preprocessing* dan di tokenisasi, setiap token dilewatkan ke dalam pretrained BERT untuk menghasilkan embedding berdimensi tinggi yang merepresentasikan konteks semantik kata dalam kalimat.



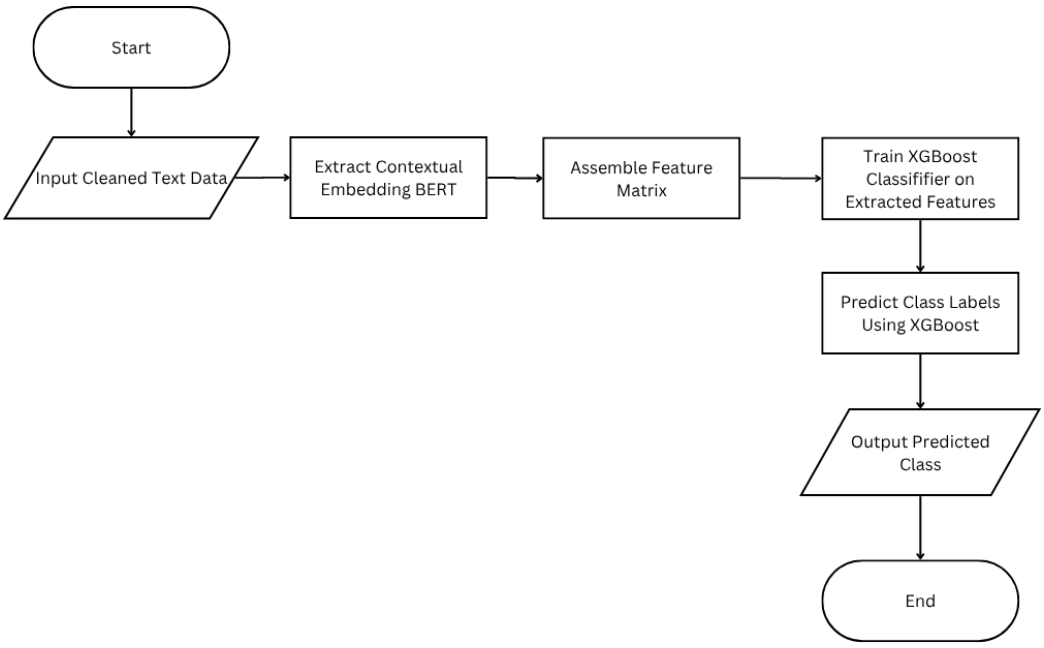
Gambar 3.6. Alur Proses Algoritma BERT + RF

Selanjutnya, seluruh himpunan *embedding* dikumpulkan menjadi matriks fitur yang menjadi input bagi model *Random Forest*. *Random Forest* membangun puluhan atau bahkan ratusan *decision tree* secara paralel, di mana setiap pohon dilatih pada subset acak data dan subset acak fitur untuk meningkatkan keberagaman. Keputusan akhir diambil melalui mekanisme *majority voting* antar pohon, sehingga model mampu menangkap interaksi non-linier antar fitur dan menahan *overfitting* lebih baik daripada single tree.

Pada fase pelatihan, hyperparameter Random Forest seperti jumlah estimator (`n_estimators`), kedalaman maksimum (`max_depth`), dan rasio fitur per split (`max_features`) dioptimalkan menggunakan teknik *Randomized Search Cross Validation*. Setelah model terpilih, proses prediksi menghitung vote antarpohon untuk menentukan label sentimen "*Pro", "Contra",* dan *"Neutral*" pada setiap artikel uji. Evaluasi performa dilakukan dengan metrik *accuracy, precision, recall*, dan *F1-score*, sehingga memperlihatkan seberapa baik *Random Forest* memanfaatkan *embedding BERT* dalam memisahkan ketiga kelas sentimen.

3.1.2.4. Biddirectional Encorder Representations from Transformers dan Random Forest

Sebagai langkah awal pada proses *Bidirectional Encorder Representations from Transformers* (BERT) dan *XGBoost*, dimulai dengan *preprocessing* dan tokenisasi korpus berita dan pada setiap token melalui proses BERT. Dari lapisan akhir BERT diekstrak pada *embedding* yang merefleksikan konteks sematik dalam kalimat. Seluruh himpunan *embedding* kemudian dirangkum menjadi matriks fitur di mana setiap baris mewakili *embedding* sebuah entitas atau unit teks. Matriks fitur hasil ekstraksi BERT kemudian mulai diproses ke model *XGBoost*.



Gambar 3.7. Alur Proses Algoritma BERT + XGBoost

Model *XGBoost* sendiri merupakan algoritma *gradient boosting* yang membangun rangkaian *descision trees* secara berurutan. Setiap *decision trees* sendiri dilatih untuk memperbaiki *decision trees* sebelumnya yang kemudian digunakan untuk meminimalkan

fungsi *loss* yang telah tergradasi. Pada fase optimasi, *hyperparameter* utama (`n_estimators`), (`learning_rate`), dan (`max_depth`), diatur ulang menggunakan *RandomizedSearch Cross Validation*. Setelah model terpilih, prediksi probablititas seperti kelas "*Pro", "Contra",* dan *"Neutral*" dihitung sebagai penjumalah kontribusi disetiap keputusan. Evaluasi performa dilakukan dengan metrik *accuracy*, *precision*, *recall*, dan *F1-Score*, sehingga dapat diperlihatkan seberapa efektif XGBoost memanfaatkan embedding BERT dalam memisahkan ketiga kelas sentimen.

3.1.5 Evaluasi Model

Proses dalam tahapan evaluasi hasil dilakukan untuk mengukur sejauh mana model yang dikembangkan mampu memberikan prediksi yang akurat sesuai dengan target klasifikasi. Proses evaluasi ini penting untuk menentukan performa setiap algoritma yang digunakan dalam penelitian, yaitu *Bidirectional Encorder Representations from Transformers* (BERT) + *Support Vector Machine* (SVM), *Bidirectional Encorder Representations from Transformers* (BERT) + *Random Forest,* dan *Bidirectional Encorder Representations from Transformers* (BERT) + *XGBoost*. Adapun metode evaluasi yang digunakan meliputi beberapa metrik berikut:

1. Accuracy

   Accuracy (Akurasi) merupakan metrik evaluasi yang mengukur presentase prediksi yang akurat terhadap total data yang diuji, berikut merupakan rumus perhitungan *accuracy*:

$$Accuracy = \frac{Jumlah\ Prediksi\ Benar}{Total\ data}$$

2. Precision

   Precision mengukur proporsi prediksi positif yang benar dibandingkan dengan seluruh prediksi positif yang dibuat oleh model, Jumlah kalimat positif yang diklasifikasikan ke dalam kelas emosi dengan benar ditunjukkan sebagai *True Positive* (TP), Jumlah kalimat negatif yang diklasifikasikan sebagai positif ke dalam kelas emosi ditunjukkan sebagai *False Positive* (FP), sementara jumlah kalimat negatif yang diklasifikasikan sebagai negatif ke dalam kelas emosi ditunjukkan sebagai *False Negative* (FN), berikut merupakan rumus perhitungan precision:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

3. Recall

Recall, atau sensitivitas, mengukur sejauh mana model mampu mendeteksi semua instance yang benar-benar termasuk dalam kelas positif. Adapun rumus *recall* adalah sebagai berikut:

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

4. F1-Score

F1-Score merupakan metrik gabungan yang mengharmonisasikan *precision* dan *recall*, terutama ketika terdapat *trade-off* antara kedua metrik tersebut. Adapun rumus terkait *F1-Score*, adalah sebagai berikut:

$$F - Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

BAB 4 HASIL DAN PEMBAHASAN

Pada bab ini akan disajikan hasil eksperimen dengan pendekatan analisis sentimen pada teks berita yang kemudian di analisis secara terstruktur pada relasi entitas melalui *Knowledge Graph*, dan diperkuat melalui *Graph Convolutional Network* (GCN). Analisis lanjutan tersebut guna mengungkap pola komunitas dan kemiripan antar-entitas. Proses dilakukan secara bertahap sesuai dengan proses analisis yang telah dijalankan untuk menjawab rumusan masalah penelitian.

4.1 Persiapan Data

4.1.1 Komposisi Korpus Berita

Dataset yang digunakan dalam penelitian ini terdiri dari 1.295 artikel berita terkait teknologi pertanian yang diperoleh dari empat portal berita daring Internasional, seperti *BBC*, *The Guardian*, *The Independent*, dan *Reuters* yang terbit pada rentang tahun 2013-2025. Proses pengumpulan dataset berupa artikel berita menggunakan metode scraping dan manual. Metode pengumpulan *scraping* dikumpulkan menggunakan *API Key* pada beberapa portal berita terkait. Proses pengumpulan artikel berita pada teknik *scraping* menggunakan kata kunci seperti, *"AgriTech",* publikasi artikel antara 2010-2024. Dataset yang terkumpul sekitar 6.300 artikel berita, dimana portal berita yang menggunakan teknik *scraping* adalah *BBC* dan *The Guardian*. Sejalan dengan pengumpulan artikel berita tersebut banyak artikel berita yang tidak lolos tahap filtering oleh penulis, terlebih pada berita yang tidak membahas adanya teknologi pertanian ataupun tema pertanian didalamnya.

```
import requests

# API Key
API_KEY = "API Key penulis"
# The Guardian API URL
BASE_URL = "https://content.guardianapis.com/search"
# More precise agricultural technology keywords
keywords = [
    "AgriTech", "Smart Farming", "AI in Agriculture", "IoT in Agriculture",
    "Automation in Agriculture", "Drones in Farming", "Agricultural Robotics",
    "Precision Agriculture", "Digital Farming", "Big Data in Agriculture",
    "Remote Sensing in Agriculture", "Climate-smart Agriculture"
]
# Research time range
from_date = "2010-01-01"
to_date = "2025-03-31"
# Target number of articles
max_articles = 1000
total_articles = 0
```

Metode pengumpulan artikel selanjutnya untuk dua portal berita lainnya (*The Independent dan Reuters*) menggunakan metode pencarian manual pada API dengan hasil 195

artikel. Adapun distribusi jumlah artikel akhir dari masing-masing sumber adalah sebagai berikut:

Portal Berita	Jumlah Artikel
BBC	100 artikel
The Guardian	1.000 artikel
The Independent	75 artikel
Reuters	120 artikel

Tabel 4.1 Distribusi Dataset Berita

Nama Kolom	Deskripsi	Tipe Data
Titile	Judul artikel berita	`Str`
URL	Tautan lengkap menuju artikel asli pada situs web sumber	`Str`
Date_Published	Tanggal dan waktu publikasi artikel	`Datetime`
Author	Nama penulis, atau organisaai yang menerbitkan artikel	`Str` atau `None`
Content	Isi lengkap dari artikel dalam bentuk paragraph naratif	`Str`
News	Sumber Portal Berita	`Str`

Tabel 4.2 Deskripsi kolom Dataset Berita

Melalui proses pengumpulan data yang dilakukan, data yang diperoleh masih bersifat mentah dan tidak dapat langsung digunakan pada tahapan berikutnya. Pada beberapa artikel masih terdapat teks berita yang terlalu panjang, terlalu banyak *noise* pada artikel, dan beberapa berita masih menggunakan topik diluar digitalisasi pertanian khususnya pada pengambilan berita berupa scraping. Contoh isi dataset mentah yang diperoleh dapat dilihat pada tabel sebagai berikut:

Judul	Isi Artikel
US Agtech Capital Drought Coninous, Dairy and Solar Sectors Offer Bright Sports	June 20 (Reuters) - The U.S. AgTech sector is navigating- a challenging investment climate@. Yet, amid the funding downturn, some companies are carving out growth opportunities, particularly in the “dairy and solar sectors”.
£45M For Technology Such as Fruit Picking Robots and Cow ‘Fitbits’ On Farms	Mr Zeichner said: “That is why I’m delighted to see money getting out the door to British farmers. This £45 million will support them with technology to boost food production, profits and the rural economy.”

Tabel 4.3 Isi Artikel Berita

Sejalan dengan hal tersebut, masih banyak kondisi-kondisi yang menujukkan isi artikel perlu menggunakan proses *pre-pocessing* data untuk menyaring berita, menyeragamkan isi pada kolom *content*, dan menghapus beberapa *noise* pada kalimat sehingga dapat digunakan pada tahap analisis berikutnya.

4.1.2 Pra-pemrosesan Data

Tahapan pemrosesan data dalam penelitian dilakukan untuk membersihkan dan mempersiapkan data agar layak digunakan dalam proses analisis lanjutan. Sebelum melakukan analisis sentimen dan konstruksi *knowledge graph*, setiap dokumen berita terlebih dahulu dibersihkan dan dinormalisasi melalui rangkaian tahapan pra-pemrosesan. Langkah-langkah tersebut meliputi cleaning dan normalisasi pada karakter non-alfanumerik seperti tanda baca, konversi huruf menjadi huruf kecil (*lowecasing*), penghilangan stopwords seperti "*the*", "*and*", "*in*" (*stopwords removing*), pemecahan teks menjadi token atau kata (t*okenization*), dan pengembalian kata pada bentuk dasar (*lemmatization*). Hasilnya merupakan korpus teks yang lebih konsisten dan terstruktur, sehingga model klasifikasi maupun ekstraksi entitas pada tahap selanjutnya dapat bekerja dengan lebih akurat dan efisien terlebih pada klasifikasi sentimen dalam berita.

4.4.1.1 Normalisasi Dokumen Berita

Pada tahapan normalisasi dokumen berita, setiap artikel yang telah di *filter* sesuai dengan topik teknologi pertanian melalui proses *cleaning* untuk menghilangkan *noise* dan menyeragamkan format teks. Tahap awalan yaitu melakukan l*owercasing*, dimana semua karakter pada isi konten berita diubah menjadi huruf kecil dengan `text = tex.lower()`, dengan tujuan agar kata yang di inputkan pada sistem diperlakukan sama. Kemudian dilanjutkan proses pembuangan karakter non-alfanumerik seperti angka, simbol, dan tanda baca. Pada proses ini, pembuangan angka, persentase tetap diikutkan guna kelengkapan isi konten berita secara utuh `text = re.sub(r'[^a-z0-9\$\s]', ' ', text)`. Contoh hasil pada teks artikel berita dari *"£45 M For Technology Such as Fruit ...*" menjadi "*for technology such as fruit ...*".

Tahapan normalisasi berikutnya setalah teks sudah bersih dari beberapa *noise* adalah tokenisasai dengan `toks = text.split()` yaitu memecah kata menjadi token dan dilanjutkan penghapusan *stopword* seperti *"the", "and",* dan *"in"* serta token pendek yang kurang dari ($\leq$2 huruf). Hal tersebut menurunkan dimensi data dan menghapus kata yang kurang informatif yang nanti juga berpengaruh pada relasi antar entitas pada tahapan pembuatan tripes dan relasi. Contoh pada teks adalah sebagai berikut, kalimat "*agtech sector is navigating a challenging…*" setelah mengalami tokenisasi dan *stopwords removing* menjadi "'*agtech', 'sector', 'navigating','challenging'…*".

Akhir dari tahapan normalisasi dokumen berita yaitu tahap lematasi *(lemmatization)* dengan `toks = [lemmatizer.lemmatize(tok, get_wordnet_pos(pos)) for tok, pos`

`in pos_tags]`, untuk mengembalikan bentuk kata dasar dari token. Adapau hasil akhir dari keseluruhan proses normalisasi dokumen ditunjukkan pada tabel 4.4 sebagai berikut.

Konten sebelum di Normalisasi	Konten setelah Normalisasi
June 20 (Reuters) - The U.S. AgTech sector is navigating- a challenging investment climate@. Yet, amid the funding downturn, some companies are carving out growth opportunities, particularly in the "dairy and solar sectors".	june reuters agtech sector navigate challenging investment climate yet amid fund downturn company carve growth opportunity particularly dairy solar sector
Mr Zeichner said: "That is why I'm delighted to see money getting out the door to British farmers. This £45 million will support them with technology to boost food production, profits and the rural economy."	Mr zeichner say why delight see money get out door british farmer this million will support them technology boost food production profit rural economy

Tabel 4.4 Isi Artikel Berita Sebelum dan Sesudah di Normalisasi

4.4.1.2 Pelabelan Teks Berita melalui Natural Language Processing (NLP)

Studi terkait pelabelan sentimen dokumen berita pertanian dilakukan secara otomatis menggunakan *Natural Language Processing* (NLP) dengan *VADER*, di mana proses tersebut dilengkapi dengan tiga *lexicon domain adapted* untuk kosakata teknologi pertanian guna meningkatkan sensivitas kontekstual.

Lexicon	Keyword	Makna Sentimen	Referensi
POS_LEX	fertilizer, yield, sustainability, opportunities, transparency	Menandai peningkatan, keberlanjutan, dan peluang	(Mohr & Höhler, 2023)
NEG_LEX	insufficient, delay, challenges, costs, skepticism	Menggambarkan hambatan, penundaan, atau keraguan	(Mohr & Höhler, 2023)
NEU_LEX	transforms, report, structural, according	Isi teks tidak dapat dikategorikan dalam prositif maupun negatif	(Z. Chen et al., 2022)

Tabel 4.5 Lexicon Domain Adapted Artikel

Lexicon pertama berdasarkan penelitian yang dilakukan oleh Mohr & Hohler (2023) untuk menangkap istilah-istilah positif dan negatif yang sesuai dengan topik teknologi pertanian (Mohr & Höhler, 2023), sedangkan lexicon kedua mengadopsi istilah netral dari Zhang et al.,

(2022) untuk mengurangi neutral false (Z. Chen et al., 2022). Proses pelabelan mengikuti mekanisme dua tahapan, yaitu compound score VADER yang dikombinasikan dengan pencocokan keyword lexicon untuk menetapkan label “Pro”, “Contra”, dan “Neutral”. Untuk menjamin kualitas pelabelan, validasi manual terhadap 120 sample menegaskan keandalan metode ini sebagai dasar ground truth bagi ekseprimen klasifikasi selanjutnya.

Dengan demikian, Lexicon Domain yang telah didefinisikan pada tabel 4.5 kemudian diterapkan pada two stage labeling. Di mana terdapat pemberian skor polaritas keseluruhan dengan memanfaatkan SentimentIntesityAnalyzer dari library VADER. Setiap dokumen berita diubah menjadi skor compound yang mencerminkan polaritas gabungan (-1 hingga +1), yang kemudian dicocokan dengan lexicon dengan kategori sebagai berikut:

- Jika skor ≥ 0.05 dan dokumen memuat satu atau lebih kata dari POS_LEX, maka label dikategorikan “Pro”
- Jika skor ≤ 0.05 dan dokumen memuat satu atau lebih kata dari NEG_LEX, maka label dikategorikan “Contra”
- Dokumen yang tidak memenuhi kriteria Pro atau Contra secara numerik maupun kontekstual dikategorikan sebagai Neutral

Hasil dari pencocokan skor pada *lexicon* memenuhi syarat untuk dikategorikan di beberapa sentimen, dengan kata lain dokumen hanya diberi label ekstrem “*Pro*” dan “*Contra*” jika terbukti numerik sesuai dengan skor *VADER* dan kontekstual dari *keyword agritech*. Sementara itu, apabila terdapat dokumen yang tidak terbukti kontekstual sesuai pada skor maka dibiarkan “*Neutral*”. Berdasarkan hasil pelabelan, distribusi tiga kelas sentimen ditampilka pada Tabel 4.6. sebagai berikut:

Label	Jumlah	Presentase (%)
Pro	433	63.5
Contra	156	22.9
Neutral	93	13.6

Tabel 4.6 Distribusi Sentimen Artikel

Distribusi sentimen memperlihatkan bahwa mayoritas artikel sekitar 63,5% memuat *framming* yang mendukung “*Pro*” sebagai adopsi teknologi pertanian, sementara opini negatif atau “*Contra*” hanya 22,9% dan 13,6% teks bersifat netral. Kecenderungan dominasi label “*Pro*” ini kemungkinan mencerminkan sikap optimism media terhadap inovasi teknologi pertanian selama periode pengamatab. Sebaliknya porsi “*Contra*” relatif rendah yang menunjukkan

bahwa pemberitaan jarang menyoroti kekhawatiran atau hambatan implementasi teknologi. Adanya 14% dokumen "*Neutral*" sendiri menandakan sejumlah artikel bersifat informatif atau deskriptif tanpa muatan opini yang jelas. Untuk memperjelas mekanisme *two stage labeling*, berikut contoh dokumen artikel beserta skor *VADER* dan kata kunci yang mempengaruhi label.

Label	Konten Artikel	Compound	Keyword Match
Pro	climate change pose grow threat banana production farmer report great transparency	0.75	`transparency`
Contra	note send along memo usda say review cause undue delay direct aid individual exempt process	-1.00	`delay`

Tabel 4.6 Hasil Pelabelan Dokumen

Pada dataset pelabelan, dipilih dua dokumen yang merepresantatifkan pelabelan *pro* maupun *contra* dengan skor *compound* tertinggi yang memenuhi syarat *keywords* pada domain "*transparency*", dan "*delay*". Pemilihan berdasarkan dua tahapan, di mana dokumen harus memiliki skor polaritas gabungan *VADER* di atas ≥ 0.005 atau dibawah ≤ 0.05. Kemudian untuk syarat yang kedua adalah memuat minimal satu kata dari *lexicon domain adapted*. Dengan demikian, label "*Pro*", dan "*Contr*a" hanya diberikan ketika kedua kondisi numerik dan kontekstual terpenuhi, sedangkan dokumen berita yang tidak memenuhi salah satu dari keduanya dikategorikan sebagai label "*Neutral*". Adapun contoh tersebut menegaskan efektivitas skema *two stage labeling* dalam menahan *false positive* merupakan langkah yang tepat sebagai penentuan skema labeling pada artikel berita.

4.4.1.3 Klasifikasi Topik dalam Dokumen Berita

Setelah melalui tahapan dalam klasifikasi sentimen pada artikel berita, setiap dokumen melalui tahapan klasifikasi topik. Pada tahap klasifikasi topik, setiap dokumen berita ditempatkan ke dalam salah satu dari enam kategori utama berdasarkan kemunculan *keywords* dan bobot komponen menggunakan *library NMF* pada represntasi *TF-IDF*. Keenam topik dirujuk dari Wolfert et al., (2017), dan Kamilaris et al., (2017) yang dijabarkan pada Tabel 4.7. Pendekatan tersebut memungkinkan pemetaan secara tajam terhadap kecenderungan liputan media teknologi pertanian sekaligus menyediakan kerangka empiris sebagai analisis lebih lanjut pada distribusi sentimen maupun analisis korelasi dengan relasi antar-entitas pada proses selanjutnya.

Topik	Kategori Topik (Label)	Lima Keywords Teratas	Referensi
0	Food Security & Risk Management	drone, ministry, country, official, security	(Wolfert et al., 2017)
1	Climate & Environmental Monitoring	climate, emission, carbon, agriculture, methane	(Kamilaris et al., 2017)
2	Big Data Analytical & Clous Platforms	robot, use, drone, human, technology	(Wolfert et al., 2017)
3	Economic & Family Farming	farmer, farm, food, tax, crop	(Kamilaris et al., 2017)
4	Biotech & Crop Genetic	gene, china, gene edit, edit, crop	(Kamilaris et al., 2017)
5	Market Access & Trade Dynamics	australian, market, price, labour, cut	(Wolfert et al., 2017)

Tabel 4.7 Kategori Topik pada Dokumen Berita

Kemudian, setelah pemberian label, dihitung frekuensi dokumen per topik untuk melihat distribusi keseluruhan yang hasilnya adalah sebagai berikut:

Kategori Topik (Label)	Jumlah Artikel	Presentase (%)
Food Security & Risk Management	170	28
Climate & Environmental Monitoring	143	23.6
Big Data Analytical & Clous Platforms	133	21.9
Economic & Family Farming	101	16.6
Biotech & Crop Genetic	90	14.8
Market Access & Trade Dynamics	45	7.4

Tabel 4.8 Distribusi Topik Dokumen Berita

Pada hasil kategorisasi topik artikel, topik *Food Security & Risk Management* mendominasi sekitar 28%, hal ini mencerminkan banyaknya laporan terkait ketahanan pangan dan manajemen reskio terlebih pada iklim maupun pada bidang agroteknologi dibuktikan dengan korelasi topik *Climate & Environmental Monitoring* yang mencapai angka 23.6% dan *Big Data Analytics & Cloud Platforms* sekitar 21.9% yang menempati porsi besar mengindikasikan minat media pada isu-isu terkait. Topik *Market Access & Trade Dynamics* memiliki klasifikasi topik relative rendah dengan 7.4%, hal tersebut dapat dimungkinkan pada kecendurungan media dalam liputan perdagangan lebih tersegmentasi atau memiliki pengaruh unsur-unsur lainnya seperti *keywords* yang menempal lebih banyak pada topik-topik tertentu.

Untuk lebih memperjelas kembali terkait mekanisme klasifikasi topik dalam artikel berita, berikut contoh aplikasi ke dataset, yang memuat tiga sample berita dari hasil *filtering* dengan menampilka contoh *top 3 keywords* yang digunakan dalam artikel. Pada tabel dicantumkan tiga contoh topik dengan tema *Climate & Environmental Monitoring*, *Big Data Analytical & Cloud Platforms*, serta *Biotech & Corp Genetics*.

Topik	Judul	Isi Artikel	Top 3 Keywords
Climate & Environmental Monitoring	The Future of Bananas Is Under Threat	"climate change pose grow threat banana production farmers …"	sustainability, fertilizer, crop
Big Data Analytical & Cloud Platforms	Us Agtech Capital Drought Continues, Dairy and Solar Sectors Offer Bright Spots	"agtech sector navigate challenging headwin…"	agtech, data, platform
Biotech & Crop Genetics	Australian Trial of Gene-Edited Wheat Aims For 10% Bigger Yields	"researchers trial gene editing in rice to enhance drought… …"	gene, biotech, edit

Tabel 4.9 Klasifikasi Topik pada Dokumen Berita

Sebagai acuan dalam klasfikasi dan analisis lebih lanjut, pemetaan topik pada tahap pra-permrosesan dataset berikut berhasil menggambarkan keragaman tema liputan agritech dalam korpus berita berdasarkan segmentasi tema menyeluruh. Sehingga menyediakan landasan empiris yang kuat untuk analisis lanjutan mengenai dinamika media dalam menyoroti masalah pada sektor pertanian.

4.2 Evaluasi Klasifikasi Sentimen Dokumen Berita

Pada bagian ini, dilakukan evaluasi kinerja berbagai metode klasifikasi sentimen pada kumpulan dokumen berita yang telah melalui pra-pemrosesan dataset sebelumnya. Tahapan evaluasi klasifikasi sentimen dibagi menjadi dua pendekatan namun memiliki tujuan yang sama yaitu melihat performa dari model mana yang memiliki hasil terbaik dari segi akurasi dan *f1-Score* sebagai klasifikasi sentimen dokumen berita teknologi pertanian. Langkah pertama menguji coba model *BERT* dengan Fine Tuning hyperparameter yang kemudian sebagai *embedding*, *BERT* dipadukan pada *SVM*, *Random Forest*, dan *XGBoost*. Akhirnya dari keempat model kemudian dilihat mana yang memiliki performa terbaik. Hasil performa terbaik diuji melalui *meta leaner GradienBoost* dan RandomizedSearch pada parameter untuk memvalidasi model terbaik dengan uji coba *hyperparameter* terbaik masing-masing model. Adapun uji *paired test* juga digunakan untuk membandingkan hasil *base leaner* terbaik tanpa *stacking* dengan rata-rata hasil *stacking* .

4.2.1 Konfigurasi Model Dokumen Berita

Konfigurasi model dokumen berita dilakukan melalui eksperimen dokuman tahap awal dengan membandingkan model *BERT* sebagai *base* untuk *embedding* terhadap *SVM*, *Random Forest*, dan *XGBoost* yang digunakan sebagai *classifier* klasifikasi sentimen. Adapun model yang digunakan, untuk mempermudah identifikasi konfigurasi, diberikan kode konfigurasi

model pada Tabel 4.10 sebagai acuan dalam penjelasan analisis hasil konfigurasi maupun eksplorasi lebih lanjut.

Model	Kode
BERT Fine-Tune	BF
BERT + SVM	BS
BERT + Random Forest	BR
BERT + XGBoost	BX

Tabel 4.10 Kode Model Sentimen

Dalam penelitian ini, pemilihan hyperparameter pada setiap model dasar dilakukan secara sistematis untuk memastikan performa optimal.

Model	**Hyperparameter**	**Nilai**
BERT Fine-Tune	Learning rate	2e-5
	Epochs	8
	Batch Size	16
	Random State	42
BERT + SVM	C	1e-2, 1e-1, 1, 10, 100,1e3
	Kernel	[Poly, rbf, linear]
	gamma	[scale, auto, 1e-3, 1e-2, 1e-1,1]
BERT + Random Forest	n_estimator	[100, 300, 500, 800, 1000]
	max_depth	[None, 10, 20, 30, 50]**;**
	leaf	[1, 2, 5]
	split	[2, 5, 10]
BERT + XGBoost	n_estimator	[100, 300, 500, 800, 1000]
	max_depth	[3, 5, 7, 9]
	Learning rate	[0.01, 0.05, 0.1, 0.2]

Tabel 4.12 Hyperparameter Utama Model Sentimen

Pada *BERT-Fine Tune* ditetapkan nilai *learning rate* 2e-5 setelah dilakukan melalui parameter *learning rate* 3e-5, 1e-5, dan 5e-6 untuk hasil terbaik diperoleh menggunakan *learning rate* 2e-5 (2×10^{-5}), sedangkan jumlah *epoch* sendiri melalui 8 epoch setelah 10 dan 20 epoch memiliki hasil dibawah 8 epoch dan memiliki *running time* sekitar 42 hingga 50 menit pada hasil eksperimen sebelumnya, sedangkan untuk *Batch Size* menggunakan 16, berdasarkan praktik terbaik pada *literature fine-tuning model pre-trained*. Sementara itu, *hyperparameter* pada SVM menggunakan *setting parameter* dalam rentang *C* dari 10^{-2} hingga 10^{3}, dan dipertimbangkan *kernel poly, linear, rbf*. Pada *Random Forest* sendiri penggunaan *hyperparameter estimator* pada rentang 10 – 1000, dengan kedalaman maksimal 50. Pada *XGBoost* memiliki *estimator* yang sama dengan *Random Forest* dan kedalaman maksimal 9.

Sistem pencarian parameter terbaik menggunakan *RandomizedSearch* dengan rentang parameter yang telah dideklarasikan sebelumnya.

Setelah menetapkan ruang pencarian pada hyperparameter, tahap selanjutnya adalah mengevaluasi kombinasi terbaik melalui eksperimen *RandomizedSearch* dengan uji coba disetiap fold pada setiap prosesnya. Hasil eksplorasi diorganisis berdasarkan kode model, yang akan ditunjukkan pada tabel uji eksplorasi hyperparameter. Metrik utama yang ditampilkan pada tabel uji coba adalah *Accuracy* dan *F1-macro*, sehingga dapat diamati seberapa besar peningkatan performa yang dicapai dibandingkan konfigurasi default. Adapun pada pengujian hyperparameter model dokumen berita, disajikan ringkasan hasil uji coba sebagai berikut:

Model	***Hyperparameter***				**Accuracy (%)**	**F1-macro (%)**
	Epoch	**Learning Rate**	**Batch Size**	**Random State/C/ n_estimators/max_depth**		
BF - 01	8	2e-5	16	Seed: 42	68.0	60.2
BF – 02	8	2e-5	16	Seed: 42	71.0	59.5
BF – 03	8	2e-5	16	Seed: 42	71.0	60.3
BF – 04	8	2e-5	16	Seed: 42	66.0	53.9
BF - 05	8	2e-5	16	Seed: 42	65.0	50.2
BS - 01	8	2e-5	16	Seed: 42; kernel: rbf**;** γ: scale**;** C:0.1	97.1	96.1
BS - 02	8	2e-5	16	Seed: 42; kernel: rbf**;** γ:0.001**;** C:1000	94.2	91.9
BS - 03	8	2e-5	16	Seed: 42; kernel: rbf**;** γ: 0.001**;** C:1000	72.1	54.7
BS - 04	8	2e-5	16	Seed: 42; kernel: rbf**;** γ: scale**;** C:0.1	96.3	95.4
BS - 05	8	2e-5	16	Seed: 42; kernel: rbf**;** γ: scale**;** C:0.1	96.3	94.8
BR – 01*	8	2e-5	16	Seed: 42; n_estimators:100**;** max_depth: None	**97.8**	**96.9**
BR - 02	8	2e-5	16	Seed: 42; n_estimators:100**;** max_depth: None	94.2	92.4
BR - 03	8	2e-5	16	Seed: 42; n_estimators:100**;** max_depth:50	75.7	59.3
BR - 04	8	2e-5	16	Seed: 42; n_estimators:1000**;** max_depth:20	97.1	96.1
BR - 05	8	2e-5	16	Seed: 42; n_estimators:100**;** max_depth: None	97.1	96.5

BX - 01	8	2e-5	16	Seed: 42; n_estimators:1000; max_depth:7	95.6	93.6
BX - 02	8	2e-5	16	Seed: 42; n_estimators:500; max_depth:9	92.7	89.8
BX - 03	8	2e-5	16	Seed: 42; n_estimators:1000; max_depth:9	74.3	61.0
BX - 04	8	2e-5	16	Seed: 42; n_estimators:800; max_depth:3	98.5	98.5
BX - 05	8	2e-5	16	Seed: 42; n_estimators:100; max_depth:7	96.3	96.1

Tabel 4.13 Eksplorasi Hyperparameter menggunakan RandomizedSearch

Eksplorasi *hyperparameter* yang digunakan pada model, model BF menggunakan *Cross Validation* (BF 01 – 05) mempertahankan penggunaan *learning rate* 2×10^{-5}, 8 epoch dan batch size 16, performa BF hanya berkisar antara 65% hingga 71% pada *accuracy*. Pada *fold 2* dan 3 (BF – 02 dan BF – 03) sedikit mengungguli varian lain dengan *accuracy* 71% menegaskan bahwa *fine-tuning* pada *BERT* murni mencapai platu tanpa dukungan *classifier* eksternal.

Selanjutnya integrasi embedding *BERT* dengan *SVM* melalui RandomizedSearch pada *Cross Validation* memperlihatkan peningkatan jauh dari sebelumnya menggunakan *BERT* tunggal. BS – 1 dengan kernel rbf, gamma scale, dan C 0.1 memunculkan hasil tertinggi accuracy 97.1% dan F1-Macro mencapai 96.1%. Hal tersebut juga memunculkan hasil yang seimbang pada BS 04 dan BS – 05 dengan gamma scale serta C 0.1 yang memiliki hasil 96% pada accuracy. Sedangkan penggunaan C terlampau tinggi sekitar 1000 mengalami penurunan performa, khususnya BS – 03 yang hanya mencapai 72.1 % untuk accuracy, yang mengindikasi overfitting pada beberapa fold.

Sedangkan pada hasil BR sendiri, fold pertama (BR – 01) dengan estimator 100 dan kedalaman (depth) memberikan performa tinggi dengan accuracy 97.8% dan F1-macro 96.9%. Dari beberapa eksplorasi parameter pada RandomizedSeacrh disetiap fold, menambah jumlah estimator pohon hingga 1000 pada BR – 04 hanya menurunkan accuracy menjadi 97.1% menunjukkan *diminishing returns* pada penambahan estimator jika dibandingkan dengan BR - 01. Sementara itu, pada BR – 03 menambah varian max_depth (kedalaman) sekitar 50, justru menurun drastis ke 75.7%, menandakan overfitting ketika lapisan pohon terlalu dalam.

Akhirnya, kombinasi model BX mencapai puncak performa BX – 04 dengan 800 pohon estimator dan kedalaman sekitar 3 dengan accuracy serta F1-macro 98.5%. Kedalaman pohon yang lebih kecil, jika melihat uji eksplorasi pada BR dan BX terbukti mengoptimalkan boosting.

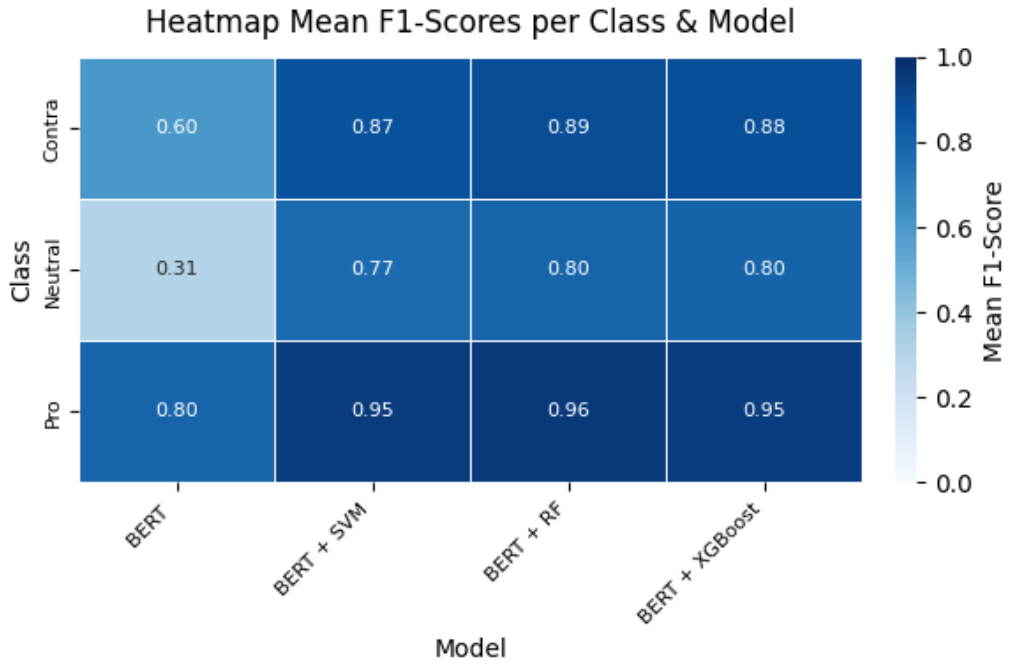
Varian kedalaman 0 pada BX – 03 juga memiliki kasus yang serupa dengan sebelumnya menurun hingga 74.3% pada accuracy. Hal itu menandai overfitting, meskipun BX-02 dengan estimator 500 dan kedalaman 9 tetap mempertahankan accuracy diatas 90%, menunjukkan ketangguhan XGBoost terhadap variasi parameter.

Model	Accuracy (%)	Std Accuracy	F1_Score (%)	Std F1_Score
BF	68.0	0.025	57.0	0.046
BS	91.0	0.096	87.0	0.160
BR	**92.0**	0.084	87.0	0.146
BX	91.0	0.088	**88.0**	0.137

Tabel 4.14 Performa Model Sentimen

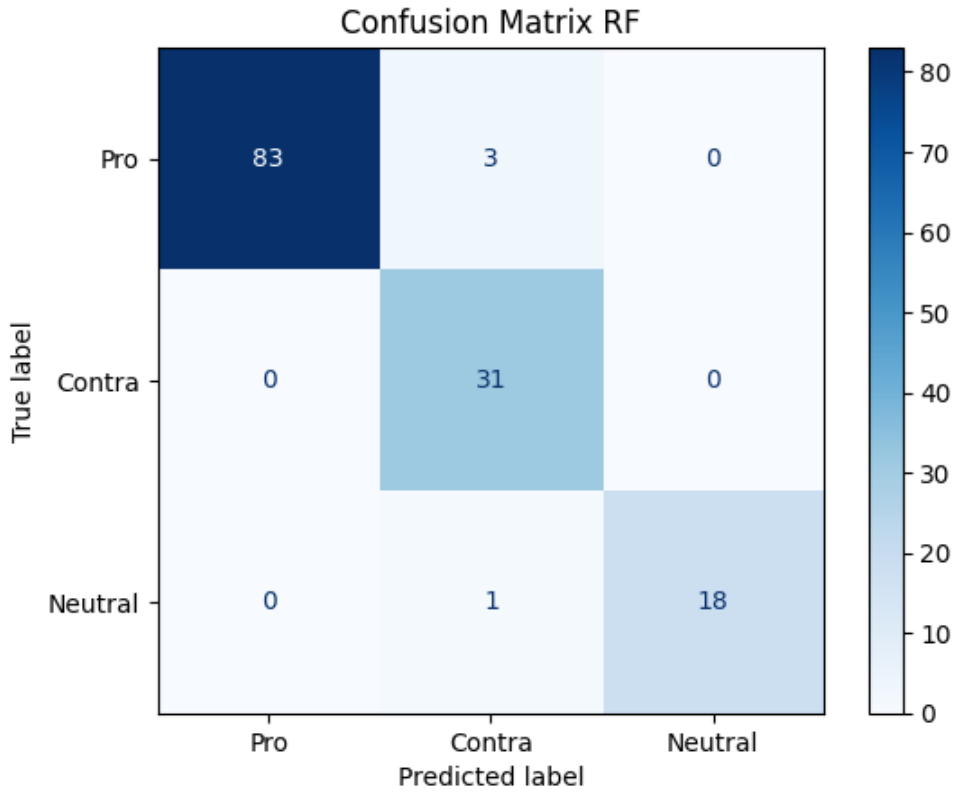
Berdasarkan konfigurasi *hyperparameter* terbaik diatas, keempat model dasar (BF, BS, BR, BX) kemudian dievaluasi secara menyulurh menggunakan *Cross Validation*, yang menghasilkan hasil rata-rata keseluruhan model pada beberapa parameter. Tabel 4.14 menyajikan metrik rata-rata *Accuracy* dan *F1-score* serta standar deviasi keduanya yang akan dijadikan acuan untuk analisis lebih lanjut pada distribusi sentimen maupun topik dokumen berita, yang kemudian diuji *meta leaner* untuk hasil model terbaik.

Hasilnya, model BF hanya mencapai rata-rata accuracy sekitar 68% dan F1-Score 57%, sedangkan BS, BR, dan BX memiliki hasil rata-rata yang hampir sama sekitar 91%-92% untuk accuracy dan 87%-88% untuk F1-Score, dengan BR unggul di ketiga model meskipun tidak terlalu besar sekitar 92% dan untuk F1-Score sendiri BX memiliki score tertinggi sekitar 88%. Standar deviasi yang lebih rendah pada BR dan BX mengindikasikan konsistensi lebih baik antar fold.



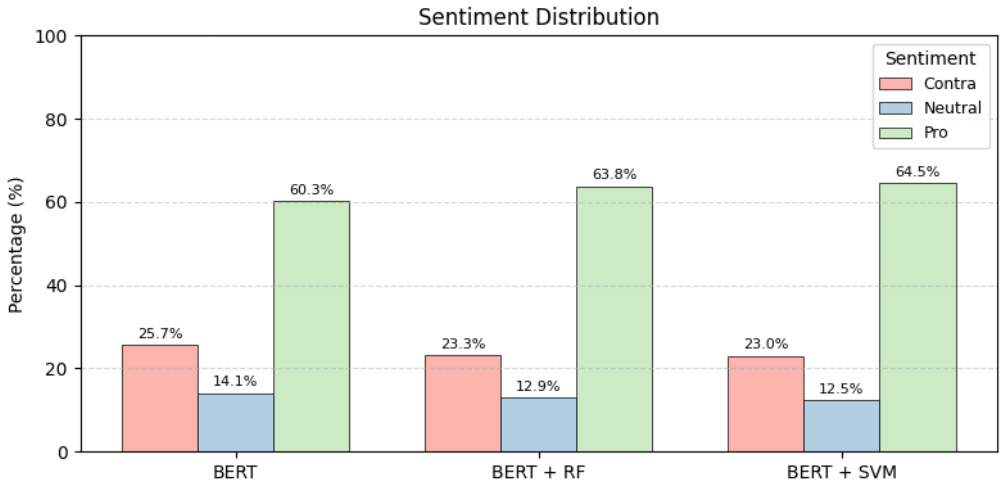
Gambar 4.1. Heatmap Mean F1-Score Model Sentimen Dokumen Berita

Untuk menggali lebih dalam distribusi performa pada masing-masing kelas sentimen seperti *Pro, Contra*, dan *Neutral*, divisualisasikan mean F1-Score per kelas dari setiap model dalam bentuk heatmap. Visualisasi ini akan membantu mengidentifikasi kekuatan dan kelemahan tiap model dasar dalam mengangai ketidakseimbangan dan kompleksitas kelas sentimen. Pada visualisasi heatmap sendiri terlihat model BR (BERT + Random Forest) selain unggul pada kelas *accuracy* juga unggul dalam klasifikasi sentimen khusunya dalam klasifikasi kelas *Pro* dengan hasil 96% sedikit unggul dengan BS dan BX sekitat 95%. Keseluruhan dalam kelas *Contra* memiliki hasil tertinggi dengan 89% pada klasifikasi. Disini, secara keseluruhan BR memiliki performa tertinggi dalam kelas rata-rata *f1-score* yang dapat digunakan sebagai indikasi model terbaik pada analisis model. Sejalan dengan hal tersebut, untuk memeriksa pola kesalahan prediksi, ditampilkan *confusion matrix* dari model BR-01 yang merupakan model dengan tingkat accuracy terbaik sebelumnya pada Gambar 4.2 guna melihat distribusi sentimen dan validasi persebaran sentimen dalam model.



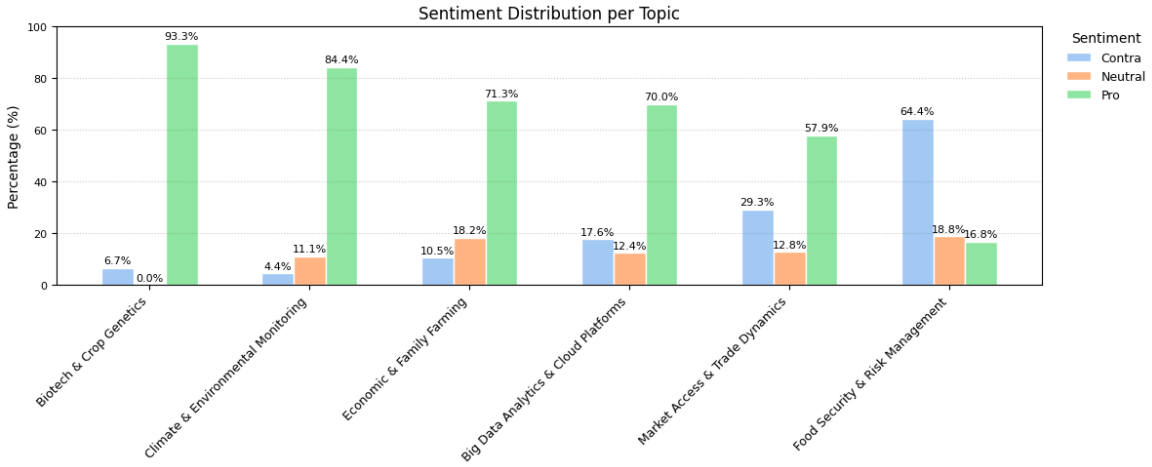
Gambar 4.2. Confussion matrix BERT + RF

Analisis hasil *confussion matrix* BR sendiri hampir seluruh dokumen Pro dan Contra diklasifikasikan benar. Dari 86 dokumen Pro dan 31 dokumen Contra terklarifikasi dengan benar. Sementara 1 dari 19 dokumen Neutral tidak tepat diklasifikasikan ke arah Contra. Tingginya diagonal pada matriks ini menegaskan keandalan BR-01 dalam membedakan sentimen ekstrem. Dengan demikian, meskipun BR – 01 memiliki accuracy mencapai 97%, perbaikan threshold atau augmentasi data untuk kelas Neutral dapat lebih meningkatkan keseimbangan performa. Kemudian, pada tingkat distribusi sentimen ketiga model terbaik seperti BS, BR, dan BX dapat divisualisasikan sebagai berikut pada Gambar 4.3.



Gambar 4.3. Distribusi Sentimen Model Terbaik Dokumen Berita

Distribusi presentase prediksi ketiga sentimen yang memiliki performa terbaik dibandingkan untuk analisis klasifikasi lebih lanjut pada layer sentimen. Terlihat bahwa ketiganya cenderung terfokus pada kelas Pro dengan presentase sekitar 60% - 65% dari semua dokumen. Sedangkan kesesuaian dengan hasil pada ground truth, kelas Neutral memiliki rata-rata sekitar 15% dan Contra mendekati 23% hingga 26%. Hal tersebut mengindikasikan bias korpus atau bias model ke sentimen positif. Selain segmentasi dokumen berita dengan ketiga model terindikasi pada bias kelas pro, distribusi awal pada dokumen berita juga terindikasi bias pada sentimen Pro. Maka dapat diambil kesimpulan bahwa berita bertajuk positif memberi gambaran optimisme tentang adopsi teknologi pada sektor pertanian. Dengan demikian, korpus berita ini dapat dijadikan rujukan empiris untuk memperkuat argumen bahwa perkembangan teknologi pertanian selaras dengan kebutuhan modernisasi produksi saat ini. Fakta bahwa sentimen positif begitu konsisten muncul menegaskan keyakinan pemangku kepentingan terhadap peran teknologi sebagai pendorong utama efisiensi, produktivitas, dan keberlanjutan dalam praktik pertanian kontomporer. Untuk memperkuat distribusi sentimen bukan hanya membandingkan pada model klasifikasi. Diberikan juga kategori topik yang dikolerasi secara langsung pada label sentimen.



Gambar 4.4. Distribusi Sentimen pada Topik Dokumen Berita

Untuk melengkapi pemahaman konteks tematik, Gambar 4.4 menyajikan distribusi sentimen pada setiap topik dokumen berita. Dari hasil visualisasi kategori per topik sendiri domain *Biotech & Crop Genetics* dan *Climates & Environmental Monitoring* didominasi dengan sentimen Pro, dimana kelas berikut sangat dominan sekitar 93.3% dan 84.4%. Sedangkan topik *Food Security & Risk Management* justru lebih di dominasi oleh sentimen Contra sekitar 64.4% mencerminkan kecemasan akan risiko dalam keamanan pangan. Variasi ini menggambarkan bahwa optimisme terhadap teknologi pertanian tidak merata di semua sub-bidang, sehingga strategi komunikasi dan kebijakan perlu disesuaikan dengan karakteristik pada masing-masing topik.

Selain menggunakan metode berikut, dilakukan eksperimen menggunakan 390 dataset artikel berita yang telah di proses bersih dimana data tersebut telah di analisis sentimen metode *balancing* data dengan *thresehold* sekitar 0.1 untuk dataset. Kemudian diberikan juga *balancing* pada *ground truth* untuk menyamakan distribusi kelas menggunakan deklarasi target kelas seperti berikut: Pro: 170; Contra: 130; Neutral:90. Proses *balancing* berikut digunakan pada dataset awal dalam distribusi kelas sentimen, maupun distribusi kelas target pada masing-masing model. Metode berikut sebagai metode yang digunakan untuk persebaran distribusi sentimen yang lebih rata dan tidak terlihat dominan pada kelas-kelas tertentu.

Pada eksperimen yang dilakukan, BERT menggunakan hyperparameter epoch:5; batch size: 16; dan learning rate: 3e-5, yang digunakan sebagai base model dan dibandingkan untuk kedua model lainnya menggunakan SVM dan RF. Untuk hasil yang di dapat dari hasil eksperimen berikut adalah, accuracy dari BF meningkat menjadi 85%, untuk BR masih mendominasi sekitar 95%, sedangkan BS ada di peringkat kedua dengan 93%.

Sedangkan pada distribusi sentimen ketiga kelas, Pro tetap menjadi dominan utama dengan presentasi hasil 42.9%, sedangkan Contra 33.7%, dan Neutral 23.4%. Pada hasil tingkat distribusi sentimen di beberapa topik, tingkat Pro memiliki dominasi di empat topik utama seperti "*Smart Farming*", "*Policy& Market*", "*Climate Change*", serta "*AI & Automation*". Untuk Contra berhasil menempati presentase pertama pada topik terkait "*Other*", serta "*Infrastructure & Supply Chain*", dan pada kelas Neutral mendominasi topik "*IoT & Drones*" serta "*Sustainability*". Jika dibandingkan dengan hasil model yang digunakan persebaran berikut memberikan distribusi yang lebih seimbang dari sebelumnya yang condong pada distribusi Pro di keseluruhan kelas.

Eksperimen lain yang dilakukan memiliki hasil accuracy lebih tinggi dan distribusi sentimen yang lebih rata dibanding menggunakan model yang digunakan. Namun, pada hasil menggunakan metode balancing berikut memiliki keterbatasan pada kesesuaian dataset murni yang memiliki validasi kelas analisis sentimen. Pada penelitian sebelumnya, maupun penelitian lanjutan eksperimen berikut masih dapat digunaakan apabila tujuan dari penelitian didasarkan mengukur tingkat accuracy menggunakan machine learning, namun jika digunakan sebagai menghitung pada tingkat kesesuaian dataset murni sebagai langkah awal untuk pemetaan topik yang benar-benar muncul dalam artikel berita seperti penelitian yang dilakukan, penulis merasa masih kurang relevan dengan kondisi yang sebenarnya. Sehingga pada akhirnya, penulis memilih metode dengan model machine learning turning hyperparameter menggunakan randomizedsearch tanpa set balancing kelas pada dataset sebagai metode dasar pada penelitian.

Serangkaian eksplorasi konfigurasi tingkat awal dokumen berita mengkonfirmasi bahwa model ensemble berbasis BERT dengan Classifier Random Forest (BR) memberikan kinerja klasifikasi sentimen yang unggul dan konsisten. Distribusi sentimen mayoritas positif memperkuat pandangan optiomisme terhadap adopsi inovasi pertanian. Meskipun keberagaman tema memerlukan pendekatan secara terfokus, namun hasil yang didapat tidak jauh berbeda dengan keseuaian sentimen pada layer-layer topik. Dengan pemahaman mendalam mengenai kekuatan dan batasan model dasar ini, selanjutnya akan dibahas mengenai uji coba hasil terbaik model dasar pada uji coba *stacking meta learner* yang mengkombinasikan ketiga model terbaik untuk melihat akurasi hasil model awal apakah memang memiliki performa terbaik dibandingkan model lainnya, serta apakah hasil *stacking* sendiri mendapatkan hasil klasifikasi yang lebih *robust*.

4.2.2 Eksplorasi Ensemble Meta Learning

Eksplorasi *ensemble meta learning* diuji menggunakan teknik *stacking* dengan memadukan kekuatan model *machine learning*, yang mana pada tahap ini menggunakan ketiga

model base learner terbaik model awal sebelumnya, yaitu BERT + SVM (BS), BERT + Random Forest (BR). Dan BERT + XGBoost (BX). Model-model tersebut melalui pendekatan *stacking* memanfaatkan prediksi probabilitas atau fitur *embedding* dasar yang dimasukkan pada *meta learner Gradient Boosting*, sehingga diharapkan dapat memperbaiki kelemahan klasifikasi maupun melihat seberapa jauh model setelah diuji menggunakan *meta learner* apakah hasil dari BERT + RF sebagai *base model* masih memiliki performa yang terbaik seperti pada tahap sebelumnya. Secara garis besar, pada tahapan ini akan mencangkup eksplorasi *stacking meta learner GradientBoosting* dengan beberapa *hyperparameter base learner*, uji coba *hyperparameter* terbaik *meta learner* dengan *base learner*, *validation curve* pada hasil *stacking* menggunakan *hyperparameter* terbaik, uji *paired t-test* pada hasil *stacking* dengan hasil *base model* terbaik.

Pada ekseperimen menggunakan stacking , ketiga base learner terbaik BERT + SVM (BS), BERT + Random Forest (RF), dan BERT + XGBoost (BX) digunakan untuk menghasilkan prediksi probabilitas pada data validasi setiap fold. Proses pengujian dibagi menjadi dua level, dimana level tersebut dapat diuraikan sebagai berikut:

- Level-0 (Base Learner): Masing-masing model dilatih pada 5 train folds dan mengasilkan prediksi probabilitas untuk fold yang tersisa. Kemudian hasil probabilitas pada level ini oleh ketiga model dikumpulkan menjadi tiga fitur baru pada model.
- Level-1 (Meta Learner): Langkah selanjutnya, fitir probabilitas pada level-0 digunakan sebagai masukan bagi meta learner yang pada pengaplikasiannya menggunakan Gradient Boosting. Model meta learner dilatih untuk mempelajari pola kesalahan dan korelasi antar base learner.

Untuk menghindari data lekage, proses diulang secara 5-fold menggunakan *nested cross validation*, yaitu setiap split menghasilkan set level-0 dan level-1 yang terpisah, sehingga *meta learner* hanya melihat prediksi *base learner* pada data yang belum perbah dilatih oleh model *base learner* sebelumnya.

Model	**Hyperparameter**	**Nilai**
Gradient Boosting	Learning rate	uniform [0.001, 0.2]
	Max_depth	randint [1,10]
	n_estimators	randint [50, 1000]
BERT + SVM	C	uniform [0.1, 100]
	Kernel	[Poly, rbf, linear, sigmoid]
	gamma	[scale, auto]

BERT + Random Forest	n_estimator	[100, 1000]
	max_depth	[None, 10, 20, 30, 50]**;**
	split	[2, 5, 10]
XGBoost	n_estimator	[100, 1000]
	max_depth	[3, 5, 7, 10]
	Learning rate	[0.01, 0.05, 0.1]

Tabel 4.15 Hyperparamater Model Evaluasi Meta Learner

Komponen	**Model**	**Kode**
Model	BERT Fine-Tune	BF
	BERT + SVM	BS
	BERT + Random Forest	BR
	BERT + XGBoost	BX
Jenis Layer	Base Learner	BL
	Stacking	ST

Tabel 4.16 Kode Model Stacking

Sebagai eksplorasi konfigurasi model secara menyeluruh, tiap base learner dan meta learner diberi ruang pencarian hyperparameter yang tersaji pada Tabel 4.15. Mengenai hyperparameter tersebut, pada meta learner Gradien Boosting mengguankan rentang learning rate [0.001 – 0.2] dimana pada deklarasi didefinisakan melalui pengambilan uniform, dengan kedalaman pohon diskrit menggunakan randint rentang 1 hingga 10, dan jumlah estimator sekitar 50 hingga 1.000 menerapkan fungsi dari randint. Penetapan nilai hyperparameter diberikan menyeluruh pada semua model untuk diuji coba lebih lanjut sebelum masuk pada tingkat per layer dan dihasilkan hasil hyperpamater terbaik pada model dan hasil stacking untuk di eksplorasi lebih lanjut pada tahap selanjutnya.

Model	**Base Learner Param**	**Accuracy (%)**	**F1-macro (%)**
BL - BS – 01*	C=37.554; γ=scale; kernel=poly	**91.4**	**86.8**
BL - BS – 03	C=15.7019; γ=scale; kernel=poly	91.4	86.8
BL - BS – 05	C=60.2115; γ=auto; kernel=poly	91.4	86.8
BL - BS – 07	C= 83.3443; γ=auto; kernel=rbf	90.8	86.4
BL - BS - 10	C= 43.2945; γ=scale; kernel=poly	91.4	86.6
BL - BR - 01	n_est=1000; min_split=2; max_depth=50		87.8

BL - BR – 03	n_est= 100; min_split=10; max_depth=20	91.92	88.15
BL - BR – 05	n_est= 100; min_split=5; max_depth=20	92.0	88.1
BL - BR – 07*	n_est= 800; min_split=10; max_depth=10	**92.3**	**88.4**
BL - BR - 09	n_est= 1000; min_split=5; max_depth=50	92.0	88.0
BL - BX – 03	n_est=500; max_depth=7; lr=0.05	91.1	86.9
BL - BX – 04*	n_est=800; max_depth=7; lr=0.10	**91.1**	**87.0**
BL - BX – 06	n_est=300; max_depth=5; lr=0.10	91.0	86.9
BL - BX – 08	n_est=800; max_depth=3; lr=0.05	91.1	86.9
BL - BX - 10	n_est=300; max_depth=7; lr=0.05	91.0	86.8

Tabel 4.17 Hasil Pengujian Awal Base Learner

Pengujian tahap awal diawali dengan pengujian tingkat layer 0 pada base learner, di mana beberapa layer melalui proses tune pada RandomizedSearch dengan n_iter = 10 untuk mencari hyperparemter pada proses yang akan dimasukkan dan di uji kembali pada layer tingkat 1. Setiap kandidat hyperparameter di evaluasi melalui Cross Validation, lalu konfigurasi terbaik disimpan untuk masing-masing learner (Tabel 4.17). Pengujian ini diuji kembali dengan hyperparameter ruang yang lebih baik dan dalam sehingga memberikan model machine learning belajar lebih banyak pada ruang hyperparameter. Selain hal tersebut, proses pipeline menggunakan pipeline dasar model sebelumnya agar menyelaraskan model dan memang layak diuji dengan pipleline yang sama menggunakan fiture clone untuk menduplikasi pipline tanpa mengubah dan mengadaptasi hasil dari model sebelumnya. Pada tabel terlihat bahwa dari kesepuluh iterasi yang dilakukan 5 iterasi dilaporkan oleh peneliti pada masing-masing hasil model. BL–BS–01, BL–BS-03, BL–BS-05, dan BL–BS–10 konstan menggunakan kernel poly dan C 15 hingga 60, sedangkan pada hyperparameter gamma mengggunakan scale atau auto memberikan akurasi tertinggo sekitar 91.4% pada iterasi pertama (BL-BS-01). Sedangkan pada model BL-BR sendiri, mencapai akurasi tertinggi disbanding model yang lain dengan 92.3% melalui iterasi ke 7 (BL–BR–07), dan F1-macro sekitar 88.4% menggunakan estimator 800, split 10, dan kedalaman (depth) sekitar 10. Pada model terakhir yang diuji, BL-BX sendiri dengan 10 iterasi menetapkan iterasi keempat (BL-BX-04) menjadi hasil terbaik dengan akurasi sekitar 91.1% dengan f1-macro sekitar 87%. Setelah base learner terkonfigurasi, keluaran

probablilitasnya digabung menjadi fitur baru untuk meta learner. Hasil tuning dan performa meta learner kemudian disajika di Tabel 4.16.

Model	Hyperparameter			Accuracy (%)	F1-Macro (%)
	Learning rate	Estimator	Depth		
ST – 2	0.120370	264	3	90.8	86.2
ST – 5	0.0051169	393	2	91.1	86.4
ST – 6	0.167489	435	6	91.0	86.2
ST – 7	0.037365	210	5	90.4	85.6
ST – 9	0.087389	524	1	90.9	86.3
ST – 15	**0.0419325**	**389**	**2**	**91.6**	**87.2**
ST – 17	0.100035	255	3	90.7	86.0
ST – 18	0.0792121	921	2	90.4	86.0
ST – 19	0.152072	615	6	90.8	86.0
ST - 20	0.110342	526	6	90.6	85.5

Tabel 4.18 Hasil Pengujian Meta Learner terhadap Base Learner

Pada hasil pengujian menggunakan Meta Learner, diuji dengan menggunakan `n_iter = 20` pada *RandomizedSearch* guna memperoleh ruang hyperparameter terbaik dan memberikan hasil terbaik. Adapun ke-20 iterasi yang dilakukan memakan waktu running time jika digabungkan pada layer 0 dan layer 1 adalah 65 menit. Adapun pengujian juga sempat dilakukan menggunakan meta learner logistic dengan hasil running time yang lebih singkat sekiat 17 menit menggunakan metode yang sama, namun pada hasil tidak jauh berbeda dengan hasil terbaik accuracy sekitar 90% dan pada f1-macro terbaik adalah 84%. Namun, uji coba pada GradientBoost sendiri mencapai angka 91.6% dan 87.2% pada hasil f1-macro terbaik di iterasi ke-15 (ST -15). Untuk hyperparameter terbaik menggunakan kedalaman pohon (max_depth) 2, learing rate 0.041 serta, estimator terbaik 389. Kemudian, hasil parameter terbaik pada tingkat layer dihasilkan pada Tabel 4.19 untuk diuji ulang secara eksplisit pada eskplorasi hyperparameter.

Model	Hyperparameter Terbaik	F1-macro (%)	Accuracy (%)
BL – BS	kernel: poly; C: 37.554; γ: scale	86.8	91.4
BL – BR	n_estimators: 800; min_samples_split: 10; max_depth: 10	88.0	92.3
BL – BX	n_estimators: 300; max_depth: 5; learning rate: 0.1	87.0	91.0
ST	n_estimators: 389; max_depth: 2; learning rate: 0.00419	87.2	91.6

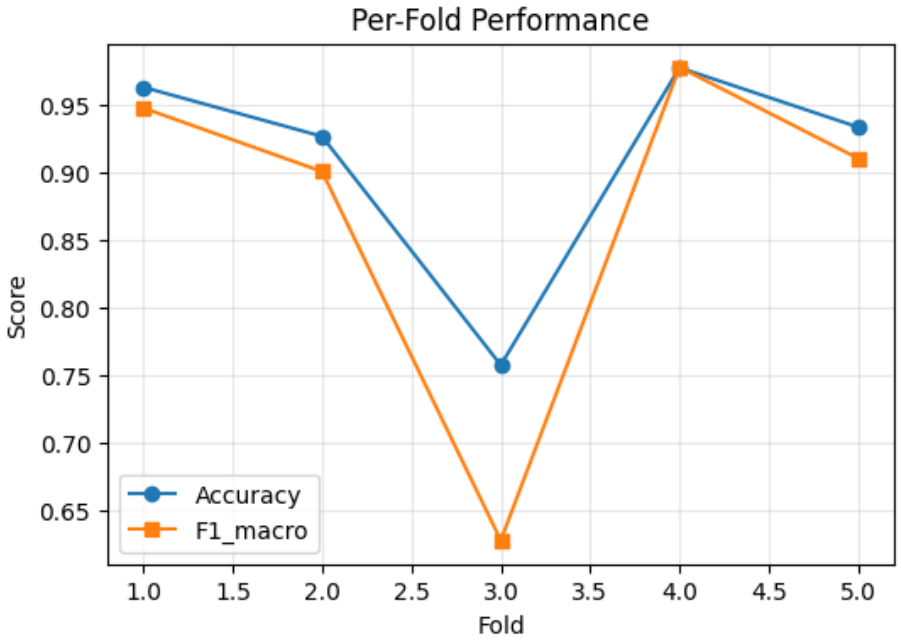
Tabel 4.19 Hyperparameter Terbaik pada Stacking

Setelah penentuan hyperparameter terbaik pada keseuruhan model, pengujian dilakukan dengan memasang parameter-parameter tersebut untuk diproses pada stacking menggunakan meta learner Gradient Boosting, dan diuji melalui Cross Validation. Hasil ringkasan performa akhir pada data validasi tiap kelas disajikan pada Tabel 4.20.

Kelas	Performa		
	Precision (%)	Recall (%)	F1-score (%)
Pro	97	97	97
Contra	90	84	87
Neutral	86	95	90
Accuracy		90	

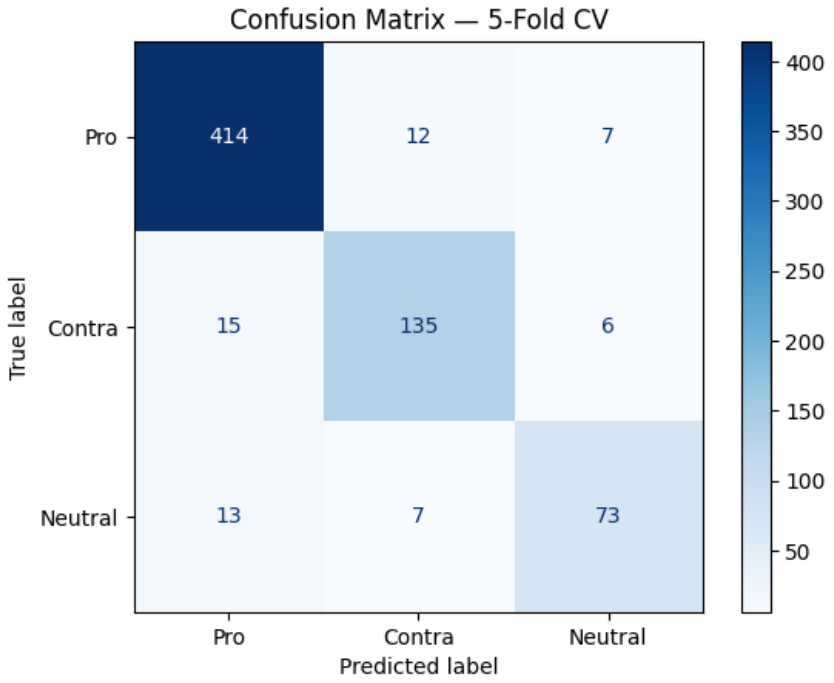
Tabel 4.20 Eksplorasi Hyperparamter Terbaik

Apabila melihat hasil eksplorasi parameter tersebut pada proses stacking. Terlohat bahwa kelas Pro mencapai precision, recall, dan F1-Score dengan baik sekitar 97%, sedangkan kelas Contra memperoleh F1-Score 87%. Dengan recall lebih rendah pada 84%, dan kelas Neutral mencatat recall tertinggi sekitar 95%, meskipun pada precision sedikit lebih rendah sekitar 86%. Secara keseluruhan, akurasi agregat berada di angka 90%.



Gambar 4.5. Grafik Performa F1-Score dan Accuracy Eksplorasi Hyperparameter

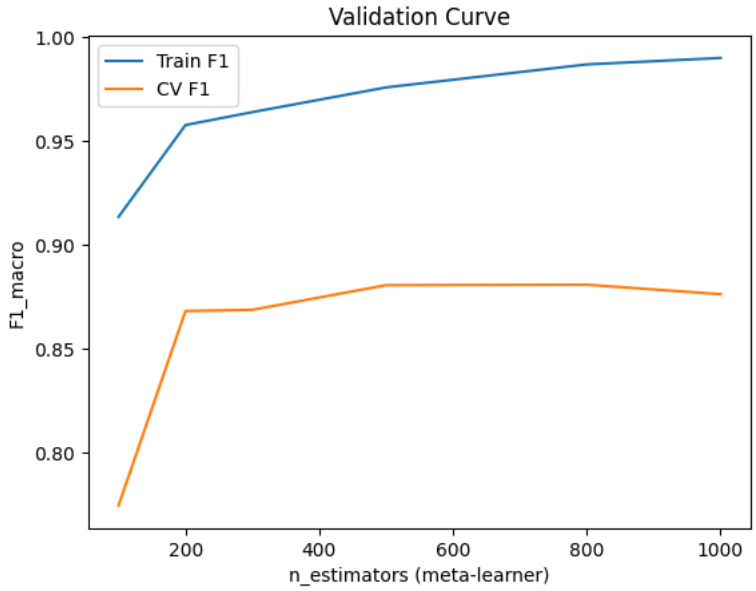
Sebagai penguatan pada performance metrik, maka divisualisasikan grafik performa per-fold terlebih pada metric accuracy dan f1-macro. Garis biru menunjukkan fluktuasi akurasi stabil di fold 1 dan 2, sedangkan pada fold 3 mengalami penurunan mencapai titik 60%, kemudian pada fold berikutnya naik secara drastic menuju angka 90% hingga fold terakhir. Garis orange tidak jauh berbeda dengan line biru pada accuracy, memberikan pengetahuan bahwa tingkat performa f1-macro juga mengalami naik turun yang sama pada proses.



Gambar 4.6. Confussion Matrix Stacking

Gambar 4.6 untuk menujukkan distribusi pengujian pada sentimen, memperlihatkan confusion matrix agregat hasil stacking . Dimana mayoritas kesalahan prediksi terjadi pada perbandingan kontra dan neutral seperti yang terlihat bahwa terindikasi 15 dokumen kontra masuk dalam diprediksi pro, dan 7 dokumen diprediksi pada konra, menandakan bahwa terdapat intrepretasi tantangan pada model bahkan di ditingkat stacking untuk memprediksi antar sentimen kontra maupun netral.

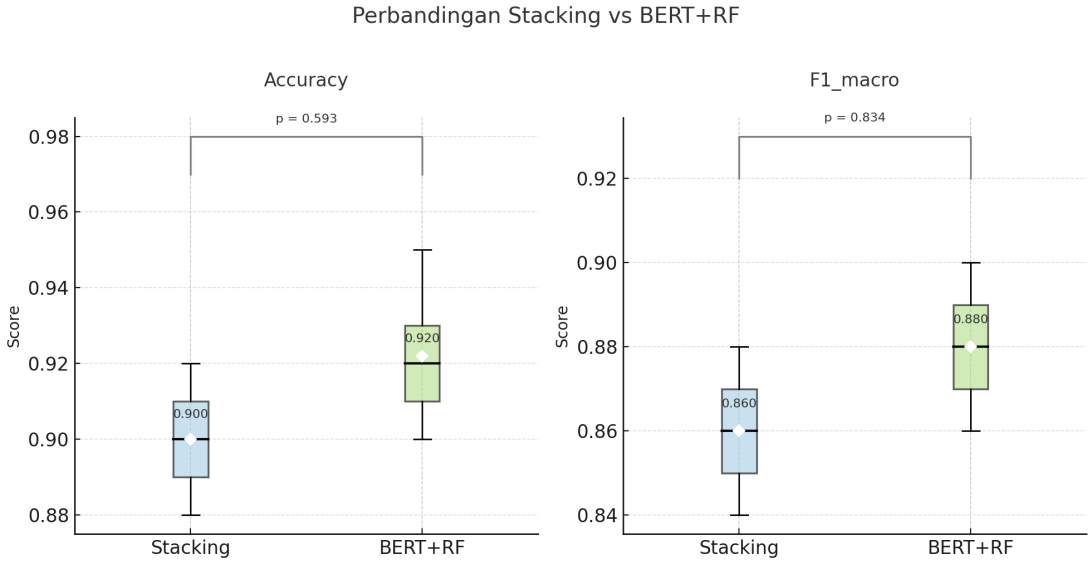
Untuk menguji performa model meta learner Gradient Boosting, digunakan pengujian pada validation curve yang digunakan untuk menilai bagaimana perform model meta learner berubah ketika jumlah estimator (`n_estimator`) divariasikan, sekaligus memantau potensi overfitting. Dengan membandingkan kurva f1_macro pada data pelatihan (*train*) dengan data validasi (*Cross Validation*), kita dapat menentukan titik keseimbangan, sehingga akhirnya dapat disimpulkan pada langkah selanjutnya bahwa menambah estimator tidak lagi secara signifikan memperbaiki generalisasi.



Gambar 4.7. Grafik Validation Curve Model Stacking

Pada Gambar 4.7, kurva berwarna biru (line train) memperlihatkan peningkatan f1-macro pada data pelatihan seiring bertambahnya jumlah estimator, mencapai hampir 99% pada estimator ke 1000. Sebaliknya, kurva orange (line CV) pada data validasi mulai stabil mendatar setelah sekitar 300 estimator, dengan nilai tertinggi 88% pada rentang 400 hingga 600 estimator. Perbedaan yang melebar antara kedua kurva di titik yang lebih tinggi menunjukkkan potensi overfitting jika terlalu banyak pohon. Oleh karena itu, pilihan `n_estimators` di kisaran 400-

600 memberikan *trade-off* terbaik antara akurasi train, dan generalisasai CV, sehingga penulis menetapkan 389 estimator iterasi 15 (ST-15) sebagai konfigurasi optimal untuk meta learner.



Gambar 4.8. Hasil Uji Coba Paired Test

Adapun pengujian selanjutnya diuji melalui paired test pada metric accuracy dan f1-macro di masing-masing model untuk memastikan apakah peningkatan performa yang diperoleh dari ensembling stacking benar-benar lebih baik daripada baseline BERT + RF. Uji ini membandingkan skor accuracy dan f1-macro pada masing-masing fold dari kedua model karena setiap fold menghasilkan sepasang observasi stacking dengan BERT + RF, agar perbedaan performa dapat diuji secara statsistik. Tujuan dari uji coba berikut adalah menilai apakah selisih rata-rata performa tersebut cukup besar dan dapat dikatakan signifikasn atau ahnya merupakan variasi kebetilan antar fold. Hasil uji paired t-test, sebagaimana ditunjukkan di Gambar 4.8, dengan keteraangan hasil sebagai berikut:

- Accuracy, pada uji coba t-stastistic = - 0.58; p-value = 0.593
- F1-macro, pada uji coba t-statistic = -0.22, p-value = 0.834

Kedua p-value jauh diatas ambang $\alpha = 0.05$, sehingga tidak ada bukti bahwa *stacking* secara signifikan mengungguli atau tertinggal dari BERT + RF pada metrik manapun. Menginat komplesitas implementasi *stacking* yang lebih tinggi dan sifat real-time requirement pada tahap konstruksi Knowledge Graph *pada* tahap selanjutnya, penulis memutuskan untuk tetap menggunakan BERT + RF sebagai model utama dalam pipeline berikutnya. Singkatnya meski secara teori *stacking* dapat menggabungkan kekuatan beberapa model, dalam praktik eksperimen ini hasilnya masih belum dapat meningkatkan performa. Hal ini terindikasi awal karena prediksi *base learner* yang serupa dimana base model dari ketiga model classifier SVM,

RF, dan XGBoost banyak berbagi pola kesalahan pada output BERT, sehingga meta learner kesulitan menemukan sinyal tambahan. Dan indikasi kedua pada Baseline BERT + Random Forest memiliki strong baseline, sehingga sudah sangat kuat dan dirasa peningkatan akan jauh lebih sulit dicapai hanya dengan stacking . Model BERT dengan classifier Random Forest menawarkan keseimbangan terbaik antara akurasi, stabil, cepat dalam inferensi, dan mudah diintegrasikan untuk mengekstrak entitas dan relasi probablitik yang akan diolah menjadi node dan edge dalam Knowledge Graph. Dengan demikian, modul sentiment stance pada Knowldege Graph akan menggunakan keluaran label dan probabilitas dari BERT + Random Forest yang memastikan konsistensi performa sekaligus efisiensi komputasi pada skala data yang lebih besar.

4.3 Pemetaan Relasi Entitas dan Analisis Graf

Setelah modul sentimen dianalisis dan diuji pada tingkat performa dan distribusi secara menyeluruh. Penetapan model terbaik sebelumnya menggunakan model hibrida machine learning berupa BERT dengan Random Forest, akan diproses lebih lanjut pada proses pemetaan relasi entitas pada mapping Knowledge Graph. Analisis pemetaan relasi menggunakan dua pendekatan, yang pertama ekstraksi dan pemetaan relasi ke format triples dimana setiap pasang entitas diekstrak dari teks dan dihubungkan dengan relasi sesuai probabilita sentimen model yang kemudian hasilnya berupa kumpulan triple seperti (agritech, relation, food) dan diolah pada dataset graf. Sedangkan pendekatan analisis knowledge graph kedua menggunakan Graph Convolutional Network (GCN) pada embedding adjacency matrix dan fitur node antar entitas untuk analisis lebih lanjut pada graph untuk mengungkap pola komunitas. Centrality, dan klster relasi. Dengan demikian, Knowledge graph tidak hanya merekam relasi sentien antar entitas, tetapi juga menyediakan kerangka untuk analisis graf lanjutan.

4.3.1 Pembentukan Knowledge Graph

4.3.1.1. Ekstraksi Entitas dan Pembentukan Triples

Pada tahap ini, teks dokumen berita yang telah tersentimen pada model hibrida BERT dan Random Forest diproses oleh pipeline `spaCy` yang diperkuat library `EntityRuler` dan PhraseMatcher untuk mengklasifikasi bahwa istilah domain dapat terdeteksi sebagai entitas yang benar. Setiap mention entitas kemudian diberi label seperti [`CROP`, `TECHNOLOGY`, `LIVESTOCK`, `POLICY`, `EVENT`, `METRIC`, `INFRASTRUCTURE`, `ENVIRONMENT`, `BUSINESS`] dengan beberapa keywords teknologi pertanian yang dimunculkan pada deklarasi awal untuk memfokuskan jenis domain label. Hasil proses pembentukan entitas dan pelabelan melalui domain ditunjukan pada Tabel 4.21 dimana

"*farmer*" terdeteksi sebagai `PEOPLE`, dan "*industries*" sebagai domain `BUSINESS`, kemudian terdapat "*China*" sebagai doman `LOC` dan "*soybearn*" sebagai object masuk dalam domain `CROP`.

Artikel asli	subject	subject_label	object	object_label
Farmer say many industries embrace challenge work suistainably	Farmer	PEOPLE	indutries	BUSINESS
china moves to curb soyberan imports	China	LOC	soybean	CROP

Tabel 4.21 Hasil Pembentukan Entitas

Selanjutnya, `DependencyMatcher` mengekstraksi pola subjek-predikat-objek maupun SVO dengan preposisi. Lemma dari kata kerja seperti "*embrace*", "*move*", "*use*" kemudian dipetakan ke dalam bentuk kanonis lewat variabel yang telah didefinisikan menggunakan `REL_CANON`, sehingga relasi seperti beberapa kata kerja sebelumnya menjadi bagian dari relasi. Setiap triples kemudian dibangun dalam format `subject, relation, object`, yang kemudian dimasukkan dalam `triple_text` untuk di analisis lebih lanjut pada tahapan *filtering* topik teknologi pertanian maupun proses pembentukan knowledge graph. Hasil triples keseluruhan terdapat 768 triples dari total pembentukan triples, proses pencocokan karakter, *keywords*, dan pelabelan, serta *filtering* tahap awal pembentukan *NER* dan *RE*.

Artikel asli	subject	subject label	relation	object	object label	Triple text
farmer say many industries embrace challenge work sustainably	farmer	PEOPLE	embrace	industries	BUSINESS	farmer embrace industries
southeast asia uses poultry to meet demand	southeast asia	LOC	user	poultry	LIVESTOCK	southeast asia use poultry

Tabel 4.22 Hasil NER dan RE

Setelah memperoleh keseluruhan kumpulan triples dari fase NER dan RE, langkah berikutnya adalah filtering untuk mengekstrak hanya relasi yang relevan dengan domain agritech. Dalam proses filtering tersebut, label entitas memilih hanya baris di mana subject_label atau object_label termasuk dalam kategori `["TECHNOLOGY","CROP","LIVESTOCK","POLICY","EVENT","METRIC","INFRASTRUCTURE"`

,"ENVIRONMENT","BUSINESS"], yang telah dideklarasikan pada awal pembentukan triples menempel keywords terkait teknologi pertanian. Kemudian selain sesuai dengan label, terdapat filtering menyertakan triple dimana teks subjek atau objek mengandung token seperti "*precsion agriculture*", "*smart farming*", "*soil sensor*", dengan pencocokan substring sederhana pada kolom `subject` dan `object`. Hasilnya, dari total 768 triples yang telah terbentuk, terfilter menjadi 55 triples yang fokus pada konteks teknologi pertanian sebagai pembangunan graph yang bersih dan sesuai dengan topik yang diangkat.

Original Sentence	Subject	Relation	Object	Subject Label	Object Label	Triple Text
Precision agriculture improves crop yield.	precision agriculture	improve	crop yield	TECHNOLOGY	CROP	precision agriculture improve crop yield
Farmers deploy drone mapping to monitor fields.	drone mapping	monitor	fields	TECHNOLOGY	ENVIRONMENT	drone mapping monitor fields
"Soil moisture sensors optimize irrigation.	soil moisture sensors	optimize	irrigation	TECHNOLOGY	INFRASTRUCTURE	soil moisture sensors optimize irrigation

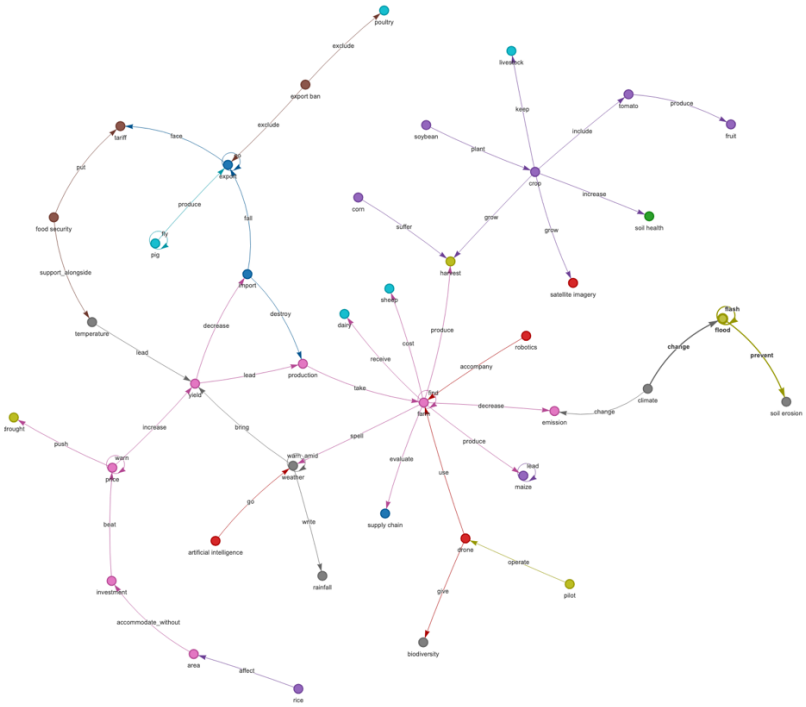
Tabel 4.23 Hasil Akhir Triples

4.3.1.2. Pembentukan Knowledge Graph

Setelah diperolah dokumen triples kumpulan berita teknologi pertanian, tahap berikutnya merupakan mapping dan pembangunan struktur graph. Di mana pada graph dibangun berdasarkan tiap subject dan object dari triple menjadi node, sedangkan nilai relation menentukan label edge yang menghubungkan kedua node tersebut.

Pembentukan knowledge graph pada tahapan dasar menggunakan library `NetworkX` untuk menginisialisasi graph kosong pada `nx.DiGraph()`, kemudian penambahan node unik dari kolo subject dan object, penambahan edge yang sesuai dengan atribut pada dataset relation, dan sentiment. Hasilnya adalah sebuah directed graph di mana setiap agritech terbuhung menurut konteks relasi sentimen yang dihasilkan pada model sentimen BERT + Random Forest sebelumnya dalam memvisualisasikan sentimen dalam entitas terkait. Pada

pembangunan graph pada tingkat dasar, maka dihasilkan hasil komunitas dan visualisasi pada Gambar 4.9.



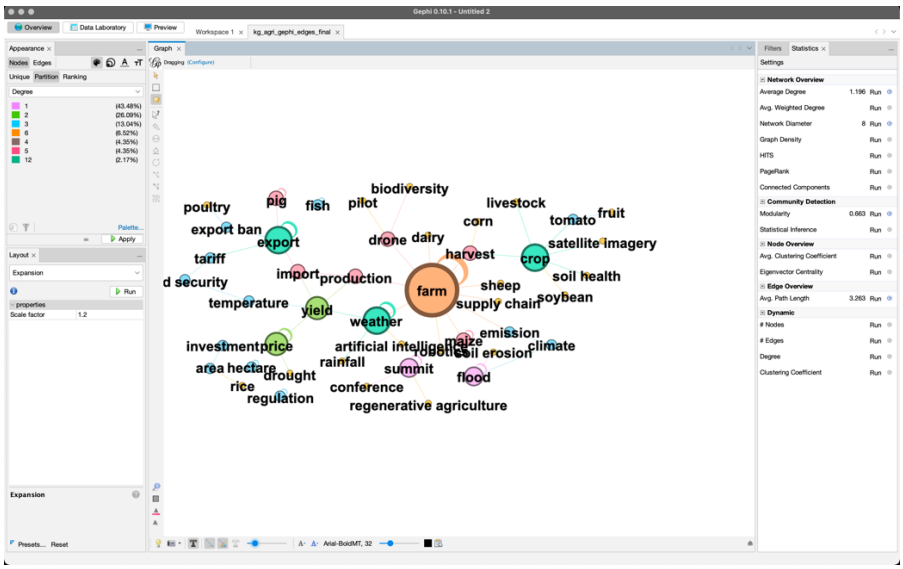
Gambar 4.9. Pembangunan Knowledge Graph dengan NetworkX

Dari pembangunan knowledge graph tersebut, maka diberikan hasil untuk klasifikasi community pada tiap entitas sebagai berkut:

```
    Community 0 (6 nodes):
       ['food security', 'import', 'production', 'tariff', 'temperature', 'yield']

    Community 1 (4 nodes):
       ['climate', 'emission', 'flood', 'soil erosion']

    Community 2 (3 nodes):
       ['artificial intelligence', 'rainfall', 'weather']

    Community 3 (9 nodes):
       ['biodiversity', 'dairy', 'drone', 'farm', 'maize', 'pilot', 'robotics',
'sheep', 'supply chain']

    Community 4 (4 nodes):
       ['export', 'export ban', 'pig', 'poultry']

    Community 5 (9 nodes):
       ['corn', 'crop', 'fruit', 'harvest', 'livestock', 'satellite imagery', 'soil
health', 'soybean', 'tomato']

    Community 6 (5 nodes):
       ['area', 'drought', 'investment', 'price', 'rice']
```

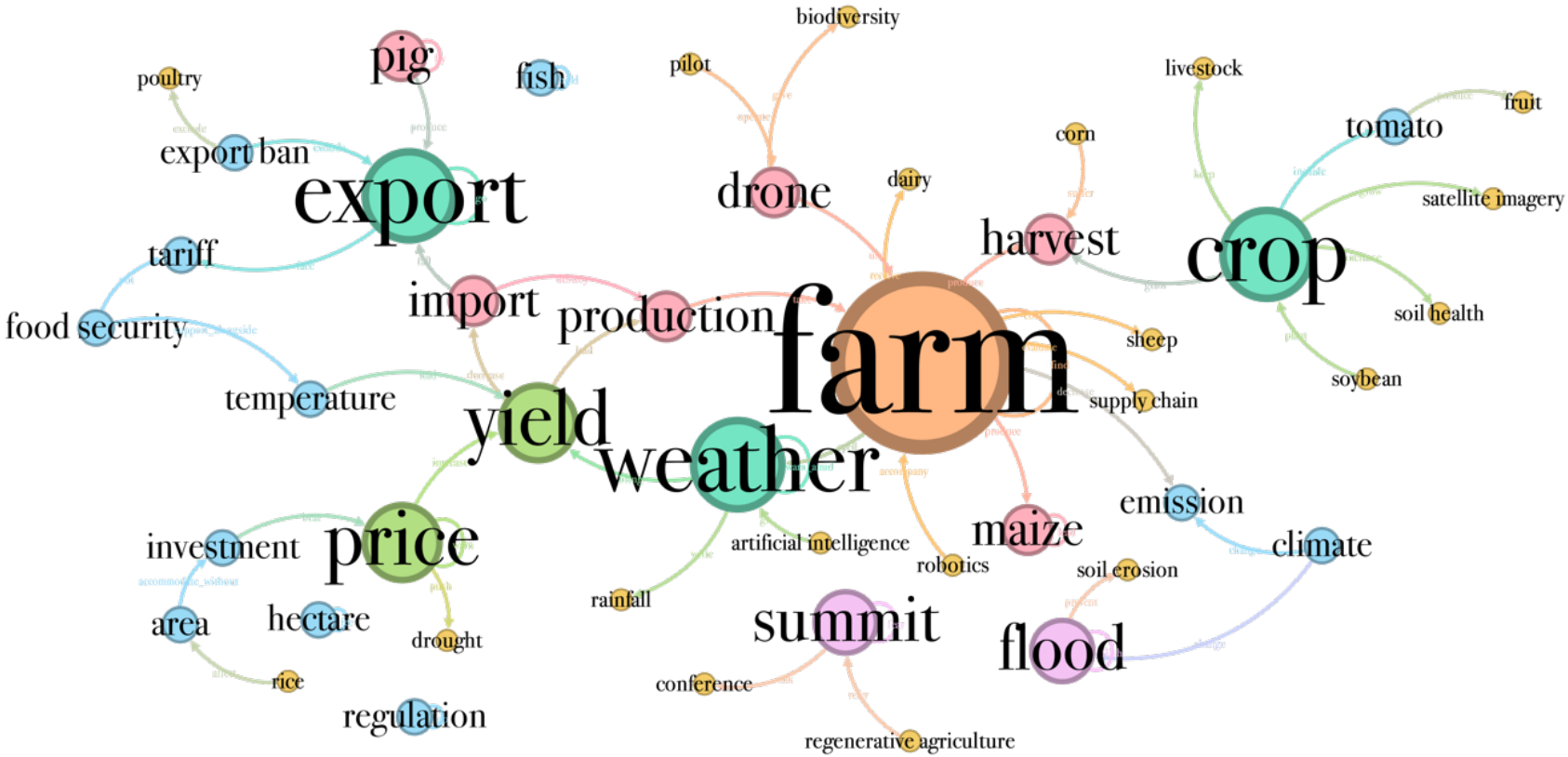
Pada tahapan visualisasi knowledge, penulis memanfaatkan panel Overview di Gephi 0.10.1 untuk mengkalsifikasikan beberapa komponen dalam *knowledge graph* yang sebelumnya telah di extract pada format `.csv` pada tahapan filtering sebelum visualisasi pada `pyVis`.

Metadata *Node* dan *Egde* meload edge list (`source, target, relation`) beserta atribut sentimen dan topik. Atribut tersebut dikonversi menjadi kolom *Node* seperti `entity_type`, dan *Edge* untuk `relation_type`, dan *polarity*. Statistik dasar perhitungan seperti connected components, diameter, dan `degree distribution` dilakukan melalui Statistics pada *Networrk Overviwe.*



Gambar 4.10. Laman Gephi 0.10.1

Visualisasi Awal Algoritma **Expansion** dijalankan dengan setting `scale factors 1.2` untuk memisahkan komponen secara organic. Warna simpul diatur berdasarkan jenis entitas seperti `PERSON, ORG, EVENT`, sedangkan ketebalan sisi mempresentasikan frekuennsi relasi. Pada Interpretasi hasil awal menampakkan beberapa relasi, seperti istilah "*farm*", "*weather*", dan "*corp*", yang memiliki degree tinggi, menandakan perannya sebagai pusat pada entitas dan hubungan.



Gambar 4.11. Knowledge Graph Agritech

4.3.2 Representasi dan Analisis Graph Convolutional Network (GCN)

Dalam tahapan analisis pada unsur-unsur yang ada di dalam knowledge graph, proses biasanya dapat di analisis pada tahapan dasar pembentukan knowledge graph dengan `NetworkX`. Namun, pada penelitian kali ini, untuk mendapatkan analisis yang lebih jelas dan mampu memahami relasi-relasi antar entitas dalam berita, penulis mengadaptasi proses analisis menggunakan GCN sebagai tahapan analisis lebih lanjut dan terstruktur. Pada tahapan awal sendiri terdiri pada tahapan representasi adalah mengektrak embedding untuk tiap node menggunakan Graph Convolutional Network (GCN). Pada proses ini, dimanfaatkan struktur koneksi antar entitas yang dinyatakan dalam *adjency matrix* serta fitur awal node, dalam hal ini *one hot encoding* berdasarkan label doman seperti `TECHNOLOGY`, `CROP`, atau `BUSINESS`. GCN kemudian mempelajari bobot progasi informarasi diantara tetangga tiap node melalui lapisan konvolusi. Lapisan pertama menyerap pola konektivitas lokal. Sedangkan lapisan kedua memperluas konteks hingga mencakup pola global dalam subgraph. Setelah dilakukan pelatihan, GCN menghasilkan vektor berdimensi yang menggambarkan posisi dan peranan setiap entitas dalam Knowledge Graph, yang kemudian vektor-vektor tersebut ini yang selanjutnya digunakan sebagai analisis komunitas dan clustering pada tahap analisis.

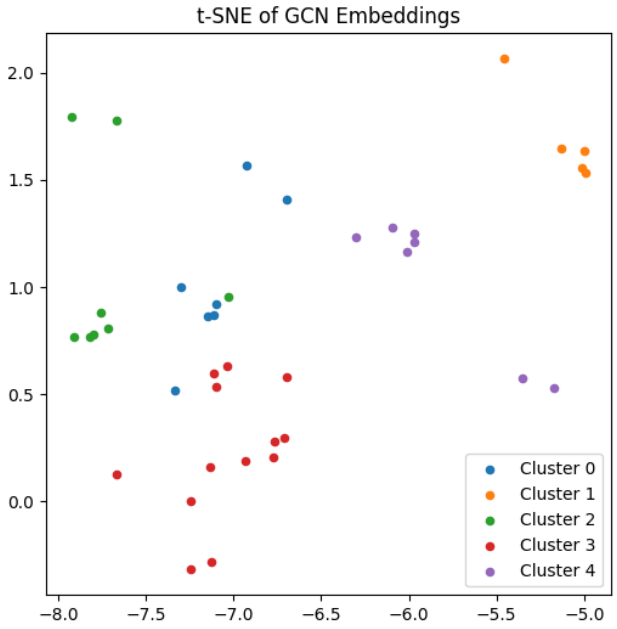
4.3.2.1. Pelatihan GCN denga K-Means dan Louvain

Sejalan dengan analisis knowledge graph menggunakan GCN yang telah di ekstrak sebelumnya, langkah berikutna adalah melakukan clustering menggunakan algoritma K-Means untik mengidentifikasi kelompok topik utama dalam Knowledge Graph. Pada eksperimen ini, jumlah klaster dipilih berdsarkan elbow meyhod dan domain agrtech yang kita dokuskan. Hasil clustering menunjukkan pembagian lima klaster dengan ukuran klaster, anggota tiap cluster sebagai berikut masing-masing sebagai berikut:

Klaster	Jumlah Anggota	Anggota
0	7 node	*"Price", "drought", "pig", "investment", "biodiversity", "area", "soybearn"*
1	5 node	*"Harvest", "artificial intelligence", "weather", "rainfall", "corn"*
2	8 node	*"Climate", "flood", "dairy", "piloy", "drone", "emission", "sheep", "supply chain"*
3	13 node	*"Import", "export", "food security", "temperature", "yield", "export ban", "poltry", "production", "tarif", "tomato", "fruit", "livestock", "rice"*
4	7 node	*"Farm", "crop", "robotics", "maize", "satellite imagery", "soil erosion", "soil health"*

Tabel 4.24 K-mean Klastering

Untuk memvisualisasikan pemisahan klaster, embedding direduksi ke dua dimensi menggunakan t-SNE. Gambar 4.12 memperlihatkan bahwa tiap klaster membentuk gugus yang relative terpisah, misalnya klaster yang didominasi oleh entitas environment seperti "*climate*", "*floor"*, dan klaster teknologi seperti "*drone"*, "*robotics"* dimana interpretasi masing-masing klaster membantu mengungkap suvdomain spesifik.

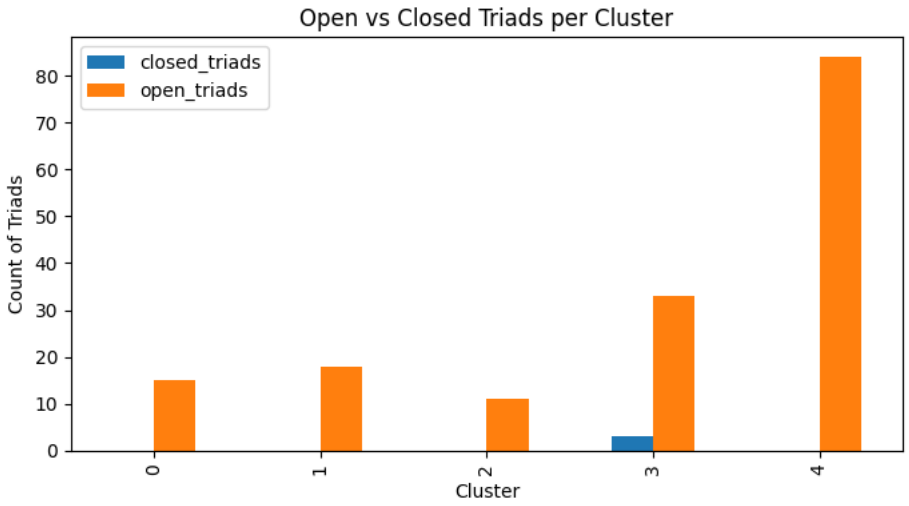


Gambar 4.12. t-SNE Cluster K-Mean

Setelah adanya analisis clustering, dimunculkan adanya hasil analisis berupa *open triad* serta *close triad*, analisis berikut memiliki tujuan untuk mengevaluasi struktur triadic pada setiap kelompok node dengan menghitung jumlah open triad yang saling terhubung dalam suatu komunitas maupun antar entitas (jumlah node terhubung kurang dari tiga node), sedangkan closed triad merupakan jumlah tiga node yang saling terhubung penuh. Dari perhitungan tersebut mengindikasikan sejauh mana node-node dalam satu klaster sudah membentuk keterhubungan atau masih banyak celah relasi yang dapat di eksplorasi. Pada ekseprimen ini setiap klaster di evaluasi dengan menghitung semua kombinasi tripled node dan memeriksa apakah ketiga pasang hubungan didalamnya terdapat closed maupun open triad. Hasil analisis dirangkum dalam Tabel 4.25.

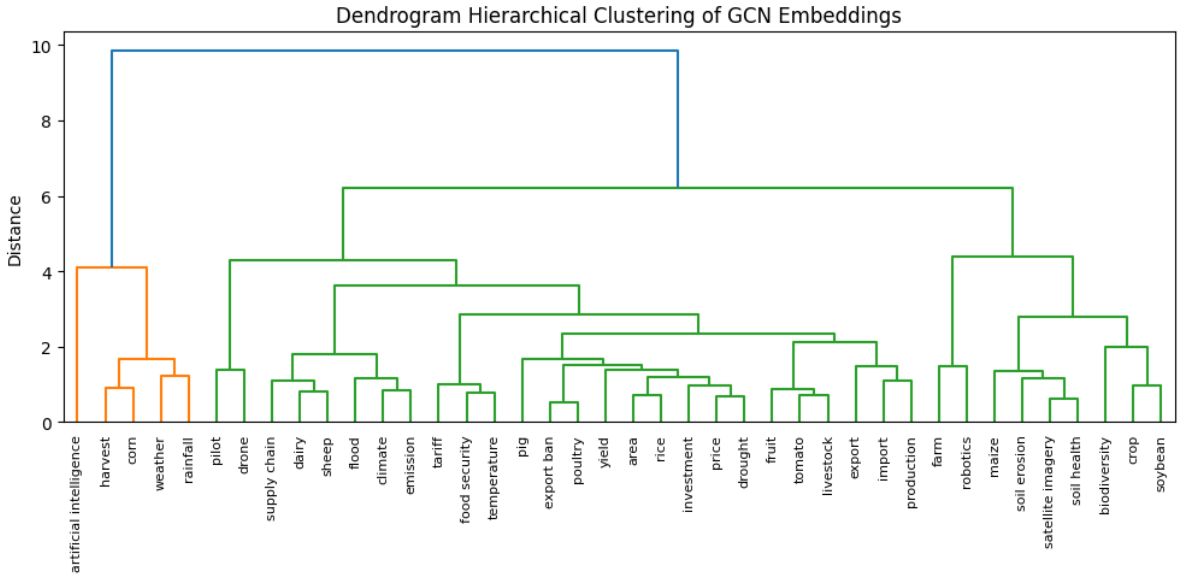
Klaster	Close Triads	Open Triads
0	0	15
1	0	18
2	0	11
3	3	33
4	0	84

Tabel 4.25 K-mean Open triad dan Close Triad



Gambar 4.13 K-Mean Open Triad dan Close Triad

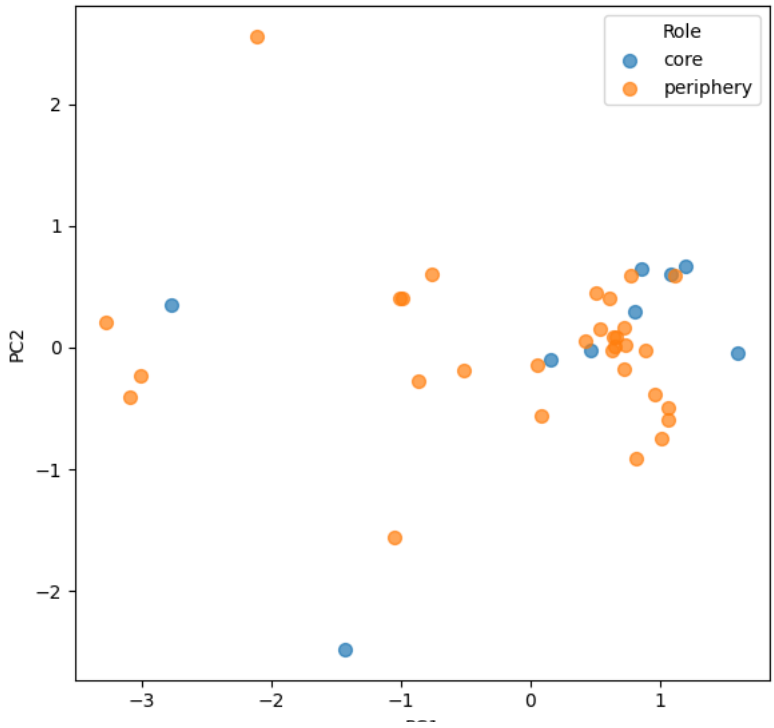
Dari tabel diatas terlihat bahwa mayoritas klaster kususnya klaster 0,1,2, dan 4 memiliki closed triad berniali nol, yang artinya belum ada closed triangles yang ada pada klastering tersebut. Hanya klaster 3 yang mencatatat tiga close triad, yang berarti menunjukkan sedikit peringkatan kohesi. Dominasi open triad menandakan masih banyak potensi relasi baru antar entitas yang belum dibentuk, sehingga struktur graph di dalam cluster masih longgar. Temuan tersebut mengindikasikan dua hal, yang pertama, bahwa pembagian klaster dengan K-Mean belum menangkap semua ikatan lokal yang dapat memperkaya knowledge graph, kemudian indikasi kedua masih terdapat ruang untuk menambahkan edges baru msialnya menghubungkan entitas yang saat ini hanya terhubung secara berantai tanpa adanya indikasi hubungan relasi sepenuhnya. Oleh karena itu, diperlukan pengecekan struktur graph lanjutan melalui beberapa analisis seperti dendogram hirarki, dan community profiling.



Gambar 4.14 Hirarki Dendogram K-Means

Setelah mengidentifikasi dominasi pada open triad pada setiap cluster, dilakukan pemeriksaan lebih mendalam terhadp pola konektivitas dengan beberapa metode analisis graf menggunakan dendogram hierarki. Dengan membuat dendogram dari embedding GCN, terlihat bahwa meski K-Means membagi node menjadi lima kelompok terpisah, sejumlah sub-klaster masih tergabung dalam branch besar yang sama. Hal ini menegaskan bahwa pembagian awal belum sepenuhnya memisahkan node-node dengan kedeatan historis tinggi, sehingga beberapa komunitas inti masih tercampur dengan entitas lain di tingkat hirarki yang lebih tinggi.

Kemudian untuk melakukan pengecekan ulang, juga dilakukan pengecekan pada community *role profiling* untuk menghitung profil peran setiap node, dimana perhitungan ini digunakan sebagai *core* atau *periphery*. Berdasarkan metrik *k-core decomposition*. Hasilnya sendiri menunjukkan 9 dari 40 node yang bertatus *core*, sedangkan sisanya sekitar 31 node berstatus *periphery*. Distribusi ini mengindikaasikan bahwa sebagian besar entita berada jauh pada jaringan, dan memperkuat temuan banyaknya open triad dan rendahnya kohesi lokal.



Gambar 4.15 Core dan Periphery K-Means

Untuk menemukan potensi relasi baru yang dapat menutup open triad, penulis menerapkan pendekatan *link suggestion* berdasarkan kemiripan embedding. Misalnya node "*import*" memiliki lima kandidat terkait "*export*", "*production*", "*temperature*","*fruit*","*rice*", maupun link suggestion yang lain, yang tecantum dalam Tabel 4.26. berikut ini.

Entity	Link Suggestion
Import	['export', 'production', 'temperature', 'fruit', 'rice']
Export	['poultry', 'export ban', 'import', 'area', 'drought']
Climate	['flood', 'emission', 'dairy', 'sheep', 'supply chain']
Flood	['climate', 'emission', 'dairy', 'poultry', 'sheep']
Price	['drought', 'area', 'investment', 'rice', 'export']
Drought	['price', 'area', 'rice', 'export', 'poultry']
Food security	['temperature', 'tariff', 'fruit', 'livestock', 'import']
Temperature	['food security', 'tariff', 'import', 'fruit', 'export']
farm	['robotics', 'maize', 'biodiversity', 'crop', 'soil health']
dairy	['sheep', 'emission', 'supply chain', 'climate', 'flood']

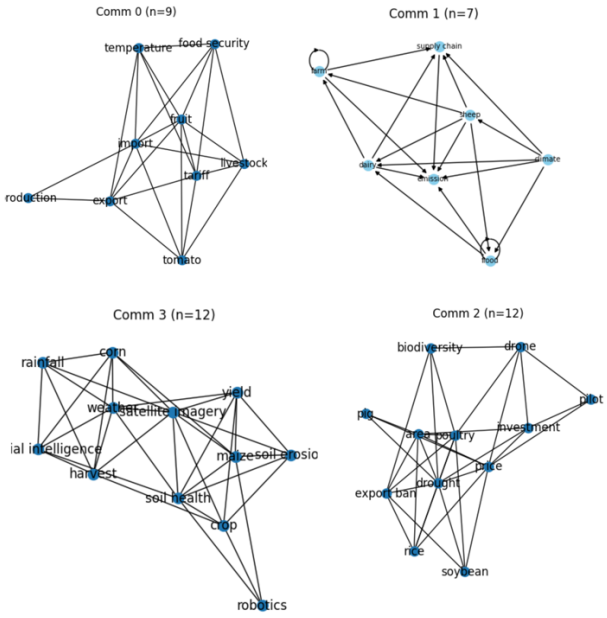
Tabel 4.26 Link Suggestion Knowledge Graph

Keseluruhan analisis dendogram yang masih memperlihatkan branch gabungan, mayoritas node di *periphery*, dan daftar calon link baru mengkonfirmasi bahwa struktur knowledge graph pasca K-Means masih memiliki banyak celah relasi. Dengan begitu, perlu dicari metode clustering yang lebih adaptif terhadap modularitas dan kohesi lokal. Untuk analisis lebih jauh, maka diperluka *re-embedding knowledge graph* pada GCN dan penggabungan dataset link suggsestion untuk analisis lebih lanjut menggunakan Louvain. Louvain digunakan sebagai eksplorasi komunitas yang lebih jauh dan diharapkan mampu menutup open triad yang tersisa.

Berdasarkan temuan pada analisis K-Mean, struktur knowledge graph setelah penambahan rekomendasi relasi dilakukan *re-embedding* pada graf dan diperkaya menjadi 149 edges dari 95 edges, sehingga graf terbaru mencerminkan potensi relasi antar entitas yang lebih lengkap. Pada langkah berikutnya, algorita Louvain diterapkan dan memberikan hasil empat komunitas baru dengan modularity sekitar 0.457.

Komuitas	Ukuran Anggota	Anggota
0	9 node	*“export”, “food security”, “fruit”, “import”, “livestock”, “production”, “tarif”, “temperature”, “tomato”*
1	7 node	*“Climate”,”dairy”,”emission”,”farm”, “flood”, “sheep”, “supply chain”*
2	12 node	*“area”,”biodiversity”,”drone”,”drought”,”export ban”, “investment”,”pig”,”pilot”, “poyltry”, “price”, “rice”, “soybean”*
3	12 node	*“artificial intelligence”,” “corn”,”crop”,”harvest”,”maize”,”rainfall”,”robotic”, “satellite imagery”,”soil erosion”,”soil health”,”weather”,”yield”*

Tabel 4.27 Daftar Komunitas Member Louvain

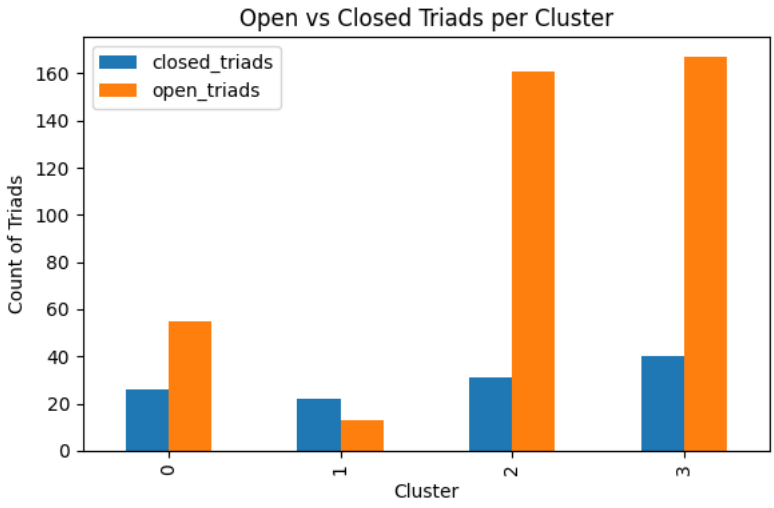


Gambar 4.16 Visualisasi Komunitas Louvain

Visualisasi topologi tiap komunitas menunjukkan bahwa Louvain berhasil membentuk klaster yang lebih padat pada setiap subgraph yang memperlihatkan peningakat close triad dan keterhubungan ganda antar node, dimana open triad meskipun masih mendominasi pada beberapa bagian lebih berkurang.

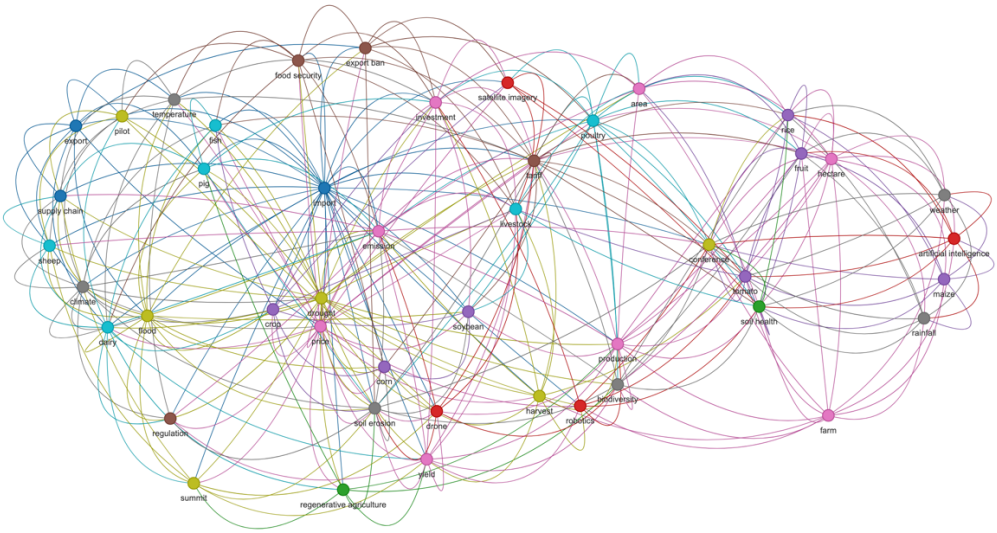
Komunitas	Close Triads	Open Triads
0	26	55
1	22	13
2	32	161
3	40	167

Tabel 4.28 Distribusi Triad Louvain



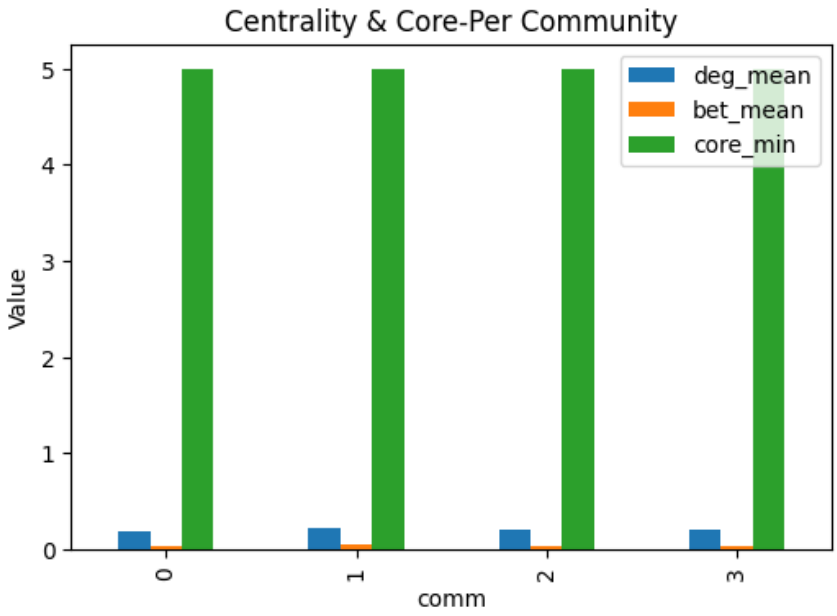
Gambar 4.17 Distribusi Triad Louvain

Keadaan yang terdapat dalam kedua hasil diatas cukup baik, di mana hal tersebut menandakan struktur komunitas ynag lebih kohesif dibandingkan dengan klastering pada K-Means. Sedangkan pada *connected component* terindikasi terdapat hanya 1 pada graf, yang memastikan seluruh node terhubung dalam satu kerangka komunitas besar, namun terbagi dalam empat sub-komunitas.



Gambar 4.18 Louvain Knowledge Graph

Lebih lanjut, analisis *centrality* dan *core per community* pada model graf bertujuan untuk mengukur seberapa sentral masing-masing node berada dalam jaringan lokal dan global. Degree centrality memberikan gambaran tentang jumlah koneksi langsung yang dimiliki node. Semakin tinggi nilai *degree*, maka semakin banyak relasi langsung yang diikutinya. Sementra itu*, betweennes centrality* mengindikasikan peran node sebagai penghubung informasi antara subgraph yang berbeda, node dengan nilai betwenness tinggi memfasilitasi aliran informasi lintas komunitas.



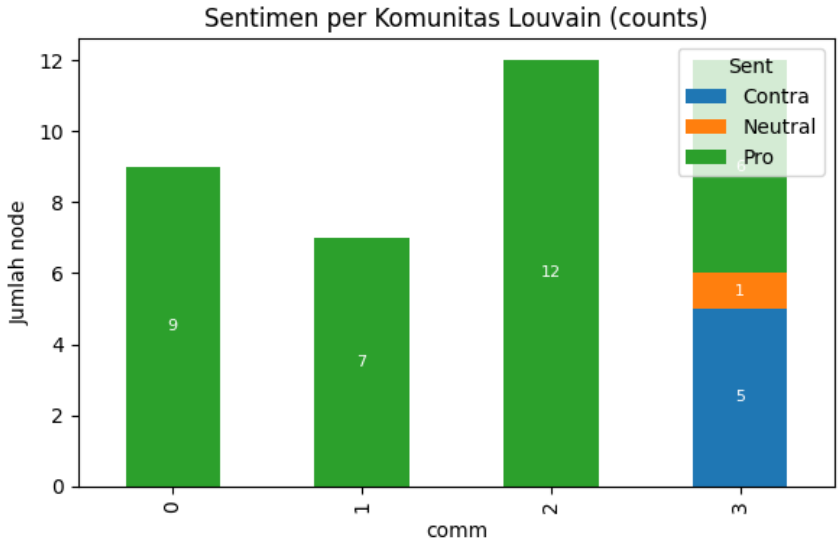
Gambar 4.19 Centrality dan Core Setiap Komunitas

Gambar 4.58 diatas, terlihat bahwa nilai deg_mean pada keempat komunitas berkisar antara 0.19 hingga 0.22, menandakan bahwa rata-rata setiap node memiliki sekitar 20% keterhubungan dibandingkan dengan ukuran komunitasnya. Nilai `bet_mean` relatif rendah sekitar 0.02 hingga 0.04, yang mengindikasikan hanya sedikit node yang benar-benar menjadi penghubing utama antar komunitas. Konfigurasi ini sesuai dengan tujuan knowledge graph untuk menonjolkan beberapa "hub" informasi sambil menjada agar sebagaian besar entitas tetap terfokus dalam sub-domain masing-masing. Sedangkan nilai `core_min` sama-sama memiliki nilai 5 pada tiap komunitas mengkonfirmasi bahwa setiap cluster memilii minimal lima entitas yang tergabung dalam k-core inti sebuah struktur kohesif di mana setiap node terhubung sedikitnya lima tetangga dalam komunitas. Hal ini menunjukkan bahwa meskipun perbedaan ukuran 7 hingga 12 node, semua komunitas berhasil mempertahankan tingkat kohesi minimal yang diperlukan untuk analisis lanjutan.

Komunitas	Contra	Neutral	Pro
0	0	0	9
1	0	0	7
2	0	0	12
3	5	1	6

Tabel 4.29 Distribusi Sentimen tiap Komunitas

Selanjutnya, distribusi sentimen per komunitas sendiri dianalisis untuk memahami kecenderungan entitas dalam dokumen berita pada sentimen. Pada Tabel 4.29, terlihat bahwa tiga komunitas pertama, yaitu komunitas 0, 1, 2 didominasi oleh Pro, di mana masing-masing memiliki nilai 9 untuk komunitas 0, komunitas 1 dengan 7 sentimen Pro, dan 12 sentimen pada komunitas 2. Sedangkan Komunitas 3 menampilkan distribusi sentimen yang lebih beragam, dengan terdapat 5 node Contra, 1 Neutral, dan 6 Pro. Keberadaan proporsi kontra yang signifikan mengindikasikan bahwa subgraph ini mencakup diskusi kritis atau masalah kontroversi, seperti bahasan topik yang cenderung ekstrem.



Gambar 4.20 Distribusi Sentimen tiap Komunitas

Secara keseluruhan, analisis distribusi sentimen menegaskan bahwa komunitas-komunitas dalam knowledge graph tidak hanya terpisah secara tropical, tetapi juga secara emosional. Beberapa komunitas menunjukkan persepsi positif yang konsisten, sementara yang lain menangani isu-isu kompleks dengan respon yang beragam,

Sebagai analisis lanjutan pada knowledge graph, dilakukan analisis tiap komunitas untuk memahami eksplorasi distribusi berita teknologi pertanian berdasarkan topik yang telah diklasifikasikan untuk mengidentifikasi kecenderungan sentimen di dalamnya. Adapun tujuan dari analisis berikut untuk menilai sejauh mana media memberitahukan inovasi pertanian serta dampak framing tersebut terhadap proses adopsi teknologi di sektor pertanian masa depan.

Komunitas	Top-Keyword Paling Sering Muncul
0	*"food security", "export", "import", "temperature", "fruit"*

Komunitas	Top-Keyword Paling Sering Muncul
1	*"dairy", "emission", "climate", "supply chain", "sheep"*
2	*"price", "drought", "soybean", "investment", "drone"*
3	*"rainfall", "corn", "yield", "crop", "soil health"*

Tabel 4.30 Top-Keywords Entitas dalam Komunitas

Setiap komunitas yang terdeteksi oleh algoritma Louvain di knowledge graph dikaitkan kembali dengan klasifikasi topik yang ada pada dokumen berita sebagai validasi terkait tema yang digunakan pada analisis sentimen memang benar mencerminkan relasi antar entitas pada knowledge graph pada akhirnya. Langkah awal dalam klasifikasi dan kolerasi antar topik dengan sentimen adalah dengan mencetak entitas yang sering muncul pada tiap komunitas untuk menemukan pola hubungan pada topik yang telah di klasifikasikan sebelumnya.

Pada komunitas 0, yang didominasi keyword seperti *"food security"*, *"export"*, *"import"*, *"temperature"*, *"fruit"*, mencerminkan koterkaitan dengan topik *Food Security & Risk Management*. Di mana seluruh beritapada distribusi sentimen sebelumnya pada knowledge graph mengarah pada klasifikasi sentimen Pro, menegaskan bahwa media menyoroti adopsi teknologi pertanian sebagai solusi utama untuk menjaga ketahanan pangan global, sekaligus meminimalkan risiko terkait perubahan iklim dan flukuasi pasar.

Sedangkan pada komunitas 1, dengan keyword unggulan yang dihitung pada beberapa relasi adalah *"dairy"*, *"emission"*, *"climate"*, *"supply chain"*, *"sheep"*, menyoroti topik *Climate & Environmental Monitoring* serta *Maket Access & Trade Dynamics*, di mana pemberitaan menekankan teknologi monitoring emisi dan inovasi dalam rantai pasok peternakan.

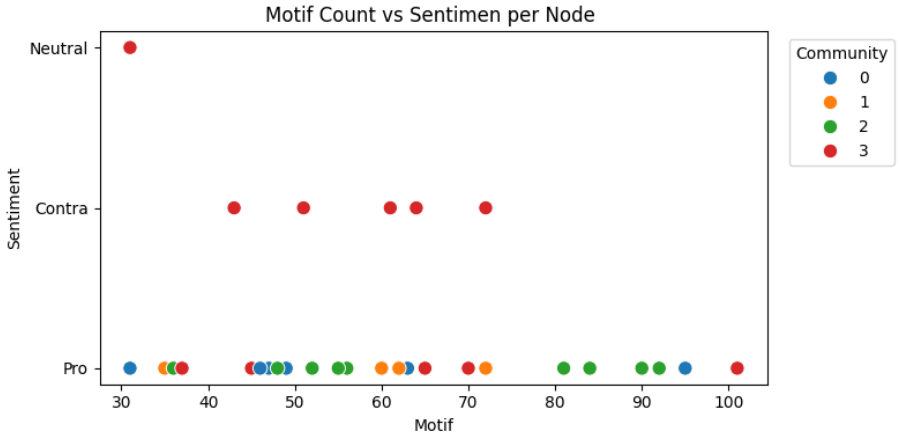
Komunitas 2 sendiri mengelompokkan keywords *"price"*, *"drought"*, *"soybean"*, *"investment"*, *"drone"*, yang mana topik tesebut lebih mengarah pada Economic & Farming serta *Big Data Analytics & Cloud Plaforms*. Isu harga komoditas dan kekeringan diangat sebagai pemberitaan yang menarik sebagai investasi pada platform digital misalnya *drone* atau sensor berbasis *cloud* untuk meningkatkan efisiensi produksi dan ketanan petani skala kecil,

Kemudian pada komunitas terakhir yang memiliki distribusi sentimen terbagi rata, berfokus pada *"rainfall"*, *"corn"*, *"yield"*, *"crop"*, *"soil health"*, merupakan persimpangan antara *Biotech & Crop Genetic* dan *Economic & Family Farming*, namun tetap menampilkan dinamika sentimen paling beragam untuk tiga kelas. Di satu sisi, pemberitaan berperspektif pro menitikberatkan pada inovasi, sedangkan sudut pandang kontra mengangkat kekhawatiran mengenai potensi

risiko ekologi dan meningkatnya ketergantungan pada teknologi. Temuan ini menggambarkan adanya keseimbangan antara optimisme terhadap kemajuan teknologi dengan wacana kritis yang mempertanyakan dampak jangka panjangnya. Secara keseluruhan, media cenderung memosisikan teknologi pertanian dalam kerangka yang positif menghubungkan setiap topik dengan solusi digital dan inovasi mutakhir namun tetap menyediakan ruang bagi perdebatan kritis, khususnya terkait isu lingkungan dan keberlanjutan.

Node	Komunitas	Sentimen	Motif
Pilot	2	Pro	36
Poultry	2	Pro	90
Export ban	2	Pro	65
Rainfall	3	Contra	43
price	2	Pro	84

Tabel 4.31 Top-Node dalam Motif Komunitas



Gambar 4.21 Distribusi Sentimen pada Motif

Selanjutnya, untuk menggali pola hubungan mikro di dalam setiap komunitas, dilakukan motif analisis yaitu perhitungan frekuensi subgraf tiga-node yang paling sering muncul pada setiap entitas. Motif-motif ini mengungkap klaster-klaster kecil yang memperkuat narasi tertentu. Sebagai contoh, node '*poultry*' dan '*price*' di Komunitas 2 memiliki motif count yang sangat tinggi (90 dan 84), menunjukkan bahwa isu peternakan unggas dan harga komoditas sering muncul dalam subgraf yang padat cocok dengan framing Pro pada topik *Economic & Family Farming* dan *Big Data Analytics & Cloud Platforms*. Sebaliknya, node '*rainfall*' di Komunitas 3, dengan motif count 43 dan sentimen Contra, terletak dalam subgraf yang memprakarsai diskusi risiko iklim dan memberi dasar yang kuat terkait framing kritis pada topik *Climate & Environmental Monitoring*. Dengan demikian, motif analysis memperkuat wawasan kita bahwa media menyoroti inovasi teknologi Pro dalam klaster-klaster

tersegmentasi, namun tetap menyediakan bagi narasi kritis Contra pada simpul-simpul tertentu yang menyorot isu keberlanjutan.

Secara keseluruhan rangkaian eksperimen yang meliputi klasifikasi sentimen menggunakan BERT dengan SVM, Random Forest, dan XGBoost, kemudian eksplorasi ensemble stacking meta leaning dan analisis pemetaan artikel berita dalam knowledge graph telah memberikan gambaran komprehensif tentang bagaimana media memberitakan teknologi pertanian. Pertama, model BERT + RF terbukti unggil dengan akurasi dan F1-macro tertinggi dengan 97% dan 96.9%, meskipun pada *stacking* meta leaner mampu meningkatkan performa terutama pada kelas Neutral. Kedua, distribusi sentimen pada tiap topik menunjukkan dominasi narasi Pro do mayoritas tema dari *Food Security* hingga *Big Data Analytics* yang mencerminkan optimisme media terhadap adopsi teknologi. Kemudian temuan ketiha, klatering komunita Louvain mengaska keterkaitan antara enitas berita dan topik klasifkasi, sementara motof analysis memetakan node-node yang merefleksikan ttik-titik kritid dan solutif.

Sejalan dengan tujuan penelitian ini dirumuskan, temuan berikut bukan hanya memvalidasi kerangka sentimen per topik dalam berita, namun juga membuka ruang diskusi mengenai framing media. Di satu sisi, framing berita dan isi konten berita mempengaruhi adaptasi teknologi dengan menekankan inovasi dan efisiensi, disisi lain menggarisbawahi kekhwatiran ekologis serta tantanagn keberlanjutan jangka panjang. Hal tersebut dapat dijadikan implikasi kesimpulan secara teoritis pada analisis selanjutnya, dan rekomenasi strategi komunikasi teknologi pertanian yang lebih seimbang.

BAB 5 KESIMPULAN DAN SARAN

Pada bab ini disajikan rangkuman temuan dari seluruh tahapan analisis sentimen pada berita teknologi pertanian guna menjawab rangkaian rumusan masalah, beserta rekomendasi untuk penyempurnaan model di penelitian selanjutnya.

5.1 Kesimpulan

Berdasarkan serangkaian eksperimen dalam analisis sentimen berita teknologi pertanian dengan Knowledge Graph dan Machine Learning, dapat dirangkum beberapa temuan utama sebagai berikut:

1. Analisis sentimen berbasis machine learning pada dokumen berta tingkat awal menggunakan BERT Fine-Tune sebagai base learner kemudian dikombinasikan dengan Support Vector Machine (SVM), Random Forest, dan XGBoost sendiri memiliki konfigurasi terbaik pada BERT + Random Forest dengan rata-rata accuracy 92.4% dan F1-macro 88%. Sedangkan F1 pada BR-01 mencapai titik tertinggi pada Pro dengan f1-score 98%.
2. Pada eksperimen lainnya, analisis sentimen machine learning dokumen tingkat awal menggunakan teknik *balancing* data pada tiga kelas label "*Pro*", "*Contra*", dan "*Neutral*" hasilnya adalah kombinasi BERT + Random Forest tetap unggul dengan accuracy 95%, disusul pada BERT + SVM pada 93% untuk accuracy, dan pada BERT Fine-tuned dengan accuracy 85%.
3. Pada metode *balancing* 390 dataset artikel berita, menggunakan *balancing class* pada distribusi class ground truth dan setiap model. Adapun balancing berikut menggunakan class pro sekitar 170 dataset, contra 130, dan neutral 90, kemudian diberikan thresehold sekitar 0.10. Meskipun dari tingkat accuracy mengalami peningkatan, namun pada dataset hasil ketiga kelas yang terbagi rata tidak sama dengan dataset asli pada dataset yang telah diuji dan di validasi secara manual. Sebaliknya, jika mencari hasil terbaik dengan ukuran accuracy, maka metode berikut dapat digunakan untuk penelitian lebih lanjut, dan sebagai tingkat pengukuran distribusi kelas sentimen yang dapat tersebar di keseluruhan kelas.
4. Skema stacking Gradient Boosting meningkatkan F1-macro keseluruhan menjadi 87.2 % dengan perbaikan khusus pada kelas Neutral, meskipun pada hasil rata-rata performa f1-macro masih dibawah hasil base learner pada BERT + Random Forest

sebagai model dengan hasil terbaik pada tahap awal. Selaras dengan performa tersebut accuracy juga tidak dapat mengungguli model base learner dengan hanya 91% pada uji stcaking. Pada uji paired t-test, accuracy hanya mendapat angka t-statistic: -0.58.

5. Skema stacking pada Gradient Boosting dan meta learner LogisticRegression yang gagal mengindikasikan dua hasil temuan, dimana penggunaan BERT sebagai base model untuk model SVM, RF, dan XGBoost belajar kesalahan yang sama sehingga meta learner kesulitan menemukan sinyal tambahan. Dan indikasi kedua pada Baseline BERT + Random Forest sendiri sudah memiliki performa terbaik sehingga model meta learner maupun model lainnya kesulitan dalam mencapai performa yang lebih baik maupun sama dengan baseline tersebut.
6. Pada analisis sentimen yang ada pada distribusi topik pada dokumen berita, sentimen Pro mendominasi sekitar 62% pada topik *Food Security & Risk Management* serta *Big Data & Cloud Platforms*, menandakan optimisme media terhadap solusi sigital untuk ketahanan pangan. Sedangkan sentimen Contra paling tinggo pada topik Market Access & Trade Dynamics, terutama terkait volatilitas harga dan risko ekspor-impor. Sebaliknya Neutral berperan sebagai penyampai konteks kebijakan dan perkembangan regulasi.
7. Pada pemetaan relasi antar entitas, K-Means meskipun belajar banyak pada adjency metric dari embedding GCN, belum mampu memberikan klaster yang baik pada beberapa entitas sehingga dalam kasus knowledge graph terdapat banyaknya distribusi open triad yang memberikan celah pada beberapa relasi untuk tidak saling berhubungan.
8. Analisis knowledge graph pada GCN memberikan analisis yang cukup baik dan mampu memprediksi temuan seperti link suggestion, seperti pada kasus klastering K-Means sebelumnya. Sehingga dapat merumuskan relasi yang tidak terlihat, hal ini mampu menjawab analisis knowledge graph pada GCN teruji dalam menilai kesalahan dibanding analisis knowledge graph secara langsung pada grafik.
9. Pada pemetaan relasi enitas melalui knowledge graph, algoritma Louvain memisahkan empat komunitas utama yang selaras dengan enam topik klasifikasi. Di mana pada komunitas 0 berisikan topik yang condong pada *Food Security*, dan komunitas 2 terkorelasi *Economic & Family Farming*.
10. Pada analisis motif sentimen menujukkan node dengan hitungan motif >80 hampir seluruhnya bermuatan sentimen Pro, sedangkan node bermotif 40-60 menampilkan kergaman opini "*Pro*", "*Contra*", dan "*Neutral*". Hal ini menegaskan bahwa intensitas relasi entitas ikut mempengaruhi bias framing media.

11. Implikasi empiris yang diperoleh, media terlalu arus cenderung memposisikan teknologi pertanian dalam kerangka positif dengan mengkaitkannya dengan efisiensi, produktivitas, dan mitigasi iklim. Namun, ruang diskusi pada hasil tetap hadir dengan adanya isu risiko ekologi, ketergantuang pertain terhadap teknologi, serta fluktuasi pasar.

5.2 Saran

Berdasarkan temuan pada kesimpulan sebelumnya, berikut beberapa saran yang diberikan untuk pengembangan penelitian di masa mendatang, yaitu sebagai berikut:

1. Pada hasil stcaking menggunakan GradientBoosting pada meta leaner dirasa masih kurang mampu untuk mengungguli base learner, sehingga perlu menggunakan meta leaner yang mampu menangkap pola sederhana dan non-linear sekaligus. Atau penambahan fitur dengan confidence score dan probabilitas kelas dari base leaner ke meta leaner, jadi bukan hanya menguji pada label prediksi.
2. Menyeimbangkan pemberitaan terkait sentimen Contra maupun Neutral, sebagai penyeimbang artikel berita bertajuk sentimen Pro, sehingga distribusi sentimen pada dokumen berita merata dan tidak terlalu optimis pada konten bertajuk positif.
3. Analisis lanjutan dengan memperluas korpus dan menambahkan link suggestion pada model louvian agar relasi yang tidak terlihat muncul, sehingga hasil mapping pada graph semakin baik dan terdapat temuan-temuan baru terkait tema yang muncul pada artikel berita.

DAFTAR PUSTAKA

Ancín, M., Pindado, E., & Sánchez, M. (2022). New trends in the global digital transformation process of the agri-food sector: An exploratory study based on Twitter. *Agricultural Systems*, *203*. https://doi.org/10.1016/j.agsy.2022.103520

Breiman, L. (2001). *Random Forests* (Vol. 45).

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (n.d.). *New Avenues in Opinion Mining and Sentiment Analysis*. http://converseon.com

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, Z., Jin, J., & Li, M. (2022). Does media coverage influence firm green innovation? The moderating role of regional environment. *Technology in Society*, *70*. https://doi.org/10.1016/j.techsoc.2022.102006

Chen, Z., Yang, H., Ye, P., Zhuang, X., Zhang, R., Xie, Y., & Ding, Z. (2024). How does the perception of informal green spaces in urban villages influence residents' complaint Sentiments? a Machine learning analysis of Fuzhou City, China. *Ecological Indicators*, *166*. https://doi.org/10.1016/j.ecolind.2024.112376

Christopher D. Manning, P. R. and H. S. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

Christopher M. Bishop. (2006). Pattern recognition and machine learning. In *Information Science and Statistics*. Springer New York, NY. https://link.springer.com/book/9780387310732

Cimiano, P., & Paulheim, H. (2016). Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods. In *Semantic Web* (Vol. 0). IOS Press. http://www.geonames.org/

Cortes, C., Vapnik, V., & Saitta, L. (1995). Support-Vector Networks Editor. In *Machine Leaming* (Vol. 20). Kluwer Academic Publishers.

Cortes, C. , & V. V. (1995). Support-vector networks. *Machine Learning*, 273–297.

Costola, M., Hinz, O., Nofer, M., & Pelizzon, L. (2023). Machine learning sentiment analysis, COVID-19 news and stock market reactions. *Research in International Business and Finance*, *64*. https://doi.org/10.1016/j.ribaf.2023.101881

Daza, A., González Rueda, N. D., Aguilar Sánchez, M. S., Robles Espíritu, W. F., & Chauca Quiñones, M. E. (2024). Sentiment Analysis on E-Commerce Product Reviews Using Machine Learning and Deep Learning Algorithms: A Bibliometric Analysisand Systematic Literature Review, Challenges and Future Works. In *International Journal of Information Management Data Insights* (Vol. 4, Issue 2). Elsevier B.V. https://doi.org/10.1016/j.jjimei.2024.100267

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. http://arxiv.org/abs/1810.04805

Duryea, E., Ganger, M., & Hu, W. (2016). Exploring Deep Reinforcement Learning with Multi Q-Learning. *Intelligent Control and Automation*, *07*(04), 129–144. https://doi.org/10.4236/ica.2016.74012

Fletcher, R., Kueng, L., Kleis Nielsen, R., Selva, M., & Suárez, E. (2020). *Journalism, Media and Technology Trends and Predictions 2020*. https://doi.org/10.60625/risj-ryxt-ja51

Goldberg, Y., & Levy, O. (2014). *word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method*. http://arxiv.org/abs/1402.3722

Hao, J., Pei, L., He, Y., Xing, Z., & Weng, Y. (2024). TCKGCN: Graph convolutional network for aspect-based sentiment analysis with three-channel knowledge fusion. *Neurocomputing*, *600*. https://doi.org/10.1016/j.neucom.2024.128163

Havrlant, L., & Kreinovich, V. (2017). A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation). *International Journal of General Systems*, *46*(1), 27–36. https://doi.org/10.1080/03081079.2017.1291635

Hermida, A., Paulussen, S., Quandt, T., & Reich, Z. (2011). *The active recipient: Participatory journalism through the lens of the Dewey-Lippmann debate*. https://www.researchgate.net/publication/264003338

Hogan, A., Blomqvist, E., Cochez, M., D'Amato, C., Melo, G. De, Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., Ngomo, A. C. N., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys*, *54*(4). https://doi.org/10.1145/3447772

Holton, M., Riley, M., & Kallis, G. (2023). Keeping on[line] farming: Examining young farmers' digital curation of identities, (dis)connection and strategies for self-care through social media. *Geoforum*, *142*. https://doi.org/10.1016/j.geoforum.2023.103749

Ittefaq, M., Zain, A., Arif, R., Ala-Uddin, M., Ahmad, T., & Iqbal, A. (2025). Global news media coverage of artificial intelligence (AI): A comparative analysis of frames, sentiments, and trends across 12 countries. *Telematics and Informatics*, *96*. https://doi.org/10.1016/j.tele.2024.102223

Jalal, N., Mehmood, A., Choi, G. S., & Ashraf, I. (2022). A novel improved random forest for text classification using feature ranking and optimal number of trees. *Journal of King Saud University - Computer and Information Sciences*, *34*(6), 2733–2742. https://doi.org/10.1016/j.jksuci.2022.03.012

Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2020). *A Survey on Knowledge Graphs: Representation, Acquisition and Applications*. https://doi.org/10.1109/TNNLS.2021.3070843

Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2022). A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(2), 494–514. https://doi.org/10.1109/TNNLS.2021.3070843

Jurafsky, D. , & M. J. H. (2019). *Speech and Language Processing (3rd ed.)*. Pearson.

Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. In *Computers and Electronics in Agriculture* (Vol. 143, pp. 23–37). Elsevier B.V. https://doi.org/10.1016/j.compag.2017.09.037

Kipf, T. N., & Welling, M. (2016). *Semi-Supervised Classification with Graph Convolutional Networks*. http://arxiv.org/abs/1609.02907

Lin, S. Y., Kung, Y. C., & Leu, F. Y. (2022). Predictive intelligence in harmful news identification by BERT-based ensemble learning model with text sentiment analysis. *Information Processing and Management*, *59*(2). https://doi.org/10.1016/j.ipm.2022.102872

Lindgren, E., Lindholm, T., Vliegenthart, R., Boomgaarden, H. G., Damstra, A., Strömbäck, J., & Tsfati, Y. (2022). Trusting the Facts: The Role of Framing, News Media as a (Trusted) Source, and Opinion Resonance for Perceived Truth in Statistical Statements. *Journalism and Mass Communication Quarterly*. https://doi.org/10.1177/10776990221117117

Liu, B. (2022). Sentiment analysis and opinion mining. *Sprnger Nature*.

Loughran, T. , & M. B. (2011). *When is a liability not a liability? Textual analysis, dictionaries*.

Luo, M., & Mu, X. (2022). Entity sentiment analysis in the news: A case study based on Negative Sentiment Smoothing Model (NSSM). *International Journal of Information Management Data Insights*, *2*(1). https://doi.org/10.1016/j.jjimei.2022.100060

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011

Mesut, B., Başkor, A., & Buket Aksu, N. (2023). Role of artificial intelligence in quality profiling and optimization of drug products. In *A Handbook of Artificial Intelligence in Drug Delivery* (pp. 35–54). Elsevier. https://doi.org/10.1016/B978-0-323-89925-3.00003-4

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*. http://arxiv.org/abs/1301.3781

Mohr, S., & Höhler, J. (2023). Media coverage of digitalization in agriculture - an analysis of media content. *Technological Forecasting and Social Change*, *187*. https://doi.org/10.1016/j.techfore.2022.122238

Montomoli, J., Romeo, L., Moccia, S., Bernardini, M., Migliorelli, L., Berardini, D., Donati, A., Carsetti, A., Bocci, M. G., Wendel Garcia, P. D., Fumeaux, T., Guerci, P., Schüpbach, R. A., Ince, C., Frontoni, E., Hilty, M. P., Alfaro-Farias, M., Vizmanos-Lamotte, G., Tschoellitsch, T., … Colak, E. (2021). Machine learning using the extreme gradient boosting (XGBoost) algorithm predicts 5-day delta of SOFA score at ICU admission in COVID-19 patients. *Journal of Intensive Medicine*, *1*(2), 110–116. https://doi.org/10.1016/j.jointm.2021.09.002

Mulyani, Y. P., Saifurrahman, A., Arini, H. M., Rizqiawan, A., Hartono, B., Utomo, D. S., Spanellis, A., Beltran, M., Banjar Nahor, K. M., Paramita, D., & Harefa, W. D. (2024). Analyzing public discourse on photovoltaic (PV) adoption in Indonesia: A topic-based sentiment analysis of news articles and social media. *Journal of Cleaner Production*, *434*. https://doi.org/10.1016/j.jclepro.2023.140233

Pang, B., & L. L. (2008). *Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval*.

Pérez-Mesa, J. C., García Barranco, M. [a]C, Serrano Arcos, M. [a]M, & Sánchez Fernández, R. (2023). Agri-food crises and news framing of media: an application to the Spanish greenhouse sector. *Humanities and Social Sciences Communications*, *10*(1). https://doi.org/10.1057/s41599-023-02426-y

Ramos, J. (n.d.). *Using TF-IDF to Determine Word Relevance in Document Queries*.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*(1), 93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). *A Primer in BERTology: What we know about how BERT works*. http://arxiv.org/abs/2002.12327

Russell, S. , & N. P. (2020). *Artificial Intelligence: A Modern Approach (4th ed.)* (4th Edition).

Rust, N. A., Jarvis, R. M., Reed, M. S., & Cooper, J. (2021). Framing of sustainable agricultural practices by the farming press and its effect on adoption. *Agriculture and Human Values*, *38*(3), 753–765. https://doi.org/10.1007/s10460-020-10186-7

Sammut, C. m, & Webb, G. I. (2017). TF-IDF weighting. *Encyclopedia of Machine Learning and Data Mining*.

Siegrist, M., & Hartmann, C. (2020a). Consumer acceptance of novel food technologies. In *Nature Food* (Vol. 1, Issue 6, pp. 343–350). Springer Nature. https://doi.org/10.1038/s43016-020-0094-x

Siegrist, M., & Hartmann, C. (2020b). Consumer acceptance of novel food technologies. In *Nature Food* (Vol. 1, Issue 6, pp. 343–350). Springer Nature. https://doi.org/10.1038/s43016-020-0094-x

Smairi, N., Abadlia, H., Brahim, H., & Chaari, W. L. (2024). Fine-tune BERT based on Machine Learning Models For Sentiment Analysis. *Procedia Computer Science*, *246*(C), 2390–2399. https://doi.org/10.1016/j.procs.2024.09.531

Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (n.d.). *Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank*. Association for Computational Linguistics. http://nlp.stanford.edu/

Sun, C., Huang, L., & Qiu, X. (n.d.). *Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence*. https://github.com/uclmr/jack/tree/master

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). *Lexicon-Based Methods for Sentiment Analysis*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017a). *Attention Is All You Need*. http://arxiv.org/abs/1706.03762

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017b). *Attention Is All You Need*. http://arxiv.org/abs/1706.03762

Wan, Y., Chen, Y., Lin, J., Zhong, J., & Dong, C. (2024). A knowledge-augmented heterogeneous graph convolutional network for aspect-level multimodal sentiment analysis. *Computer Speech and Language*, *85*. https://doi.org/10.1016/j.csl.2023.101587

Wester, J., Turffs, D., McEntee, K., Pankow, C., Perni, N., Jerome, J., & Macdonald, C. (2023). Agriculture and downstream ecosystems in Florida: an analysis of media discourse. *Environmental Science and Pollution Research*, *30*(2), 3804–3816. https://doi.org/10.1007/s11356-022-22475-1

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big Data in Smart Farming – A review. In *Agricultural Systems* (Vol. 153, pp. 69–80). Elsevier Ltd. https://doi.org/10.1016/j.agsy.2017.01.023

Wu, Z., Chen, F., Long, G., Pan, S., Zhang, C., & Yu, P. S. (2019). *A Comprehensive Survey on Graph Neural Networks*. https://doi.org/10.48550/arXiv.1901.00596

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2019). *A Comprehensive Survey on Graph Neural Networks*. https://doi.org/10.1109/TNNLS.2020.2978386

Xu, J., She, S., & Liu, W. (2022). Role of digitalization in environment, social and governance, and sustainability: Review-based study for implications. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.961057

Yadav, V., & Bethard, S. (2019). *A Survey on Recent Advances in Named Entity Recognition from Deep Learning models*. http://arxiv.org/abs/1910.11470

Zhang, L., Wang, S., & Liu, B. (2022). *Deep Learning for Sentiment Analysis : A Survey*.

Zhong, W., Xu, J., Tang, D., Xu, Z., Duan, N., Zhou, M., Wang, J., & Yin, J. (n.d.). *Reasoning Over Semantic-Level Graph for Fact Checking*. https://demo.allennlp.org/

Zhu, Z., Wang, L., Gu, D., Wu, H., Janfada, B., & Minaei-Bidgoli, B. (2023). Is Prompt the Future? A Survey of Evolution of Relation Extraction Approach Using Deep Learning and Big Data. *International Journal of Information Technologies and Systems Approach*, *16*(1). https://doi.org/10.4018/IJITSA.328681

LAMPIRAN

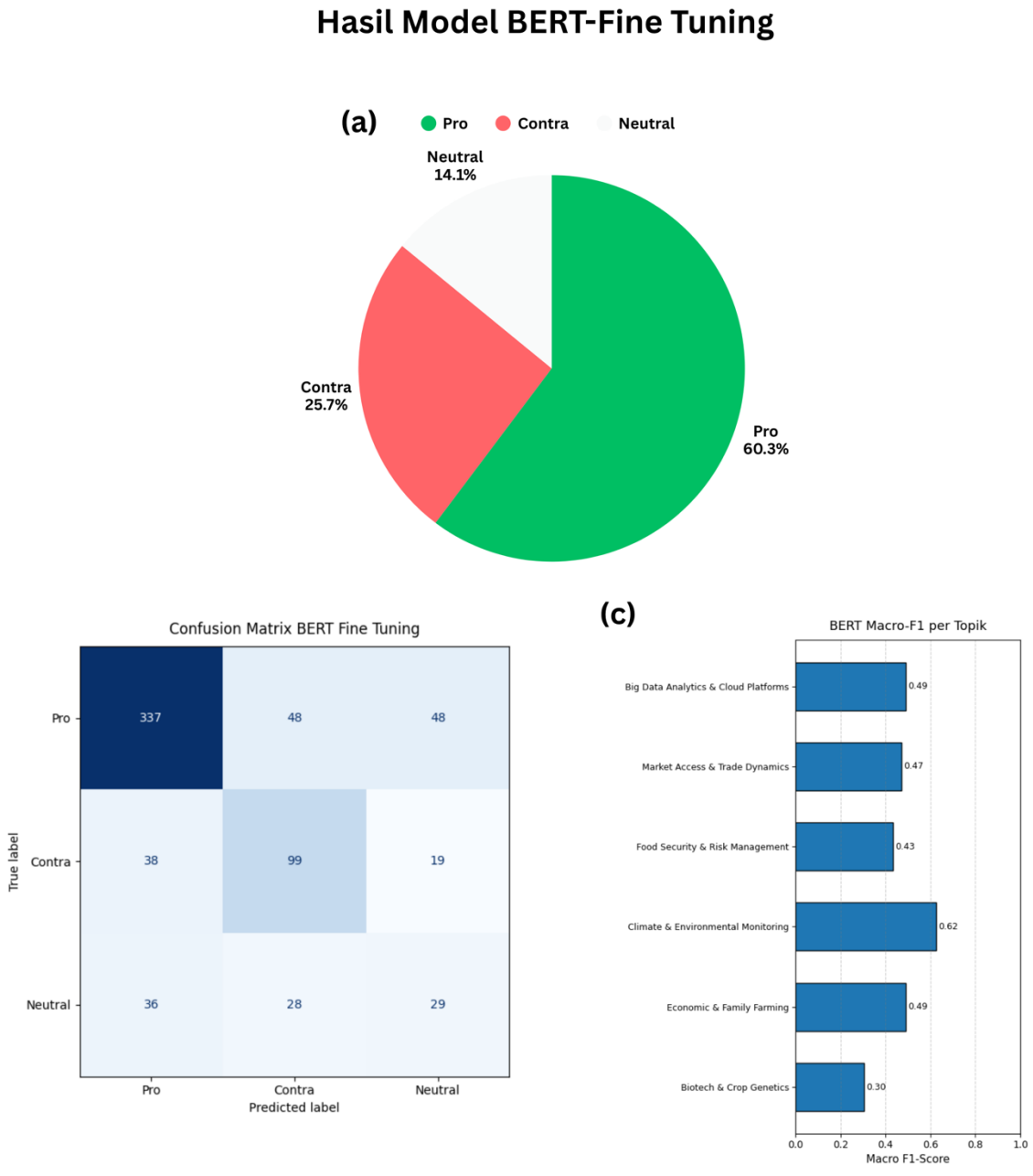
Lampiran 1: Script Metode Penelitian

Github Link:

https://github.com/elisanursm/data_fp/blob/main/Try%203/Try_label/try5_final.ipynb

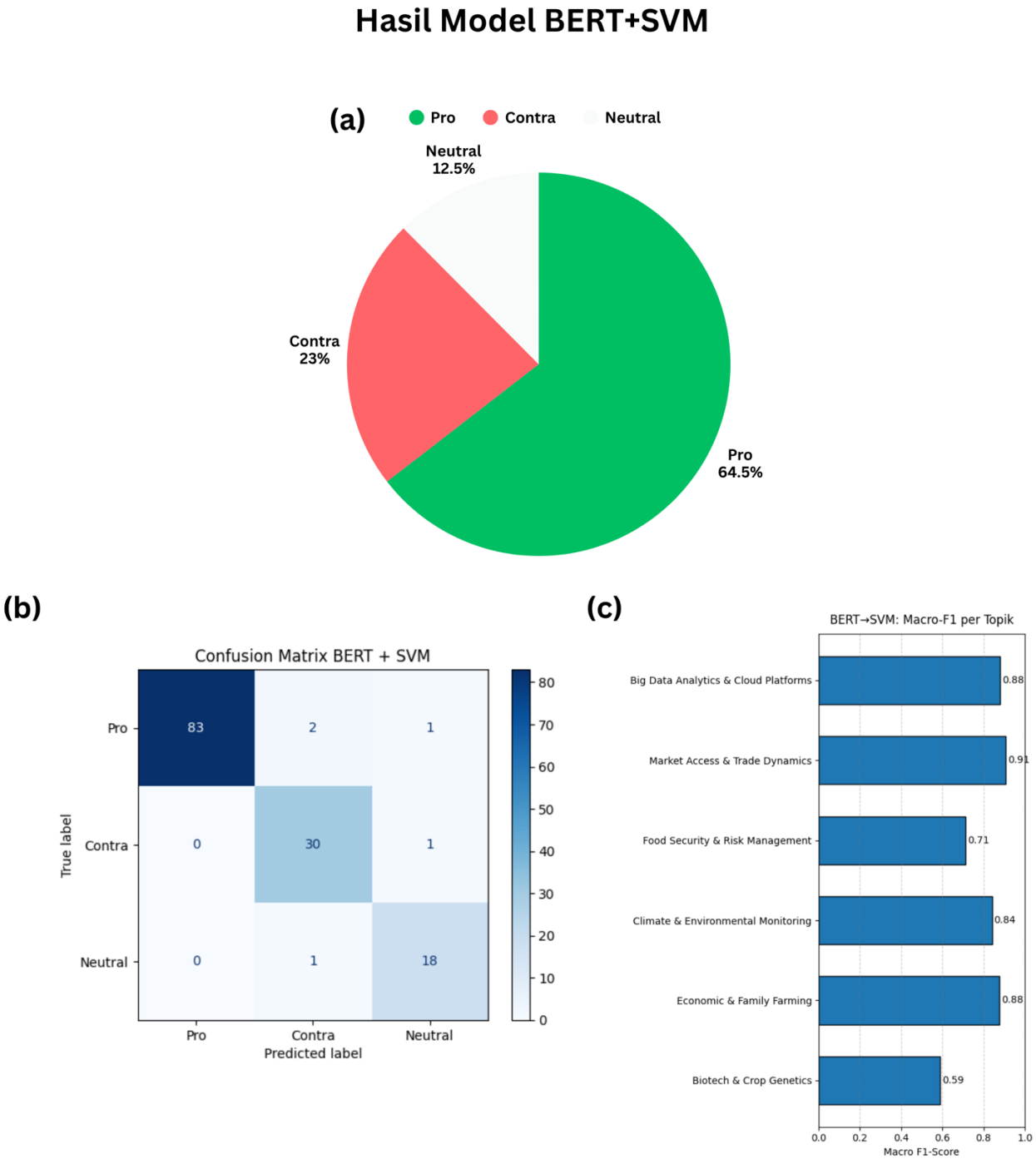
Lampiran 2: Hasil Analisis Sentimen Dokumen Berita Tingkat Awal

*) Visualisasi Hasil Model BERT-Fine Tuning Dokumen Berita



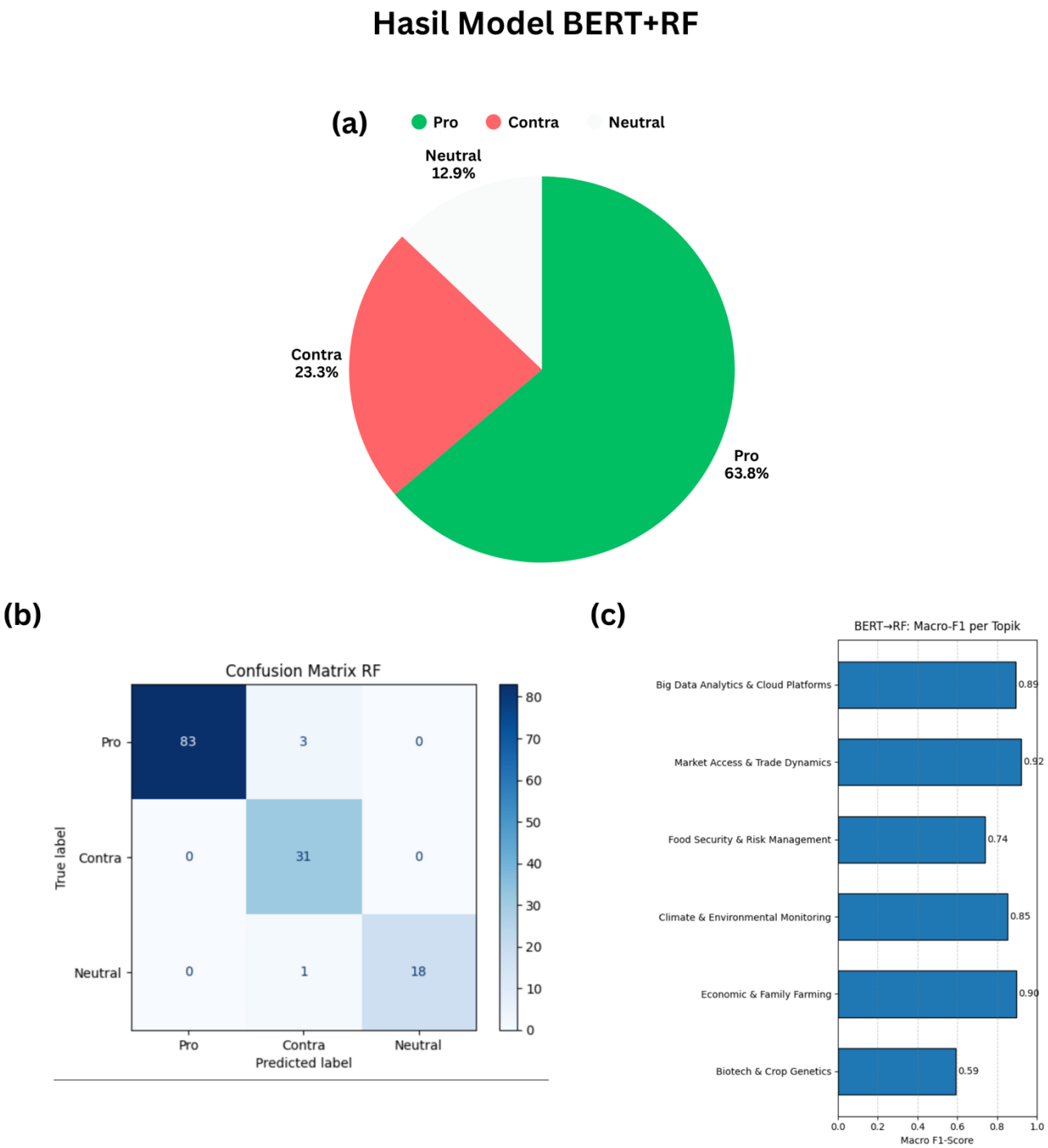
Gambar Lampiran 1. Hasil Model BERT (a) Distribusi Sentimen BF; (b) Confussion Matrix BF; (c) Distribusi Topik BF

*) Visualisasi Hasil Model BERT+SVM Dokumen Berita



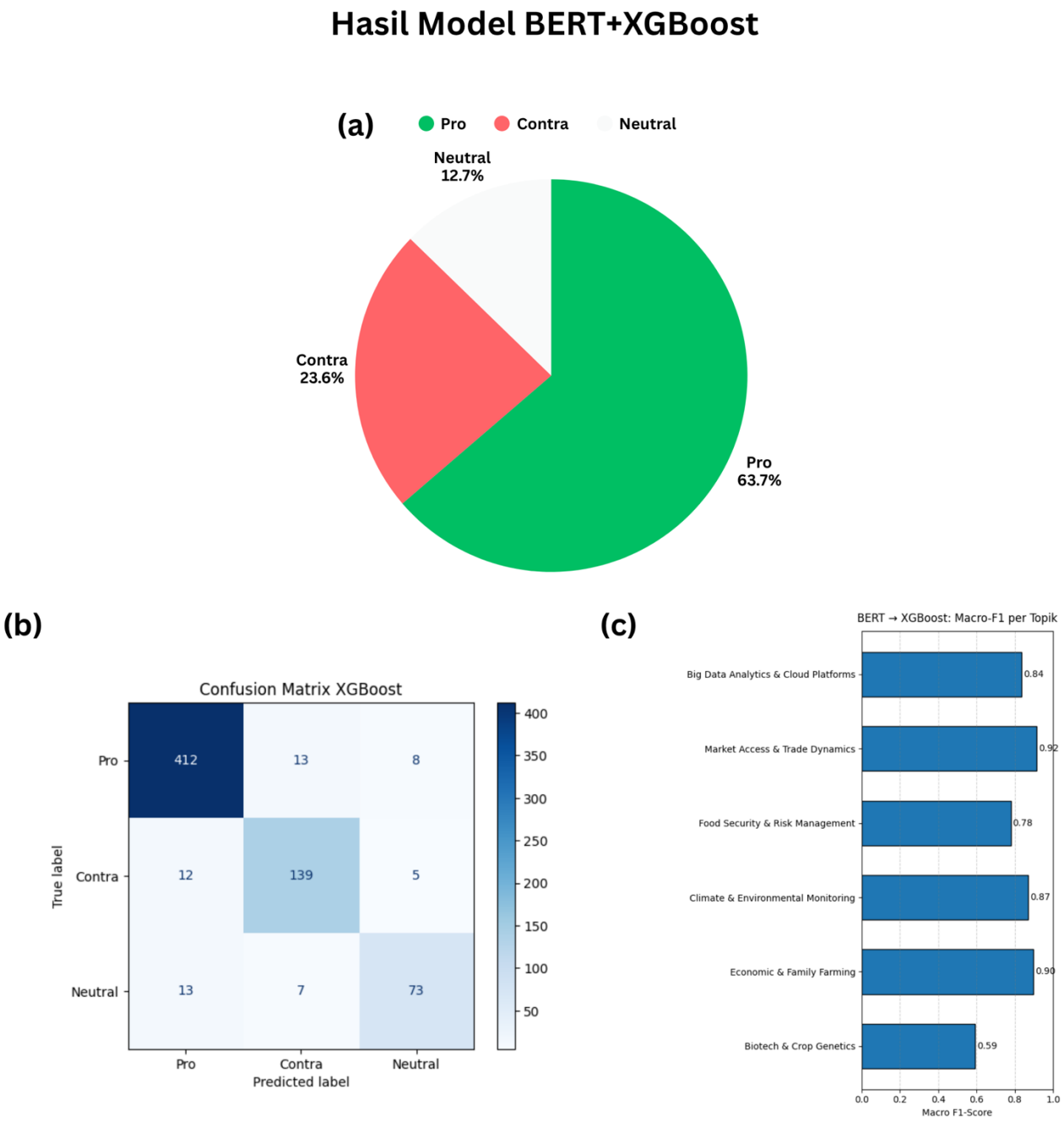
Gambar Lampiran 2. Hasil Model BERT+SVM (a) Distribusi Sentimen BS; (b) Confussion Matrix BS; (c) Distribusi Topik BS

*) Visualisasi Hasil Model BERT+Random Forest Dokumen Berita



Gambar Lampiran 3. Hasil Model BERT+RF (a) Distribusi Sentimen BR; (b) Confussion Matrix BR; (c) Distribusi Topik BR

*) Visualisasi Hasil Model BERT+XGBoost Dokumen Berita



Gambar Lampiran 4. Hasil Model BERT+XGBoost (a) Distribusi Sentimen BX; (b) Confussion Matrix BX; (c) Distribusi Topik BX

Lampiran 3: Eksperimen Hasil Analisis Sentimen Dokumen Berita

*) Hasil Eksperimen 3 Model (n=390)

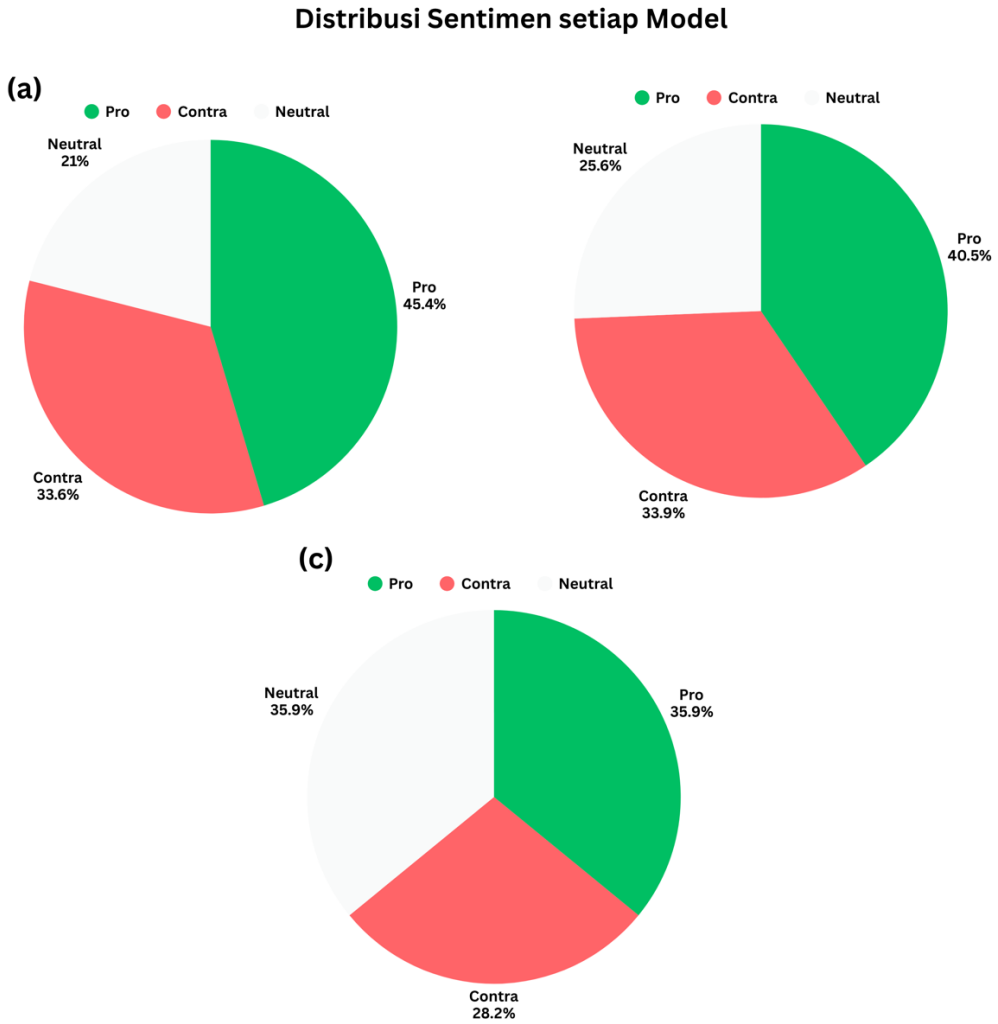
Model	*Hyperparameter*				Accuracy (%)	F1-macro (%)
	Epoch	Learning Rate	Batch Size	Random State/C/ n_estimators/max_depth		
BF - 01	5	3e-5	16	Seed: 42	88	89
BF – 02	5	3e-5	16	Seed: 42	78	76.6
BF – 03	5	3e-5	16	Seed: 42	82	81
BF – 04	5	3e-5	16	Seed: 42	88	88
BF - 05	5	3e-5	16	Seed: 42	87	86.6
BS - 01	5	3e-5	16	Seed: 42; kernel: linear; γ: scale; C:0.1	**97.4**	**97.3**
BS - 02	5	3e-5	16	Seed: 42; kernel: linear; γ: scale; C:0.1	92.3	91.3
BS - 03	5	3e-5	16	Seed: 42; kernel: linear; γ: scale; C:0.1	93.6	93
BS - 04	5	3e-5	16	Seed: 42; kernel: linear; γ: scale; C:0.1	94.9	94.3
BS - 05	5	3e-5	16	Seed: 42; kernel: linear; γ: scale; C:0.1	89.7	89
BR – 01*	5	3e-5	16	Seed: 42; n_estimators:3343; max_depth: None	96.2	95.6
BR - 02	5	3e-5	16	Seed: 42; n_estimators:3343; max_depth: None	93.6	92.4
BR - 03	5	3e-5	16	Seed: 42; n_estimators:3343; max_depth: None	94.9	92.3
BR - 04	5	3e-5	16	Seed: 42; n_estimators:3343; max_depth: None	94.9	94.3
BR - 05	5	3e-5	16	Seed: 42; n_estimators:3343; max_depth: None	94.9	95.89

*) Hasil Eksperimen Accuracy 3 Model (n=390)

Model	Accuracy (%)	Std Accuracy
BF	85	0.046
BS	93	0.029
BR	**95**	0.009

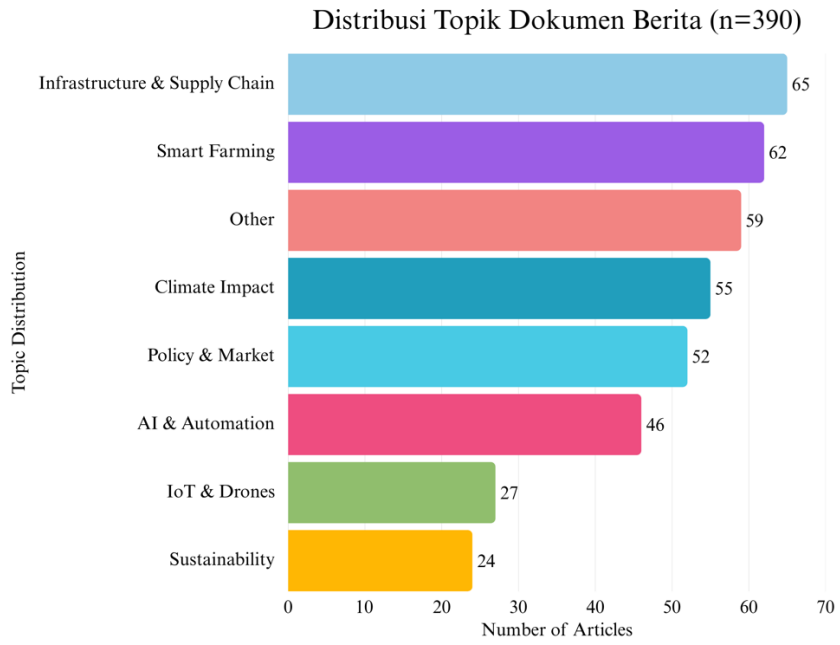
*) Distribusi Sentimen Eksperimen 3 Model (n=390)

Kelas	BF(%)	BS(%)	BR(%)
Contra	33.59	33.85	33.59
Neutral	21.03	25.64	23.59
Pro	45.38	40.51	42.82



Gambar Lampiran 5. Distribu Sentimen Model Eksperimen (a) Distribusi Sentimen BF; (b) Distribusi Sentimen BS; (c) Distribusi Sentimen BR

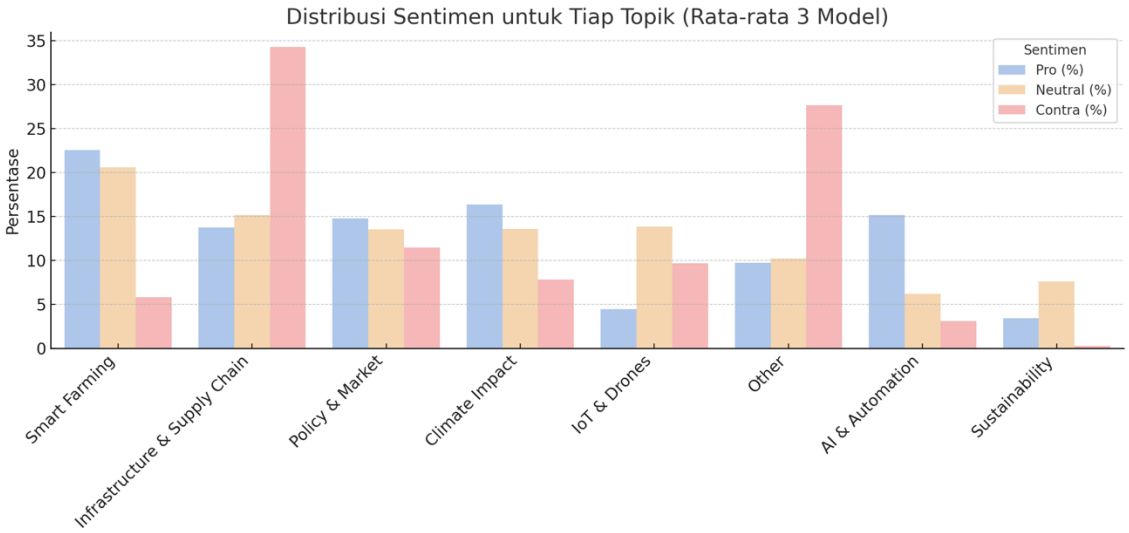
*) Distribusi Topik Dokumen Berita (n=390)



Gambar Lampiran 6. Distribusi Topik Dokumen Berita

*) Distribusi Sentimen per-Topik (n=390)

Topik	Pro (%)	Neutral (%)	Contra (%)	Dominan Sentimen
Smart Farming	22.53	20.57	5.83	Pro
Infrastructure & Supply Chain	13.73	15.13	34.27	Contra
Policy & Market	14.77	13.53	11.47	Pro
Climate Impact	16.37	13.57	7.83	Pro
IoT & Drones	4.43	13.87	9.67	Neutral
Other	9.73	10.20	27.63	Contra
AI & Automation	15.17	6.20	3.07	Pro
Sustainability	3.40	7.60	0.27	Neutral



Gambar Lampiran 7. Distribusi Sentimen Dokumen Berita per Dokumen

Lampiran 4: Hasil Stacking Model

*) Hasil Hyperparameter Base Model 10 Iterasi

Model	Base Learner Param	Accuracy (%)	F1-macro (%)
BL - BS – 01*	C=37.554; γ=scale; kernel=poly	**91.4**	**86.8**
BL – BS - 02	C=73.2994; γ=scale; kernel=linear	89.4	84.8
BL - BS – 03	C=15.7019; γ=scale; kernel=poly	91.4	86.8
BL - BS – 04	C=15.7019; γ=scale; kernel=poly	89.4	84.8
BL - BS – 05	C=60.2115; γ=auto; kernel=poly	91.4	86.8
BL - BS – 06	C=2.15845; γ=auto; kernel=sigmoid	88.1	83.2
BL - BS – 07	C= 83.3443; γ=auto; kernel=rbf	90.8	86.4
BL - BS – 08	C=18.2825; γ=scale; kernel=linear	89.4	84.8
BL - BS – 09	C=30.5242; γ=auto; kernel=linear	89.4	84.8
BL - BS - 10	C= 43.2945; γ=scale; kernel=poly	91.4	86.6
BL - BR - 01	n_est=1000; min_split=2; max_depth=50	91.9	87.8
BL - BR - 02	n_est=800; min_split=2; max_depth=50	91.9	87.8
BL - BR – 03	n_est= 100; min_split=10; max_depth=20	91.9	88.1
BL - BR - 04	n_est=100; min_split=5; max_depth=20	92.0	88.1
BL - BR – 05	n_est= 100; min_split=5; max_depth=20	92.0	88.1
BL - BR - 06	n_est=300; min_split=2; max_depth=50	91.8	87.8
BL - BR – 07*	n_est= 800; min_split=10; max_depth=10	**92.3**	**88.4**
BL - BR - 08	n_est=500; min_split=10; max_depth=50	91.9	87.9
BL - BR - 09	n_est= 1000; min_split=5; max_depth=50	92.0	88.0
BL - BR - 10	n_est=1000; min_split=2; max_depth=50	91.9	87.9
BL - BX – 01	n_est=100; max_depth=5; lr=0.05	89.7	85.3
BL - BX – 02	n_est=300; max_depth=10; lr=0.01	88,5	83.1
BL - BX – 03	n_est=500; max_depth=7; lr=0.05	91.1	86.9

BL - BX – 04*	n_est=800; max_depth=7; lr=0.10	**91.1**	**87.0**
BL - BX – 05	n_est=100; max_depth=7; lr=0.05	89.7	85.3
BL - BX – 06	n_est=300; max_depth=5; lr=0.10	91.0	86.9
BL - BX – 07	n_est=100; max_depth=7; lr=0.1	91.2	87.1
BL - BX – 08	n_est=800; max_depth=3; lr=0.05	91.1	86.9
BL - BX – 09	n_est=100; max_depth=5; lr=0.01	90	85.6
BL - BX - 10	n_est=300; max_depth=7; lr=0.05	91.0	86.8

*) Hasil Hyperparameter Stacking 20 Iterasi

Model	Hyperparameter			Accuracy (%)	F1-Macro (%)
	Learning rate	Estimator	Depth		
ST - 1	0.075908	750	8	89.5	84.1
ST – 2	0.120370	264	3	90.8	86.2
ST - 3	0.012617	422	8	88.5	82.8
ST - 4	0.121223	180	8	90	85.1
ST – 5	0.0051169	393	2	91.1	86.4
ST – 6	0.167489	435	6	91.0	86.2
ST – 7	0.037365	210	5	90.4	85.6
ST - 8	0.037365	210	5	90.4	85.6
ST – 9	0.087389	524	1	90.9	86.3
ST - 10	0.087389	524	1	90.1	86.3
ST - 11	0.103847	180	9	90.9	82.7
ST - 12	0.172988	70	7	88.4	84.9
ST - 13	0.091100	437	2	90	85.4

ST - 14	0.189440	826	2	90	85.4
ST – 15	**0.0419325**	**389**	**2**	**91.6**	**87.2**
ST - 16	0.049205	477	7	90.4	85.2
ST – 17	0.100035	255	3	90.7	86.0
ST – 18	0.0792121	921	2	90.4	86.0
ST – 19	0.152072	615	6	90.8	86.0
ST - 20	0.110342	526	6	90.6	85.5

Lampiran 5: Eksperimen Hasil Stacking menggunakan LogisticRegression

*) Hasil Hypeparameter Base Model

Model	Base Learner Param	Accuracy (%)	F1-macro (%)
BL - BS – 01*	C=37.554; γ=scale; kernel=poly	**91.48**	**86.81**
BL – BS - 02	C=73.2994; γ=scale; kernel=linear	89.43	84.78
BL - BS – 03	C=15.7019; γ=scale; kernel=poly	91.48	86.81
BL - BS – 04	C=15.7019; γ=scale; kernel=poly	89.43	84.78
BL - BS – 05	C=60.2115; γ=auto; kernel=poly	91.48	86.81
BL - BS – 06	C=2.15845; γ=auto; kernel=sigmoid	83.11	83.16
BL - BS – 07	C= 83.3443; γ=auto; kernel=rbf	90.90	86.42
BL - BS – 08	C=18.2825; γ=scale; kernel=linear	89.43	84.78
BL - BS – 09	C=30.5242; γ=auto; kernel=linear	89.43	84.78
BL - BS - 10	C= 43.2945; γ=scale; kernel=poly	91.48	86.81
BL - BR - 01	n_est=1000; min_split=2; max_depth=50	91.92	87.89
BL - BR - 02	n_est=800; min_split=2; max_depth=50	91.92	87.89
BL - BR – 03	n_est= 100; min_split=10; max_depth=20	91.92	88.15
BL - BR - 04	n_est=100; min_split=5; max_depth=20	92.34	88.81
BL - BR – 05	n_est= 100; min_split=5; max_depth=20	92.07	88.19
BL - BR - 06	n_est=300; min_split=2; max_depth=50	91.77	87.78

BL - BR – 07*	n_est= 800; min_split=10; max_depth=10	**92.07**	**88.45**
BL - BR - 08	n_est=500; min_split=10; max_depth=50	91.92	87.97
BL - BR - 09	n_est= 1000; min_split=5; max_depth=50	92.07	88.03
BL - BR - 10	n_est=1000; min_split=2; max_depth=50	91.92	87.89
BL - BX – 01	n_est=100; max_depth=5; lr=0.05	91.04	85.82
BL - BX – 02	n_est=300; max_depth=10; lr=0.01	90.75	86.43
BL - BX – 03	n_est=500; max_depth=7; lr=0.05	91.19	85.91
BL - BX – 04	n_est=800; max_depth=7; lr=0.10	91.19	85.35
BL - BX – 05	n_est=100; max_depth=7; lr=0.05	89.72	87.03
BL - BX – 06*	n_est=300; max_depth=5; lr=0.10	**91.04**	**87.03**
BL - BX – 07	n_est=100; max_depth=7; lr=0.1	88.55	83.14
BL - BX – 08	n_est=800; max_depth=3; lr=0.05	91.19	86.94
BL - BX – 09	n_est=100; max_depth=5; lr=0.01	90.11	85.57
BL - BX - 10	n_est=300; max_depth=7; lr=0.05	91.04	86.88

*) Hasil Stacking Logistic Regression

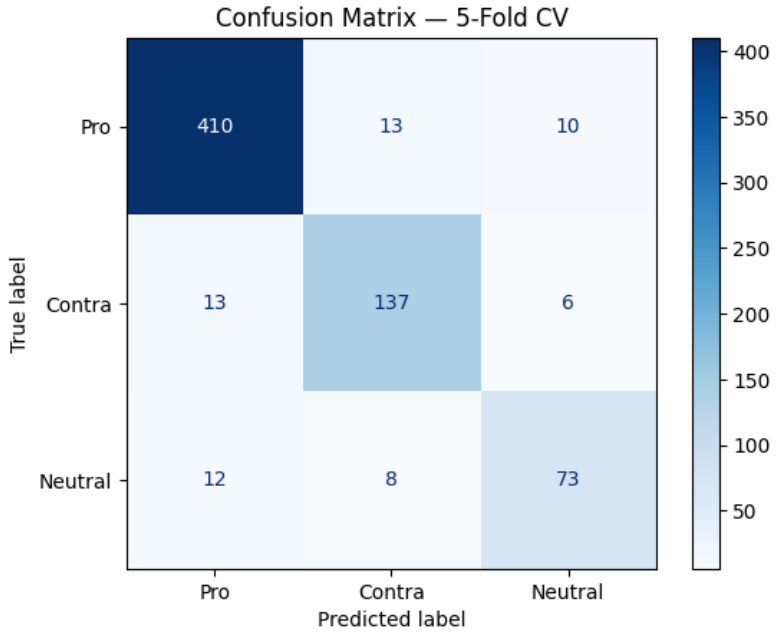
Model	**Hyperparameter**			**Accuracy (%)**	**F1-Macro (%)**
	C	**Penalty**	**Solver**		
ST - 1	3.7554	l1	Liblinear	89.29	83.86
ST – 2	7.32994	l1	Liblinear	88.85	83.36
ST - 3	1.57019	l1	Liblionear	89.43	84.48
ST - 4	**0.590836**	**l1**	**Liblinear**	**90.90**	**87.05**
ST – 5	6.02115	l1	liblinear	90.31	85.62

ST – 6	0.215845	l1	saga	89.43	86.39
ST – 7	8.33443	l1	Saga	89.43	84.61
ST - 8	1.82825	l1	Liblinear	90.31	84.42
ST – 9	3.05242	l1	Liblinear	89.29	85.77
ST - 10	4.32945	l1	1liblinear	89.43	83.86
ST - 11	6.12853	l1	Saga	9031	84.61
ST - 12	2.93145	l1	Liblinear	90.31	85.77
ST - 13	4.5707	l1	saga	89.87	85.41
ST - 14	2.00674	l1	Saga	90.02	85.59
ST – 15	5.93415	l1	Liblinear	88.85	83.36
ST - 16	6.08545	l1	Liblinear	88.85	83.25
ST – 17	0.660516	l1	Liblinear	90.90	86.93
ST – 18	9.66632	l1	Saga	89.43	85.61
ST – 19	3.05614	l1	Saga	89.87	85.41
ST - 20	6.85233	l1	Saga	90.02	85.55

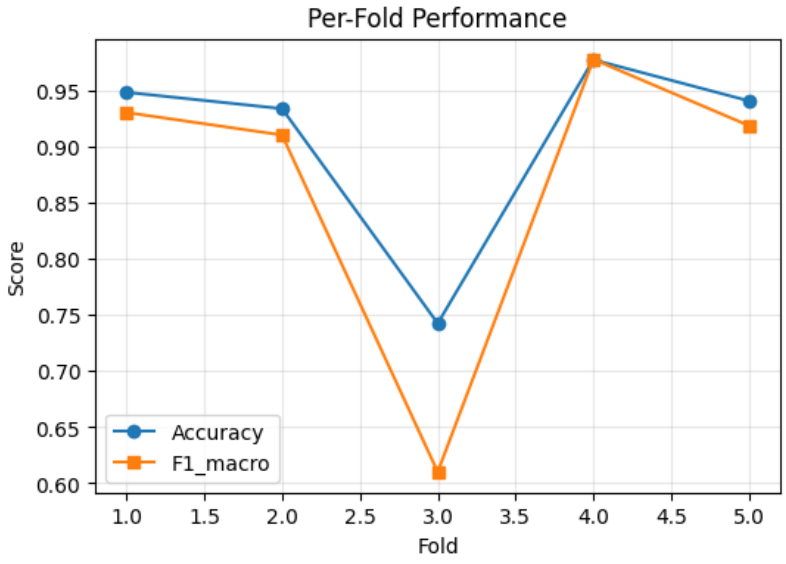
*) Hasil Eksperimen Stacking Hyperparameter Terbaik

Model	**Hyperparameter Terbaik**	**F1-macro (%)**	**Accuracy (%)**
BL – BS	kernel: poly; C: 37.554; γ: scale	86.8	91.5
BL – BR	n_estimators: 800; min_samples_split: 10; max_depth: 10	88.5	92.4
BL – BX	n_estimators: 300; max_depth: 5; learning rate: 0.1	87.0	91.0
ST	penalty: l2; solver: liblinear; C: 0.590836	87.1	90.6

Kelas	Performa		
	Precision (%)	Recall (%)	F1-score (%)
Pro	94.6	94.8	94.6
Contra	87.2	88.0	87.2
Neutral	80.4	77.6	77.2
Accuracy		91.1	



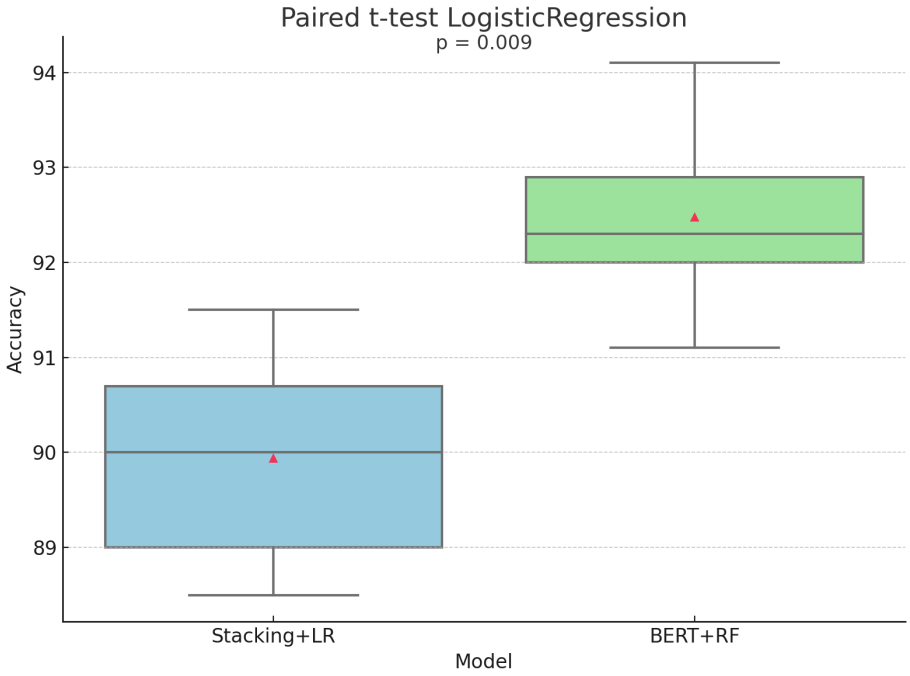
Gambar 6. Confussion Matrix Hasil Stacking Hyperparameter Terbaik



Gambar Lampiran 8. Performance per-fold Stacking Hyperparameter Terbaik

*) Hasil Eksperimen Stacking Uji T-Paired Test

Model	Accuracy (%)	Std Dev	Paired t-test (t, p)
BERT+RF	92.36	0.080	
Stacking+LR	90.90	0.084	t = -2.235; p = 0.089



Gambar Lampiran 9. Hasil Paired t-test dengan LogisticRegression

BIODATA PENULIS

Penulis lahir di Magetan, 1 Juli 1997 dan merupakan anak pertama, serta satu-satunya di dalam keluarga. Penulis telah menempuh pendidikan formal di TK Aisyah II Magetan, kemudian melanjutkan di SD Muhammadiyah 1 Magetan, SMP Negeri 1 Magetan, dan SMA Negeri 1 Magetan. Setelah menyelesaikan Pendidikan formal pada tingkat menengah atas pada tahun 2016, penulis melanjutkan pendidikan di D3 Tenik Informatika Universitas Sebelas Maret, meskipun harus terhenti di tahun 2017 karena tercantum kembali sebagai mahasiswa baru di Universitas Negeri Malang pada Program Studi Pendidikan Teknik Informatika, dan lulus pada tahun 2022. Pada saat menjadi mahasiswa, penulis memiliki ketertarikan dalam dunia Graphic Design, dan Games, serta memiliki riwayat dalam beberapa organisasi seperti Badan Eksekutif Mahasiswa Fakultas, Badan Eksekutif Mahasiswa Universitas, Koperasi Mahasiswa, dan memiliki beberapa pengalaman terkait narasumber bidang design dan kepemimpinan, maupun sebagai asisten dosen dalam pembuatan games yang berfokus pada Virtual Reality.

Setelah lulus, Penulis bekerja di Perusahaan Digital Marketing PT. Sahada Laku Utama, kemudian memutuskan untuk melanjutkan Magister di Program Studi Sistem Informasi, pada tahun 2024 pada semester genaap dengan NRP 6026232010. Dalam studinya, penulis menyadari memiliki ketertarikan dalam bidang rekayasa data, sehingga dalam menempuh studi di Magister penulis lebih banyak memahami dan mempelajari data, terlebih dalam topik machine learning. Sehingga, dalam kepenulisan penulis memilih topik berikut sebagai topik dalam penelitian oleh penulis. Di luar lingkungan kampus, penulis merupakan professional dalam bidang graphic designer, dan web developer yang aktif sejak tahun 2022. Apabila ingin berdiskusi lebih lanjut, penulis dapat dihubungi melalui email, sebagai berikut: elisanurmufida@gmai.com.