STUDI PERBANDINGAN METODE EKSTRAKSI FITUR UNTUK *TOPIC MODELING* BERBASIS ASPEK DAN SENTIMEN ANALISIS PADA ULASAN PRODUK E-COMMERCE

THESIS

Oleh:
FAHRENDRA KHOIRUL IHTADA
NIM. 230605210033

PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025

STUDI PERBANDINGAN METODE EKSTRAKSI FITUR UNTUK *TOPIC MODELING* BERBASIS ASPEK DAN SENTIMEN ANALISIS PADA ULASAN PRODUK E-COMMERCE

THESIS

Diajukan kepada:
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Magister Komputer (M.Kom)

Oleh:
FAHRENDRA KHOIRUL IHTADA
NIM. 230605210033

PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025

STUDI PERBANDINGAN METODE EKSTRAKSI FITUR UNTUK *TOPIC MODELING* BERBASIS ASPEK DAN SENTIMEN ANALISIS PADA ULASAN PRODUK E-COMMERCE

THESIS

Oleh:
FAHRENDRA KHOIRUL IHTADA
NIM. 230605210033

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 4 Juni 2025

Pembimbing I,	Pembimbing II,
Dr. Cahyo Crysdian	Dr. Zainal Abidin, M.Kom
NIP. 19740424 200901 1 008	NIP. 19760613 200501 1 004

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

Dr. Cahyo Crysdian
NIP. 19740424 200901 1 008

STUDI PERBANDINGAN METODE EKSTRAKSI FITUR UNTUK *TOPIC MODELING* BERBASIS ASPEK DAN SENTIMEN ANALISIS PADA ULASAN PRODUK E-COMMERCE

THESIS

Oleh:
FAHRENDRA KHOIRUL IHTADA
NIM. 230605210033

Telah Dipertahankan di Depan Dewan Penguji Thesis
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Magister Komputer (M.Kom)
Tanggal: 4 Juni 2025

Susunan Dewan Penguji

Penguji I	: Dr. Irwan Budi Santoso M. Kom NIP. 19770103 201101 1 004	( )
Penguji II	: Dr. M. Amin Hariyadi, MT NIP. 19670018 200501 1 001	( )
Pembimbing I	: Dr. Cahyo Crysdian NIP. 19740424 200901 1 008	( )
Pembimbing II	: Dr. Zainal Abidin, M.Kom NIP. 19760613 200501 1 004	( )

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

Dr. Cahyo Crysdian
NIP. 19740424 200901 1 008

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	: Fahrendra Khoirul Ihtada
NIM	: 230605210033
Program Studi	: Magister Informatika
Fakultas	: Sains dan Teknologi

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

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

Malang, 4 Juni 2025
Yang membuat pernyataan,

FAHRENDRA KHOIRUL IHTADA
NIM. 230605210033

MOTTO

“Jangan gagal hanya karena tidak tahu. Jadi, ayo cari tahu”

“Step by step to the next step”

HALAMAN PERSEMBAHAN

Dengan segenap rasa syukur *alhamdulillah rabbil 'alamin*, saya persembahkan hasil karya Tesis ini kepada:

1. Kedua orang tua dan saudara saya, yakni Bapak Harun, Ibu Lilis, Mbak Anindya, dan Adik Aliya, yang telah membimbing, mendoakan, menyayangi, dan mendukung dalam setiap langkah yang saya lalui. Terimakasih banyak atas semuanya yang telah diberikan sehingga bisa menyelesaikan tesis ini.
2. Dosen saya, Bapak Dr. Cahyo Crysdian , yang telah membimbing saya sejak skripsi hingga tesis ini. Terima kasih atas ilmu dan wawasan yang membangun membentuk pola pikir saya, baik dalam akademik maupun kehidupan.
3. Teman-teman seperjuangan Fast-Track yang telah menjadi rekan belajar dan saling mendukung sepanjang perjalanan ini.
4. Sahabat-sahabat masa kuliahku Rizha, Yoga, Zul, dan Nabila, terima kasih telah mewarnai perjalanan ini dengan canda tawa dan semangat. Terima kasih juga atas segala dukungan, bantuan, serta "*roasting*"-annya saat aku mulai malas-malasan, itu semua sangat berarti bagiku.
5. Beasiswa Riset BAZNAS 2024 yang telah memberikan dukungan penuh dalam penyelesaian penelitian ini.

Semoga Allah senantiasa membimbing kita ke jalan yang lurus dan memudahkan setiap langkah dalam mewujudkan mimpi-mimpi kita. Aamiin ya Rabbal 'Alamiin.

KATA PENGANTAR

Assalamualaikum wr wb.

Puji syukur penulis panjatkan kehadirat Allah Swt. atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan thesis ini yang berjudul "Evaluasi Ulasan Produk E-Commerce Berbasis Aspek *Topic Modeling* Dan *Sentiment Analysis* Menggunakan *Deep Neural Network*" dengan tepat waktu.

Dalam penulisan skripsi ini, penulis menyadari banyak pihak yang terlibat baik dalam proses membimbing penulisan dan juga memberikan semangat dan dukungan moral atau material. Untuk itu, penulis ingin menyampaikan terima kasih kepada:

1. Prof. Dr. H.M. Zainuddin, MA., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Hariani, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Cahyo Crysdian, selaku Ketua Program Studi Magister Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang, serta Dosen Pembimbing I dalam tesis ini dari awal hingga akhir.
4. Dr. Zainal Abidin, M.Kom selaku Dosen Pembimbing II yang telah memberikan banyak dukungan dan bimbingan dalam penulisan tesis ini.
5. Dr. Irwan Budi Santoso M.Kom selaku Penguji I yang telah memberikan banyak saran untuk menyelesaikan tesis ini.
6. Dr. M. Amin Hariyadi, M.T selaku Penguji II yang telah memberikan banyak saran untuk menyelesaikan tesis ini.

7. Seluruh Dosen dan Jajaran Staf Program Studi Magister Informatika yang telah memberikan banyak bantuan dalam tesis ini.
8. Kedua orang tua, Bapak Harun Munif dan Ibu Lilis Sarmini serta Kakak Anindya dan Adik Aliya yang telah memberikan dukungan kepada penulis untuk menyelesaikan tesis ini.
9. Teman-teman Angkatan 2023 Magister Informatika yang telah memberikan banyak bantuan baik material maupun dukungan intelektual, semangat, serta motivasi dalam menyelesaikan penyusunan tesis ini.
10. Semua pihak yang tidak dapat disebutkan satu per satu, termasuk responden yang telah memberikan kontribusi, saran, dan dukungan dalam perjalanan penulisan tesis ini.

Akhir kata, penulis mengakui bahwa penulisan pada tesis ini masih banyak kekurangan.

Wassalamualaikum Warahmatullahi Wabarakatuh.

Malang, 4 Juni 2025

Penulis

DAFTAR ISI

PERNYATAAN KEASLIAN TULISAN ..... **v**
MOTTO ..... **vi**
HALAMAN PERSEMBAHAN ..... **vii**
KATA PENGANTAR ..... **viii**
HALAMAN PERSETUJUAN ..... **x**
DAFTAR ISI ..... **x**
DAFTAR GAMBAR ..... **xii**
DAFTAR TABEL ..... **xiv**
ABSTRAK ..... **xv**
ABSTRACT ..... **xvi**
الملخص ..... **xvii**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Pernyataan Masalah 6
1.3 Tujuan Penelitian 6
1.4 Batasan Masalah 7
1.5 Manfaat Penelitian 7
BAB II STUDI PUSTAKA **8**
2.1 Pentingnya Ulasan *E-commerce* 8
2.2 Tantangan Klasifikasi Aspek dalam Ulasan E-commerce 10
2.3 Kerangka Teori 13
BAB III DESAIN DAN IMPLEMENTASI **20**
3.1 Kerangka Konsep 20
3.2 Prosedur Penelitian 21
3.3 Akuisisi Data 21
3.4 Desain Sistem 24
3.5 *Data Cleaning* 25
3.5.1 *Case Folding* 25
3.5.2 *Punctuation Removal* 26
3.5.3 *Tokenization* 26
3.5.4 *Stopword Removal* 27
3.5.5 *Stemming* 27
3.6 Term Frequency – Inverse Class Frequency (TF-ICF) 28
3.7 Word2Vec 30
3.8 IndoBERT 32
3.9 Neural Network 36
3.9.1 Neural Network pada *Aspect Topic Modeling* 37
3.9.2 Neural Network pada *Sentiment Analysis* 43
3.9.3 Proses *Training* 45
3.10 Eksperimen 51
3.11 Evaluasi 52
BAB IV MODEL NEURAL NETWORK BERBASIS TF-ICF **55**
4.1 Desain Sistem 55

4.2 Ekstraksi Fitur TF-ICF 56
4.3 Topologi Neural Network pada Model TF-ICF-NN 58
4.3.1 Perhitungan Jumlah Node pada Hidden Layer 58
4.3.2 Topologi Neural Network berbasis TF-ICF untuk *Aspect Topic Modelling* 59
4.3.3 Topologi Neural Network berbasis TF-ICF untuk *Sentiment Analysis* 60
4.4 Hasil pada *Aspect Topic Modeling* TF-ICF-NN 61
4.5 Hasil pada *Sentiment Analysis* TF-ICF-NN 64
4.6 Kesimpulan Model TF-ICF-NN 65
BAB V MODEL NEURAL NETWORK BERBASIS WORD2VEC 67
5.1 Desain Sistem 67
5.2 Word2Vec Embedding 68
5.3 Topologi *Neural Network* pada Word2Vec-NN 70
5.3.1 Perhitungan Jumlah Node pada Hidden Layer 70
5.3.2 Topologi Neural Network berbasis Word2Vec untuk *Aspect Topic Modelling* 71
5.3.3 Topologi Neural Network berbasis Word2Vec untuk *Sentiment Analysis* 72
5.4 Hasil pada *Aspect Topic Modeling* Word2Vec-NN 73
5.5 Hasil pada *Sentiment Analysis* Word2Vec-NN 75
5.6 Kesimpulan Model Word2Vec-NN 77
BAB VI MODEL NEURAL NETWORK BERBASIS INDOBERT 79
6.1 Desain Sistem 79
6.2 IndoBERT Embedding 80
6.3 Topologi *Neural Network* pada IndoBERT-NN 82
6.3.1 Perhitungan Jumlah Node pada Hidden Layer 82
6.3.2 Topologi Neural Network berbasis IndoBERT untuk *Aspect Topic Modeling* 83
6.3.3 Topologi Neural Network berbasis IndoBERT untuk *Sentiment Analysis* 84
6.4 Hasil pada *Aspect Topic Modeling* IndoBERT-NN 86
6.5 Hasil pada *Sentiment Analysis* IndoBERT-NN 88
6.6 Kesimpulan Model IndoBERT-NN 90
BAB VII PEMBAHASAN 92
BAB VIII KESIMPULAN 105
7.1 Kesimpulan 105
7.2 Saran 106
DAFTAR PUSTAKA 108

DAFTAR GAMBAR

Gambar 1.1. Grafik proyeksi pengguna *e-commerce* Indonesia 1
Gambar 2.1. Kerangka Teori 14
Gambar 3.1. Kerangka Konsep 20
Gambar 3.2. Prosedur Penelitian 21
Gambar 3.3. Desain Sistem 24
Gambar 3.4. IndoBERT *Input* 33
Gambar 3.5. Kerangka IndoBERT 34
Gambar 3.6. IndoBERT output untuk klasifikasi 35
Gambar 3.7. Topologi desain neural network pada *aspect topic modeling* 38
Gambar 3.8. *Feedforward* neural network pada *aspect topic modeling* 39
Gambar 3.9. Topologi desain neural network pada *sentiment analysis* 44
Gambar 3.10. *Feedforward* neural network pada analisis sentimen 44
Gambar 3.11. *Backpropagation* neural network 46
Gambar 3.12. *Flowchart* proses *training* 51
Gambar 4.1. Desain Sistem Model TF-ICF-NN 55
Gambar 4.2 Topologi Neural Network dengan Input Vektor TF-ICF untuk *Aspect Topic Modeling* 59
Gambar 4.3 Topologi Neural Network dengan Input Vektor TF-ICF untuk *Sentiment Analysis* 60
Gambar 4.6. *Loss* pada *Training* Model TF-ICF-NN *Aspect Topic Modeling* 61
Gambar 4.7. *Confusion Matrix* pada Model TF-ICF-NN *Aspect Topic Modeling* 62
Gambar 4.8. *Loss* pada *Training* Model TF-ICF-NN *Sentiment Analysis* 64
Gambar 4.9. *Confusion Matrix* pada Model TF-ICF-NN *Sentiment Analysis* 64
Gambar 5.1. Desain Sistem Model Word2Vec-NN 67
Gambar 5.2 Topologi Neural Network dengan Input Vektor Word2Vec untuk *Aspect Topic Modeling* 71
Gambar 5.3 Topologi Neural Network dengan Input Vektor Word2Vec untuk *Sentiment Analysis* 72
Gambar 5.5. *Loss* pada *Training* Model Word2Vec-NN *Aspect Topic Modeling* 73
Gambar 5.6. *Confusion Matrix* pada Model Word2Vec-NN *Aspect Topic Modeling* 74
Gambar 5.7. *Loss* pada *Training* Model Word2Vec-NN *Sentiment Analysis* 75
Gambar 5.8. *Confusion Matrix* pada Model Word2Vec-NN *Sentiment Analysis* 76
Gambar 6.1. Desain Sistem Model IndoBERT-NN 79
Gambar 6.2 Topologi Neural Network dengan Input Vektor IndoBERT embedding untuk *Aspect Topic Modeling* 83
Gambar 6.3 Topologi Neural Network dengan Input Vektor IndoBERT untuk *Sentiment Analysis* 85
Gambar 6.4. *Loss* pada *Training* Model IndoBERT-NN *Aspect Topic Modeling* 86
Gambar 6.5. *Confusion Matrix* pada Model IndoBERT-NN *Aspect Topic Modeling* 86
Gambar 6.6. *Loss* pada *Training* Model IndoBERT-NN *Sentiment Analysis* 88
Gambar 6.7. *Confusion Matrix* pada Model IndoBERT-NN *Sentiment Analysis* 88
Gambar 7.1. Distribusi Ulasan Berdasarkan Kategori Aspek 92
Gambar 7.2. Distribusi Ulasan Berdasarkan Kategori Sentimen 93
Gambar 7.3. Perbandingan Performa Model pada *Aspect Topic Modeling* 95

Gambar 7.4. Perbandingan Hamming Loss Semua Model pada *Aspect Topic Modeling* .97
Gambar 7.5. Perbandingan Performa Model pada *Sentiment Analysis* 98
Gambar 7.6. *Confusion Matrix* Semua Model pada *Sentiment Analysis*.....100

DAFTAR TABEL

Tabel 3.1. Performa Setiap Model Kerangka Teori 14
Tabel 3.2. Aspek ulasan *e-commerce* 22
Tabel 3.3. Sampel data ulasan *e-commerce* pada kasus *aspect topic modeling* 22
Tabel 3.4. Sentimen ulasan *e-commerce* 23
Tabel 3. 5. Sampel data ulasan *e-commerce* pada kasus analisis sentimen 23
Tabel 4.1. Sampel Ulasan Sebelum dan Sesudah *Preprocessing* 56
Tabel 4.2. Hasil Ekstraksi Fitur Menggunakan TF-ICF 57
Tabel 4.3 Hasil Perhitungan Neuron pada Model TF-ICF-NN 58
Tabel 4.5. Metrik Performa pada Model TF-ICF-NN *Aspect Topic Modeling* 62
Tabel 4.6. Metrik Performa pada Model TF-ICF-NN *Sentiment Analysis* 65
Tabel 5.1. Hasil Embedding Menggunakan Word2Vec 69
Tabel 5.2. Hasil Perhitungan Neuron pada Model Word2Vec-NN 70
Tabel 5.3. Metrik Performa pada Model TF-ICF-NN *Aspect Topic Modeling* 74
Tabel 5.4. Metrik Performa pada Model Word2Vec-NN *Sentiment Analysis* 76
Tabel 6.1. Sampel Ulasan Sebelum dan Sesudah *Preprocessing* 80
Tabel 6.2. Hasil Embedding Menggunakan IndoBERT 81
Tabel 6.3 Hasil Perhitungan Neuron pada Model IndoBERT-NN 82
Tabel 6.5. Metrik Performa pada Model TF-ICF-NN *Aspect Topic Modeling* 87
Tabel 6.6. Metrik Performa pada Model IndoBERT-NN *Sentiment Analysis* 89

ABSTRAK

Ihtada, Fahrendra Khoirul. 2025. **Studi Perbandingan Metode Ekstraksi Fitur Untuk *Topic Modeling* Berbasis Aspek Dan Sentimen Analisis Pada Ulasan Produk *E-Commerce*.** Tesis. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Cahyo Crysdian, MCS (II) Dr. Zainal Abidin, M.Kom.

Kata kunci: *E-commerce, Aspect-Based Topic Modeling, Sentiment Analysis, IndoBERT, Deep Neural Network.*

Perkembangan *e-commerce* yang pesar dalam beberapa tahun terakhir menyebabkan peningkatan signifikan terhadap jumlah ulasan produk dari pelanggan. Ulasan ini tidak hanya sebatas *feedback* biasa, dalam sudut pandang pelanggan, ulasan pelanggan sebelumnya dijadikan referensi dalam pengambilan keputusan pembelian . Ulasan juga penting bagi penjual untuk mengevaluasi kualitas produk dan layanan pada toko di e-commerce mereka. Penelitian ini mengusulkan pendekatan *aspect-based topic modeling* dan *sentiment analysis* terhadap ulasan *e-commerce* menggunakan arsitektur *Deep Neural Network*. Tiga teknik representasi fitur dibandingkan, yaitu TF-ICF, Word2Vec, dan IndoBERT. Hasil eksperimen menunjukkan bahwa IndoBERT memberikan performa terbaik dalam kedua tugas yang diuji. Pada klasifikasi aspek, model ini mencapai *accuracy* 81,39%, *precision* 91,49%, *recall* 91,25%, *f1-score* 91,37%, serta *hamming loss* terendah sebesar 0,0819. Sementara itu, pada klasifikasi sentimen, diperoleh *f1-score* sebesar 0,6008. Secara keseluruhan, penelitian ini menunjukkan bahwa pemodelan berbasis aspek dan analisis sentimen pada ulasan *e-commerce* dapat memberikan pemahaman yang lebih dalam terhadap persepsi pelanggan, serta menegaskan keunggulan model berbasis representasi kontekstual seperti IndoBERT dalam mengekstraksi informasi bermakna dari teks berbahasa Indonesia.

ABSTRACT

Ihtada, Fahrendra Khoirul. 2025. **Comparative Study of Feature Extraction Methods for Aspect-Based Topic Modeling and Sentiment Analysis on E-Commerce Product Reviews.** Thesis. Department of Informatics Engineering Faculty of Science and Technology Maulana Malik Ibrahim State Islamic University Malang. Supervisor: (I) Dr. Cahyo Crysdian, MCS (II) Dr. Zainal Abidin, M.Kom.

The growth of e-commerce in recent years has led to a significant increase in the number of product reviews from customers. The product reviews are not just ordinary feedback, in the customer's point of view, previous customer reviews are used as a reference in making purchasing decisions. The product reviews are also important for sellers to evaluate the quality of products and services in their e-commerce stores. This research proposes an aspect-based topic modeling and sentiment analysis approach to e-commerce reviews using Deep Neural Network architecture. Three feature representation techniques are compared, namely TF-ICF, Word2Vec, and IndoBERT. The experimental results show that IndoBERT provides the best performance in both tasks tested. In aspect classification, the model achieved 81.39% accuracy, 91.49% precision, 91.25% recall, 91.37% f1-score, and the lowest hamming loss of 0.0819. Meanwhile, in sentiment classification, an f1-score of 0.6008 was obtained. Overall, this study shows that aspect-based modeling and sentiment analysis on e-commerce reviews can provide a deeper understanding of customer perceptions, and confirms the superiority of contextual representation-based models such as IndoBERT in extracting meaningful information from Indonesian text.

Keywords: *E-commerce, Aspect-Based Topic Modeling, Sentiment Analysis, IndoBERT, Deep Neural Network.*

الملخص

الاهتدى، فهرندرا خير . 2025. *دراسة مقارنة لطرق استخراج الملامح لنمذجة المواضيع القائمة على الجوانب وتحليل المشاعر في مراجعات منتجات التجارة الإلكترونية*. البحث الجامعي. قسم الهندسة المعلوماتية ، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: د. جاهيو كريسديان، الماجستير. المشرف الثاني: د. زين العابدين، الماجستير.

الكلمات الرئيسية:*.الشبكة العصبية العميقة ،IndoBERT ،التجارة الإلكترونية، نمذجة الموضوع المستندة إلى الجوانب، تحليل المشاعر*

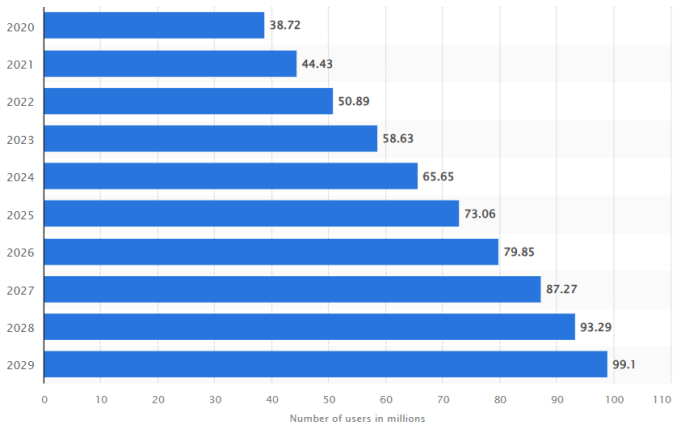
أدى نمو التجارة الإلكترونية في السنوات الأخيرة إلى زيادة كبيرة في عدد مراجعات المنتجات من العملاء .مراجعات المنتجات ليست مجرد تعليقات عادية، فمن وجهة نظر العميل، تُستخدم مراجعات العملاء السابقة كمرجع في اتخاذ قرارات الشراء .مراجعات المنتجات مهمة أيضًا للبائعين لتقييم جودة المنتجات والخدمات في متاجر التجارة الإلكترونية الخاصة بهم .يقترح هذا البحث نهجًا قائمًا على نمذجة الموضوعات TF-IDF وتحليل المشاعر لمراجعات التجارة الإلكترونية باستخدام بنية الشبكة العصبية العميقة .تتم مقارنة ثلاث تقنيات لتمثيل السمات، وهي يوفر أفضل أداء في كلتا المهمتين اللتين تم IndoBERT تُظهر النتائج التجريبية أن .IndoBERTو ،Word2Vec و ،ICF اختبارهما .في تصنيف الجوانب، حقق النموذج دقة 81.39 %ودقة 91.49 %ودقة 91.49 %واستدعاء 91.25 %و91.37 %من f1 0.6008 وأقل خسارة في التضمين بلغت 0.0819 .وفي الوقت نفسه، في تصنيف المشاعر، تم الحصول على درجة f1 درجة .تبلغ بشكل عام، تُظهر هذه الدراسة أن النمذجة القائمة على الجوانب وتحليل المشاعر على مراجعات التجارة الإلكترونية يمكن أن يوفر فهمًا في استخراج معلومات ذات مغزى من IndoBERT أعمق لتصورات العملاء، ويؤكد تفوق النماذج القائمة على التمثيل السياقي مثل .النص الإندونيسي

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam beberapa tahun terkahir, kegiatan jual beli yang popular saat ini adalah *e-commerce*. *E-commerce* sendiri adalah singkatan dari *electronic commerce*, yang merujuk pada pembelian dan penjualan produk atau layanan secara online melalui platform atau situs web. Beberapa contoh platform *e-commerce* di Indonesia diantaranya yaitu Tokopedia, Shopee, Bukalapak, Blibli, dan Lazada. *E-commerce* merubah perilaku konsumen dalam berbelanja dengan berbagai kemudahan yang ada. Dengan *e-commerce*, kegiatan belanja konsumen tidak lagi terbatas ruang dan waktu, karena konsumen dapat berbelanja dimana dan kapan saja.



Gambar 1.1. Grafik proyeksi pengguna *e-commerce* Indonesia

Kemudahan yang ditawarkan *e-commerce* membuat konsumen beralih dari jual konvensional ke jual beli *online*. Hal ini dapat dilihat dari peningkatan jumlah pengguna *e-commerce* tiap tahunnya. Menurut laporan Statista, jumlah pengguna di pasar *e-commerce* di Indonesia diperkirakan akan terus meningkat antara tahun

2024 dan 2029 dengan total 33,5 juta pengguna (+51,03%). Setelah peningkatan selama sembilan tahun berturut-turut, indikator ini diperkirakan akan mencapai 99,1 juta pengguna dan mencapai puncaknya pada tahun 2029. Khususnya, jumlah pengguna pasar *e-commerce* terus meningkat selama beberapa tahun terakhir (Statista Research Department, 2024).

Dengan trend *e-commerce* yang terus berkembang ini, jumlah ulasan juga semakin meningkat. Pertumbuhan ini menunjukkan bahwa pelanggan semakin mengandalkan ulasan untuk membuat keputusan pembelian. Menurut Chen *et al*., (2022), hampir 60% konsumen menelusuri ulasan setidaknya sekali seminggu. Biasanya, pelanggan akan menelusuri ulasan terlebih dahulu untuk memastikan kualitas produk dan kegunaannya (Zhao, 2020). Berdasarkan survei yang dilakukan oleh Mardianti *et al*., (2018), sebanyak 93,5% responden mengatakan bahwa mereka selalu menggunakan fitur ulasan sebelum melakukan pembelian produk. Dari hal tersebut dapat diambil kesimpulan bahwa ulasan berperan penting dalam lingkup *e-commerce*.

Banyaknya ulasan yang ada menyimpan informasi yang berharga, bagi pelanggan, mereka kesulian untuk memilah ulasan yang dapat membantu mereka untuk memilih produk yang mereka inginkan (Yanti *et al*., 2024). Disisi lain, penting juga bagi para pelaku bisnis di *e-commerce* untuk membaca ulasan sehingga mereka dapat merekomendasikan produk terbaik mereka atau menemukan keluhan dari pelanggan untuk meningkat pelayanan dan reputasi mereka (Gurav *et al*., 2023). Oleh karena itu, sangat penting mengolah ulasan ulasan yang banyak ini secara efisien. Pemilahan ulasan dengan cara manual akan memakan banyak waktu

dan tenaga. Tidak hanya itu, proses pemilahan secara manual sangat berpotensi akan kesalahan subjektif. Dengan demikian diperlukan sistem yang dapat mengklasifikasikan ulasan secara otomatis kedalam beberapa kategori ulasan untuk diekstrak dan mendapatkan informasi berharga didalamnya.

Beberapa penelitian telah mengeksplorasi klasifikasi ulasan *e-commerce* di Indonesia dengan menggunakan berbagai metode. Imron *et al*., (2023), meneliti IndoBERT untuk mendeteksi aspek aspek dalam ulasan dari Bukalapak, mendeteksi kategori kategori seperti layanan, pengemasan, kualitas, harga, dan akurasi dalam dataset yang terdiri dari 3114 ulasan. Hasilnya menunjukkan efektifitas IndoBERT mencapai akurasi yang signifikan, mereka juga membandingkan kinerjanya dengan dua model neural network seperti Convolutional Neural Network (CNN) dan Long Short-Term Memory (LSTM). Model CNN berhasil mengungguli LSTM dengan akurasi masing masing 94,86% dan 88,92%. Namun, peneliti menyampaikan tantangan adanya tantangan dengan ulasan yang terdiri dari Bahasa campuran dan butuhnya komputasi yang tinggi. Nasiri & Budi, (2019), menganalisis deteksi aspek ulasan dari mobile app menggunakan dataset 64113 ulasan dari Google Play Store dan Apple App Store. Mereka mengusulkan Gated Recurrent Unit (GRU), dengan model CNN dan mencapai akurasi 71, 3 % dnegan lebih dari separuh aspek mendapatkan *f*1-*score* melebihi 80%. Meski demikian, mereka menyoroti perlunya analisi lebih lanjut untuk meningkatkan identifikasi aspek individual.

Sistem klasifikasi untuk ulasan cenderung menggunakan *binary classification*, dimana output hanya dikategorikan kedalam satu aspek atau kelas

saja (Lunardi *et al*., 2016). Metode ini kurang efektif karena banyak ulasan yang diberikan oleh pelanggan tidak hanya membahas satu aspek saja, tetapi bisa saja melibatkan dua atau aspek sekaligus dalam satu ulasan (Deniz *et al*., 2022). Untuk mengatasi keterbatasan pendekatan klasifikasi biner yang tidak mampu menangkap multi-aspek dalam satu ulasan, penelitian ini mengusulkan *topic modeling* dengan beberapa aspek terfokus. Klasifikasi multilabel memungkinkan setiap ulasan dikategorikan kedalam beberapa aspek yang mencerminkan fakta bahwa ulasan sering kali membahas lebih dari satu topik atau aspek (Khlifi *et al*., 2023). Pendekatan ini memastikan bahwa setiap aspek yang disebutkan dalam ulasan diindentifikasi secara akurat. Penerapannya pada ulasan *e-commerce* di Indonesia masih banyak belum dieksplorasi terlepas dari pertumbuhan industri *e-commerce* di Indonesia.

Penelitian ini mengusulkan *topic modelling* berbasis aspek untuk ulasan *e-commerce*, dengan mengkategorikannya ke dalam tiga aspek utama: “produk”, “layanan pelanggan”, dan “pengiriman”. Aspek produk menggambarkan kepuasan pelanggan dengan kualitas, performa, dan kesesuaian produk dengan deskripsi yang diberikan. Aspek layanan pelanggan mencakup interaksi antara pelanggan dengan pelayanan yang diberikan, seperti keramahan, responsivitas, dan hal lainnya yang terkait dengan pelayanan pelanggan. Sementara itu, aspek pengiriman melibatkan aspek kecepatan pengiriman, kondisi barang saat diterima, dan ketepatan waktu pengiriman

Selain mengklasifikasikan ulasan ke dalam tiga aspek utama, penelitian ini juga menerapkan analisis sentimen pada masing-masing aspek yang teridentifikasi.

Setelah ulasan berhasil dikelompokkan berdasarkan aspek produk, pengiriman, atau layanan pelanggan, sistem akan menganalisis polaritas sentimen (positif, netral, negatif) yang terkandung dalam setiap bagian ulasan yang relevan dengan aspek tersebut. Pendekatan ini memungkinkan pelaku bisnis untuk tidak hanya mengetahui aspek apa saja yang dibahas konsumen, tetapi juga memahami bagaimana persepsi mereka terhadap masing-masing aspek tersebut. Namun, penerapan strategi *topic modeling* dengan multi-aspek dan polaritas (sentimen) masih jarang dieksplorasi pada ulasan *e-commerce* terlepas dari pertumbuhan industri *e-commerce* di Indonesia.

Untuk mencapai tujuan tersebut, penelitian ini membandingkan beberapa metode *feature extraction* seperti Term Frequency – Inverse Class Frequency (TF-ICF), Word2Vec, dan IndoBERT. Selanjutnya, penelitian ini menggunakan Neural Network (NN) dengan *deep architecture* untuk melakukan klasifikasi aspek dan sentimen. Neural Network (NN) dipilih sebagai *classifier* dalam penelitian ini karena kemampuannya menangkap pola kompleks. Studi oleh Chhabra *et al*., (2024) membuktikan keunggulan NN (khususnya arsitektur LSTM) dalam analisis sentimen ulasan *e-commerce*, mencapai akurasi 92.3% pada dataset Amazon, jauh melampaui model tradisional seperti SVM. Untuk kasus teks berbahasa Indonesia, penelitian oleh Koto *et al*., (2021) memperlihatkan bahwa fine-tuning IndoBERT dengan NN menghasilkan akurasi 88.5%, membuktikan kompatibilitas tinggi antara arsitektur transformer dan NN.

Dengan pendekatan ini, diharapkan model dapat mengidentifikasi aspek dan sentimen yang relevan dari setiap ulasan secara akurat. Dalam Al-Qur'an Surah Al-Qomar ayat 49, Allah SWT berfirman:

إِنَّا كُلَّ شَيْءٍ خَلَقْنَٰهُ بِقَدَرٍ

"Sesungguhnya Kami menciptakan segala sesuatu menurut ukuran" (QS. Al-Qomar 54:49).

Menurut tafsir Ibnu Katsir, surah Al-Qomar ayat 49 menjelaskan bahwa Allah SWT telah menentukan ukuran masing-masing untuk makhluk-Nya dan Dia memberi petunjuk kepada semua makhluk-Nya. Sebagaimana kandungan ayat tersebut, penelitian ini juga didasarkan pada parameter dan ukuran dalam mengklasifikasi ulasan ke dalam aspek dan sentimen. Pendekatan berbasis ukuran ini tidak hanya menjamin akurasi teknis, tetapi juga sejalan dengan nilai Islam tentang keadilan dalam mengevaluasi informasi.

1.2 Pernyataan Masalah

1. Faktor apa saja yang mempengaruhi kinerja model dalam melakukan klasifikasi ulasan produk *e-commerce*?
2. Pendekatan apa yang paling optimal untuk melakukan klasifikasi aspek dan sentiment pada ulasan produk *e-commerce*?

1.3 Tujuan Penelitian

1. Mengetahui faktor apa saja yang mempengaruhi kinerja model dalam melakukan klasifikasi ulasan produk *e-commerce*.

2. Mengetahui pendekatan yang paling optimal untuk melakukan klasifikasi aspek dan sentimen pada ulasan produk *e-commerce*.

1.4 Batasan Masalah

1. Data yang digunakan adalah ulasan-ulasan produk dari pelanggan platform *e-commerce* Tokopedia dari berbagai kategori produk. *Ground truth* dataset didapatkan dari *crowdsource* yang melibatkan responden umum.
2. Bahasa yang digunakan pada dataset penelitian ini hanya terbatas pada Bahasa Indonesia.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat dimanfaatkan oleh:

1. Industri *e-commerce*, model klasifikasi aspek ulasan yang efisien dapat membantu mereka dalam memahami kebutuhan pelanggan mengevaluasi produk, meningkatkan pelayanan pelanggan dan mengoptimalkan sistem pengiriman yang ada.
2. Pelaku bisnis *e-commerce*, mereka dapat mengidentifikasi serta menganalisis ulasan yang memerlukan perhatian khusus, seperti ulasan yang berkaitan dengan produk yang sering dibeli, *feedback* dalam layanan pelanggan, atau masalah pengiriman barang. Hal ini akan membantu mereka dalam meningkatkan kualitas layanan mereka dan meningkatkan reputasi mereka di mata pelanggan.

.

BAB II

STUDI PUSTAKA

2.1 Pentingnya Ulasan *E-commerce*

Dalam platfrom *e-commerce*, terdapat kolom ulasan pelanggan untuk digunakan pelanggan menyampaikan kepuasan dan pengalamannya setelah berbelanja. Ulasan merupakan salah satu bentuk rating atau penilaian dari *e-commerce* (Kaushik *et al*., 2018). Saat akan membeli suatu barang, konsumen memiliki kebiasaan untuk mencari informasi bagaimana bentuk fisik atau produk barangnya secara langsung, kemudian membandingkan harga dan kualitasnya dari toko satu ke toko lain. Namun, hal ini berbeda ketika konsumen berbelanja melalui *e-commerce*. Konsumen memiliki keterbatasan untuk merasakan langsung bentuk fisik dari produk ataupun bagaimana pelayanan yang diberikan oleh penjual (Yanti *et al*., 2024). Disinilah ulasan berperan bagi konsumen untuk mendapatkan informasi-informasi tersebut sebelum melakukan pembelian. Dengan ulasan, mereka dapat mengumpulkan informasi tentang barang tersebut dari pengalaman orang lain yang sudah membeli barang tersebut sebelumnya (Payne & Howes, 2013).

Ulasan pelanggan yang merupakan opini atau pengalaman konsumen terhadap produk atau layanan telah menjadi faktor krusial dalam proses pengambilan keputusan pembelian (Amien *et al*., 2021). Fitur ulasan yang disediakan *e-commerce* memungkinkan calon pembeli untuk memperoleh Gambaran yang lebih komprehensif mengenai kualitas produk dan kredibilitas

penjual. Berdasarkan survei oleh (Power Reviews, 2023). hampir seluruh pembeli *online* (99,75%), mengaku membaca ulasan setidaknya sesekali, dan mayoritas (91%) bahkan selalu secara teratur melakukannya. Angka-angka ini menggambarkan bahwa ulasan telah menjadi rujukan utama bagi konsumen sebelum memutuskan untuk bertransaksi. Dengan kata lain, ulasan tidak hanya sekedar komentar, tetapi telah berevolusi menjadi alat yang sangat efektif untuk mempengaruhi perilaku konsumen.

Tidak hanya berisi informasi berharga bagi konsumen, ulasan juga bermanfaat bagi pelaku bisnis *e-commerce*. Ulasan pelanggan merupakan bentuk dari *electronic word of mouth* (WOM) atau promosi dari mulut ke mulut yang merujuk pada konten buatan pengguna yang diposting disitus online (Almana & Mirza, 2013). Melalui ulasan, pelaku bisnis *e-commerce* dapat memperoleh *feedback* langsung dari konsumen mengenai produk atau layanan yang mereka tawarkan. Ulasan-ulasan ini memiliki informasi yang berharga tentang berbagai aspek dari *e-commerce*. Dalam ulasan tersebut, pelanggan sering kali memberikan *insight* tentang kualitas produk, layanan pelanggan, dan juga pengiriman dari pengalaman yang mereka dapatkan. Ulasan juga menyimpan sentimen yang mencerminkan tingkat kepuasan atau ketidakpuasan pelanggan (Rayhan *et al*., 2017). Analisis sentimen terhadap aspek-aspek spesifik dalam ulasan memungkinkan pelaku bisnis untuk mengidentifikasi secara tepat area yang perlu diperbaiki (misalnya: sentimen negatif dominan pada aspek "pengiriman") sekaligus memperkuat keunggulan yang sudah ada (sentimen positif pada aspek " produk"). Dengan mengekstrak informasi-informasi tersebut, pelaku bisnis dapat

mengetahui aspek mana saja yang perlu ditingkatkan dimasa depan sehingga dapat meningkatkan kepuasan pelanggan, daya saing bisnis, dan meningkatkan reputasi mereka dimata pelanggan (Noori, 2021).

2.2 Tantangan Klasifikasi Aspek dalam Ulasan E-commerce

Klasifikasi aspek ulasan bertujuan untuk mengidentifikasi bagian-bagian spesifik (aspek) yang dibahas dalam sebuah ulasan, seperti produk, layanan pelanggan, atau pengiriman. Dengan memahami aspek aspek ini, pelaku bisnis dapat menggali preferensi konsumen secara lebih mendalam, sehingga dapat digunakan untuk meningkatkan kualitas produk atau layanan, serta membantu pengambilan keputusan pembelian bagi konsumen (Ismet *et al*., 2022). Karena umumnya ulasan berbentuk teks, maka klasifikasi aspek ulasan termasuk dalam kategori tugas klasifikasi teks. Teknik ini merupakan salah satu bidang *natural language processing* (NLP) yang memiliki penerapan seperti analisis sentimen, klasifikasi dokumen, deteksi spam, hingga klasifikasi ulasan.

Namun, tantangan muncul karena ulasan pengguna sering kali memuat lebih dari satu topik atau aspek dalam satu entitas teks. Contohnya: "Produknya berkualitas, tetapi pengiriman sangat lambat dan layanan pelanggan tidak responsif.". Ulasan ini mencakup tiga aspek: produk, pengiriman, dan layanan pelanggan dimana kasus ini tidak bisa ditangani dengan klasifikasi aspek *single-label* karena hanya menghasilkan satu output (Sajid *et al*., 2023). Jika pendekatan klasifikasi hanya menghasilkan satu kategori utama, informasi penting lain yang turut dibahas berisiko diabaikan, sehingga menurunkan akurasi analisis pada ulasan (Mcilroy *et al*., 2015). Oleh karena itu, diperlukan metode yang mampu

mengidentifikasi semua aspek relevan dalam sebuah ulasan secara simultan. Tak hanya aspek, polaritas (sentimen) terhadap masing-masing aspek juga perlu ditelaah. Dalam contoh di atas, sentimen terhadap "produk" bersifat positif, sementara "pengiriman" dan "layanan pelanggan" bersifat negatif.

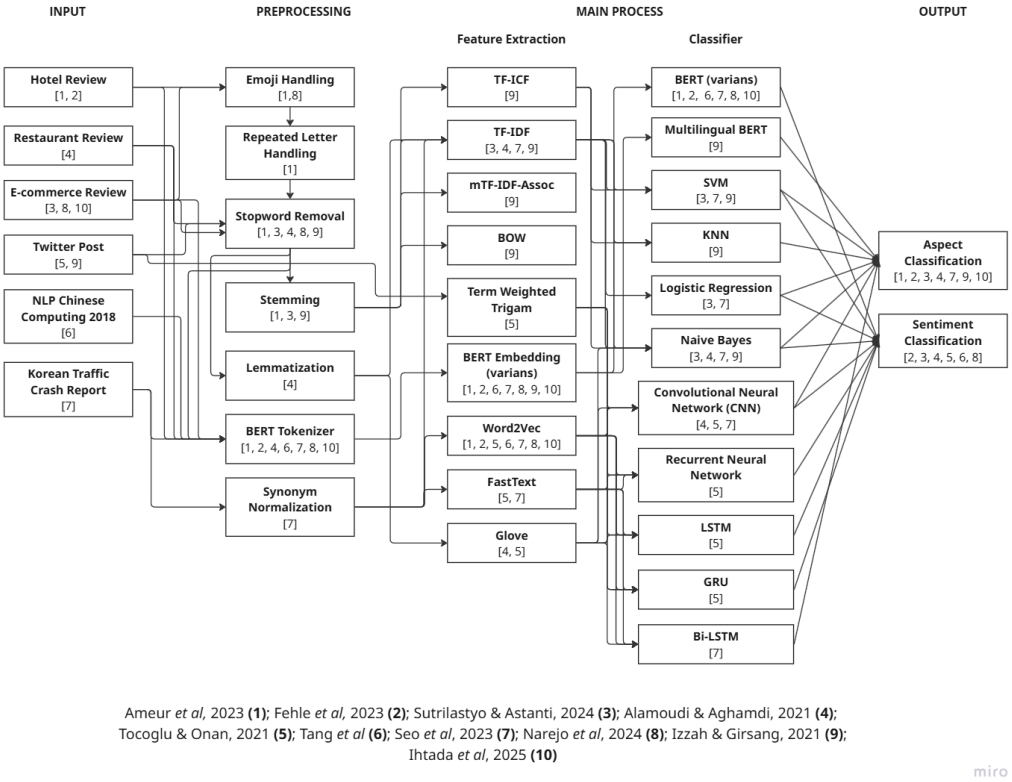
Beberapa penelitian telah dilakukan untuk mengatasi tantangan klasifikasi aspek ulasan dan mengeksplorasi potensi pendekatan multi-label. Kumar *et al*., (2019), melakukan penelitian yang berfokus pada pengembangan model klasifikasi sentimen multilabel berbasis aspek pada ulasan pengguna obat Abilify (Aripiprazole). Model ini bertujuan untuk mengklasifikasikan ulasan pengguna ke dalam satu kumpulan sentimen aspek atau *aspect term sentiments* (ATS) berdasarkan efektivitas, kemudahan penggunaan, dan kepuasan. Mereka menggunakan dataset yang didapat dari WebMD mencakup 1722 contoh ulasan pengguna, masing-masing dengan rating. Penelitian tersebut menerapkan TF-IDF untuk representasi fitur atau pembobotan kata disertai dengan penggunaan seleksi fitur Information Gain untuk memilih fitur khusus dan fitur sharing dari label. Persoalan klasifikasi multilabel diubah menjadi single-label dengan menggunakan tiga metode: Binary Relevance (BR), Classifier Chains (CC), dan Label Powerset (LP). Penelitian ini menggunakan metode klasifikasi dasar diantaranya Naïve Bayes (NB), Decision Tree (DT), dan Support Vector Machine (SVM). Model klasifikasi menunjukkan performa yang cukup baik, terutama model SVM yang paling unggul dibandingkan DT dan NB baik ketika menggunakan seleksi fitur ataupun tidak. Pada metode BR, SVM menghasilkan *hamming loss* yang rendah sebesar 27.0%, rata-rata nilai *f*1-*micro* mencapai hasil yang lebih baik sebesar

82.5% pada metode BR, sementara metode LP hanya memberikan performa yang lebih cepat dalam membangun dan menguji model. Fehle *et al*., (2023), melakukan penelitian mengenai analisis sentimen berbasis aspek atau *aspect-based sentiment analysis* (ABSA) multilabel pada ulasan hotel Jerman. Ulasan diklasifikasikan kedalam beberapa label pada tiap kategori: aspek (hotel, lokasi, makanan dan minuman, layanan, ruangan), atribut (umum, harga, ketenangan, kebersihan, gaya), serta polaritas (positif, negatif, netral). Dataset penelitian ini diperoleh dari TripAdvisor yang mencakup 1.512 ulasan pengguna mengenai hotel-hotel di Regensburg, Jerman. Tugas ABSA ditangani sebagai persoalan klasifikasi multi-label dengan menggunakan model *pre-trained* transfromer, khususnya model berbasis BERT. Model ini mencapai micro F1-score hingga 91% untuk klasifikasi kategori anotasi aspek dan 81% untuk anotasi aspek dan atribut pada dataset. Ririn dan Astanti, (2024) menerapkan teknik klasifikasi multilabel untuk mengkategorikan kebutuhan pelanggan selama fase konseptual pada pengembangan produk baru. Teknik klasifikasi multilabel didasarkan pada "kriteria desain produk" dan "sentimen ulasan". Kerangka kerja yang diusulkan diuji dengan menggunakan data set ulasan pakaian wanita dari situs *e-commerce* yang diunduh dari www.kaggle.com. Teknik klasifikasi multilabel didasarkan pada kriteria desain produk dan sentimen ulasan. Pendekatan klasifikasi multilabel pada penelitian ini memanfaatkan beberapa metode machine learning: Linear Regression (LR), Support Vector Machine (SVM), dan Complement Naïve Bayes (Complement NB). Penelitian ini juga menggunakan Synthetic Minority Oversampling Technique (SMOTE) untuk mengatasi ketidakseimbangan kelas dalam set data.

Hasil menunjukkan menunjukkan bahwa model yang diusulkan dapat mengkategorikan ulasan pelanggan. Model LR memberikan performa yang unggul dibandingkan SVM dan Complement NB dengan nilai *accuracy* 64% pada prediksi kriteria desain produk dan *accuracy* 78% untuk prediksi sentimen. Penelitian lainnya dilakukan oleh (Ameur *et al*., 2023) berfokus pada pengembangan model deteksi kategori aspek atau *aspect category detection* (ACD) multilabel untuk ulasan hotel Arab. Model ini bertujuan untuk mengklasifikasikan ulasan pengguna ke dalam beberapa kelas aspek: hotel, ruangan, fasilitas kamar, fasilitas umum, layanan, lokasi, makanan dan minuman. Dataset ini terdiri dari 1.839 contoh pelatihan dan 452 contoh pengujian ulasan hotel Arab dari dataset Arab SemEval-2016. Setiap ulasan dianotasi dengan aspek, atribut, dan polaritas. Pendekatan yang diusulkan memanfaatkan model AraBERT untuk ekstraksi fitur dan menggabungkannya dengan Dynamically Weighted Loss Function (DWLF) untuk menangani distribusi data yang tidak seimbang. Model mencapai f1-*score* sebesar 67,3%, mengungguli metode-metode *state-of-the-art* untuk deteksi aspek pada dataset ulasan hotel Arab.

2.3 Kerangka Teori

Bagian ini membahas berbagai teori dari penelitian terdahulu untuk mengeksplorasi jenis-jenis teks yang digunakan sebagai input. Selain itu, dibahas pula teknik-teknik *preprocessing* yang diterapkan untuk mempersiapkan data, metode ekstraksi fitur yang digunakan, serta berbagai pendekatan *machine learning* hingga *deep learning* beserta hasil yang diperoleh. Kerangka teori untuk klasifikasi teks dapat dilihat pada Gambar 2.1.



Gambar 2.1. Kerangka Teori

Tabel 3.1. Performa Setiap Model Kerangka Teori

No	Metode	Peneliti	Precision	Recall	F1-score
1	AraBERT	Ameur *et al*. (2023)	-	-	67.30%
2	German BERT (deepset-gbert-large)	Fehle *et al*. (2023)	-	-	91.00%
3	TF-IDF+ Logistic Regression	Sutrilastyo & Astanti (2023	61.00%	63.00%	62.00%
4	ALBERT	Alamoudi & Alghamdi (2021)	98.30%	98.30%	98.30%
5	FastText trigram	Onan & Tocoglu (2021)	-	-	-
6	BERT-MSAUC	Tang *et al*. (2020)	-	-	62.00%
7	BERT	Seo *et al*. (2023)	-	-	98.19%
8	EEBERT	Narejo *et al*. (2024)	99.21%	99.21%	99.21%
9	mTF-IDF-Assoc + SVM	Izzah & Girsang (2021)	-	-	-
10	IndoBERT + LSTM	Ihtada *et al.* (2025)	96.20%	95.20%	95.70%

Penelitian oleh Ameur *et al.* (2023) mengusulkan pendekatan berbasis AraBERT untuk deteksi kategori aspek (ACD) pada ulasan hotel berbahasa Arab. Studi ini ditujukkan untuk mengatasi kompleksitas morfologi bahasa Arab dan ketidakseimbangan data pada dataset SemEval-2016. Menanggapi masalah-masalah tersebut, studi ini menerapkan teknik preprocessing khusus, diantaranya normalisasi teks, *handling* emoji, *stemming* dan Dynamic Weighted Loss Function (DWLF) untuk menyeimbangkan kelas minor. Hasilnya, model mencapai *f*1-*score* 67.3%, mengungguli metode sebelumnya seperti IndyLSTM (58.05%).

Fehle *et al*. (2023) juga melakukan penelitian serupa dengan dataset ulasan hotel Jerman dari TripAdvisor dengan 5000 kalimat teranotasi. Sedikit berbeda dengan penelitian oleh Ameur *et al*. (2023), studi ini berfokus untuk mengeksplorasi ulasan bahasa Jerman dengan ABSA dalam klasifikasi multilabel. Beberapa model dari G-BERT, varian BERT untuk dataset berbahasa Jerman, diterapkan untuk menyelesaikan kasus ini. Didapatkan hasil bahwa model deepset-gbert-large mengungguli model G-BERT yang lain dengan *f*1-*score* 91.00% untuk klasifikasi kategori aspek dan *f*1-*score* 81.00% untuk gabungan kategori aspek dan polaritas (sentimen).

Sutrilastyo & Astanti (2023) melakukan penelitian untuk mengkategorikan kebutuhan pelanggan dalam pengembangan produk baru menggunakan analisis sentimen dan klasifikasi multi-label berbasis *supervised learning*. Fokus utama penelitian adalah pada tahap konseptual pengembangan produk, dimana ulasan pelanggan dari *e-commerce* dianalisis dengan dua label, yaitu "kriteria desain produk" (*aesthetic*, *usefulness*, *performance and quality*, *extrinsic*) dan "sentimen

ulasan" (positif, netral, negatif). Penelitian ini menerapkan TF-IDF untuk mengekstraksi fitur-fitur dari ulasan. Dari beberapa model klasifikasi yang digunakan, Logistic Regression memberikan performa terbaik dengan *precision* 61%, *recall* 63%, dan *f*1-*score* 62% mengungguli performa SVM dan Naive Bayes. Alamoudi dan Alghamdi (2021) menganalisis ulasan restoran dari platform Yelp dengan menggunakan teknik analisis sentimen dan *aspect-based sentiment analysis* (ABSA). Tujuannya adalah membandingkan tiga pendekatan klasifikasi sentimen, yaitu *machine learning*, *deep learning*, dan *transfer learning*. serta mengusulkan metode baru berbasis *semantic similarity* untuk ekstraksi aspek secara *unsupervised*. Hasilnya menunjukkan bahwa model ALBERT mencapai performa tertinggi dengan nilai *precision*, *recall*, dan *f*1-*score* sebesar 98,30% untuk klasifikasi biner, sementara metode ekstraksi aspek yang diusulkan mencapai akurasi 83,04%.

Onan dan Tocoglu (2021) mengusulkan kerangka kerja berbasis *deep learning* untuk mengidentifikasi sarkasme dalam teks, khususnya pada platform media sosial. Metode yang digunakan menggabungkan *neural language model* berbobot term (menggunakan skema pembobotan Inverse Gravity Moment atau IGM) dengan arsitektur Stacked Bidirectional LSTM. Penelitian ini mengevaluasi berbagai model *word embedding*, diantaranya word2vec, fastText, GloVe, dan fungsi pembobotan term, serta model *deep learning*. Hasil tertinggi didapatkan oleh model fastText dengan pembobotan trigam dengan akurasi mencapai 95.30%.

Tang *et al*. (2020) mengusulkan model BERT-MSAUC (BERT for Multi-Label Sentiment Analysis in Unbalanced Code-Switching Text) untuk menganalisis

sentimen multi-label pada teks *code-switching*. Mereka memodifikasi model BERT untuk mengatasi tiga masalah utama, yaitu deteksi emosi dalam teks multi-bahasa (terutama Cina-Inggris), klasifikasi multi-label di mana beberapa emosi dapat muncul bersamaan, dan menangani distribusi data emosi yang tidak seimbang. Eksperimen pada dataset NLPCC 2018 menunjukkan bahwa BERT-MSAUC mencapai performa lebih baik dibandingkan model sebelumnya, dengan *f*1-*score* sebesar 62%, terutama dalam mendeteksi emosi yang memiliki sampel lebih sedikit.

Seo *et al*. (2023) mengusulkan pendekatan klasifikasi teks berbasis model BERT untuk menganalisis deskripsi kecelakaan lalu lintas yang tidak terstruktur. Penelitian ini berfokus pada tiga tugas klasifikasi utama: (1) identifikasi manuver kendaraan sebelum kecelakaan, (2) lokasi tabrakan pada kendaraan, dan (3) penentuan pelaku kecelakaan. Beberapa model *machine learning* dan *deep learning* diimplementasikan untuk menangani klasifikasi pada studi ini, seperti Logistic Regression, SVM, CNN, hingga BERT. Didapatkan hasil bahwa model BERT menunjukkan kinerja yang superior dengan *f*1-*score* 98.19%, lebih tinggi dibandingkan model lainnya.

Penelitian oleh Narejo *et al*. (2024) memperkenalkan EEBERT (Emoji-Enhanced BERT), sebuah model berbasis arsitektur BERT yang diperluas dengan emoji *embedding* untuk analisis sentimen pada ulasan produk Amazon. Model ini mengatasi keterbatasan analisis sentimen tradisional dengan mengintegrasikan emoji sebagai fitur penting dalam memahami konteks emosional dari teks. Evaluasi menunjukkan bahwa model ini mencapai performa yang tinggi, dengan *precision*,

recall, dan *f*1-*core* mencapai 99.21%, terutama untuk kategori sentimen seperti "Dissatisfaction" dan "Side Effect".

Studi oleh Izzah & Girsang (2021) mengusulkan mTF-IDF-Assoc, sebuah skema pembobotan istilah (*term weighting*) yang dimodifikasi untuk klasifikasi teks pada dataset berita pendek dari Twitter. Metode ini menggabungkan konsep Document Length (DL) untuk menormalisasi frekuensi istilah, Inverse Document Frequency (IDF), dan nilai asosiasi (Assoc) berbasis *confidence* untuk mengukur kekuatan istilah dalam setiap kelas. Metode ini dibandingkan dengan beberapa skema pembobotan lain, termasuk TF-ICF (Inverse Class Frequency) yang juga merupakan metode *supervised* dan diimplementasikan sebagai salah satu *baseline* untuk mengevaluasi kinerja mTF-IDF-Assoc. Hasil eksperimen menunjukkan bahwa mTF-IDF-Assoc yang dikombinasikan dengan SVM mencapai akurasi tertinggi 82.32%, mengungguli TF-ICF dan skema pembobotan lainnya seperti TF-IDF dan TF-Assoc.

Ihtada *et al*. (2025) mengusulkan pendekatan klasifikasi multi-label untuk ulasan *e-commerce* dalam bahasa Indonesia menggunakan model IndoBERT. Fokus penelitian adalah mengkategorikan ulasan ke dalam tiga aspek utama: produk, layanan pelanggan, dan pengiriman. Penelitian ini membandingkan tiga arsitektur model, yaitu end-to-end IndoBERT, IndoBERT-CNN, dan IndoBERT-LSTM, untuk menentukan performa terbaik dalam klasifikasi multi-label. Penelitian ini mengkategorikan ulasan ke dalam tiga aspek (produk, layanan pelanggan, dan pengiriman) dan mengevaluasi tiga model: end-to-end IndoBERT, IndoBERT-CNN, dan IndoBERT-LSTM. Analisis mendalam terhadap *threshold*

(0.3–0.99) mengungkap bahwa nilai 0.6 memberikan keseimbangan terbaik antara presisi dan recall, dengan IndoBERT-LSTM mencapai *f*1-*score* tertinggi sebesar 95.7%.

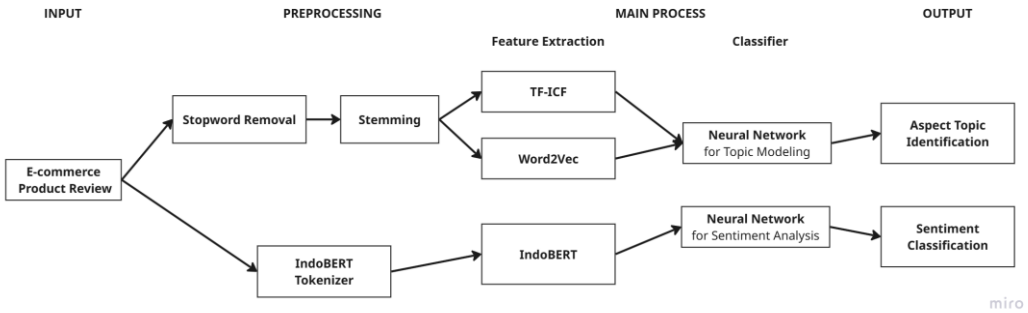
Berdasarkan tinjauan literatur di atas, dapat disimpulkan bahwa pendekatan *deep learning* berbasis transformer seperti BERT dan variannya (misalnya IndoBERT, AraBERT, atau G-BERT) telah terbukti efektif dalam tugas klasifikasi teks pada ulasan produk *e-commerce*. Beberapa penelitian menunjukkan bahwa teknik ekstraksi fitur berbasis *pretrained language model* (seperti Word2Vec, IndoBERT) cenderung mengungguli metode tradisional seperti TF-IDF atau TF-ICF, terutama dalam menangani kompleksitas linguistik dan ketidakseimbangan data. Namun, optimalisasi *feature extraction* dan pemilihan *classifier* (seperti neural network, LSTM, atau CNN) tetap memegang peranan krusial dalam meningkatkan performa model.

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Kerangka Konsep

Kerangka konsep pada penelitian ini akan digambarkan pada Gambar 3.1. Tahapan terdiridari *input*, *preprocessing, main process*, dan *output*.



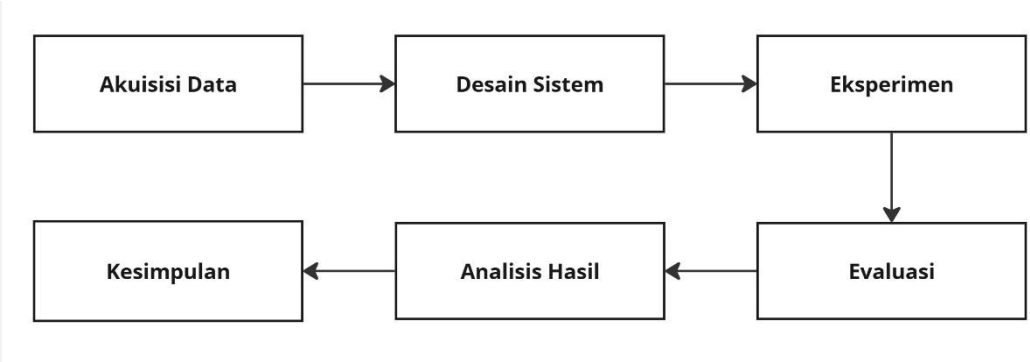
Gambar 3.1. Kerangka Konsep

Input penelitian ini menggunakan Ulasan produk *e-commerce*. Tahapan pertama dilakukan data cleaning terlebih tahulu pada data teks ulasan untuk membersihkan *noise* dan membuat data lebih mudah diproses oleh sistem. Data cleaning meliputi *case folding, punctuation removal, tokenization*, dan *stemming*. Setelah dilakukan data cleaning, digunakan beberapa metode feature extraction untuk membandingkan performa dari masing masing metode yang menjadi tujuan dari penelitian ini. Beberapa feature extraction yang digunakan meliputi Term Frequency – Inverse Class Frequency (TF-ICF), word2Vec, dan Bidirectional Encoder Representations from Transformer (BERT). Output dari ketiga metode tersebut berupa vektor numerik yang nantinya akan dikirim ke input layer pada

neural network sebagai *classifier*. Output classifier berupa *aspect topic classification dan sentiment.*

3.2 Prosedur Penelitian

Prosedur penelitian merupakan tahapan tahapan ketika melakukan penelitian. Pada penelitian ini tahapan pertama yang dilakukan adalah akuisisi data, selanjutnya membuat rancangan rancangan desain sistem yang akan dibangun. Dari desain sistem yang dibuat diimplementasikan dengan beberapa eksperimen. Hasil dari eksperimen dilakukan evaluasi dan analisis dengan tujuan untuk melihat apakah percobaan eksperimen berhasil, memahami hasil, dan sebagai referensi percobaan yang lebih baik di penelitian selanjutnya. Prosedur penelitian ini dapat dilihat pada Gambar 3.2.



Gambar 3.2. Prosedur Penelitian

3.3 Akuisisi Data

Pada penelitian ini data yang digunakan diambil dari Kaggle. Data berupa ulasan produk Tokopedia dari berbegai kategori produk, mulai dari *fashion*, elektronik, olahraga, pertukangan, handphone, dan lain lain. Data yang diambil

masih berupa data mentah tanpa label aspek dan sentimen. Oleh karena itu, perlu dilakukan pelabelan terlebih dahulu pada data yang diambil.

Dalam penelitian ini akan digunakan 4002 data ulasan yang telah dilabeli. Data tersebut akan menjadi dasar untuk melatih dan menguji model klasifikasi yang akan dikembangkan pada penelitian ini. Pelabelan dilakukan dengan melakukan survei kepada orang-orang. Dengan survei ini, setiap data ulasan di kategorikan ke dalam kategori aspek topik dan juga sentimen. Kategori aspek terdiri dari tiga, yaitu aspek produk, pelayanan pelanggan, dan pengiriman. Kategori aspek ini, setiap ulasan bisa memiliki lebih dari satu aspek.bisa dua bahkan tiga sekaligus. Penjelasan dari masing masing aspek ditampilkan pada Tabel 3.2.

Tabel 3.2. Aspek ulasan *e-commerce*

Aspek	Deskripsi
Produk	Kepuasan pelanggan terhadap kualitas, performa, dan kesesuaian produk dengan deskripsi yang diberikan
Layanan Pelanggan	Interaksi antara pelanggan dengan penjual, keramahan dan kecepatan respon dari penjual, dan penanganan keluhan
Pengiriman	Kecepatan pengiriman, kondisi barang saat diterima, dan ketepatan waktu pengiriman

Untuk mendapatkan pemahaman yang lebih detail, sampel data pada kasus *aspect topic modeling* ditunjukkan pada Tabel 3.3.

Tabel 3.3. Sampel data ulasan *e-commerce* pada kasus *aspect topic modeling*

No	Review	Aspek
1	barangnya sampai dengan selamat Pengirimannya cepat cuma sehari. stiknya sudah dipakai buat nostalgia main harvest moon berjalan dengan sempurna haha thanks ^_^	Produk, Pengiriman
2	Barang sudah diterima mudah 2 an bisa berfungsi bagus dan awet. terima kasih	Produk
3	Thanks gan barangnya sudah diterima dengan baik.	Pengiriman
4	lumayanlah tapi analog kiri stiknya gak berfungsi padahal sudah saya setting di emulatornya tapi analog kanannya bisa.	Produk

5	sesuai gambar dan deskripsi. harga termurah se tokopedia :)	Produk
6	Alhamdulillah barang sudah sampai Pelayanan cepat pengiriman cepat Cukup puas	Layanan Pelanggan, Pengiriman

Kategori lainnya adalah kategori sentimen. Dalam kategori sentimen terdiri dari tiga kelas, yaitu negatif, netral, dan positif. Penjelasan dari masing-masing sentimen yang digunakan dalam penelitian ini ditunjukkan pada Tabel 3.4.

Tabel 3.4. Sentimen ulasan *e-commerce*

Setimen	Deskripsi
Negatif	Ulasan menyatakan kata-kata negatif dan mengandung kekecewaan dari pelanggan
Netral	Ulasan yang hanya menyatakan fakta tanpa sentiment jelas atau menyatakan positif dan negatif secara bersamaan
Positif	Ulasan menyatakan kata kata positif dan mengandung kepuasan dari pelanggan

Untuk mendapatkan pemahaman yang lebih detail, sampel data pada kasus analisis sentimen ditunjukkan pada Tabel 3.5.

Tabel 3. 5. Sampel data ulasan *e-commerce* pada kasus analisis sentimen

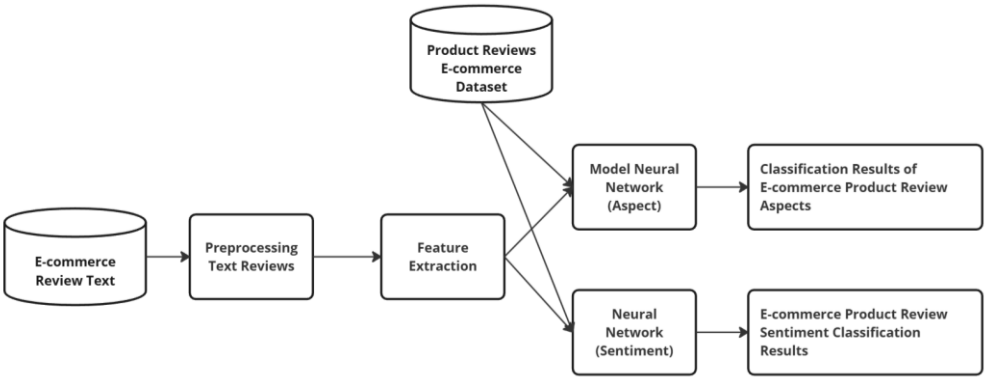
No	Review	Sentimen
1	barangnya sampai dengan selamat Pengirimannya cepat cuma sehari. stiknya sudah dipakai buat nostalgia main harvest moon berjalan dengan sempurna haha thanks ^_^	Positif
2	Barang sudah diterima mudah 2 an bisa berfungsi bagus dan awet. terima kasih	Positif
3	Thanks gan barangnya sudah diterima dengan baik.	Positif
4	lumayanlah tapi analog kiri stiknya gak berfungsi padahal sudah saya setting di emulatornya tapi analog kanannya bisa.	Netral
5	sesuai gambar dan deskripsi. harga termurah se tokopedia :)	Netral
6	barang sudah sampai dan langsung ditest. Button arah bawah tidak berfungsi. Padahal deskripsi ditest sebelum dikirim.	Negatif

Berdasarkan pemaparan sampel dan label diatas, Pada penelitian melakukan dua kasus, yaitu klasifikasi *aspect topic* dan *analysis sentiment*. Data ulasan yang

dipakai dua kasus tersebut adalah sama, namun dengan label yang berbeda sesuai dengan label dari kasus masing masing.

3.4 Desain Sistem

Desain penelitian ini diilustrasikan dalam Gambar 3.1 yang terdiri dari beberapa tahapan, yaitu input ulasan produk *review e-commerce*, *text preprocessing* yang terdiri dari *data cleaning* dan *feature extraction*, *classifier*, dan juga output berupa ulasan produk *e-commerce* terklasifikasi.



Gambar 3.3. Desain Sistem

Pada Gambar 3.3 digambarkan bahwa ada perbedaan pada *feature extraction* IndoBERT, dimana pada IndoBERT ini tanpa melalui *data cleaning*. Hal ini dikarenakan pada IndoBERT sendiri sudah mengatasi teks mentah dengan tokenizer khusus yang telah dilatih sebelumnya. Tokenizer IndoBERT mampu menangani berbagai bentuk teks, termasuk huruf besar-kecil, tanda baca, serta karakter khusus tanpa perlu preprocessing tambahan seperti *stopword removal* atau *stemming*.

Dalam penelitian ini, penggunaan IndoBERT tanpa *preprocessing data cleaning* bertujuan untuk mempertahankan informasi sebanyak mungkin dari teks

asli sehingga model dapat belajar pola yang lebih kompleks dan mendalam dari data ulasan *e-commerce*.

3.5 *Data Cleaning*

Data cleaning merupakan proses merubah data mentah menjadi data yang bersih dengan memeriksa nilai yang hilang, data *noisy*, dan juga ketidakkonsitenan. Hal ini dilakukan agar membantu meningkatkan akurasi model klasifikasi (Uysal & Gunal, 2014).

Data cleaning pada data teks merupakan salah satu bagian dari *text preprocessing*. Data ulasan produk *e-commerce* pada penelitian ini di ekstrak dengan melewati beberapa proses, yaitu *case folding, tokenization, stopword removal*, daan diakhiri dengan *stemming*.

3.5.1 *Case Folding*

Case Folding adalah tahapan *text preprocessing* yang memiliki tujuan untuk mengubah semua huruf dalam dokumen menjadi huruf kecil (Jaman & Abdulrohman, 2019). Hal ini dilakukan untuk membuat teks dalam dokumen menjadi lebih seragam dalam penulisan huruf, dengan begitu memudahkan dalam pemrosesan teks secara keseluruhan. Huruf besar dan huruf kecil pada teks diduga tidak memiliki kontras makna yang begitu tinggi. Didalam *case folding*, semua huruf besar (*uppercase*) didalam dokumen diubah menjadi huruf kecil (*lowercase*), sedangkan untuk huruf kecil tetap tidak berubah.

Sebagai contoh, jika ada teks “Ini Merupakan Contoh Dari *Case Folding*” setelah melewati proses *case folding*, teks tersebut akan berubah menjadi “ini

merupakan contoh dari *case folding*". Dengan begitu, hasil teks menjadi seragam dan perbedaan huruf besar dan huruf kecil tidak lagi berpengaruh dalam pemrosesan teks tersebut.

3.5.2 *Punctuation Removal*

Punctuation removal atau Penghapusan tanda baca adalah tahapan dalam *text preprocessing* dimana menghapus semua tanda baca, simbol, *link*, dan emoji di dalam teks dokumen. *Punctuation removal* memiliki tujuan untuk membersihkan elemen-elemen yang tidak diperlukan lagi agar teks yang akan diproses menjadi lebih fokus pada isi kontennya dan lebih mudah untuk dilakukan pemrosesan pada tahap selanjutnya (Banks *et al*., 2018).

Sebagai contoh, jika ada teks "Hari ini, cuaca sangat cerah! Apakah kamu ingin pergi ke taman?" setelah melewati proses *punctuation removal*, teks tersebut akan berubah menjadi "Hari ini cuaca sangat cerah Apakah kamu ingin pergi ke taman". Dalam contoh ini, tanda baca koma, tanda seru, tanda tanya dihapus dari teks. Setelah penghapusan, teks menjadi lebih sederhana dan hanya berfokus pada kata tanpa ada tanda baca yang mengganggu.

3.5.3 *Tokenization*

Tokenization adalah tahapan dalam *text preprocessing* dimana teks yang panjang dipecah menjadi unit-unit yang lebih kecil. Unit-unit kecil ini disebut "token" (Jaman & Abdulrohman, 2019). *Tokenization* memiliki tujuan untuk memecah teks menjadi unit-unit yang lebih mudah di proses. Dengan begitu, pemrosesan dapat dilakukan dengan lebih efisien pada tahapan selanjutnya.

Sebagai contoh, jika ada teks “Saya suka belajar *machine learning*” setelah melewati proses *tokenization*, teks akan berubah token-token yang mempresentasikan unit-unit kecil dari teks tersebut, seperti ["Saya", "suka", "belajar", "*machine*", "*learning*"].

3.5.4 *Stopword Removal*

Stopword removal adalah tahapan dalam *text preprocessing* dimana kata-kata yang umum dan yang memiliki sedikit arti seperti “dan”, “di”, “dari”, “yang”, dan yang lain sebagainya dihapus dari teks dokumen (Sarica & Luo, 2020). *Stopword* biasanya muncul dalam jumlah besar tanpa bergantung pada bahasan tertentu. *Stopword removal* memiliki tujuan untuk meningkatkan fokus pemrosesan pada kata-kata yang penting sehingga meringankan pemrosesan pada tahap lebih lanjut (Banks *et al*., 2018).

Sebagai contoh teks "Saya sedang belajar tentang ilmu komputer di universitas." Setelah melakukan penghapusan *stopword*, kata-kata seperti "saya", "sedang", "tentang", dan "di" yang umumnya tidak memberikan banyak makna spesifik dapat dihilangkan. Hasilnya, teks yang tersisa adalah "belajar ilmu komputer universitas".

3.5.5 *Stemming*

Stemming adalah proses dalam *text preprocessing* dimana kata-kata diubah menjadi bentuk kata dasarnya dengan menghapus akhiran dan awalan pada kata awal tersebut. *Stemming* memiliki tujuan untuk mengurangi variasi kata-kata yang maknanya mirip menjadi seragam dan dianggap sama. Dengan begitu, akan

mengurangi dimensi data teks dan meningkatkan pemrosesan menjadi lebih efektif (Purnamasari *et al*., 2020).

Sebagai contoh, kata "berlari", "berlarian", dan "berlarilah" setelah menjalani proses stemming akan menjadi kata dasar "lari". Proses ini membantu mengelompokkan kata-kata yang memiliki dasar kata yang sama sehingga pemrosesan teks dapat lebih efektif.

3.6 Term Frequency – Inverse Class Frequency (TF-ICF)

Dalam penelitian ini, Term Frequency – Inverse Class Frequency (TF-ICF) merupakan tahapan yang sangat penting dalam proses ekstraksi fitur dari teks. Setelah melalui tahap *text preprocessing*, TF-ICF berperan dalam mengubah teks menjadi vektor numerik yang merepresentasikan tingkat kepentingan kata dalam kumpulan data.

TF-ICF merupakan variasi dari Term Frequency – Inverse Document Frequency (TF-IDF) yang telah banyak digunakan dalam penelitian *text mining* (Hakim *et al*., 2014). Jika pada TF-IDF suatu kata dinilai penting berdasarkan kemunculannya dalam dokumen, maka pada TF-ICF tingkat kepentingan suatu kata dihitung berdasarkan kemunculannya dalam kelas (bukan dokumen). TF-ICF lebih sesuai digunakan dalam klasifikasi teks, terutama dalam kasus *multilabel classification*, karena mempertimbangkan distribusi kata di seluruh kategori (*classes*) dibandingkan hanya dalam dokumen individual.

Konsep utama TF-ICF jika suatu kata muncul di sedikit kelas, maka kata tersebut dianggap lebih penting dan memiliki nilai TF-ICF yang tinggi. Jika suatu

kata muncul di banyak kelas, maka kata tersebut dianggap kurang informatif dan memiliki nilai TF-ICF yang rendah.

TF-ICF terdiri dari dua komponen utama, yaitu Term Frequency (TF) dan Inverse Class Frequency (ICF). TF-IDF terdiri dari dua komponen utama, yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF). TF mengukur frekuensi suatu kata dalam dokumen dan dihitung menggunakan Persamaan 3.1:

$$TF_{(t,d)} = \frac{N_{t,d}}{\sum_{i=0}^{n} N_{i,d}} \tag{3.1}$$

Keterangan:
t = kata yang dihitung
d = dokumen tempat kata tersebut dihitung
$N_{t,d}$ = jumlah kemunculan kata *t* dalam dokumen *d*
$\sum_{i=0}^{n} N_{i,d}$= total jumlah kata dalam dokumen *d*

ICF menilai seberapa penting suatu kata di semua kelas dan dihitung dengan Persamaan 3.2:

$$ICF_{(t)} = log \frac{|C|}{\sum c_t \in C} \tag{3.2}$$

Keterangan:
|*C*| = total semua kelas
c_t = kelas *c* yang mengandung kata *t*
$\sum c_t \in C$ = jumlah kelas *c* yang mengandung kata *t*

TF-ICF dihitung dengan mengalikan nilai TF dan IDF seperti pada Persamaan 3.3:

$$TFICF_{(t,d)} = TF_{(t,d)} \times ICF_{(t,)} \tag{3.3}$$

Output dari TF-ICF berupa matriks numerik. Dengan TF-ICF, penelitian ini dapat mengidentifikasikan kata-kata yang unik pada dataset. Matriks numerik yang dihasilkan dari TF-IDF menjadi dasar untuk analisis selanjutnya.

3.7 Word2Vec

Word2Vec adalah salah satu model *embedding* kata yang populer dalam pemrosesan bahasa alami (NLP) yang dikembangkan oleh Mikolov *et al.* pada 2013 (Mikolov, Chen, *et al*., 2013)**.** Model ini mengubah kata-kata menjadi vektor numerik (*word embeddings*) dalam ruang berdimensi tinggi, sehingga kata-kata dengan makna atau konteks serupa akan memiliki representasi vektor yang mirip (Mikolov, Sutskever, *et al*., 2013). Keunggulan utama Word2Vec terletak pada kemampuannya menangkap hubungan semantik dan sintaksis antar kata secara efektif (Jansen, 2017). Contohnya, jika kita punya kata *king* (raja), lalu mengganti unsur *man* (pria) dengan *woman* (wanita), hasilnya mendekati *queen* (ratu). Selain itu, Word2Vec juga mampu mengidentifikasi kesamaan makna antara kata-kata seperti "cepat" dan "cekat", meskipun memiliki bentuk yang berbeda.

Word2Vec menggunakan dua arsitektur utama, yaitu *Continuous Bag of Words* (CBOW) dan *Skip-gram*. CBOW memprediksi kata target berdasarkan konteks sekitarnya, cocok untuk data yang relatif kecil, sementara *Skip-gram* memprediksi kata-kata konteks dari kata target, dan lebih efektif untuk dataset besar. Kedua arsitektur ini memanfaatkan jaringan saraf dangkal yang dilatih menggunakan teknik *negative sampling* atau *hierarchical softmax* untuk meningkatkan efisiensi komputasi (Jansen, 2017).

Pada penelitian ini, model Word2Vec yang akan digunakan adalah model skip-gram, dimana model menerima sebuah kata target (w_i) dan mencoba memprediksi kata-kata konteks (w_o) di sekitarnya dalam jendela tertentu. Untuk setiap kata dalam korpus, model membentuk pasangan data pelatihan. Probabilitas kata konteks terhadap kata target dihitung dengan menggunakan fungsi *softmax* pada Persamaan 3.4:

$$P(w_o|w_i) = \frac{\exp\left(v'^{T}_{w_o} v_{w_i}\right)}{\sum_{w=1}^{W} \exp\left(v'^{T}_{w} v_{w_i}\right)} \tag{3.4}$$

Keterangan:
v_{w_i} adalah vektor kata input
v_{w_o} adalah vektor kata output atau kata konteks

Dalam skip-gram, untuk meningkatkan rata rata log probabilitas dari semua pasangan kata (target dan konteks) dalam korpus digunakan fungsi pada Rumus 3.5:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c<j<c, j\neq 0} \log P\left(w_{t+j}|w_t\right) \tag{3.5}$$

Keterangan:
T adalah jumlah total kata dalam korpus
c adalah ukuran jendela konteks

Dalam penelitian ini, digunakan model Word2Vec pretrained yang dilatih menggunakan Wikipedia Bahasa Indonesia. Model ini dikenal dengan nama idwiki200, karena menggunakan dimensi *embedding* sebesar 200. Model ini dilatih dengan menggunakan arsitektur Skip-Gram, sehingga mampu menangkap hubungan kata dalam konteks yang luas dan relevan dengan Bahasa Indonesia formal. Karena Wikipedia menggunakan bahasa yang baku dan informatif, maka

kosakata dalam idwiki200 umumnya terdiri dari kata-kata baku yang sesuai dengan kaidah Bahasa Indonesia.

Word2Vec menghasilkan vektor kata, tetapi dalam penelitian ini dibutuhkan representasi kalimat untuk masukan ke dalam model klasifikasi. Oleh karena itu, digunakan metode sederhana namun efektif, yaitu:

$$Vektor\ kalimat = \frac{1}{n}\sum_{i=1}^{n} Vec(w_i) \tag{3.6}$$

Keterangan:
n adalah jumlah kata dalam kalimat ulasan
w_i adalah kata ke-i dalam kalimat ulasan
$Vec(w_i)$ adalah hasil vektor Word2Vec dari kata w_i

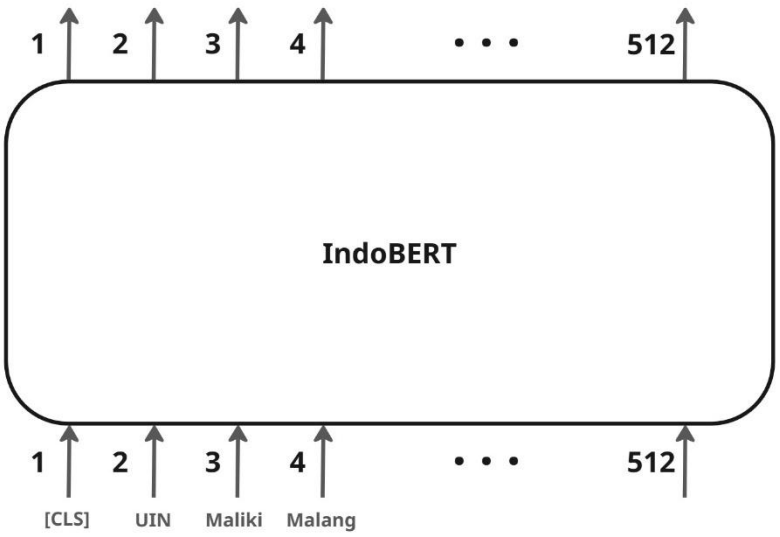
Dengan cara ini, setiap ulasan produk diubah menjadi sebuah vektor tetap yang mewakili makna rata-rata dari seluruh kata dalam ulasan tersebut. Teknik ini disebut juga sebagai *mean pooling* atau *average embedding*, dan merupakan salah satu pendekatan yang efisien untuk membentuk representasi dokumen berdasarkan Word2Vec.

3.8 IndoBERT

IndoBERT adalah model BERT yang dilatih menggunakan teknik *pre-training* Masked Language Modeling (MLM) dan Next Sentence Prediction (NSP) pada dataset bahasa Indonesia. Keberadaan IndoBERT telah membuka kemungkinan baru untuk memahami bahasa Indonesia yang kompleks, sehingga memungkinkan berbagai implementasi NLP seperti analisis sentimen, pemrosesan teks, dan ekstraksi informasi menjadi lebih canggih dan dapat diandalkan (Forry Kusuma & Chowanda, 2023). Penelitian ini menggunakan model *pre-trained*

IndoBERT dengan sebuah *benchmark* baru yang disebut *Indonesian* NLP (IndoNLU) yang diusulkan oleh (Wilie *et al*., 2020). Model ini dilatih menggunakan dataset Indo4B dengan lebih dari 4 miliar kata dan 250 juta kalimat. Selain dilatih dengan dataset dan *benchmark* yang lebih luas, model ini juga memberikan performa yang lebih baik dibandingkan model IndoBERT sebelumnya yang menggunakan IndoLEM sebagai *benchmark* (Kartika *et al*., 2023). Hal ini mendorong penulis untuk mengimplementasikan dan menyempurnakan IndoBERT dari *benchmark* IndoNLU dalam penelitian ini.

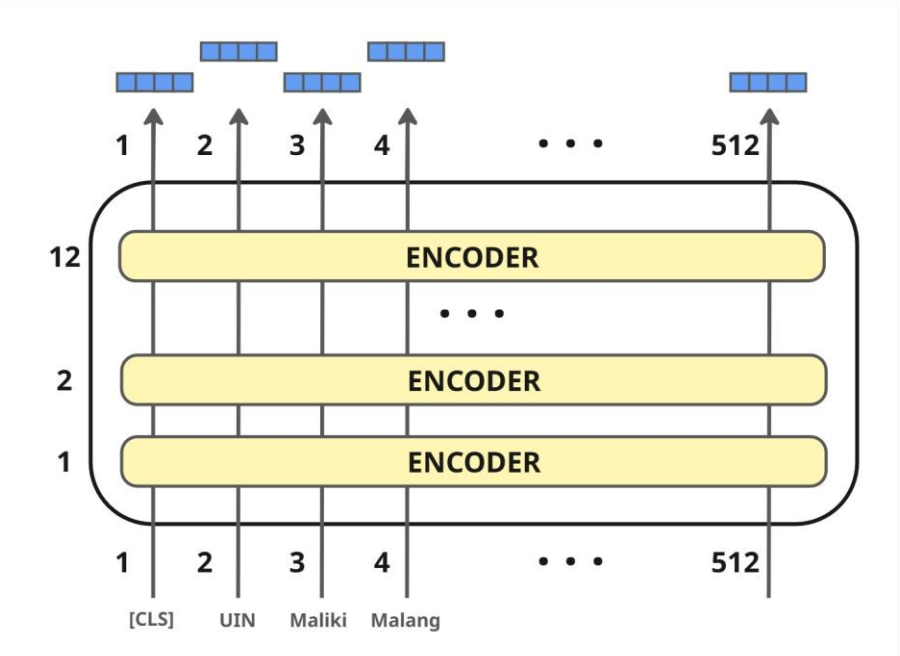
Pada model BERT termasuk IndoBERT, sebelum memberikan input, teks mentah akan melalui proses tokenisasi menggunakan IndoBERT Tokenizer, yang ditunjukkan pada Gambar 3.4.



Gambar 3.4. IndoBERT *Input*

Maksimal token yang dapat dimasukkan ke dalam IndoBERT adalah 512 token. Pada proses tokenisasi ini, akan ditambahkan tanda khusus seperti “[CLS]” dan “[SEP]” [18]. Tanda khusus ini seperti instruksi untuk mulai membaca (untuk

[CLS]) dan berhenti membaca (untuk [SEP]) memungkinkan model BERT untuk fokus pada setiap kalimat. Selanjutnya, kalimat-kalimat tersebut akan diubah menjadi token kata, seperti memecah kalimat yang panjang menjadi beberapa kata (Imron *et al*., 2023b). Untuk memahami arti dari kata-kata tersebut, IndoBERT memiliki kamus khusus yang dibuat dengan menggunakan metode Sentence-Piece. Untuk kata-kata yang belum pernah ditemui sebelumnya atau kata-kata baru, kata-kata tersebut akan dipecah menjadi bagian-bagian yang lebih kecil untuk memahami maknanya. Selanjutnya, data ulasan yang telah dilabeli akan diubah menjadi vektor numerik menggunakan *word embedding* dari IndoBERT (Tanaka *et al*., 2020). Vektor numerik hasil embedding kemudian diproses melalui arsitektur internal BERT yang dapat dilihat pada Gambar 3.5.

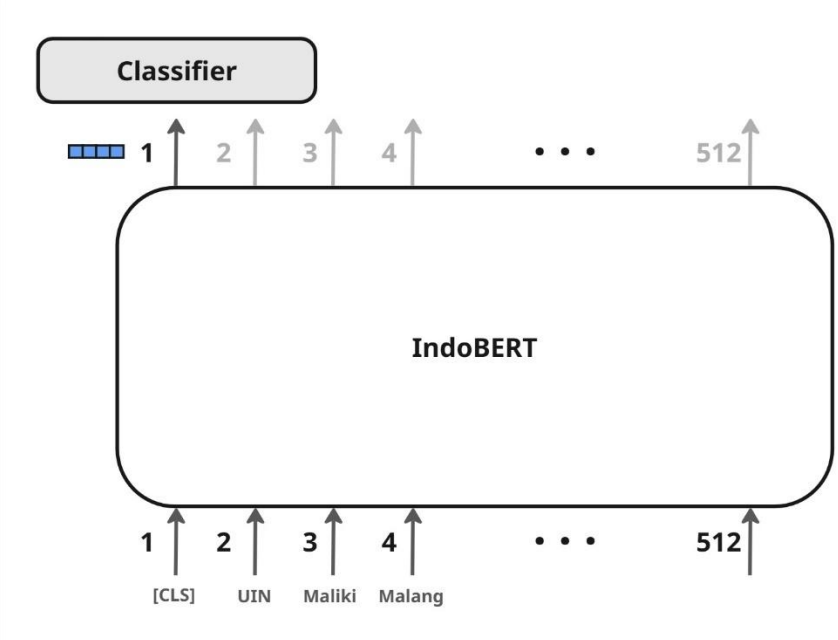


Gambar 3.5. Kerangka IndoBERT

Arsitektur ini terdiri dari multiple encoder layers (12 encoder untuk BERT-base/IndoBERT-base atau 24 encoder untuk BERT-large). Setiap encoder

menerapkan mekanisme self-attention yang memungkinkan setiap token "memperhatikan" semua token lainnya dalam *sequence*, sehingga menghasilkan representasi yang mempertimbangkan konteks dari kedua arah (*bidirectional*).

Output dari mekanisme self-attention kemudian diproses melalui feed-forward neural network dan diteruskan ke encoder berikutnya. Pada tahap ini, model melakukan *forward pass* melalui seluruh lapisan Transformer, dengan setiap layer secara progresif menyempurnakan representasi setiap token berdasarkan konteksnya dalam kalimat. Proses ini berulang melalui semua layer encoder, sehingga informasi kontekstual semakin diperkaya di setiap layer.



Gambar 3.6. IndoBERT output untuk klasifikasi

Setelah melewati semua lapisan transformer, setiap token pada sequence input akan menghasilkan output vektor berdimensi 768. Sesuai pada Gambar 3.6, Untuk tugas klasifikasi teks, representasi vektor dari token [CLS] yang berada di posisi pertama sequence diambil sebagai representasi keseluruhan input dan

dimasukkan ke *classifier* untuk menghasilkan prediksi akhir. Representasi vektor dari token-token lainnya dapat digunakan untuk tugas-tugas tingkat token seperti *named entity recognition* atau *question answering*.

Berdasarkan alur kerja diatas terlihat bagaimana IndoBERT mengolah informasi kontekstual secara *bidirectional*, menghasilkan representasi vektor yang kaya akan informasi semantik yang dapat dimanfaatkan untuk berbagai tugas pemrosesan bahasa alami, khususnya untuk bahasa Indonesia dalam kasus IndoBERT.

3.9 Neural Network

Neural network adalah model komputasi yang meniru cara kerja otak. *artificial neuron* saling terhubung membentuk neural network. Sama seperti otak, neural network menerima input, terjadi proses internal seperti aktivasi neuron, dan menghasilkan output. Pada neural network terdapat beberapa bagian didalamnya. Input layer digunakan untuk menerima informasi dari nilai input. Hidden layer berlokasi di antara input layer dan output layer yang memiliki fungsi memetakan pola informasi internal. Output layer merupakan hasil dari informasi yang telah diproses (Mijwil *et al*., 2019). Pada neural network dilakukan perhitungan hasil kali antara input dan bobot yang sesuai, kemudian fungsi aktivasi diterapkan untuk mendapatkan output dari layer sebelum dan mengirimkannya sebagai input pada layer berikutnya (Sharma *et al*., 2020).

Semua layer pada neural network memiliki node dan setiap node memiliki bobot yang berfungsi dalam memproses informasi dari layer satu ke layer berikunya (Sharma *et al*., 2020). Dalam neural network juga ada parameter tambahan yaitu

bias, bias ditambahkan pada setiap node. Ini memungkinkan model untuk belajar pola yang lebih kompleks. Dalam perhitungan, bias dijumlahkan dengan hasil perkalian input dan bobot, dengan adanya bias neural network dapat menghasilkan hasil yang lebih akurat dan dapat diandalkan dalam berbagai tugas pembelajaran mesin.

Topologi neural network pada penelitian ini memiliki jumlah hidden layer sebanyak tiga. Jumlah node pada masing masing hidden layer ditentukan berdasarkan Persamaan 3.7 (Rachmatullah *et al*., 2020).

$$Nh = \frac{\sqrt{1 + 8N_i}}{2} - 1 \tag{3.7}$$

Keterangan:
Nh adalah jumlah node pada hidden layer
N_i adalah jumlah node input

Pada penelitian ini, pemilihan jumlah layer dan node dibuat dengan mengikuti beberapa alasan dasar:

a. Input layer : Panjang menyesuaikan output vektor *feature extraction*.

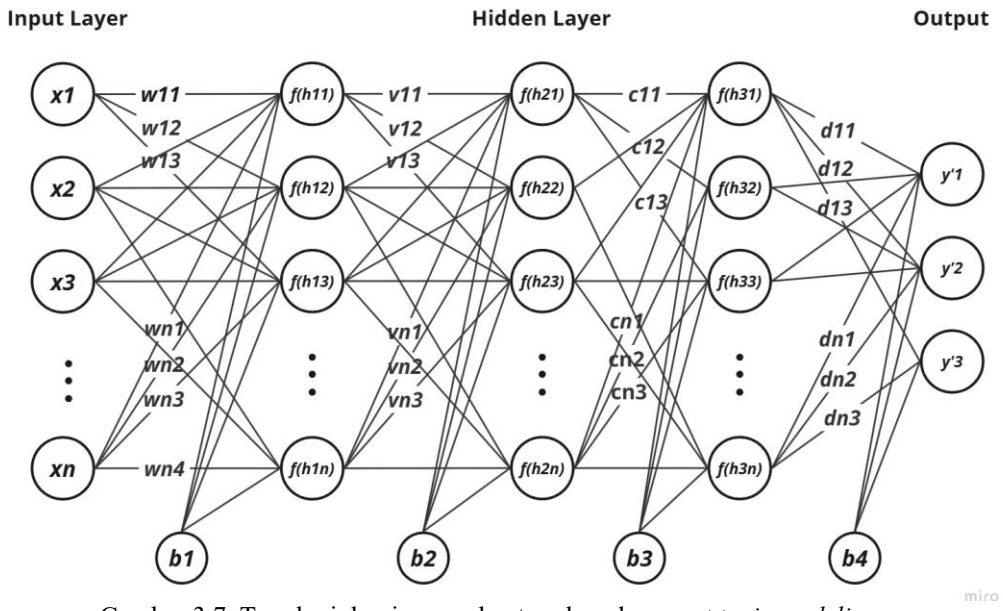
b. Hidden layer : Jumlah node menyesuaikan hasil dari Persamaan 3.7.

c. Output layer : Menggunakan 3 node karena output dari sistem ini adalah tiga kelas aspek dan tiga kelas sentiment.

3.9.1 Neural Network pada *Aspect Topic Modeling*

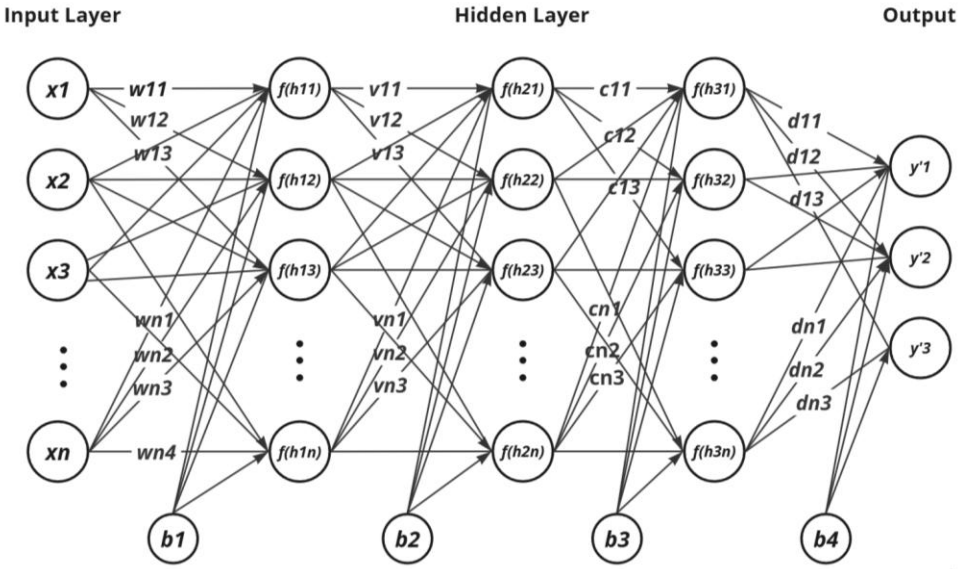
Setelah mengetahui bagaimana cara kerja neural network secara umum, bagian ini akan menjelaskan bagaimana model tersebut digunakan pada kasus *aspect topic modeling*. Pada kasus ini, model tidak hanya memprediksi satu label saja, tetapi bisa lebih dari satu label aspek dalam satu ulasan, dimana terdapat 3

aspek yang difokuskan yaitu produk, layanan pelanggan, dan pengiriman. Topologi neural network pada kasus *aspect topic modeling* ini disajikan pada Gambar 3.7.



Gambar 3.7. Topologi desain neural network pada *aspect topic modeling*

Berdasarkan Gambar 3.7, topologi tersebut menggambarkan susunan layer dan jumlah node pada masing-masing layer, mulai dari input layer, tiga hidden layer, hingga output layer. Penentuan jumlah node pada hidden layer didasarkan pada perhitungan menggunakan Persamaan 3.7. Selanjutnya, proses menghitung input hingga ke layer output dilakukan dengan mengalirkan data melalui setiap node secara bertahap. Proses ini dikenal sebagai feedforward, dan alurnya ditunjukkan pada Gambar 3.8.



Gambar 3.8. *Feedforward* neural network pada *aspect topic modeling*

Ketika data dimasukkan ke dalam neural network, setiap neuron dalam layer input menerima sinyal dan mengalikannya dengan bobotnya, kemudian menambahkan nilai bias. Sinyal-sinyal ini kemudian diteruskan ke layer berikutnya, di mana proses yang sama terjadi lagi, hingga akhirnya mencapai layer output.

Berikut adalah langkah proses yang terjadi selama *feedforward*:

1. Pertama dilakukan perhitungan pada hidden layer 1.

$$h_{1i} = b_{1i} + \sum_{j} w_{ji} x_i \tag{3.8}$$

Keterangan:
h_{1i} adalah node hidden layer 1 indeks ke *i.*
b_{1i} adalah nilai bias pada hidden layer 1 indeks ke *i.*
$\sum_j w_{ji} x_i$ adalah total dari perkalian bobot *(weight)* hidden layer 1 dan input layer.

Persamaan ini menunjukkan bahwa perhitungan pada hidden layer 1 melibatkan perkalian *w* dan *x*, kemudian ditambahkan bias b_1 sebagai bias pada layer pertama.

2. Hasil dari hidden layer 1 dilakukan proses aktivasi menggunakan fungsi aktivasi ReLU.

$$a_i = f(h_i) \tag{3.9}$$

$$a_{1i} = max(0, a_i) \tag{3.10}$$

Keterangan:
a_{1i} adalah nilai aktivasi dari nilai hidden layer 1 indeks ke *i.*
$max()$ adalah fungsi aktivasi ReLU.

Tahap ini diproses melalui fungsi aktivasi ReLU. Hasilnya adalah matriks a_1 yang berisi nilai aktivasi untuk setiap node pada hidden layer 1.

3. Setelah didapatkan nilai hidden layer 1, dilanjutkan perhitungan pada hidden layer 2.

$$h_{2i} = b_{2i} + \sum_j v_{ji} h_{1i} \tag{3.11}$$

Keterangan:
h_{2i} adalah node hidden layer 2 indeks ke *i*.
b_{2i} adalah nilai bias pada hidden layer 2 indeks ke *i*
$\sum_j v_{ji} h_{1i}$ adalah total dari perkalian bobot (*weight)* hidden layer kedua dan hidden layer pertama.

4. Hasil dari hidden layer 2 dilakukan proses aktivasi menggunakan fungsi aktivasi ReLU.

$$a_i = f(h_i) \tag{3.12}$$

$$a_{2i} = max(0, a_i) \tag{3.13}$$

Keterangan:
a_{2i} adalah nilai aktivasi dari nilai hidden layer 2 indeks ke *i.*
$max()$ adalah fungsi aktivasi ReLU.

Tahap ini diproses melalui fungsi aktivasi ReLU. Hasilnya adalah matriks a_2 yang berisi nilai aktivasi untuk setiap node pada hidden layer 2.

5. Setelah didapatkan nilai hidden layer 2, dilanjutkan perhitungan pada hidden layer 3.

$$h_{3i} = b_{3i} + \sum_j c_{ji} h_{2i} \tag{3.14}$$

Keterangan:
h_{3i} adalah node hidden layer 3 indeks ke *i*.
b_{3i} adalah nilai bias pada hidden layer 3 indeks ke *i*
$\sum_j c_{ji} h_{2i}$ adalah total dari perkalian bobot (*weight)* hidden layer 3 dan hidden layer 2.

6. Hasil dari hidden layer 3 dilakukan proses aktivasi menggunakan fungsi aktivasi ReLU.

$$a_i = f(h_i) \tag{3.15}$$

$$a_{3i} = max(0, a_i) \tag{3.16}$$

Keterangan:
a_{3i} adalah nilai aktivasi dari nilai hidden layer 3 indeks ke *i.*
$max()$ adalah fungsi aktivasi ReLU.

Tahap ini diproses melalui fungsi aktivasi ReLU. Hasilnya adalah matriks a_3 yang berisi nilai aktivasi untuk setiap node pada hidden layer 3.

7. Setelah didapatkan nilai hidden layer 3, dilanjutkan perhitungan pada output layer.

$$y'_i = b_{4i} + \sum_j d_{ji} h_{3i} \tag{3.17}$$

Keterangan:
y'_i adalah node output layer indeks ke *i*.
b_{4i} adalah nilai bias pada output layer indeks ke *i*
$\sum_j c_{ji} h_{3i}$ adalah total dari perkalian bobot (*weight)* output layer dan hidden layer 3.

8. Hasil output layer dilakukan proses aktivasi menggunakan fungsi *sigmoid*. Fungsi *sigmoid* digunakan karena output *aspect topic modeling* berupa *multilabel categorical.*

$$Sigmoid(y'_i) = \frac{1}{1 + e^{-y'_i}} \tag{3.18}$$

$$y' aspect\ topic\ modeling = Sigmoid(y'_i) \tag{3.19}$$

Keterangan:
y'_i adalah output neuron ke-*i*
y' adalah hasil output dari model neural network.

9. Hasil proses fungsi aktivasi *sigmoid* merupakan hasil output. Langkah selanjutnya adalah menghitung nilai error menggunakan *binary cross-entropy* untuk *aspect topic modeling*.

$$BCE = -\frac{1}{N}\sum_{j=1}^{N}[y_{ij}\ log\left(y'_{ij}\right) + (1 - y_{ij}) log(1 - y'_{ij})] \tag{3.20}$$

Keterangan:
y_{ij} adalah output target *ground truth* untuk kelas ke-*i* dari data ke-*j*.
y'_{ij} adalah output prediksi dari model untuk kelas ke-*i* dari data ke-*j*.

10. Setelah hasil output dari fungsi sigmoid, selanjutnya diterapkan *threshold* sebesar 0.5 untuk menentukan apakah suatu kelas dianggap aktif atau tidak dalam satu ulasan. Kelas dianggap aktif jika nilai probabilitas lebih dari atau sama dengan 0.5, dan dianggap tidak aktif kurang dari 0.5.

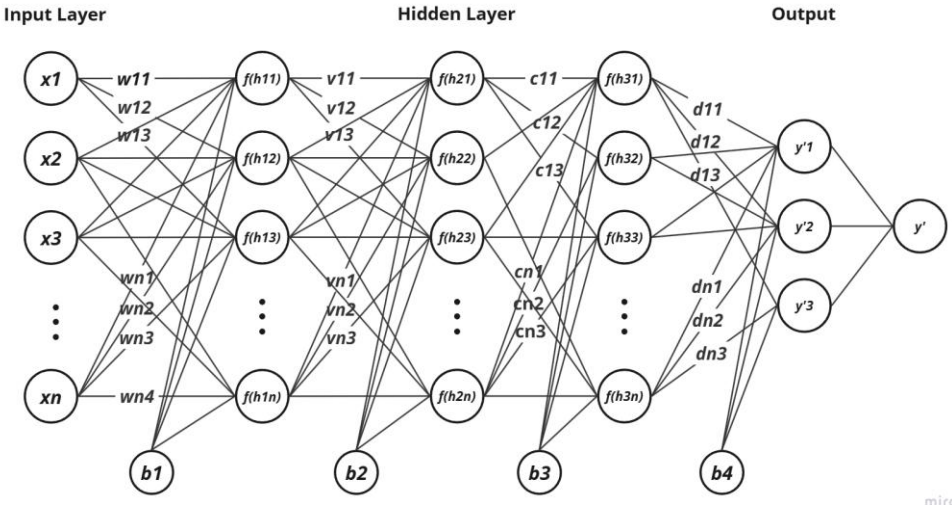
$$y'_{ij} = \begin{cases} 1, & jika\ y'_{ij} \geq 0.5 \\ 0, & jika\ y'_{ij} < 0.5 \end{cases} \quad (3.21)$$

Keterangan:
y'_{ij} adalah output akhir berupa label biner 0 atau 1 pada kelas ke-i dari data ke-j

Persamaan-persamaan di atas menunjukkan bagaimana nilai aktivasi neuron dihitung pada setiap layer dalam neural network. Dengan menghitung nilai aktivasi neuron pada setiap layer dan dilakukan *threshold* pada hasil output, maka dapat dihasilkan nilai output yang merupakan prediksi kelas dari data input.

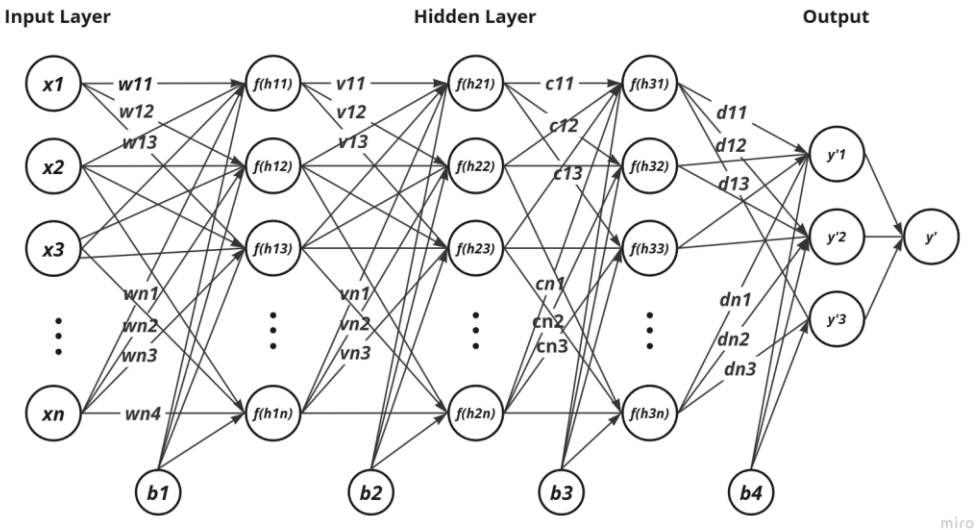
3.9.2 Neural Network pada *Sentiment Analysis*

Berbeda dengan *aspect topic model*, model neural network pada kasus analisis sentimen menghasilkan output berupa 3 kelas sentimen (*multiclass*) yang terdiri dari positif, netral, dan negatif. Desain topologi neural network pada kasus analisis sentimen dapat dilihat pada Gambar 3.9.



Gambar 3.9. Topologi desain neural network pada *sentiment analysis*

Berdasarkan Gambar 3.9, topologi neural network pada kasus analisis sentimen terdiri atas input layer, 3 hidden layer, dan output layer. Namun di akhir arsitektur, hanya ada 1 kelas yang paling dominan yang terpilih sebagai output. Seperti pada topologi sebelumnya, penentuan jumlah node pada hidden layer juga didasarkan pada perhitungan menggunakan Persamaan 3.7. Selanjutnya, alur untuk proses *feedforward* ditunjukkan pada Gambar 3.10.



Gambar 3.10. *Feedforward* neural network pada analisis sentimen

Proses feedforward pada topologi ini mengikuti alur yang sama dengan neural network untuk aspect topic modeling, khususnya dari tahap 1 hingga 8. Perbedaan utama terletak pada perhitungan di layer output. Dalam analisis sentimen, output layer diolah menggunakan fungsi aktivasi softmax, yang dipilih karena sifatnya yang sesuai untuk klasifikasi *multiclass*.

$$Softmax(y'_i) = \frac{\exp(y'_i)}{\sum_{j=1}^{3} \exp(y'_j)} \tag{3.22}$$

$$y' = Softmax(y'_i) \tag{3.23}$$

Keterangan:
y'_i adalah output neuron ke-*i*
y' adalah hasil output dari model neural network.

Output yang dihasilkan dari fungsi aktivasi softmax kemudian digunakan sebagai dasar perhitungan error dengan menerapkan *cross-entropy loss*.

$$CE = -\sum_{i=1}^{C} y_i \, log\,(y_i') \tag{3.24}$$

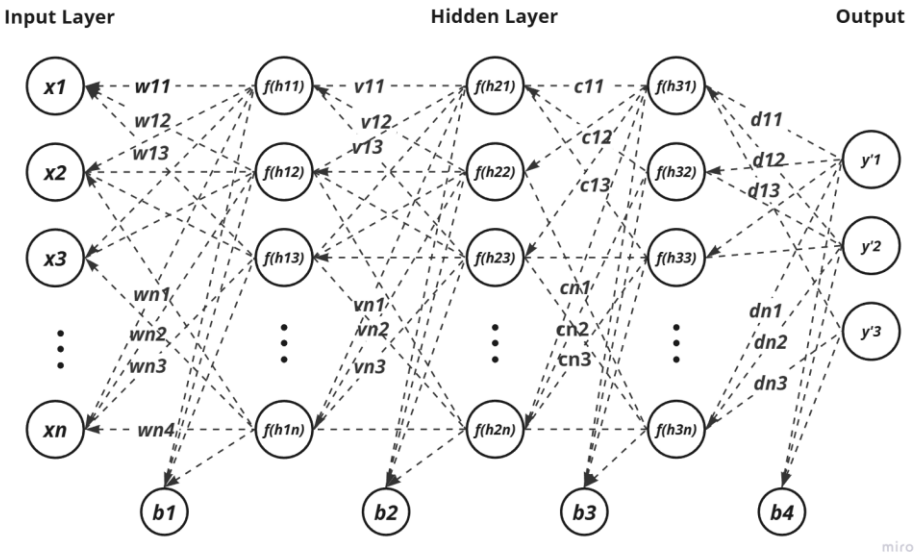
Keterangan:
y adalah output target *ground truth.*
y' adalah output prediksi dari model.

3.9.3 Proses *Training*

Untuk melakukan klasifikasi aspek ulasan produk *e-commerce*, model neural network harus dilatih (*training*) terlebih dahulu agar dapat mengidentifikasi kelas-kelas yang sesuai dengan *ground truth* yang ada (Mijwil *et al*., 2019). Tujuan proses *training* adalah mendapatkan nilai bobot dan bias yang optimal pada neural network agar dapat sesuai dengan kasus klasifikasi ulasan produk *e-commerce*.

Sebelum melakukan *training*, terdapat beberapa nilai parameter yang diperlukan, diantaranya *epoch* dan *learning rate*. Nilai dari masing-masing parameter tersebut tidak berubah (konstan) seiring berjalannya proses *training*. Sementara itu, nilai bobot dan bias pada neural network akan berubah seiring berjalannya proses *training*.

Pada proses *training*, input yang digunakan adalah kalimat dari data *training* yang telah melewati tahap *preprocessing* dan feature extraction, dimana data teks yang sebelumnya berupa teks mentah diubah menjadi vektor numerik hasil dari TF-ICF, Word2Vec, ataupun BERT embedding. Nilai dari vektor numerik hasil *feature extraction* ini dimasukkan ke dalam input layer. Kemudian, menjalani *training* (*feedforward* dan *backpropagation*) sebanyak jumlah *epoch* yang telah ditentukan. Alur dari proses *backpropagation* ditunjukkan pada Gambar 3.10.



Gambar 3.11. *Backpropagation* neural network

Neural network menginisiasi nilai awal parameter, bobot, bias, *learning rate*. Neural network mengawali proses *training* dengan menggunakan nilai awal

tersebut dan melakukan pembaruan bobot dan bias disetiap iterasi atau *epoch*. Proses menentukan seberapa besar bobot dan bias yang akan diubah dapat dilakukan dengan metode *backpropagation* (Abiodun *et al*., 2019). Proses *backpropagation* dimulai dengan menentukan error dari output *feedforward*. Error ini kemudian disebarkan dari ujung layer output ke layer input per-layer demi layer. Setelah itu, bobot dan bias diperbarui dengan memperhitungkan error yang diterima dan *learning rate* yang telah ditentukan. Proses ini dilakukan terus menerus hingga nilai output mendekati nilai actual (Li *et al*., 2012). Berikut adalah langkah proses yang terjadi selama *backpropagation*:

1. Menghitung error, turunan bobot, dan turunan bias pada output layer.

$$\delta_{output} = (y - y') \tag{3.25}$$

$$\frac{dL}{dD} = \delta_{output}(h_3)^T \tag{3.26}$$

$$\frac{dL}{db_4} = \delta_{output} \tag{3.27}$$

Keterangan:
δ_{output} adalah error pada output layer.
y adalah target output.
$\frac{dL}{dD}$ adalah turunan bobot pada output layer.
$\frac{dL}{db_4}$ adalah turunan bias pada output layer.
$(h_3)^T$ adalah nilai *transpose* hidden layer 3.

Error atau kesalahan pada output layer dihitung dan diteruskan ke layer sebelumnya untuk menghitung kesalahan bobot dan bias pada output layer.

2. Menghitung error, turunan bobot, dan turunan bias pada hidden layer 3.

$$\delta_3 = D^T * \delta_{output} * \sigma'(a_3) \tag{3.28}$$

$$\frac{dL}{dC} = \delta_3 h_2 \tag{3.29}$$

$$\frac{dL}{db_1} = \delta_3 \tag{3.30}$$

Keterangan:
δ_3 adalah error pada hidden layer 3.
$\frac{dL}{dC}$ adalah turunan bobot pada hidden layer 3.
$\frac{dL}{db_3}$ adalah turunan bias pada hidden layer 3.
D^T adalah nilai *transpose* matriks bobot antara hidden layer 3 dan output layer.
$\sigma'()$ adalah turunan dari fungsi aktivasi ReLU pada hidden layer 3.
h_2 adalah nilai hidden layer 2.

Error dari output layer digunakan untuk mencari error pada hidden layer 3. Selanjutnya hasil dari error hidden layer 3 digunakan untuk menghitung turunan dari bobot dan bias pada hidden layer 3. Dengan begitu penyesuaian yang diperlukan untuk bobot dan bias hidden layer 3 berdasarkan apa yang terjadi pada output layer.

3. Menghitung error, turunan bobot, dan turunan bias pada hidden layer 2.

$$\delta_2 = C^T * \delta_3 * \sigma'(a_2) \tag{3.31}$$

$$\frac{dL}{dV} = \delta_2 h_1 \tag{3.32}$$

$$\frac{dL}{db_2} = \delta_2 \tag{3.33}$$

Keterangan:
δ_2 adalah error pada hidden layer 2.
$\frac{dL}{dV}$ adalah turunan bobot pada hidden layer 2.
$\frac{dL}{db_2}$ adalah turunan bias pada hidden layer 2.
C^T adalah nilai *transpose* matriks bobot antara hidden layer 2 dan hidden layer 3.
$\sigma'()$ adalah turunan dari fungsi aktivasi ReLU pada hidden layer 2.
h_1 adalah nilai hidden layer 1.

Error dari hidden layer 3 digunakan untuk mencari error pada hidden layer 2. Selanjutnya hasil dari error hidden layer 2 digunakan untuk menghitung turunan dari bobot dan bias pada hidden layer 2. Dengan begitu penyesuaian yang diperlukan untuk bobot dan bias hidden layer 2 berdasarkan apa yang terjadi pada hidden layer 3.

4. Menghitung error, turunan bobot, dan turunan bias pada hidden layer 1.

$$\delta_1 = V^T * \delta_2 * \sigma'(a_1) \tag{3.34}$$

$$\frac{dL}{dW} = \delta_1 x \tag{3.35}$$

$$\frac{dL}{db_1} = \delta_1 \tag{3.36}$$

Keterangan:
δ_1 adalah error pada hidden layer 1.
$\frac{dL}{dW}$ adalah turunan bobot pada hidden layer 1.
$\frac{dL}{db_1}$ adalah turunan bias pada hidden layer 1.
V^T adalah nilai *transpose* matriks bobot antara hidden layer 1 dan hidden layer 2.
$\sigma'()$ adalah turunan dari fungsi aktivasi ReLU pada hidden layer 1.
x adalah nilai input.

Error dari hidden layer 2 digunakan untuk mencari error pada hidden layer 1. Selanjutnya hasil dari error hidden layer 1 digunakan untuk menghitung turunan dari bobot dan bias pada hidden layer 1. Dengan begitu penyesuaian yang diperlukan untuk bobot dan bias hidden layer 1 berdasarkan apa yang terjadi pada hidden layer 2.

5. Langkah terakhir yaitu *update* parameter bobot dan bias.

$$D^t = D^{t-1} - \alpha \frac{dL}{dD} \tag{3.37}$$

$$b_4{}^t = b_4{}^{t-1} - \alpha \frac{dL}{db_4} \tag{3.38}$$

$$C^t = C^{t-1} - \alpha \frac{dL}{dC} \tag{3.39}$$

$$b_3{}^t = b_3{}^{t-1} - \alpha \frac{dL}{db_3} \tag{3.40}$$

$$V^t = V^{t-1} - \alpha \frac{dL}{dV} \tag{3.41}$$

$$b_2{}^t = b_2{}^{t-1} - \alpha \frac{dL}{db_2} \tag{3.42}$$

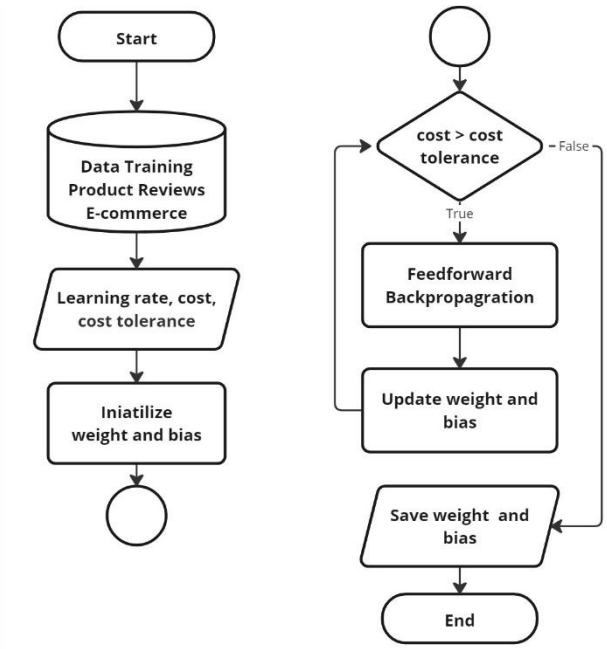
$$W^t = W^{t-1} - \alpha \frac{dL}{dW} \tag{3.43}$$

$$b_1{}^t = b_1{}^{t-1} - \alpha \frac{dL}{db_1} \tag{3.44}$$

Keterangan:
α adalah laju pembelajaran (*learning rate*).
$\frac{dL}{dD}$ adalah turunan bobot pada output layer.
$\frac{dL}{db_4}$ adalah turunan bias pada output layer.
$\frac{dL}{dC}$ adalah turunan bobot pada hidden layer 3.
$\frac{dL}{db_3}$ adalah turunan bias pada hidden layer 3.
$\frac{dL}{dV}$ adalah turunan bobot pada hidden layer 2.
$\frac{dL}{db_2}$ adalah turunan bias pada hidden layer 2.
$\frac{dL}{dW}$ adalah turunan bobot pada hidden layer 1.
$\frac{dL}{db_1}$ adalah turunan bias pada hidden layer 1.

Persamaan *update* parameter menggunakan turunan error untuk mengarahkan perubahan bobot dan bias, dan laju pembelajaran untuk mengontrol kecepatan perubahan tersebut.

Alur dari proses *training* ditunjukkan pada Gambar 3.11. Pada proses *training*, terjadi perhitungan antara nilai prediksi model dan *ground truth*. Proses tersebut menghasilkan nilai error dari model hasil proses *training*.



Gambar 3.12. *Flowchart* proses *training*

3.10 Eksperimen

Tahap eksperimen pada penilitian ini dilakukan dengan perangkat lokal untuk melakukan klasifikasi multilabel pada ulasan produk *e-commerce*. Device yang digunakan pada penelitian ini adalah Macbook Pro M4 dengan RAM 16GB. Pemodelan *deep neural network* serta proses pelatihan dilakukan menggunakan bahasa pemrograman Python dengan *library* TensorFlow sebagai kerangka kerja utama dalam membangun dan melatih model.

Eksperimen dalam penelitian ini berfokus pada perbandingan teknik *feature extraction* serta arsitektur Neural network. Hasil dari setiap eksperimen akan dievaluasi menggunakan metrik *hamming loss, accuracy, precision, recall, dan f1-score* untuk menentukan konfigurasi terbaik dalam pemrosesan ulasan *e-commerce.*

3.11 Evaluasi

Pada penelitian ini, hasil kinerja dari model akan dievaluasi menggunakan lima metrik evaluasi yang umum digunakan dalam klasifikasi, yaitu *hamming loss, f1-score, precission, recall*, dan *accuracy.*

1. *Hamming Loss*

Hamming loss mengukur rata-rata kesalahan prediksi pada setiap label dalam suatu sampel. Metrik ini menunjukkan proporsi label yang salah diklasifikasikan, baik itu label yang seharusnya ada tetapi tidak diprediksi (*false negative*), maupun label yang tidak seharusnya ada tetapi diprediksi (*false positive*). Semakin kecil nilai *hamming loss*, semakin baik performa model. *Hamming loss* didefinisikan sebagai berikut:

$$Hamming\ Loss = \frac{1}{NL}\sum_{i=1}^{N}\sum_{j=1}^{L} 1(y_{ij} \neq y'_{ij}) \tag{3.45}$$

Keterangan:
N adalah jumlah sampel data.
L adalah jumlah label.
y_{ij} adalah label asli untuk sampel ke-*i* dan label ke-*j*.
y'_{ij} adalah label prediksi untuk sampel ke-*i* dan label ke-*j*.
$1(y_{ij} \neq y'_{ij})$ adalah fungsi indikator yang bernilai 1 jika prediksi tidak sesuai dengan label asli, dan 0 jika sebaliknya.

2. *Recall*

Recall mengukur sejauh mana model mampu menemukan semua label yang relevan dalam dataset. Nilai *recall* yang tinggi menunjukkan bahwa model mampu mengidentifikasi hampir semua label yang relevan, tetapi mungkin mengorbankan *precision* jika terdapat banyak *false positive*.

$$Recall = \frac{True\ Positive}{True\ Positive\ +\ False\ Negative} \tag{3.46}$$

3. *Precision*

Precision metrik yang penting dalam klasifikasi *multilabel* karena mengukur seberapa banyak prediksi positif yang benar dibandingkan dengan total prediksi positif yang dibuat oleh model.

$$Precision = \frac{True\ Positive}{True\ Positive\ +\ False\ Positive} \tag{3.47}$$

4. *F1-score*

F1-score adalah metrik berbasis rata-rata harmonik dari *precision* dan *recall*. *f1-score* digunakan untuk mengukur keseimbangan antara *precision* dan *recall* dalam kasus klasifikasi, terutama saat terdapat ketidakseimbangan jumlah sampel antar label.

$$F1 - Score = 2\ \times \frac{Precision\ \times Recall}{Precision + Recall} \tag{3.48}$$

5. *Accuracy*

Accuracy adalah nilai persentase prediksi benar terhadap semua data.

$$Accuracy = \frac{TP + TN}{TN + TN + FP + FN} \tag{3.49}$$

Keterangan:
TP adalah *True Positive*.
TN adalah *True Negative*.
FP adalah *False Positive*.
FN adalah *False Negative*.

Pada evaluasi sentimen, penelitian ini akan menggunakan empat metrik utama, yaitu precision, recall, F1-score, dan accuracy. Sementara itu, pada evaluasi

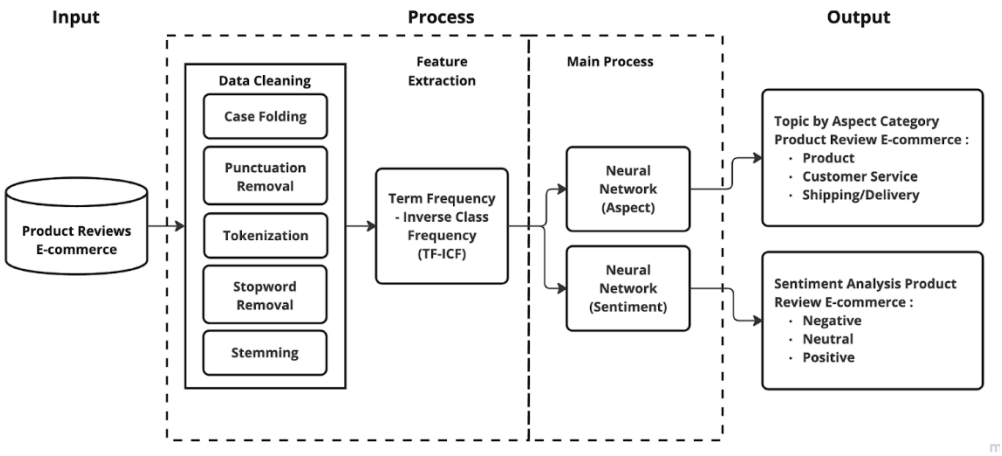
aspect topic modeling, penelitian ini menerapkan seluruh metrik tersebut ditambah hamming loss. Penambahan hamming loss pada *aspect topic modeling* diperlukan karena prediksi dan *ground truth* dapat memiliki lebih dari satu kelas (misalnya dua, tiga, atau bahkan tidak ada sama sekali). Dengan demikian, hamming loss membantu mengukur kesalahan klasifikasi pada skenario multilabel, sehingga evaluasi menjadi lebih komprehensif dan representatif terhadap karakteristik data.

BAB IV

MODEL NEURAL NETWORK BERBASIS TF-ICF

4.1 Desain Sistem

Desain sistem pada bab ini mengacu pada kerangka kerja terstruktur yang mengintegrasikan metode TF-ICF (Term Frequency - Inverse Category Frequency) dan Neural Network (NN) untuk mengkategorikan ulasan produk *e-commerce* dan menganalisis sentimen. Desain sistem dari model TF-ICF-NN ditunjukkan pada Gambar 4.1.



Gambar 4.1. Desain Sistem Model TF-ICF-NN

Alur sistem dimulai dengan menerima input data teks berupa ulasan produk dari platform *e-commerce*. Data mentah ini kemudian melalui tahap *preprocessing*. Tahap kunci sistem melibatkan ekstraksi fitur menggunakan metode TF-ICF, yang secara khusus dikembangkan untuk menghitung bobot kata dengan mempertimbangkan dua faktor utama: frekuensi kemunculan term dalam kategori tertentu (TF) dan tingkat kekhususan term terhadap kategori tersebut (ICF). Fitur

ini kemudian diproses oleh dua model NN terpisah: satu untuk klasifikasi kategori aspek dan satu lagi untuk analisis sentimen. Model NN pertama menghasilkan output kategori aspek (produk, layanan pelanggan, pengiriman). Sementara, model NN kedua menghasilkan output sentimen (negatif, netral, positif).

4.2 Ekstraksi Fitur TF-ICF

TF-ICF disini bertindak sebagai ekstraksi fitur yang nantinya akan menjadi input dari *classifier*. Sebelum melakukan transformasi TF-ICF, teks ulasan perlu menjalani *preprocessing* terlebih dahulu.

4.2.1. Hasil *Preprocessing*

Dalam penelitian ini, beberapa sampel ulasan produk e-commerce dipilih untuk mengilustrasikan proses transformasi TF-ICF. Seperti ditunjukkan pada Tabel 4.1.

Tabel 4.1. Sampel Ulasan Sebelum dan Sesudah *Preprocessing*

No	Ulasan Asli	Sesudah Preprocessing
1	Barang sudah diterima mudah 2 an bisa berfungsi bagus dan awet. terima kasih	barang terima mudah 2 an fungsi bagus awet terima kasih
2	barangnya bagus 👍 👍	barang bagus
3	Kualitas dan harga ok pengiriman cepat.	kualitas harga kirim cepat
4	Nice. Pacingnya rapi respon cepat barang juga cepat sampai :)	nice pacingnya rapi respon cepat barang cepat
…	…	…
4002	"dikirim jam 12.00 terima jam 17.00 Pengiriman cepat"	kirim jam 12 00 terima jam 17 00 kirim cepat

Hasil *preprocessing* pada Tabel 4.1 menunjukkan bahwa ulasan-ulasan tersebut telah berhasil diubah menjadi token-token kata yang memiliki makna

penting. Proses ini secara efektif menghilangkan noise dalam teks seperti kata sambung, tanda baca, dan kata-kata umum yang tidak memberikan kontribusi signifikan terhadap klasifikasi aspek maupun sentimen. Hasil preprocessing yang lebih sederhana ini akan memudahkan perhitungan dari TF-ICF.

4.2.2. Hasil Transformasi TF-ICF

Setelah *preprocessing*, setiap kata dalam ulasan ditransformasikan ke dalam representasi numerik menggunakan metode TF-ICF. Sebagaimana ditampilkan pada Tabel 4.2, perhitungan TF-ICF dilakukan dengan mengalikan Term Frequency (TF) dengan Inverse Class Frequency (ICF).

Tabel 4.2. Hasil Ekstraksi Fitur Menggunakan TF-ICF

No	**Teks Ulasan**	**Vektor TF-ICF**
1	Thanks admin, barangnya bagus	[0.0, 0.0, 0.0, … 1.81, 0.0, 0.0]
2	Good product min, mantap buat semir sepatu 👍	[1.55, 0.0, 0.0, … 0.0, 0.0, 0.0]
3	Nice. Pacingnya rapi respon cepat barang juga cepat sampai :)	[..., 1.007, 0.0, …, 0.69, 0.32, …]

Vektor TF-ICF yang dihasilkan merupakan representasi numerik dari ulasan yang berbeda dengan TF-IDF konvensional. Pada TF-ICF, pembobotan kata mempertimbangkan distribusinya di berbagai kelas aspek dan kelas sentimen, bukan distribusi di seluruh dokumen. Hal ini membuat kata-kata yang muncul di sedikit kelas aspek mendapat bobot lebih tinggi. Dalam penelitian ini, proses ekstraksi fitur menghasilkan vektor TF-ICF dengan dimensi sebesar 2.968. Representasi TF-ICF ini kemudian digunakan sebagai input untuk model Neural Network sesuai dengan alur desain sistem.

4.3 Topologi Neural Network pada Model TF-ICF-NN

4.3.1 Perhitungan Jumlah Node pada Hidden Layer

Penentuan jumlah node pada masing-masing hidden layer dilakukan menggunakan rumus matematis yang telah dijelaskan pada Bab 3, yaitu Persamaan 3.7. Persamaan ini digunakan untuk menghitung jumlah neuron berdasarkan ukuran input dari lapisan sebelumnya. Lebih detail, hasil perhitungan neuron ditunjukkan pada Tabel 4.3.

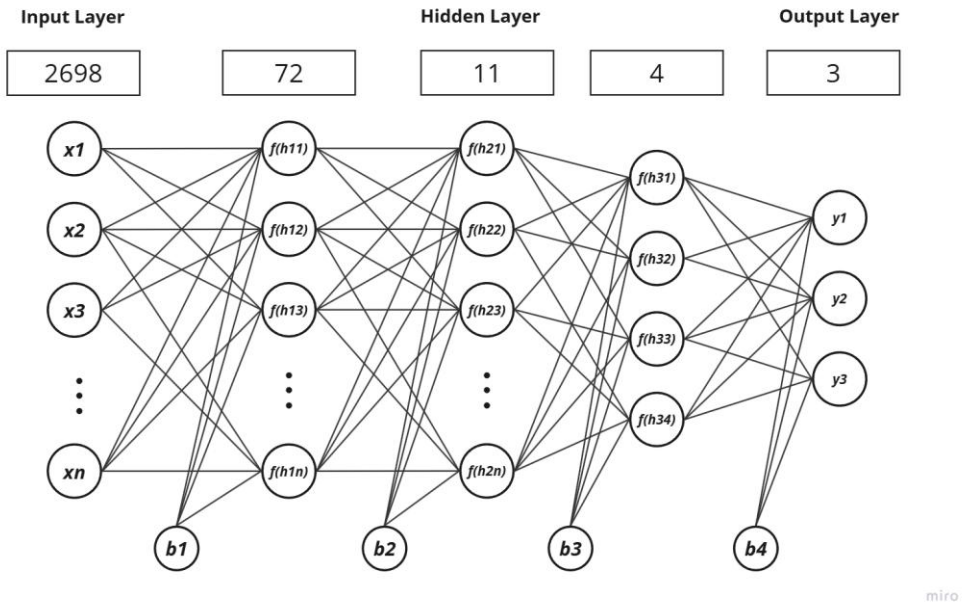
Tabel 4.3 Hasil Perhitungan Neuron pada Model TF-ICF-NN

Hidden Layer	Fitur Input (Ni)	Perhitungan Node	Jumlah Node (Nh)
1	2698	$\frac{\sqrt{1+8(2698)}}{2} - 1 = 72{,}45917233$	72
2	72	$\frac{\sqrt{1+8(72)}}{2} - 1 = 11{,}04858268$	11
3	11	$\frac{\sqrt{1+8(11)}}{2} - 1 = 3{,}727278854$	4

Perhitungan ini dilakukan secara berurutan dari layer pertama hingga ketiga, dan hasilnya digunakan secara konsisten pada kedua kasus klasifikasi, yaitu klasifikasi aspek dan analisis sentimen, karena struktur hidden layer pada kedua model bersifat identic atau sama persis. Hasil perhitungan jumlah node pada masing-masing hidden layer ditunjukkan pada Tabel 4.3. Hasil berupa bilangan desimal dibulatkan ke angka terdekat agar sesuai dengan arsitektur jaringan saraf tiruan.

4.3.2 Topologi Neural Network berbasis TF-ICF untuk *Aspect Topic Modelling*

Topologi neural network berbasis TF-ICF pada kasus klasifikasi *aspect topic* menggunakan vektor hasil transformasi TF-ICF sebagai input. Arsitektur neural network ini ditunjukkan pada Gambar 4.2.



Gambar 4.2 Topologi Neural Network dengan Input Vektor TF-ICF untuk *Aspect Topic Modeling*

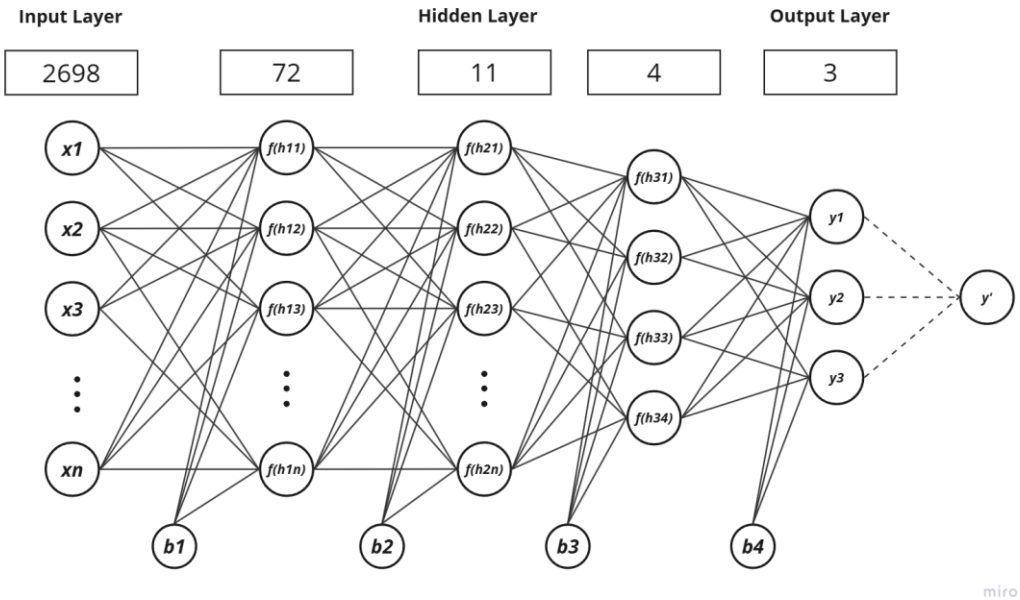
Pada Gambar 4.2 dapat dilihat bahwa panjang dimensi sebanyak 2.698 node pada input layer. Model ini memiliki tiga hidden layer yang masing-masing terdiri dari 72, 11, dan 4 node sesuai hasil perhitungan sebelumnya, serta dilengkapi dengan bias pada setiap layer. Output layer terdiri dari tiga neuron, masing-masing mewakili satu kategori aspek yang dapat muncul dalam ulasan produk, seperti aspek produk, aspek layanan pelanggan, dan aspek pengiriman.

Karena tugas klasifikasi ini bersifat multilabel, satu ulasan dapat memiliki lebih dari satu aspek. Oleh karena itu, digunakan fungsi aktivasi sigmoid pada

output layer, sehingga setiap neuron menghasilkan nilai probabilitas secara independen. Nilai ini kemudian dikonversi menjadi label akhir menggunakan threshold 0,5; output ≥ 0,5 dianggap aktif (1), dan < 0,5 dianggap tidak aktif (0). Dengan cara ini, model dapat mengidentifikasi lebih dari satu aspek dalam satu ulasan.

4.3.3 Topologi Neural Network berbasis TF-ICF untuk *Sentiment Analysis*

Topologi neural network berbasis TF-ICF pada kasus klasifikasi sentimen menggunakan vektor hasil transformasi TF-ICF sebagai input. Arsitektur neural network ini ditunjukkan pada Gambar 4.3



Gambar 4.3 Topologi Neural Network dengan Input Vektor TF-ICF untuk *Sentiment Analysis*

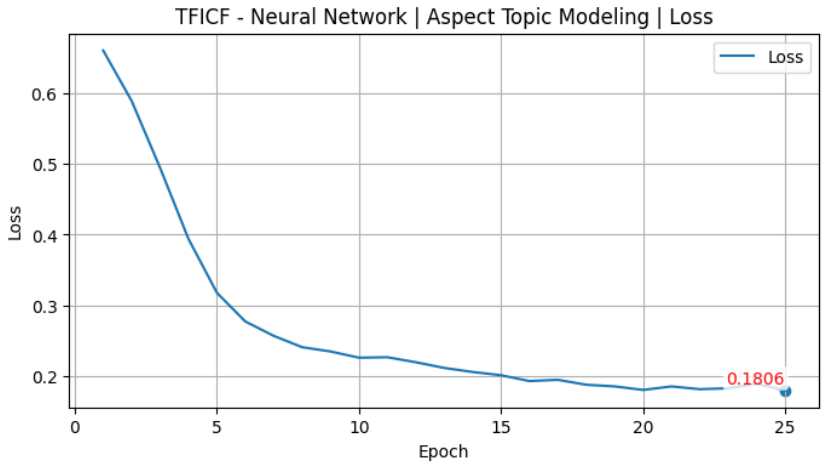
Secara umum, struktur topologi yang digunakan masih serupa, yakni terdiri dari input layer dengan 2.698 node yang merepresentasikan panjang dimensi vektor TF-ICF, tiga hidden layer dengan konfigurasi 72, 11, dan 4 neuron yang ditentukan

berdasarkan perhitungan pada Tabel 4.3, serta penggunaan bias pada setiap layer tersembunyi.

Perbedaan utama terletak pada bagian output layer. Pada kasus analisis sentimen, output layer hanya terdiri dari satu neuron, karena tugas klasifikasi yang dilakukan bersifat tunggal, Neuron output ini akan menghasilkan nilai probabilitas dari prediksi sentimen akhir yang kemudian diinterpretasikan sebagai kelas positif, netral, atau negatif melalui fungsi aktivasi yang sesuai.

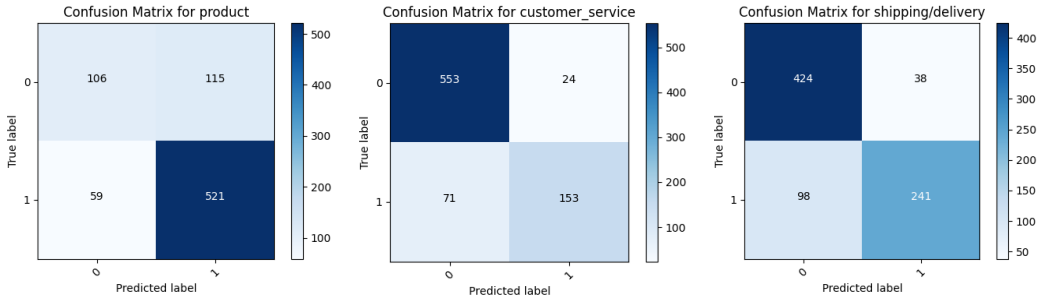
4.4 Hasil pada *Aspect Topic Modeling* TF-ICF-NN

Gambar 4.6 menunjukkan visualisasi perubahan loss selama proses pelatihan model TF-ICF-NN untuk *aspect topic modeling.*



Gambar 4.4. *Loss* pada *Training* Model TF-ICF-NN *Aspect Topic Modeling*

Pada Gambar 4.6, selama proses training, model TF-ICF-NN pada *aspect topic modeling* menunjukkan penurunan nilai loss yang signifikan. Model mencapai nilai loss terendah sebesar 0,180 pada epoch ke-25. Grafik loss mengindikasikan proses training berjalan dengan baik dan model telah mencapai korvergen.



Gambar 4.5. *Confusion Matrix* pada Model TF-ICF-NN *Aspect Topic Modeling*

Hasil *confusion matrix* model TF-ICF-NN pada data *testing* ditunjukkan pada Gambar 4.7. Pada aspek produk, model memprediksi dengan benar sebanyak 521 data yang termasuk kategori aspek produk (*true positive*) dan 106 data yang bukan aspek produk (*true negative*), sementara 174 data diprediksi secara tidak tepat (115 *false positive* dan 59 *false negative*). Untuk aspek ulasan pelanggan, model mengidentifikasi 553 *true negative* dan 153 *true positive*, dengan 95 prediksi yang tidak tepat. Sementara itu, confusion matrix untuk aspek pengiriman menunjukkan 424 *true negative* dan 241 *true positive*, dengan kesalahan prediksi sebanyak 136 data. Nilai-nilai performa dari model TF-ICF-NN pada *aspect topic modeling* ini kemudian ditunjukkan pada Tabel 4.5.

Tabel 4.4. Metrik Performa pada Model TF-ICF-NN *Aspect Topic Modeling*

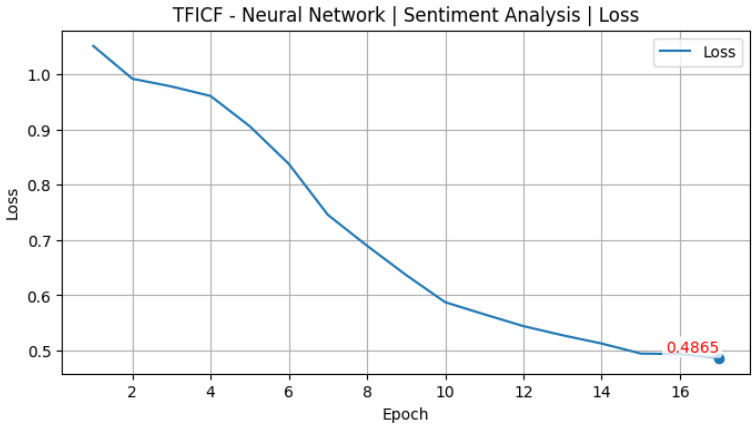
Model	**Accuracy**	**Precision**	**Recall**	**F1-score**	**Hamming Loss**
TF-ICF-NN *aspect topic modeling*	0,604245	0,849129	0,764075	0,799978	0,168539

Model TF-ICF-NN pada tugas *aspect topic modeling* menunjukkan performa yang cukup baik dengan precision sebesar 84,91%, yang mengindikasikan bahwa sebagian besar prediksi aspek yang dilakukan oleh model

adalah benar. Namun, nilai recall sebesar 76,41% menunjukkan bahwa masih ada sejumlah aspek yang relevan namun belum berhasil terdeteksi oleh model. Nilai F1-score sebesar 79,99% mencerminkan keseimbangan yang moderat antara precision dan recall. Sementara itu, accuracy berada di angka 60,42%, yang cenderung rendah mengingat konteks tugas multilabel, dan hamming loss sebesar 0,1685 menunjukkan tingkat kesalahan prediksi per label yang masih cukup signifikan.

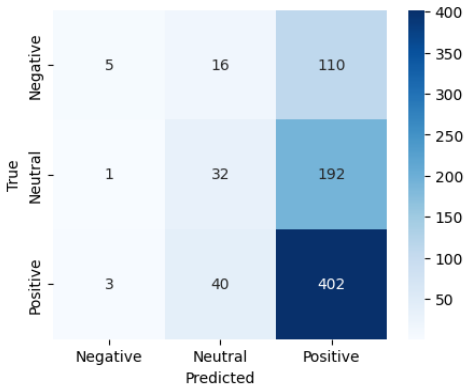
4.5 Hasil pada *Sentiment Analysis* TF-ICF-NN

Gambar 4.8 menunjukkan visualisasi perubahan loss selama proses pelatihan model TF-ICF-NN untuk *sentiment analysis*.



Gambar 4.6. *Loss* pada *Training* Model TF-ICF-NN *Sentiment Analysis*

Pada Gambar 4.8, selama proses training, model TF-ICF-NN pada *sentiment analysis* menunjukkan penurunan nilai loss yang signifikan. Model mencapai nilai loss terendah sebesar 0,4865 pada epoch ke-17. Grafik loss mengindikasikan proses training berjalan dengan baik dan model telah mencapai korvergen.



Gambar 4.7. *Confusion Matrix* pada Model TF-ICF-NN *Sentiment Analysis*

Hasil *confusion matrix* model TF-ICF-NN untuk analisis sentimen pada data *testing* ditunjukkan pada Gambar 4.9. Untuk kategori sentimen positif, model memprediksi dengan benar sebanyak 402 data (true positive) namun salah

mengklasifikasikan 43 data (3 diprediksi sebagai negatif dan 40 sebagai netral). Pada kategori sentimen netral, model mengidentifikasi dengan tepat 32 data, sementara 193 data diprediksi secara tidak tepat. Untuk kategori sentimen negatif, model hanya mampu mengklasifikasikan dengan benar 5 data, sedangkan 126 data lainnya terprediksi secara tidak akurat. Nilai-nilai perforna dari model TF-ICF-NN pada *sentiment analysis* ini kemudian ditunjukkan pada Tabel 4.6.

Tabel 4.5. Metrik Performa pada Model TF-ICF-NN *Sentiment Analysis*

Model	Accuracy	Precision	Recall	F1-score
TF-ICF-NN *sentiment analysis*	0,548064	0,496738	0,361253	0,325213

Model TF-ICF-NN pada tugas *sentiment analysis* menunjukkan performa yang tergolong rendah. Nilai accuracy sebesar 54,81% menunjukkan bahwa lebih dari separuh prediksi model sesuai dengan label sebenarnya, namun hal ini belum mencerminkan kinerja yang memadai dalam klasifikasi sentimen. Hal ini juga terlihat dari precision yang hanya mencapai 49,67%, serta recall sebesar 36,13%, yang menandakan bahwa banyak sentimen yang tidak berhasil terdeteksi. Nilai F1-score yang rendah, yakni 32,52%, menegaskan ketidakseimbangan antara kemampuan model dalam mengidentifikasi dan memprediksi sentimen dengan benar. Temuan ini mengindikasikan bahwa representasi fitur berbasis TF-ICF kurang efektif dalam menangkap nuansa sentimen dalam teks ulasan.

4.6 Kesimpulan Model TF-ICF-NN

Berdasarkan hasil eksperimen yang telah dilakukan, model TF-ICF-NN menunjukkan performa yang bervariasi. Pada tugas *aspect topic modeling*, model TF-ICF-NN mampu menghasilkan kinerja yang cukup baik dengan nilai precision

sebesar 84,91%, recall sebesar 76,41%, dan F1-score sebesar 79,99%. Hasil ini mengindikasikan bahwa model cukup efektif dalam mengenali dan memetakan aspek-aspek utama dalam ulasan e-commerce secara simultan. Nilai hamming loss sebesar 0,1685 juga memperlihatkan bahwa tingkat kesalahan prediksi per label masih dalam batas yang dapat diterima untuk kasus multilabel. Hal ini menunjukkan bahwa representasi fitur berbasis TF-ICF memiliki kemampuan yang memadai dalam menangkap informasi topikal dari teks ulasan.

Sebaliknya, pada tugas *sentiment analysis*, model menunjukkan performa yang relatif rendah. Dengan nilai accuracy 54,81%, precision 49,67%, recall 36,13%, dan F1-score 32,52% menandakan bahwa model kesulitan dalam mengidentifikasi polaritas sentimen secara akurat. Hal ini kemungkinan disebabkan oleh karakteristik TF-ICF yang hanya berbasis frekuensi dan tidak menangkap konteks semantik atau nuansa emosional dari kata, yang penting dalam analisis sentimen. Confusion matrix juga menunjukkan ketimpangan dalam klasifikasi antar kelas sentimen, terutama pada kategori netral dan negatif yang memiliki akurasi sangat rendah.

Secara keseluruhan, dapat disimpulkan bahwa model TF-ICF-NN lebih cocok untuk tugas klasifikasi aspek (multilabel) dibandingkan dengan tugas analisis sentimen (multiclass). Hal ini menegaskan bahwa pemilihan metode representasi fitur yang sesuai sangat berpengaruh terhadap keberhasilan model dalam menyelesaikan jenis klasifikasi tertentu.

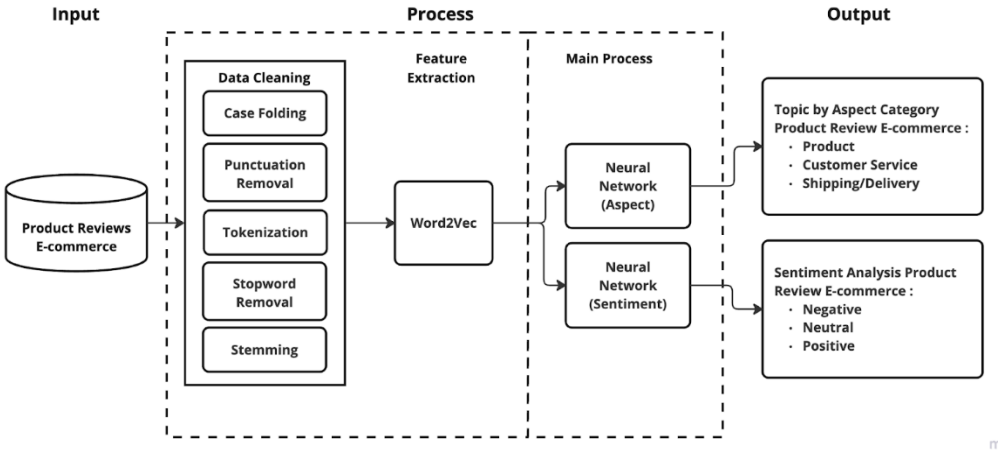
BAB V

MODEL NEURAL NETWORK BERBASIS WORD2VEC

5.1 Desain Sistem

Model Word2Vec-NN mengimplementasikan model Word2Vec sebagai word embedding yang akan digunakan untuk klasifikasi aspek dan analisis sentimen ulasan produk *e-commerce*. Penelitian ini akan menggunakan model Word2Vec pretrained yang dilatih menggunakan Wikipedia Bahasa Indonesia. Model ini dikenal dengan nama idwiki200, yang akan menghasilkan output *word embedding* berdimensi 200.

Seperti ditunjukkan pada Gambar 5.1, sistem ini terdiri dari tahap *preprocessing*, ekstraksi fitur dengan Word2Vec, dan klasifikasi menggunakan Neural Network.



Gambar 5.1. Desain Sistem Model Word2Vec-NN

Tahap *preprocessing* data ulasan mengikuti langkah-langkah yang sama seperti pada implementasi model TF-ICF sebelumnya. Teks yang telah melalui *preprocessing* selanjutnya ditransformasi menjadi vektor menggunakan Word2Vec. Pada Word2Vec, setiap kata diubah menjadi vektor angka.

Sama seperti pada desain sistem TF-ICF, setelah transformasi vektor, data diproses oleh dua model Neural Network: satu untuk klasifikasi aspek dan satu lagi untuk analisis sentimen. Model klasifikasi aspek mengkategorikan ulasan ke dalam tiga kelas: produk, layanan pelanggan, dan pengiriman. Sementara itu, model analisis sentimen mengklasifikasikan ulasan berdasarkan polaritas emosinya menjadi tiga kategori: negatif, netral, dan positif.

5.2 Word2Vec Embedding

Sebelum melakukan ekstraksi fitur menggunakan Word2Vec, teks ulasan *e-commerce* perlu melalui *preprocessing* terlebih dahulu.

5.2.1. Hasil *Preprocessing*

Serupa seperti sebelumnya, 4002 dataset ulasan *e-commerce* menjalani proses cleaning hingga stemming. Contoh hasil *preprocessing* teks ulasan dapat dilihat di Tabel 4.1 pada Bab 4. Hasil *preprocessing* ini berupa kumpulan kata-kata penting yang telah melalui proses *case folding, punctuation removal, tokenization, stopword removal,* dan *stemming*. Teks ulasan yang sudah dipreprocessing ini kemudian menjadi input untuk proses Word2Vec embedding.

5.2.2. Hasil Word2Vec Embedding

Teks yang telah melalui *preprocessing* menjadi inputan untuk proses *embedding* Word2Vec. Setiap kata dalam ulasan akan diproses oleh Word2Vec dan menghasilkan *word embedding* (vektor angka) untuk setiap kata. Sehingga dalam satu kalimat ulasan akan memiliki beberapa vektor angka berdimensi 200 yang disebut *word embedding*. Kemudian, kumpulan *word embedding* ini dilakukan *max pooling* sesuai pada Persamaan 3.6 sehingga menjadi *text embedding*. Tabel 5.1 menunjukkan beberapa contoh hasil teks ulasan yang telah direpresentasikan menjadi vektor Word2Vec.

Tabel 5.1. Hasil Embedding Menggunakan Word2Vec

No	Teks Ulasan	Vektor Word2Vec
1	Barang sudah diterima mudah 2 an bisa berfungsi bagus dan awet. terima kasih	[-0.322, 0.447, 0.061, …, -0.312, 0.199]
2	barangnya bagus 👍 👍	[-1.367, -0.767, -0.83, …, -1.157, 0.875]
3	Kualitas dan harga ok pengiriman cepat.	[-1.27, -1.089, -0.370, …, -0.405, 0.200]
4	Nice. Pacingnya rapi respon cepat barang juga cepat sampai :)	[-0.468, -0.685, 0.464, …, -0.656, 0.427]
…	…	…
5	"dikirim jam 12.00 terima jam 17.00 Pengiriman cepat"	[0.059, -0.302, 0.442, …, -0.598, -0.510]

Seperti ditunjukkan pada Tabel 5.1, setiap ulasan direpresentasikan sebagai vektor numerik dengan panjang 200. Representasi vektor ini kemudian menjadi input untuk model Neural Network dalam proses klasifikasi aspek dan analisis sentimen.

5.3 Topologi *Neural Network* pada Word2Vec-NN

5.3.1 Perhitungan Jumlah Node pada Hidden Layer

Penentuan jumlah node pada masing-masing hidden layer dilakukan menggunakan rumus matematis yang telah dijelaskan pada Bab 3, yaitu Persamaan 3.7. Persamaan ini digunakan untuk menghitung jumlah neuron berdasarkan ukuran input dari lapisan sebelumnya. Lebih detail, hasil perhitungan perhitungan neuron dapat dilihat pada Tabel 5.2.

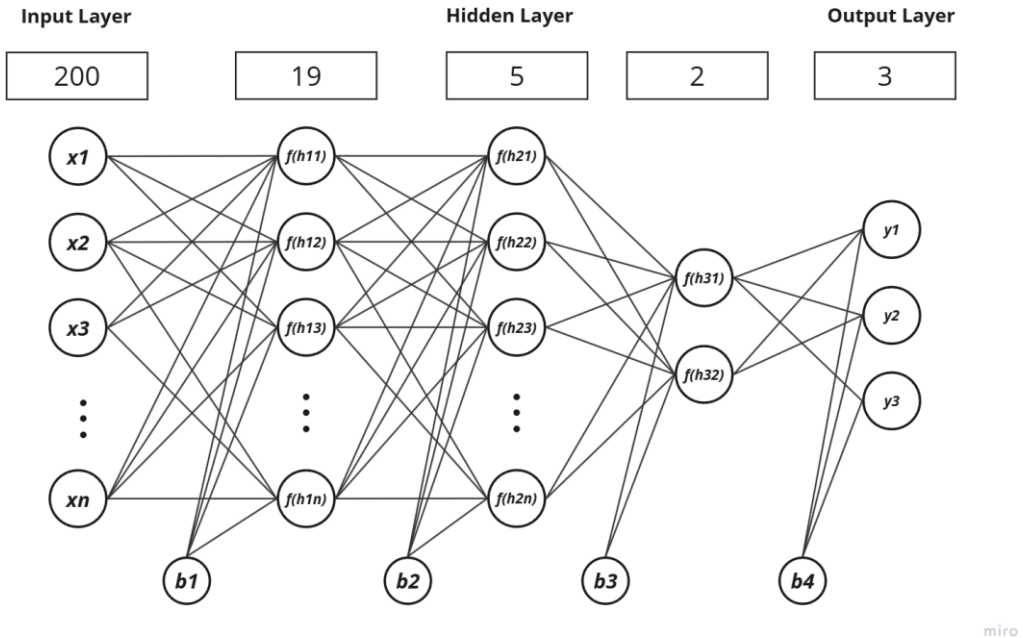
Tabel 5.2. Hasil Perhitungan Neuron pada Model Word2Vec-NN

Hidden Layer	Fitur Input (Ni)	Perhitungan	Jumlah Node (Nh)
1	200	$\frac{\sqrt{1+8(200)}}{2} - 1 = 19{,}00624902$	19
2	19	$\frac{\sqrt{1+8(19)}}{2} - 1 = 5{,}185668763$	5
3	5	$\frac{\sqrt{1+8(5)}}{2} - 1 = 2{,}259039356$	2

Perhitungan ini dilakukan secara berurutan dari layer pertama hingga ketiga, dan hasilnya digunakan secara konsisten pada kedua kasus klasifikasi, yaitu klasifikasi aspek dan analisis sentimen, karena struktur hidden layer pada kedua model bersifat identik atau sama persis. Hasil perhitungan jumlah node pada masing-masing hidden layer ditunjukkan pada Tabel 5.2. Hasil berupa bilangan desimal dibulatkan ke angka terdekat agar sesuai dengan arsitektur jaringan saraf tiruan.

5.3.2 Topologi Neural Network berbasis Word2Vec untuk *Aspect Topic Modelling*

Topologi neural network berbasis Word2Vec pada kasus klasifikasi *aspect topic* menggunakan vektor hasil transformasi Word2Vec sebagai input. Arsitektur neural network ini ditunjukkan pada Gambar 5.2.



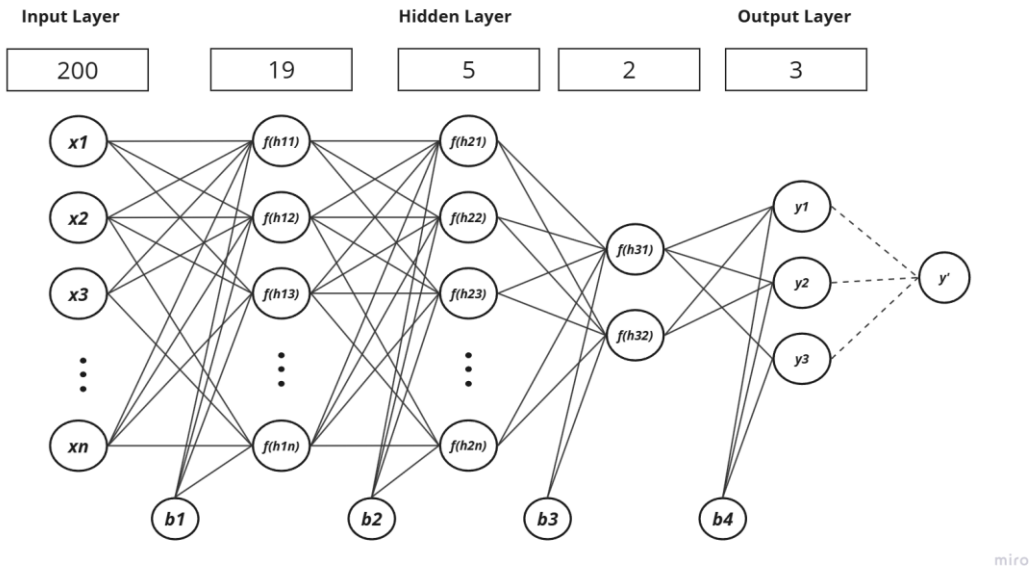
Gambar 5.2 Topologi Neural Network dengan Input Vektor Word2Vec untuk *Aspect Topic Modeling*

Pada Gambar 5.2 dapat dilihat bahwa panjang dimensi sebanyak 200 node pada input layer. Model ini memiliki tiga hidden layer yang masing-masing terdiri dari 19, 5, dan 2 node sesuai hasil perhitungan sebelumnya, serta dilengkapi dengan bias pada setiap layer. Output layer terdiri dari tiga neuron, masing-masing mewakili satu kategori aspek yang dapat muncul dalam ulasan produk, seperti aspek produk, aspek layanan pelanggan, dan aspek pengiriman.

Karena tugas klasifikasi ini bersifat multilabel, satu ulasan dapat memiliki lebih dari satu aspek. Oleh karena itu, digunakan fungsi aktivasi sigmoid pada output layer, sehingga setiap neuron menghasilkan nilai probabilitas secara independen. Nilai ini kemudian dikonversi menjadi label akhir menggunakan threshold 0,5; output ≥ 0,5 dianggap aktif (1), dan < 0,5 dianggap tidak aktif (0). Dengan cara ini, model dapat mengidentifikasi lebih dari satu aspek dalam satu ulasan.

5.3.3 Topologi Neural Network berbasis Word2Vec untuk *Sentiment Analysis*

Topologi neural network berbasis Word2Vec pada kasus klasifikasi sentimen menggunakan vektor hasil transformasi Word2Vec sebagai input. Arsitektur neural network ini ditunjukkan pada Gambar 5.3



Gambar 5.3 Topologi Neural Network dengan Input Vektor Word2Vec untuk *Sentiment Analysis*

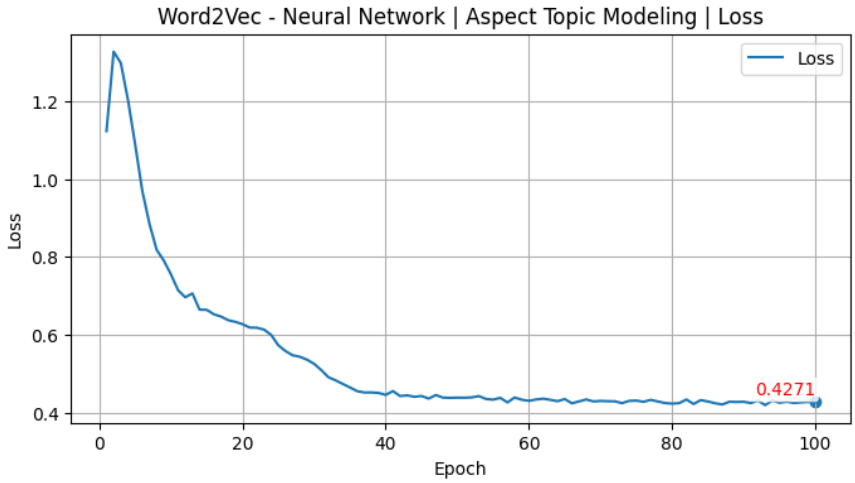
Secara umum, struktur topologi yang digunakan masih serupa, yakni terdiri dari input layer dengan 200 node yang merepresentasikan panjang dimensi vektor

Word2Vec, tiga hidden layer dengan konfigurasi 19, 5, dan 2 neuron yang ditentukan berdasarkan perhitungan pada Tabel 5.2, serta penggunaan bias pada setiap layer tersembunyi.

Perbedaan utama terletak pada bagian output layer. Pada kasus analisis sentimen, output layer hanya terdiri dari satu neuron, karena tugas klasifikasi yang dilakukan bersifat tunggal, Neuron output ini akan menghasilkan nilai probabilitas dari prediksi sentimen akhir yang kemudian diinterpretasikan sebagai kelas positif, netral, atau negatif melalui fungsi aktivasi yang sesuai.

5.4 Hasil pada *Aspect Topic Modeling* Word2Vec-NN

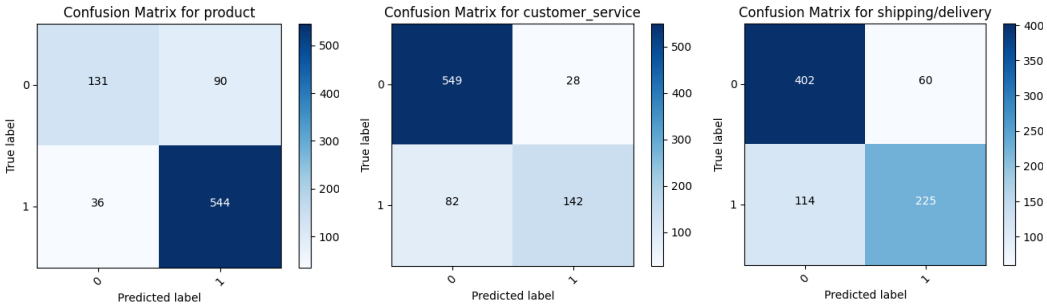
Gambar 5.5 menunjukkan visualisasi perubahan loss selama proses pelatihan model Word2Vec-NN untuk *aspect topic modeling.*



Gambar 5.4. *Loss* pada *Training* Model Word2Vec-NN *Aspect Topic Modeling*

Pada Gambar 5.5, selama proses training, model Word2Vec-NN pada *aspect topic modeling* menunjukkan penurunan nilai loss yang signifikan. Model mencapai

nilai loss terendah sebesar 0,4271 pada epoch ke-100. Grafik loss mengindikasikan proses training berjalan dengan baik dan model telah mencapai korvergen.



Gambar 5.5. *Confusion Matrix* pada Model Word2Vec-NN *Aspect Topic Modeling*

Hasil *confusion matrix* model Word2Vec-NN pada data *testing* ditunjukkan pada Gambar 5.6. Pada aspek produk, model memprediksi dengan benar sebanyak 544 data yang termasuk kategori aspek produk (*true positive*) dan 131 data yang bukan aspek produk (*true negative*), sementara 126 data diprediksi secara tidak tepat (90 *false positive* dan 36 *false negative*). Untuk aspek layanan pelanggan, model mengidentifikasi 549 *true negative* dan 142 *true positive*, dengan 110 prediksi yang tidak tepat (28 *false positive* dan 82 *false negative*). Sementara itu, *confusion matrix* untuk aspek pengiriman menunjukkan 402 *true negative* dan 225 *true positive*, dengan kesalahan prediksi sebanyak 174 data (60 *false positive* dan 114 *false negative*). Nilai-nilai performa dari model Word2Vec-NN pada *aspect topic modeling* ini kemudian ditunjukkan pada Tabel 5.3.

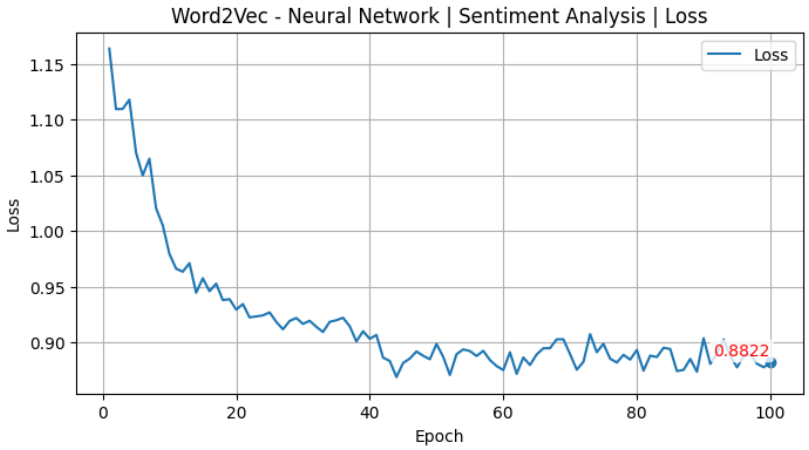
Tabel 5.3. Metrik Performa pada Model TF-ICF-NN *Aspect Topic Modeling*

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**	**Hamming Loss**
Word2Vec-NN *aspect topic modeling*	0,594257	0,827604	0,745192	0,779392	0,170620

Model Word2Vec-NN pada tugas *aspect topic modeling* menunjukkan performa yang cukup baik meski belum optimal. Dengan *accuracy* sebesar 59,43%, model mampu mengenali sebagian besar label aspek dengan benar. Nilai precision mencapai 82,76%, menunjukkan bahwa prediksi yang diberikan cukup tepat, dan recall sebesar 74,52% menandakan masih ada label yang terlewat. F1-score sebesar 77,93% mencerminkan keseimbangan antara precision dan recall, sementara hamming loss sebesar 0,1706 menunjukkan masih adanya kesalahan prediksi. Jika dibandingkan dengan model TF-ICF-NN, performa Word2Vec-NN ini relatif serupa namun belum menunjukkan peningkatan signifikan.

5.5 Hasil pada *Sentiment Analysis* Word2Vec-NN

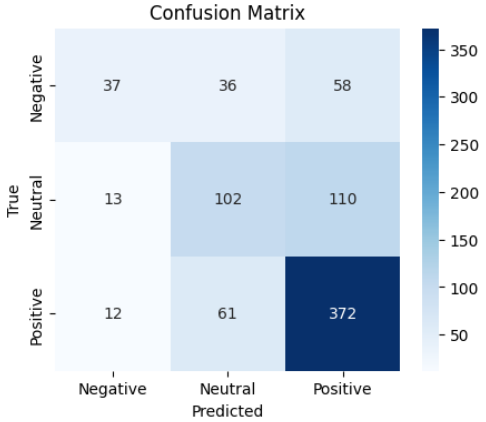
Gambar 5.7 menunjukkan visualisasi perubahan loss selama proses pelatihan model Word2Vec-NN untuk *sentiment analysis.*



Gambar 5.6. *Loss* pada *Training* Model Word2Vec-NN *Sentiment Analysis*

Pada Gambar 5.7, selama proses training, model Word2Vec-NN pada *sentiment analysis* menunjukkan penurunan nilai loss yang signifikan. Model mencapai nilai loss terendah sebesar 0,8822 pada epoch ke-100. Grafik loss

mengindikasikan proses training berjalan dengan baik dan model telah mencapai korvergen.



Gambar 5.7. *Confusion Matrix* pada Model Word2Vec-NN *Sentiment Analysis*

Hasil *confusion matrix* model Word2Vec-NN untuk analisis sentimen pada data *testing* ditunjukkan pada Gambar 5.8. Untuk kategori sentimen positif, model berhasil memprediksi dengan benar sebanyak 372 data (*true positive*) namun salah mengklasifikasikan 73 data (12 diprediksi sebagai negatif dan 61 sebagai netral). Pada kategori sentimen netral, model mengidentifikasi dengan tepat 102 data, sementara 123 data diprediksi secara tidak tepat (13 sebagai negatif dan 110 sebagai positif). Untuk kategori sentimen negatif, model mampu mengklasifikasikan dengan benar 37 data, sedangkan 94 data lainnya terprediksi secara tidak akurat (36 sebagai netral dan 58 sebagai positif). Nilai-nilai perforna dari model Word2Vec-NN pada *sentiment analysis* ini kemudian ditunjukkan pada Tabel 5.4.

Tabel 5.4. Metrik Performa pada Model Word2Vec-NN *Sentiment Analysis*

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
Word2Vec-NN *sentiment analysis*	0,637952	0,599408	0,523910	0,539960

Model Word2Vec-NN untuk tugas *sentiment analysis* memperoleh accuracy sebesar 63,80%, yang menunjukkan bahwa model dapat mengklasifikasikan sentimen secara umum dengan cukup baik. Nilai precision sebesar 59,94% dan recall sebesar 52,39% mengindikasikan bahwa model cenderung cukup tepat dalam memprediksi kelas yang benar, namun masih melewatkan sejumlah data relevan. F1-score sebesar 53,99% menunjukkan bahwa keseimbangan antara precision dan recall masih belum optimal. Jika dibandingkan dengan model TF-ICF-NN, Word2Vec-NN memberikan peningkatan performa pada seluruh metrik, menunjukkan representasi kata yang lebih baik dalam menangkap konteks sentimen.

5.6 Kesimpulan Model Word2Vec-NN

Berdasarkan hasil eksperimen yang dilakukan, model Word2Vec-NN menunjukkan kinerja yang cukup baik, meskipun belum secara signifikan melebihi model sebelumnya (TF-ICF-NN), terutama pada tugas klasifikasi aspek.

Pada tugas *aspect topic modeling*, model Word2Vec-NN mendapatkan nilai precision 82,76%, recall 74,52%, dan F1-score 77,93%. Meskipun nilai-nilai ini mencerminkan performa yang seimbang dan layak, hasilnya relatif sebanding dengan model TF-ICF-NN. Nilai hamming loss sebesar 0,1706 menunjukkan bahwa model masih melakukan kesalahan prediksi per label, meskipun dalam jumlah yang tidak terlalu besar. Dengan demikian, dapat disimpulkan bahwa penggunaan Word2Vec sebagai representasi fitur tidak memberikan keunggulan yang signifikan dalam klasifikasi aspek dibandingkan TF-ICF.

Sebaliknya, pada tugas *sentiment analysis*, model Word2Vec-NN menunjukkan peningkatan performa yang cukup jelas dibandingkan model TF-ICF-NN. Model ini berhasil mencapai accuracy 63,80%, precision 59,94%, recall 52,39%, dan F1-score 53,99%. Peningkatan ini mengindikasikan bahwa Word2Vec lebih efektif dalam menangkap konteks semantik dan nuansa sentimen dalam ulasan teks, dibandingkan dengan TF-ICF yang hanya berbasis frekuensi kata. Hasil confusion matrix juga menunjukkan distribusi prediksi yang lebih merata, meskipun kategori sentimen negatif masih sulit diklasifikasikan secara akurat.

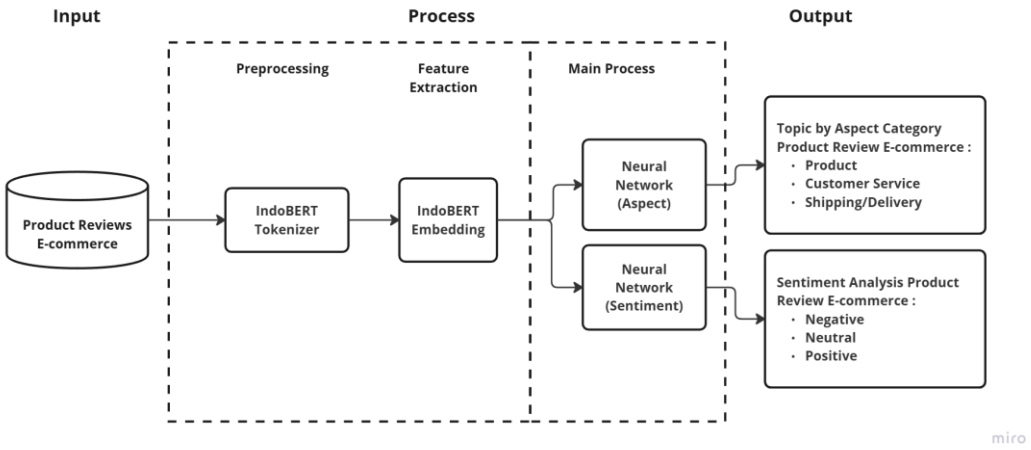
Secara keseluruhan, dapat disimpulkan bahwa model Word2Vec-NN lebih unggul dibandingkan TF-ICF-NN dalam tugas analisis sentimen, namun tidak memberikan peningkatan berarti pada klasifikasi aspek. Hal ini menunjukkan bahwa efektivitas representasi kata bergantung pada jenis tugas klasifikasi yang dijalankan. Untuk tugas-tugas yang menuntut pemahaman konteks dan makna emosional, Word2Vec merupakan pilihan yang lebih tepat dibandingkan metode berbasis frekuensi seperti TF-ICF.

BAB VI

MODEL NEURAL NETWORK BERBASIS INDOBERT

6.1 Desain Sistem

Strategi ketiga dalam penelitian ini menggunakan model IndoBERT yang dikombinasikan dengan Neural Network untuk klasifikasi aspek dan analisis sentimen ulasan produk *e-commerce*. Desain sistem dari model IndoBERT-NN ditunjukkan pada Gambar 6.1.



Gambar 6.1. Desain Sistem Model IndoBERT-NN

Sebagaimana ditunjukkan pada Gambar 6.1, desain sistem IndoBERT-NN terdiri dari tiga komponen utama: *preprocessing*, ekstraksi fitur, dan proses klasifikasi. Berbeda dengan pendekatan TF-ICF dan Word2Vec yang memerlukan serangkaian tahapan preprocessing yang kompleks, sistem IndoBERT-NN hanya menggunakan IndoBERT Tokenizer untuk *preprocessing* teks. BERT Tokenizer bertugas mengubah teks ulasan menjadi token-token yang sesuai dengan kosakata IndoBERT, termasuk penambahan token khusus seperti [CLS] di awal dan [SEP]

di akhir kalimat. Setelah tokenisasi, tahap ekstraksi fitur dilakukan dengan IndoBERT Embedding yang mengubah setiap token menjadi representasi vektor.

Pada proses klasifikasi, representasi vektor IndoBERT diumpankan ke dalam dua model Neural Network sama seperti model-model pada bab sebelumnya. Model pertama berfokus pada klasifikasi aspek, mengategorikan ulasan ke dalam tiga kategori: produk, layanan pelanggan, dan pengiriman. Model kedua melakukan analisis sentimen, mengklasifikasikan ulasan berdasarkan polaritas emosinya menjadi positif, negatif, dan netral.

6.2 IndoBERT Embedding

Pada proses ekstraksi fitur dengan menggunakan IndoBERT, teks mentah ulasan melalui tahap *preprocessing* menggunakan IndoBERT Tokenizer terlebih dahulu.

6.2.1. Hasil *Preprocessing*

Tahap *preprocessing* pada IndoBERT mencakup pemisahan teks menjadi *wordpiece* tokens, penambahan token khusus seperti [CLS] di awal dan [SEP] di akhir kalimat, serta penanganan kata yang tidak ada dalam kosakata BERT dengan memecahnya menjadi subword. Transformasi teks ulasan sesudah menjalani *preprocessing* menggunakan BERT Tokenizer ditunjukkan pada Tabel 6.1.

Tabel 6.1. Sampel Ulasan Sebelum dan Sesudah *Preprocessing*

No	Ulasan Asli	Sesudah Tokenizer
1	Barang sudah diterima mudah 2 an bisa berfungsi bagus dan awet. terima kasih	[2, 963, 1788, 783, … , 1788, 1338, 3]
2	barangnya bagus 👍 👍	[2, 963, 1305, 3]
3	Kualitas dan harga ok pengiriman cepat.	[2, 1553, 405, 2785, 972, 3]

4	Nice. Pacingnya rapi respon cepat barang juga cepat sampai :)	[2, 17514, 2905, 8152, …, 972, 3]
…	…	…
4002	"dikirim jam 12.00 terima jam 17.00 Pengiriman cepat"	[2, 2785, 727, 981, …, 2785, 972, 3]

Hasil *preprocessing* ini kemudian akan diubah menjadi representasi vektor menggunakan model IndoBERT, yang akan digunakan sebagai input untuk Neural Network dalam proses klasifikasi aspek dan analisis sentimen.

6.2.2. Hasil IndoBERT Embedding

Setelah melalui proses tokenisasi, setiap token dalam ulasan ditransformasikan menjadi representasi vektor kontekstual oleh model IndoBERT. Untuk keperluan klasifikasi, representasi dari token [CLS] diambil sebagai fitur yang merepresentasikan keseluruhan ulasan. Hasil representasi vektor IndoBERT disajikan pada Tabel 6.2.

Tabel 6.2. Hasil Embedding Menggunakan IndoBERT

No	**Teks Ulasan**	**Vektor BERT**
1	Barang sudah diterima mudah 2 an bisa berfungsi bagus dan awet. terima kasih	[1.557, 1.594, 0.251, …, 0.724, 0.067]
2	barangnya bagus 👍 👍	[1.422, 1.232, 0.507, …, 0.377, 0.848]
3	Kualitas dan harga ok pengiriman cepat.	[1.379, 1.630, 0.102, …, 0.488, -0.267]
4	Nice. Pacingnya rapi respon cepat barang juga cepat sampai :)	[1.753, 1.634, -0.263, …, 0.290, 0.676]
…	…	..
4002	"dikirim jam 12.00 terima jam 17.00 Pengiriman cepat"	[1.082, 1.411, 0.239, …, 0.197, 0.233]

Seperti ditunjukkan pada Tabel 6.2, vektor IndoBERT merepresentasikan ulasan dalam bentuk numerik multidimensi, dengan 768 dimensi. Vektor embedding IndoBERT inilah yang kemudian digunakan sebagai input untuk model Neural Network dalam proses klasifikasi aspek dan analisis sentimen ulasan produk *e-commerce*.

6.3 Topologi *Neural Network* pada IndoBERT-NN

6.3.1 Perhitungan Jumlah Node pada Hidden Layer

Penentuan jumlah node pada masing-masing hidden layer dilakukan menggunakan rumus matematis yang telah dijelaskan pada Bab 3, yaitu Persamaan 3.7. Persamaan ini digunakan untuk menghitung jumlah neuron berdasarkan ukuran input dari lapisan sebelumnya. Lebih detail, hasil perhitungan perhitungan neuron dapat dilihat pada Tabel 6.3.

Tabel 6.3 Hasil Perhitungan Neuron pada Model IndoBERT-NN

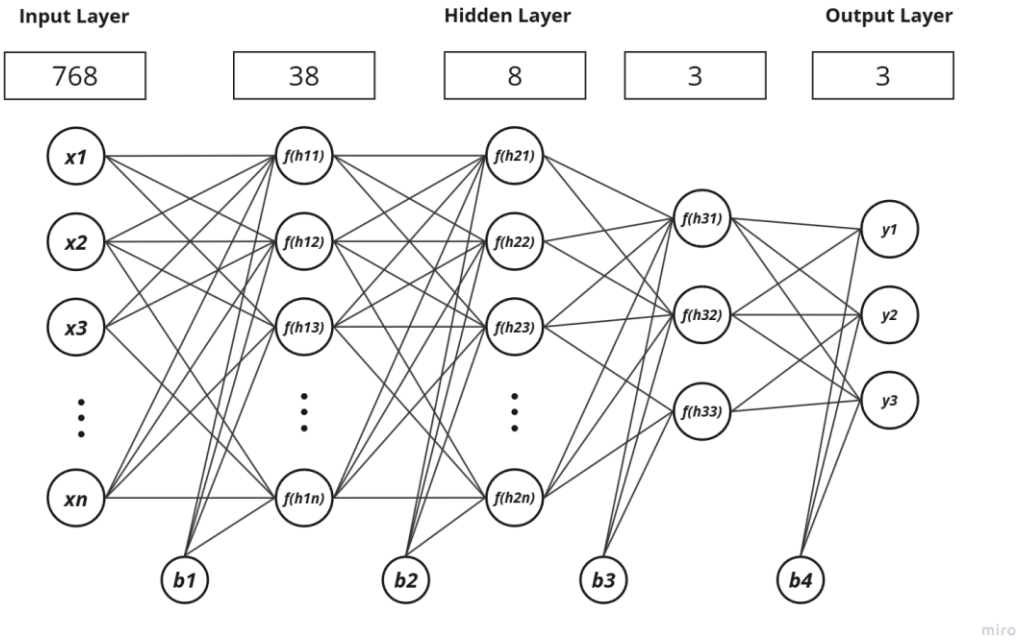
Hidden Layer	Fitur Input (Ni)	Perhitungan	Jumlah Neuron (Nh)
1	768	$\frac{\sqrt{1+8(200)}}{2} - 1 = 38{,}19502519$	38
2	38	$\frac{\sqrt{1+8(200)}}{2} - 1 = 7{,}754430329$	8
3	8	$\frac{\sqrt{1+8(200)}}{2} - 1 = 2{,}969743148$	3

Perhitungan ini dilakukan secara berurutan dari layer pertama hingga ketiga, dan hasilnya digunakan secara konsisten pada kedua kasus klasifikasi, yaitu

klasifikasi aspek dan analisis sentimen, karena struktur hidden layer pada kedua model bersifat identik atau sama persis. Hasil perhitungan jumlah node pada masing-masing hidden layer ditunjukkan pada Tabel 6.3. Hasil berupa bilangan desimal dibulatkan ke angka terdekat agar sesuai dengan arsitektur jaringan saraf tiruan.

6.3.2 Topologi Neural Network berbasis IndoBERT untuk *Aspect Topic Modeling*

Topologi neural network berbasis IndoBERT pada kasus klasifikasi aspect topic menggunakan vektor hasil transformasi IndoBERT sebagai input. Arsitektur neural network ini ditunjukkan pada Gambar 6.2.



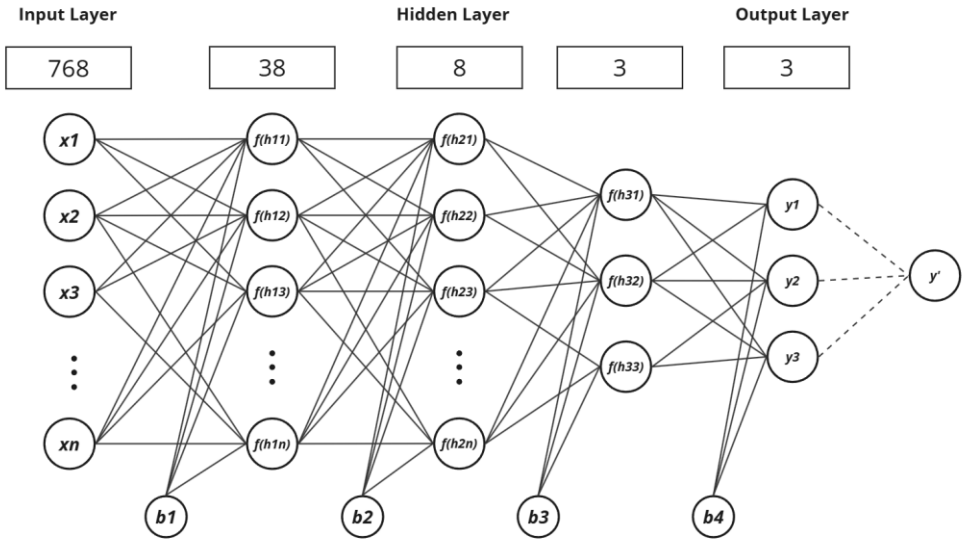
Gambar 6.2 Topologi Neural Network dengan Input Vektor IndoBERT embedding untuk *Aspect Topic Modeling*

Pada Gambar 6.2 dapat dilihat bahwa panjang dimensi sebanyak 768 node pada input layer. Model ini memiliki tiga hidden layer yang masing-masing terdiri dari 38, 8, dan 3 node sesuai hasil perhitungan sebelumnya, serta dilengkapi dengan bias pada setiap layer. Output layer terdiri dari tiga neuron, masing-masing mewakili satu kategori aspek yang dapat muncul dalam ulasan produk, seperti aspek produk, aspek layanan pelanggan, dan aspek pengiriman.

Karena tugas klasifikasi ini bersifat multilabel, satu ulasan dapat memiliki lebih dari satu aspek. Oleh karena itu, digunakan fungsi aktivasi sigmoid pada output layer, sehingga setiap neuron menghasilkan nilai probabilitas secara independen. Nilai ini kemudian dikonversi menjadi label akhir menggunakan threshold 0,5; output $\geq 0{,}5$ dianggap aktif (1), dan $< 0{,}5$ dianggap tidak aktif (0). Dengan cara ini, model dapat mengidentifikasi lebih dari satu aspek dalam satu ulasan.

6.3.3 Topologi Neural Network berbasis IndoBERT untuk *Sentiment Analysis*

Topologi neural network berbasis IndoBERT pada kasus klasifikasi sentimen menggunakan vektor hasil transformasi IndoBERT sebagai input. Arsitektur neural network ini ditunjukkan pada Gambar 6.3.



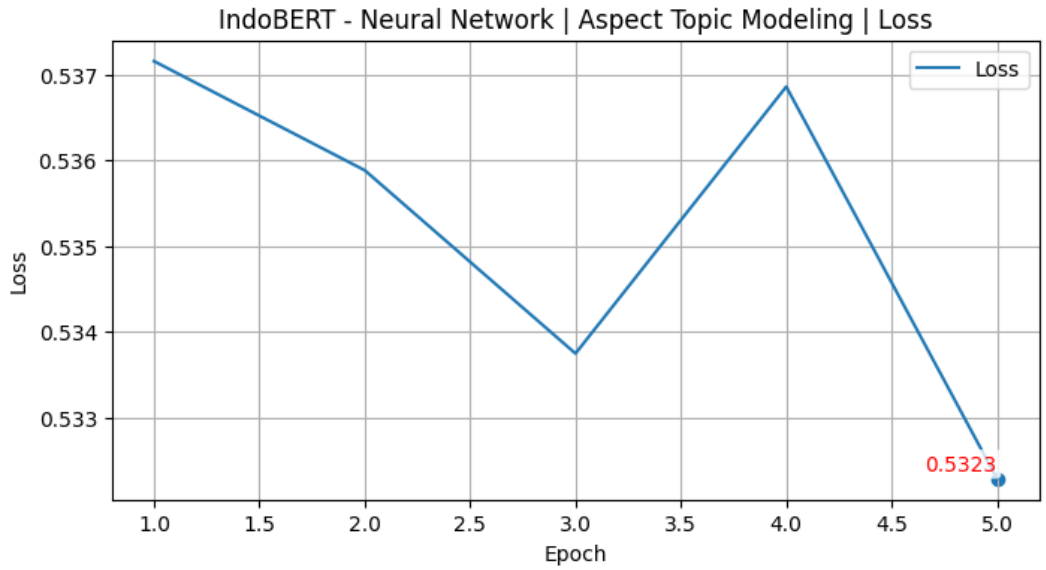
Gambar 6.3 Topologi Neural Network dengan Input Vektor IndoBERT untuk *Sentiment Analysis*

Secara umum, struktur topologi yang digunakan masih serupa, yakni terdiri dari input layer dengan 768 node yang merepresentasikan panjang dimensi vektor Word2Vec, tiga hidden layer dengan konfigurasi 38, 8, dan 3 neuron yang ditentukan berdasarkan perhitungan pada Tabel 6.3, serta penggunaan bias pada setiap layer tersembunyi.

Perbedaan utama terletak pada bagian output layer. Pada kasus analisis sentimen, output layer hanya terdiri dari satu neuron, karena tugas klasifikasi yang dilakukan bersifat tunggal, Neuron output ini akan menghasilkan nilai probabilitas dari prediksi sentimen akhir yang kemudian diinterpretasikan sebagai kelas positif, netral, atau negatif melalui fungsi aktivasi yang sesuai.

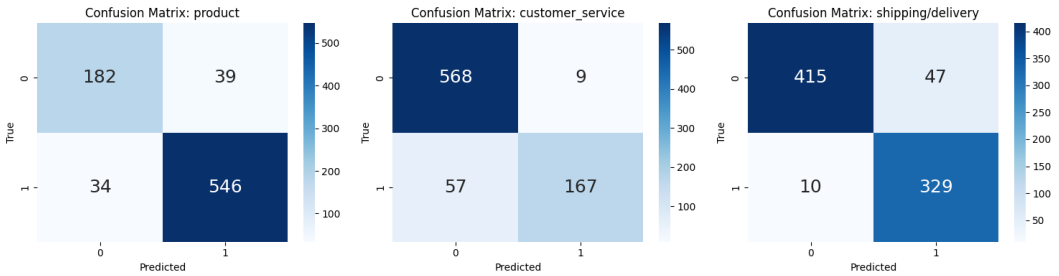
6.4 Hasil pada *Aspect Topic Modeling* IndoBERT-NN

Gambar 6.4 menunjukkan visualisasi perubahan loss selama proses pelatihan model IndoBERT-NN untuk *aspect topic modeling.*



Gambar 6.4. *Loss* pada *Training* Model IndoBERT-NN *Aspect Topic Modeling*

Pada Gambar 6.4, selama proses training, model Word2Vec-NN pada *aspect topic modeling* mencapai nilai loss terendah sebesar 0,5323 pada epoch ke-5. Pada model IndoBERT tidak memerlukan epoch yang banyak.



Gambar 6.5. *Confusion Matrix* pada Model IndoBERT-NN *Aspect Topic Modeling*

Hasil *confusion matrix* model IndoBERT-NN pada data *testing* ditunjukkan pada Gambar 6.5. Pada aspek produk, model memprediksi dengan benar sebanyak

546 data yang termasuk kategori aspek produk (*true positive*) dan 182 data yang bukan aspek produk (*true negative*), sementara 73 data diprediksi secara tidak tepat (39 *false positive* dan 34 *false negative*). Untuk aspek layanan pelanggan, model mengidentifikasi 568 *true negative* dan 167 *true positive*, dengan 66 prediksi yang tidak tepat (9 *false positive* dan 57 *false negative*). Sementara itu, *confusion matrix* untuk aspek pengiriman menunjukkan 415 *true negative* dan 329 *true positive*, dengan kesalahan prediksi sebanyak 57 data (47 *false positive* dan 10 *false negative*). Nilai-nilai performa dari model IndoBERT-NN pada *aspect topic modeling* ini kemudian ditunjukkan pada Tabel 6.5.

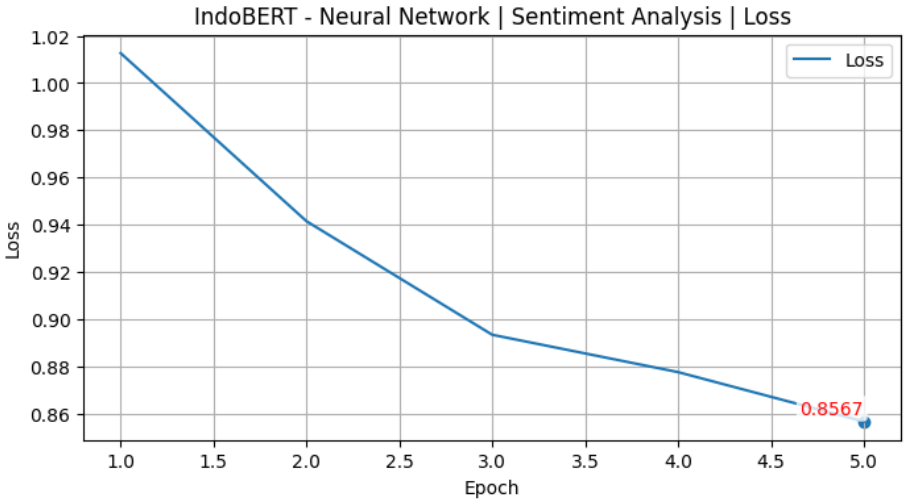
Tabel 6.4. Metrik Performa pada Model TF-ICF-NN *Aspect Topic Modeling*

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**	**Hamming Loss**
IndoBERT-NN *aspect topic modeling*	0,813983	0,914912	0,912511	0,913710	0,081981

Model IndoBERT-NN menunjukkan performa terbaik pada tugas *aspect topic modeling*, dengan accuracy mencapai 81,40% serta nilai precision (91,49%), recall (91,25%), dan f1-score (91,37%) yang sangat tinggi dan seimbang. Selain itu, nilai hamming loss yang rendah sebesar 0,0819 menegaskan kemampuan generalisasi model terhadap data multilabel. Hasil ini mencerminkan keunggulan representasi kontekstual dari IndoBERT dalam memahami makna kata-kata dalam ulasan berbahasa Indonesia, serta efektivitasnya dalam mengekstraksi informasi aspek secara lebih akurat dibandingkan pendekatan sebelumnya.

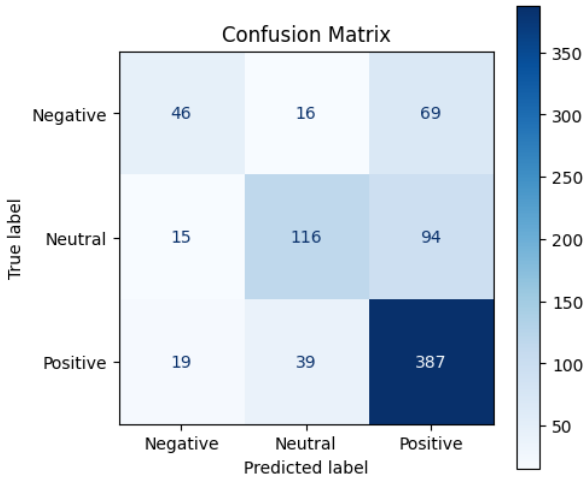
6.5 Hasil pada *Sentiment Analysis* IndoBERT-NN

Gambar 6.6 menunjukkan visualisasi perubahan loss selama proses pelatihan model IndoBERT-NN untuk *sentiment analysis.*



Gambar 6.6. *Loss* pada *Training* Model IndoBERT-NN *Sentiment Analysis*

Pada Gambar 6.6, selama proses training, model IndoBERT-NN pada *aspect topic modeling* mencapai nilai loss terendah sebesar 0,8567 pada epoch ke-5. Grafik selalu menurun menunjukkan proses training berjalan dengan baik.



Gambar 6.7. *Confusion Matrix* pada Model IndoBERT-NN *Sentiment Analysis*

Hasil *confusion matrix* model IndoBERT-NN untuk analisis sentimen pada data *testing* ditunjukkan pada Gambar 6.7. Untuk kategori sentimen positif, model berhasil memprediksi dengan benar sebanyak 387 data (*true positive*) namun salah mengklasifikasikan 58 data (19 diprediksi sebagai negatif dan 39 sebagai netral). Pada kategori sentimen netral, model mengidentifikasi dengan tepat 116 data, sementara 109 data diprediksi secara tidak tepat (15 sebagai negatif dan 94 sebagai positif). Untuk kategori sentimen negatif, model mampu mengklasifikasikan dengan benar 46 data, sedangkan 85 data lainnya terprediksi secara tidak akurat (16 sebagai netral dan 69 sebagai positif). Nilai-nilai perforna dari model IndoBERT-NN pada *sentiment analysis* ini kemudian ditunjukkan pada Tabel 6.6.

Tabel 6.5. Metrik Performa pada Model IndoBERT-NN *Sentiment Analysis*

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
IndoBERT-NN *sentiment analysis*	0,686641	0,653842	0,579536	0,600825

Model IndoBERT-NN memperoleh performa tertinggi dibandingkan model lainnya dalam klasifikasi sentimen, dengan accuracy sebesar 68,66% dan f1-score sebesar 60,08%. Meskipun metrik precision (65,38%) dan recall (57,95%) belum tergolong tinggi, nilai-nilai tersebut tetap menunjukkan adanya peningkatan signifikan dibandingkan pendekatan berbasis TF-ICF dan Word2Vec. Hasil ini menunjukkan bahwa representasi kontekstual dari IndoBERT mampu menangkap nuansa sentimen dalam bahasa Indonesia lebih baik.

6.6 Kesimpulan Model IndoBERT-NN

Model IndoBERT-NN menunjukkan performa terbaik dibandingkan dua pendekatan sebelumnya (TF-ICF-NN dan Word2Vec-NN) pada kedua tugas klasifikasi yang diuji, yaitu *aspect topic modeling* dan *sentiment analysis*.

Pada tugas *aspect topic modeling*, model IndoBERT-NN berhasil mencapai nilai metrik yang tinggi dan seimbang, dengan precision 91,49%**,** recall 91,25%**,** dan F1-score 91,37%**.** Nilai accuracy mencapai 81,40%**,** jauh lebih tinggi dibandingkan model-model sebelumnya, dan hamming loss sebesar 0,0819 menunjukkan tingkat kesalahan per label yang rendah. Hal ini menunjukkan bahwa representasi kontekstual dari IndoBERT sangat efektif dalam mengenali berbagai aspek dalam satu ulasan secara bersamaan, serta mampu melakukan generalisasi dengan baik terhadap data yang tidak dilatih.

Sementara itu, pada tugas *sentiment analysis*, IndoBERT-NN juga memberikan performa terbaik dengan accuracy 68,66%**,** precision 65,38%**,** recall 57,95%**,** dan F1-score 60,08%**.** Meskipun metrik recall masih menunjukkan adanya ruang untuk perbaikan, nilai-nilai tersebut secara keseluruhan jauh lebih baik dibandingkan dengan pendekatan berbasis TF-ICF dan Word2Vec. Hasil ini menunjukkan bahwa IndoBERT mampu memahami konteks kalimat secara lebih dalam, sehingga lebih akurat dalam menangkap polaritas dan nuansa sentimen dalam bahasa Indonesia.

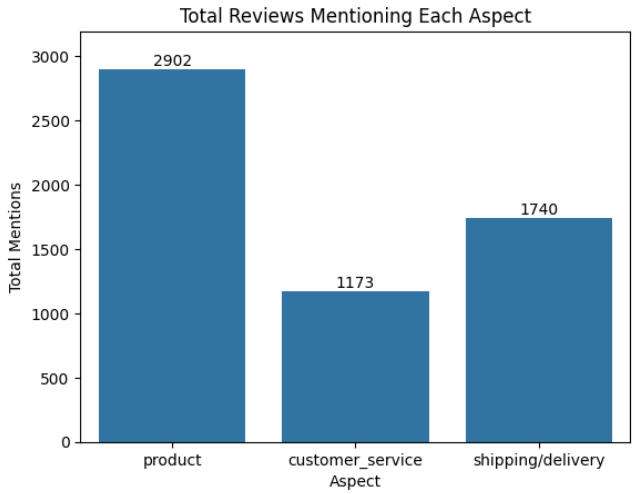
Dengan demikian, dapat disimpulkan bahwa model IndoBERT-NN merupakan pendekatan paling unggul di antara ketiga model yang dibandingkan dalam penelitian ini. Kemampuan IndoBERT dalam menangani konteks dan makna

kata secara *bidirectional* terbukti memberikan dampak signifikan terhadap peningkatan performa model, baik untuk klasifikasi aspek maupun analisis sentimen.

BAB VII

PEMBAHASAN

Sebelum melakukan analisis mendalam terhadap performa model, perlu dipahami distribusi dataset yang menjadi dasar penelitian ini. Dataset yang digunakan untuk membentuk model-model aspek dan sentimen terdiri dari 4.002 baris data yang telah dilabeli. Rincian distribusi jumlah ulasan untuk tiap-tiap aspek pada model *aspect topic modeling* ditampilkan pada Gambar 7.1.

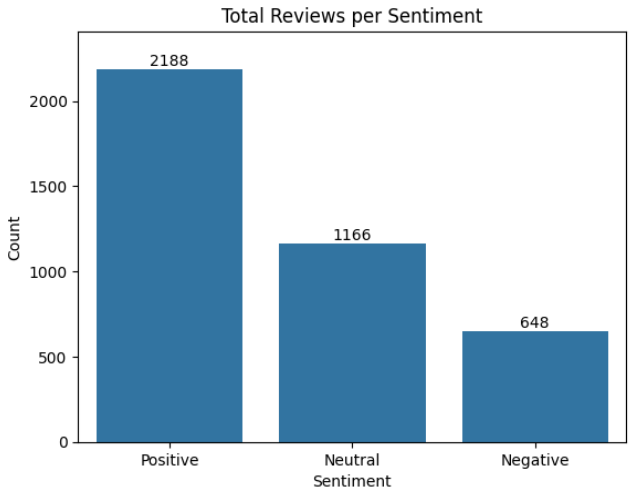


Gambar 7.1. Distribusi Ulasan Berdasarkan Kategori Aspek

Berdasarkan Gambar 7.1, terlihat bahwa dataset memiliki distribusi yang tidak seimbang antar kategori aspek. Aspek produk mendominasi dengan 2.902 ulasan, diikuti aspek pengiriman/pengantaran dengan 1.740 ulasan, dan aspek layanan pelanggan dengan 1.173 ulasan. Hal ini terjadi salah satunya karena objek dari penelitian ini adalah ulasan produk pada e-commerce, dimana e-commerce merupakan tempat jual beli sebuah produk sehingga pada umumnya pelanggan akan memberikan *feedback* atau ulasan mengenai produk yang dibeli. Kondisi

demikian menyebabkan aspek layanan pelanggan (customer service) dan aspek pengiriman (shipping/delivery) tidak akan disebutkan oleh pelanggan kecuali apabila terjadi hal yang sangat menonjol pada aspek tersebut, baik dalam konteks pengalaman yang positif maupun negatif. Sebagai contoh, pelanggan cenderung menyebutkan layanan pelangan apabila mereka mengalami pelayanan yang luar biasa atau justru sangat mengecewakan. Begitu juga dengan aspek pengiriman, umumnya hanya disebutkan jika pelanggan merasa pengiriman dari barang sangat cepat melebihi ekspektasi pelanggan atau justru terjadi keterlambatan. Oleh karena itu, frekuensi kemunculan ulasan terkait aspek layanan pelanggan dan aspek pengiriman censderung lebih sedikit dibandingkan dengan aspek produk yang hampir selalu menjadi fokus utama dalam setiap transaksi pembelian.

Sementara itu, untuk rincian distribusi ulasan pada kategori sentimen ditunjukkan melalui grafik pada Gambar 7.2.



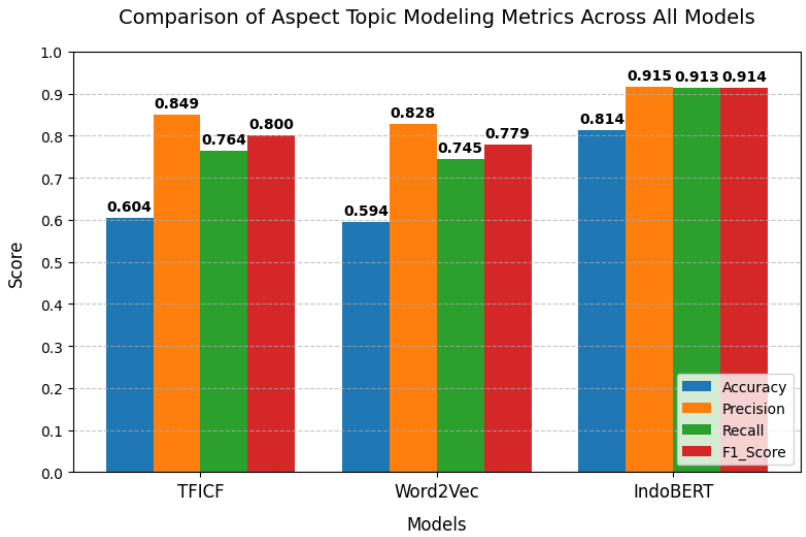
Gambar 7.2. Distribusi Ulasan Berdasarkan Kategori Sentimen

Pada Gambar 7.2 memperlihatkan distribusi sentimen dengan dominasi sentimen positif sebanyak 2.188 ulasan, diikuti sentimen netral 1.166 ulasan, dan sentimen negatif hanya 648 ulasan. Ketidakseimbangna ini merupakan fenomena umum dalam domain e-commerce. Salah satu penyebabnya adalah perilaku konsumen yang cenderung menghindari toko dengan reputasi buruk. Dalam praktiknya, ulasan negatif dari pelanggan sebelumnya sangat mempengaruhi keputusan pembelian pelanggan berikutnya dimasa depan (Amien *et al*., 2021; Stephen J. Payne & Andrew Howes, 2013). Akibatnya, toko yang sering menerima ulasan negative berpotensi mengalami penurunan penjualan dan kehilangan reputasi didalam platform e-commerce. Sebaliknya, toko yang menerima ulasan positif lebih banyak cenderung menarik lebih banyak pembeli, yang kemudian meningkatkan peluang untuk menerima ulasan positif. Mekanisme ini menciptakan siklus yang memperkuat dominan ulasan positif dalam jangka panjang.

Berdasarkan analisis distribusi kategori aspek dan sentiment, dapat disimpulkan bahwa dataset yang digunakan penelitian ini bersifat tidak seimbang atau imbalanace. Kategori aspek didominasi oleh aspek produk, sementara kategori sentiment positif. Ketidakseimbangan ini dapat memengaruhi perilaku model dalam proses pelatihan dan testing. Secara umum, model cenderung memiliki bias terhadap kelas yang dominan. Hal ini akan menyebabkan performa prediksi menurun pada kelas yang tidak dominan.

Setelah memahami distribusi data yang digunakan, langkah selanjutnya adalah mengevalusi performa masing-masing model dalam melakukan klasifikasi aspek. Penelitian ini menguji tiga pendekatan berbeda pada feature extraction, yaitu

TF-ICF, Word2Vec, dan IndoBERT. Evaluasi dilakukan dengan menggunakan lima metrik utama, yaitu *accuracy, precision, recall, f1-score,* dan *hamming loss.*



Gambar 7.3. Perbandingan Performa Model pada *Aspect Topic Modeling*

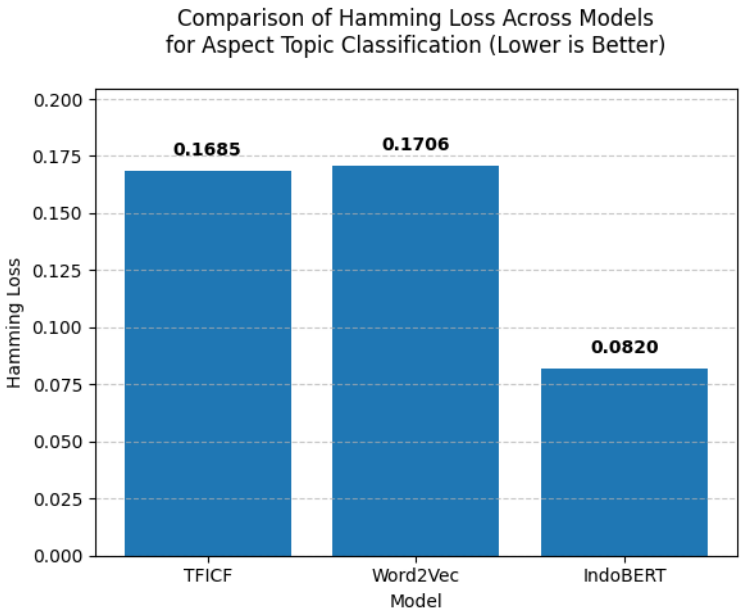
Pada Gambar 7.3 menampilkan hasil evaluasi dari ketiga model berdasarkan metrik *accuracy, precision, recall,* dan *f1-score* pada data testing. Model IndoBERT mendominasi performa yang sangat signifikan dari model lainnya. IndoBERT mendapatkan nilai accuracy 81.4%, precission 91.5%, recall 91,3%, dan f1-score sebesar 91,4%. Hal ini menunjukkan bahwa model IndoBERT memiliki representasi semantic kata kata yang baik dan daya generalisasi yang kuat, sehingga *classifier* dapat lebih mudah dan efektif.

Sebagai perbandingan, model TF-ICF dan Word2Vec menunjukkan performa yang relatif lebih rendah dibandingkan model IndoBERT pada keempat metrik tersebut. Dalam penelitian ini TF-ICF memiliki hasil precission tertinggi kedua, yaitu sebesar 84,5%. Namun, recall yang didapatkan hanya sebesar 76,41%. Selisih besar antara precission dan recall ini menunjukkan bahwa model ini

cenderung hanya memberikan prediksi positif ketika model sangat yakin, hal ini mengakibatkan model mengorbankan recall. Dalam kasus mengidentifikasi aspek, hal ini berpotensi mengakibatkan informasi penting hilang jika model tidak cukup responsif terhadap input. Model Word2vec menunjukkan pola yang serupa dengan precision 82,76% dan recall 74,51%. Model ini masih sedikit lebih rendah dibanding TF-ICF pada semua metrik. Hal ini disebabkan karena kualitas embedding dari Word2Vec bersifat umum dan mungkin kurang representatif terhadap konteks domain e-commerce dalam dataset ini.

Jika dilihat dengan seksama, semua model memiliki *accuracy* yang jauh lebih rendah dibanding metrik lainnya. IndoBERT memiliki *accuracy* sebesar 81,39% dimana metrik *precision, recall,* dan *f1-score* sudah diatas 90%. Pada TF-ICF, *accuracy* hanya 60,42% dimana metrik *precision, recall,* dan *f1-score* sudah diatas 76%. Pola yang sama juga terjadi pada Word2Vec, *accuracy* hanya mencapai 59,4% sedangkan metrik lainnya sudah diatas 74%. Konsistensi gap semua model pada metrik accuracy dengan metrik yang lain disebabkan oleh cara perhitungan accuracy yang ketat. Accuracy dalam konteks ini hanya akan menghitung prediksi benar jika seluruh kategori yang relevan dalam satu data ulasan teridentifikasi secara benar. Artinya, jika ada salah satu saja kategori yang salah teridentifikasi, maka prediksi tersebut dianggap salah sepenuhnya, meskipun sebagian kategori sudah benar. Hal ini menyebabkan nilai *accuracy* terlihat lebih rendah dibandingkan metrik *precision, recall,* dan *f1-score* dimana metrik-metrik tersebut menghitung prediksi berdasarkan per kategori atau *label-wise.* Dengan begitu, metrik *precision, recall,* dan *f1-score* menjadi lebih toleran terhadap kesalahan.

Dalam kasus pemodelan aspek ini, metrik *hamming loss* juga digunakan untuk mengevaluasi performa model secara keseluruhan. Metrik ini mengukur proporsi kesalahan prediksi terhadap jumlah total label yang ada. Semakin kecil nilai haming loss, semakin baik kinerja model dalam memprediksi label yang relevan.Visualisasi nilai hamming loss untuk masing-masing model ditampilkan pada Gambar 7.4.

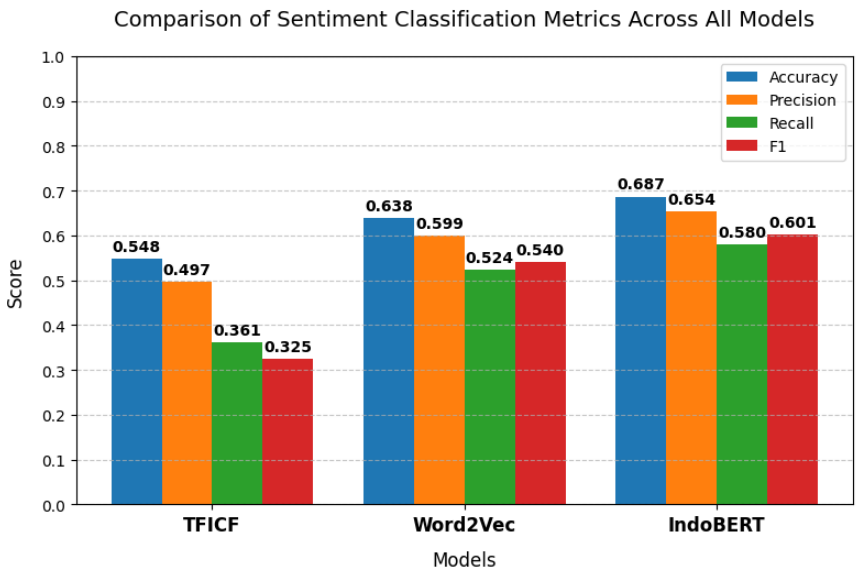


Gambar 7.4. Perbandingan Hamming Loss Semua Model pada *Aspect Topic Modeling*

Terlihat pada Gambar 7.4 model IndoBERT menunjukkan performa tebaik dengan nilai hamming loss terendah sebesar 0.0820. Nilai ini menandakan bahwa model IndoBERT melakukan kesalahan prediksi sekitar 8,2% dari total label yang seharusnya diprediksi. Sedangkan model TF-ICF dan Word2Vec memiliki nilai hamming loss yang lebih tinggi, yaitu 0.1685 dan 0.1706. Perbedaan ini menunjukkan bahwa tingkat kesalahan pada kedua model tersebut dua kali lipat dibandingkan IndoBERT, yang menunjukkan selisih performa yang cukup

signifikan. Meskipun model TF-IDF dan Word2Vec memiliki nilai *precision* dan *recall* yang cukup baik, hamming loss-nya tetap tinggi karena kesalahan kecil dalam penentuan label tetap berkontribusi terhadap kenaikan skor loss. Hal ini memperkuat temua sebelumnya bahwa pendekatan berbasi representative konstekual seperti IndoBERT lebih handal dalam memahami konteks ulasan, dibandingkan pendekatan representasi statis seperti TF-IDF dan Word2Vec.

Setelah pembahasan evaluasi performa masing-masing model dalam melakukan klasifikasi aspek. Langkah selanjutnya adalah mengevalusi performa masing model dalam melakukan klasifikasi sentimen. Sama seperti pada aspek, tiga pendekatan ektraksi fitur yang di uji dalam penelitian ini adalah TF-ICF, Word2Vec, IndoBERT. Grafik perbandingan antar model ditampilkan pada Gambar 7.5.

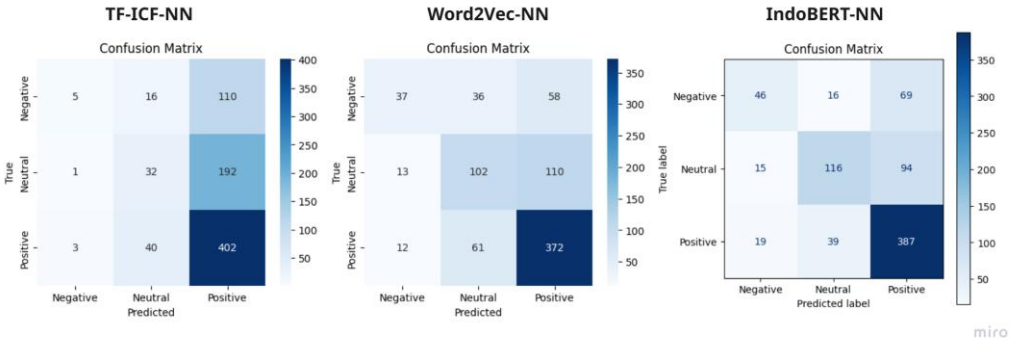


Gambar 7.5. Perbandingan Performa Model pada *Sentiment Analysis*

Pada Gambar 7.5 menunjukkan perbandingan performa antar model dalam empat metrik utama, yaitu *accuracy, precision, recall, f1-score.* Secara umum, IndoBERT kembali menunjukkan performa yang lebih unggul dibandingkan model TF-ICF dan Word2Vec. IndoBERT mencapai *accuracy* 68,66%, *precision* 65,38%, *recall* 57,95%, dan *f1-score* 60,08%. Keunggulan ini berasal dari kemampuan model dalam menangkap semantik dan konteks kata secara lebih akurat melalui *pretraining* pada *corpus* Bahasa Indonesia yang lebih besar.

Model Word2Vec juga menunjukkan performa yang cukup baik. Perbedaan performa dengan model IndoBERT tidak terlalu signifikan. Word2Vec mendapatkan *accuracy* 63,79%, *precision* 59,94%, *recall* 52,39%, dan *f1-score* 53,99%. Sebaliknya pada TF-ICF performa tertinggal cukup jauh dalam seluruh metrik. TF-ICF mendapatkan *accuracy* 54,80%, *precision* 49,67%, *recall* 36,12%, dan *f1-score* 32,52%. Hal ini menunjukkan bahwa pendekatan representasi berbasis *pre-trained embeddings* seperti Word2Vec dan IndoBERT mampu menangkap makna semantic kata dan konteks dalam kalimat ulasan dengan lebih baik dibandingkan pendekatan berbasis frekuensi pada dataset lokal seperti TF-ICF. Representasi statis seperti TF-ICF sangat bergantung pada jumlah kemunculan kata dalam dokumen dan tidak mempertimbangkan hubungan antar kata dalam kalimat. Akibatnya model menjadi kurang mampu memahami nuansa makna dari setiap ulasan. Sebaliknya, Word2Vec dilatih menggunakan korpus besar dan bersifat konstektual. IndoBERT masih unggul karena menggunakan *transformer-based architecture* dan *contextualized embeddings* yang mampu memahami konteks kata secara dinamis berdasarkan kalimat secara keseluruhan.

Namun, secara absolut, semua model masih menunjukkan performa yang tergolong rendah untuk identifikasi sentimen. Model terbaik, IndoBERT, hanya mencapai akurasi sebesar 68,66%. Salah satu penyebab utama dari rendahnya performa ini adalah distribusi data yang tidak seimbang pada dataset. Pada Gambar 7.2 terlihat bahwa kelas positif mendominasi sebagian besar data. Hal ini menyebabkan model menjadi bias dan cenderung untuk memprediksi berlebihan pada kelas positif. Hal ini terlihat pada Gambar 7.6 hasil *confusion matrix* semua model pada *sentiment analysis*.



Gambar 7.6. *Confusion Matrix* Semua Model pada *Sentiment Analysis*

Pada klasifikasi sentimen ini, dapat dilihat pada Gambar 7.6 semua model mengalami bias untuk cenderung memprediksi kelas positif. Dari ketiga model, *false negative* (FP) pada kelas positif lebih banyak daripada *true positive* (TP) pada kelas negatif maupun netral. Model TF-ICF sebanyak 110 data berlabel negatif diprediksi sebagai positif, sedangkan hanya 5 data yang berhasil diprediksi dengan benar sebagai negatif. Pola seperti ini juga terlihat pada model Word2Vec dan IndoBERT. dimana prediksi terhadap kelas positif sangatlah dominan meskipun label sebenarnya adalah netral atau negatif.

Kondisi ini menunjukkan bahwa model mengalami *over-prediction* terhadap kelaspositif akibat ketidakseimbangan distribusi kelas pada dataset. Ketika jumlah data pada satu kelas (dalam hal ini adalah positif) terlalu banyak dibanding kelas lainnya, model cenderung lebih sering memprediksi kelas dominan sehingga performa pada kelas minoritas akan turun drastis, walaupun didapatkan accuracy terlihat cukup tinggi. Hal itu menunjukkan pentingnya distribusi data yang seimbang untuk membuat model kecerdasan buatan agar mendapatkan performa yang baik tanpa terjadi bias.

Secara keseluruhan, pembahasan ini telah menjelaskan berbagai aspek penting yang mempengaruhi hasil evaluasi produk e-commerce berbasis aspek untuk topic modeling dan sentiment analysis. Dimulai dari distribusi data, baik kategori aspek maupun sentimen, terlihat bahwa dataset yang digunakan pada penelitian ini tidak seimbang, yang beruujung memberikan dampak signifikan pada performa model. Evaluasi model klasifikasi aspek menunjukkan performa IndoBERT berada diatas TF-ICF dan Word2Vec, baik dari segi metrik *accuracy, precision, recall, f1-score,* maupun *hamming loss.* Namun, selisih yang besar antara *accuracy* dengan *precision, recall,* dan *f1-score* pada semua model menunjukkan bahwa *accuracy* tidak sesuai menjadi indikator performa model pada kasus multilabel, sehingga metrik seperti *hamming loss* menjadi indikator yang lebih representatif. Sementara itu, pada klasifikasi sentiment, model berbasis Word2Vec dan IndoBERT menunjukkan performa yang relatif mirip. Mengindikasikan representasi semantik atau makna pada kata yang dimiliki embedding lebih unggul dibandingkan metode berbasis frekuensi lokal seperti TF-IDF. Namun, performa

keseluruhan pada kasus klasifikasi sentimen penelitian ini masih belum optimal dikarenakan distribusi yang tidak seimbang. Hal ini menunjukkan bahwa kualitas dari representasi fitur dan distribusi label merupakan faktor yang penting dalam membangun sistem klasifikasi berbasis aspek dan sentimen yang handal.

Hasil analisis evaluasi ulasan produk e-commerce berbasis aspek *topic modeling* dan *sentiment analysis* menggunakan *deep neural network* ini memberikan pemahaman yang lebih mendalam terhadap aspek-aspek yang menjadi perhatian utama pelanggan serta bagaimana sentimen mereka tercermin dalam ulasan yang diberikan. Dalam menentukan aspek dari suatu ulasan, terdapat indikator-indikator khusus yang menjadi ciri atau karakteristik dari masing-masing aspek tersebut. Aspek produk kepuasan pelanggan terhadap kualitas, performa, serta kesesuaian produk dengan deskripsi yang disampaikan. Aspek layanan pelanggan memiliki karakteristik ulasan yang menggambarkan pengalaman pengguna terhadap interaksi pelayanan, seperti keramahan, ketanggapan, dan kualitas bantuan yang diberikan. Aspek pengiriman ditandai dengan ulasan yang menyoroti kecepatan pengiriman, kondisi barang saat diterima, serta ketepatan waktu pengiriman. Selain itu, analisis sentimen dalam penelitian ini juga menggambarkan bagaimana sikap emosional pelanggan terhadap pengalaman berbelanja mereka, yang terbagi menjadi sentimen positif, negatif, dan netral. Sentimen positif tercermin dari ekspresi kepuasan dan pengalaman menyenangkan, sedangkan sentimen negatif menggambarkan kekecewaan atau keluhan atas produk maupun layanan. Sementara itu, sentimen netral menunjukkan adanya ulasan yang bersifat informatif tanpa mengarah pada emosi tertentu. Sejalan dengan hal ini, Al-

Qur'an menerangkan tentang karakteristik-karakteristik yang mendefinisikan sesuatu, seperti bagaimana karakteristik dari orang-orang mukmin.

ذٰلِكَ الْكِتٰبُ لَا رَيْبَ ۛ فِيْهِ ۛ هُدًى لِّلْمُتَّقِيْنَ ۙ﴿٢﴾ الَّذِيْنَ يُؤْمِنُوْنَ بِالْغَيْبِ وَيُقِيْمُوْنَ الصَّلٰوةَ وَمِمَّا رَزَقْنٰهُمْ يُنْفِقُوْنَ ۙ﴿٣﴾ وَالَّذِيْنَ يُؤْمِنُوْنَ بِمَآ اُنْزِلَ اِلَيْكَ وَمَآ اُنْزِلَ مِنْ قَبْلِكَ ۚ وَبِالْاٰخِرَةِ هُمْ يُوْقِنُوْنَ ۗ﴿٤﴾

"Kitab (Al-Quran) ini tidak ada keraguan padanya; petunjuk bagi mereka yang bertakwa (2). (yaitu) mereka yang beriman kepada yang gaib, melaksanakan salat, dan menginfakkan sebagian rezeki yang Kami berikan kepada mereka (3). dan mereka beriman kepada (Al-Quran) yang diturunkan kepadamu (Muhammad) dan (kitab-kitab) yang telah diturunkan sebelum engkau, dan mereka yakin akan adanya akhirat (4)" (QS. Al-Baqarah 2: 2-4).

Menurut Tafsir Al-Mukhtashar, orang-orang yang bertakwa adalah mereka yang beriman atau meyakini kepada perkara gaib, yaitu hal-hal yang tidak bisa ditangkap oleh panca indera dan tersembunyi, seperti keberadaan hari akhir. Mereka yang mendirikan sholat dan menunaikannya sesuai syariat, meliputi syarat, rukun, wajib, dan sunnah-sunnahnya. Mereka juga gemar menginfakkan sebagian rezeki yang diterima oleh Allah Swt. Kemudian, mereka yang beriman kepada wahyu yang Allah turunkan, serta mereka yang beriman kepada adanya akhirat. Pembahasan mengenai ciri-ciri atau karakteristik dari sesuatu juga diterangkan pada QS. Al-Baqarah ayat 2-4, yaitu mengenai ciri-ciri atau karakteristik dari orang yang bertaqwa:

اِنَّمَا الْمُؤْمِنُوْنَ الَّذِيْنَ اِذَا ذُكِرَ اللّٰهُ وَجِلَتْ قُلُوْبُهُمْ وَاِذَا تُلِيَتْ عَلَيْهِمْ اٰيٰتُهٗ زَادَتْهُمْ اِيْمَانًا وَّعَلٰى رَبِّهِمْ يَتَوَكَّلُوْنَ ۚ﴿٢﴾

"Sesungguhnya orang-orang mukmin adalah mereka yang jika disebut nama Allah, gemetar hatinya dan jika dibacakan ayat-ayat-Nya kepada mereka, bertambah (kuat) imannya dan hanya kepada Tuhannya mereka bertawakal" (QS. Al-Anfal 8: 2).

Menurut Tafsir Tahlili, Allah menjelaskan bahwa orang-orang mukmin adalah mereka yang memiliki sifat-sifat yang ada pada ayat ini. Terdapat 3 sifat yang disebutkan pada ayat ini, diantaranya: (1) jika mendengar nama Allah maka akan bergetar hatinya karena teringat keagungan dan kekuasaan-Nya, (2) mereka yang akan bertambah keimanannya jika dibacakan ayat suci Al-Qur'ran, dan (3) hanya bertaqwa kepada Allah Swt dan tidak berserah diri kepada yang lain.

Selaras dengan konsep pada QS. Al-Baqarah ayat 2-4 dan QS. Al-Anfal ayat 2 ini, orang-orang yang beriman dan orang-orang yang bertaqwa memiliki ciri-cirinya masing-masing. Begitu juga dengan aspek dan sentimen ulasan produk pada e-commerce memiliki karakteristik dan ciri-cirinya masing-masing. Karakteristik ini menjadi pedoman yang berharga bagi pelaku bisnis *e-commerce* untuk terus meningkatkan kualitas produk, layanan pelanggan, dan proses pengiriman mereka.

BAB VIII

KESIMPULAN

7.1 Kesimpulan

Evaluasi ulasan produk *e-commerce* berbasis aspek *topic modeling* dan *sentiment analysis* melakukan tiga skenario pengujian pada *feature extraction,* yaitu TF-ICF, Word2Vec, dan IndoBERT. Model dievaluasi menggunakan metrik *accuracy, precision, recall, f1-score*, dan *hamming loss*. Pada kasus identifikasi *topic modeling* berbasis aspek digunakan semua metrik, sedangkan pada kasus *sentiment analysis* tanpa *hamming loss.*

Pada kasus identifikasi topic modeling, model IndoBERT mendapatkan performa terbaik dibandingkan model lainnya dengan nilai *accuracy* 81.4%, *precission* 91.5%, *recall* 91,3%, dan *f1-score* sebesar 91,4%. Pada metrik hamming loss IndoBERT juga mendapatkan nilai terkecil dengan nilai *hamming loss* sebesar 0.0820. Hal ini menunjukkan bahwa model IndoBERT memiliki representasi semantik kata kata yang baik dan daya generalisasi yang kuat, sehingga *classifier* dapat lebih mudah dan efektif. Berhubung kasus pada tugas ini adalah multilabel, dimana setiap ulasan bisa memiliki lebih dari satu label aspek secara bersamaan, maka penggunaan metrik accuracy perlu ditinjau ulang karena tidak sepenuhnya mencerminkan performa model secara menyeluruh. Hal ini tercermin pada selisih yang besar antara metrik *accuracy* dengan metrik lainnnya pada IndoBERT dan juga model lainnya.

Pada kasus klasifikasi sentiment, model IndoBERT kembali menunjukkan performa terbaik dengan nilai *accuracy* 68,66%, *precision* 65,38%, *recall* 57,95%,

dan *f1-score* 60,08%. Namun, secara absolut, semua model masih menunjukkan performa yang tergolong rendah untuk identifikasi sentimen. Salah satu penyebab utama dari rendahnya performa ini adalah distribusi data yang tidak seimbang pada dataset. Hal ini menyebabkan model menjadi bias pada kelas yang dominan, di mana hal ini tercermin pada confusion matrix: semua model cenderung memprediksi kelas positif secara berlebihan. Akibatnya, prediksi untuk kelas negatif dan netral menjadi sangat kecil atau bahkan diabaikan. Kondisi ini membuat metrik seperti *recall* dan *f1-score* untuk kelas minoritas menjadi rendah, serta menurunkan kemampuan model dalam menangkap variasi sentimen secara menyeluruh. Untuk itu, diperlukan strategi penanganan ketidakseimbangan data, seperti oversampling, undersampling, atau penggunaan *loss function* khusus, agar model dapat belajar secara lebih adil terhadap seluruh kelas yang ada.

7.2 Saran

Penelitian evaluasi ulasan produk *e-commerce* berbasis aspek *topic modeling* dan *sentiment analysis* menggunakan deep neural network masih memiliki potensi untuk peningkatan lebih lanjut, Terdapat beberapa saran yang dapat dipertimbangkan untuk diterapkan pada penelitian yang akan mendatang:

1. Mengumpulkan data dalam jumlah yang lebih besar serta memastikan distribusinya seimbang untuk setiap kelas. Ketidakseimbangan dan keterbatasan jumlah data dapat menyebabkan model hanya menangkap pola dari kelas yang dominan, sehingga menimbulkan bias yang signifikan dan menurunkan performa model secara keseluruhan. Dengan menyamakan

jumlah data pada tiap kelas, potensi bias dapat diminimalkan sehingga model mampu melakukan pembelajaran yang lebih adil dan akurat.

2. Membuat satu model yang dapat menjalankan dua tugas sekaligus (*multi-task learning*) yaitu klasifikasi aspek dan analisis sentimen, secara bersamaan dalam satu kerangka pelatihan. Dengan pendekatan ini, model dapat lebih efisien dan potensi generalisasi model terhadap data baru meningkat. Selain itu, integrasi dua tugas dalam satu tugas juga dapat mengurangi kebutuhan komputasi yang besar.

DAFTAR PUSTAKA

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Umar, A. M., Linus, O. U., Arshad, H., Kazaure, A. A., Gana, U., & Kiru, M. U. (2019). Comprehensive Review of Artificial Neural Network Applications to Pattern Recognition. In *IEEE Access* (Vol. 7, pp. 158820–158846). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2019.2945545

Almana, A. M., & Mirza, A. A. (2013). The Impact of Electronic Word of Mouth on Consumers' Purchasing Decisions. In *International Journal of Computer Applications* (Vol. 82, Issue 9).

Ameur, A., Hamdi, S., & Yahia, S. Ben. (2023). Multi-Label Learning for Aspect Category Detection of Arabic Hotel Reviews Using AraBERT. *International Conference on Agents and Artificial Intelligence*, *2*, 241–250. https://doi.org/10.5220/0011694800003393

Amien, S., Perdana, P., Bharata Aji, T., & Ferdiana, R. (2021). Aspect Category Classification dengan Pendekatan Machine Learning Menggunakan Dataset Bahasa Indonesia (Aspect Category Classification with Machine Learning Approach Using Indonesian Language Dataset). *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi* |, *10*(3).

Ashok Kumar, J., Abirami, S., & Trueman, T. E. (2019). Multilabel aspect-based sentiment classification for abilify drug user review. *Proceedings of the 11th International Conference on Advanced Computing, ICoAC 2019*, 376–380. https://doi.org/10.1109/ICoAC48765.2019.246871

Banks, G. C., Woznyj, H. M., Wesslen, R. S., & Ross, R. L. (2018). A Review of Best Practice Recommendations for Text Analysis in R (and a User-Friendly App). *Journal of Business and Psychology*. https://doi.org/10.15139/S3/R4W7ZS

Chen, T., Samaranayake, P., Cen, X. Y., Qi, M., & Lan, Y. C. (2022). The Impact of Online Reviews on Consumers' Purchasing Decisions: Evidence From an Eye-Tracking Study. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.865702

Chhabra, D., Sah, S., Rani, R., Bisla, N., Tomar, A., & Sharma, A. (2024). Deep Learning-based Sentiment Analysis of Amazon Product Reviews. *2024 1st International Conference on Trends in Engineering Systems and Technologies (ICTEST)*, 1–7. https://doi.org/10.1109/ICTEST60614.2024.10576180

Deniz, E., Erbay, H., & Coşar, M. (2022). Multi-Label Classification of E-Commerce Customer Reviews via Machine Learning. *Axioms*, *11*(9). https://doi.org/10.3390/axioms11090436

Fehle, J., Münster, L., Schmidt, T., & Wolff, C. (2023). Aspect-Based Sentiment Analysis as a Multi-Label Classification Task on the Domain of German Hotel Reviews.

Conference on Natural Language Processing (KONVENS 2023), 202–218. https://aclanthology.org/2023.konvens-main.21/

Forry Kusuma, J., & Chowanda, A. (2023). Indonesian Hate Speech Detection Using IndoBERTweet and BiLSTM on Twitter. *JOIV : International Journal on Informatic Visualization*, *7*(3), 773–780. www.joiv.org/index.php/joiv

Gurav, Y., Ingawale, V., & Yadav, A. (2023). A STUDY ON THE IMPACT OF ONLINE PRODUCT REVIEWS ON CONSUMERS' BUYING INTENTIONS. *The Online Journal of Distance Education and E-Learning*, *11*(2). www.tojdel.net

Hakim, A. A., Erwin, A., Eng, K. I., Galinium, M., & Muliady, W. (2014). Automated Document Classification for News Article in Bahasa Indonesia based on Term Frequency Inverse Document Frequency (TF-IDF) Approach. *2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE)*.

Imron, S., Setiawan, E. I., & Santoso, J. (2023a). Deteksi Aspek Review E-Commerce Menggunakan IndoBERT Embedding dan CNN. *Journal of Intelligent System and Computation*, *5*(1), 10–16. https://doi.org/10.52985/insyst.v5i1.267

Imron, S., Setiawan, E. I., & Santoso, J. (2023b). Deteksi Aspek Review E-Commerce Menggunakan IndoBERT Embedding dan CNN. *Journal of Intelligent System and Computation*, *5*(1), 10–16. https://doi.org/10.52985/insyst.v5i1.267

Ismet, H. T., Mustaqim, T., & Purwitasari, D. (2022). Aspect Based Sentiment Analysis of Product Review Using Memory Network. *Scientific Journal of Informatics*, *9*(1), 73–83. https://doi.org/10.15294/sji.v9i1.34094

Jaman, J. H., & Abdulrohman, R. (2019). Sentiment Analysis of Customers on Utilizing Online Motorcycle Taxi Service at Twitter with the Support Vector Machine. *ICECOS 2019 - 3rd International Conference on Electrical Engineering and Computer Science, Proceeding*, 231–234. https://doi.org/10.1109/ICECOS47637.2019.8984483

Jansen, S. (2017). *Word and Phrase Translation with word2vec*. http://arxiv.org/abs/1705.03127

Kartika, B. V., Alfredo, M. J., & Kusuma, G. P. (2023). Fine-Tuned IndoBERT based model and data augmentation for indonesian language paraphrase identification. *Revue d'Intelligence Artificielle*, *37*(3), 733–743. https://doi.org/10.18280/ria.370322

Kaushik, K., Mishra, R., Rana, N. P., & Dwivedi, Y. K. (2018). Exploring reviews and review sequences on e-commerce platform: A study of helpful reviews on Amazon.in. *Journal of Retailing and Consumer Services*, *45*, 21–32. https://doi.org/10.1016/j.jretconser.2018.08.002

Khlifi, G., Jenhani, I., Messaoud, M. Ben, & Mkaouer, M. W. (2023). Multi-label Classification of Mobile Application User Reviews Using Neural Language Models. *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*. https://doi.org/10.1007/978-3-031-45608-4_31

Koto, F., Lau, J. H., & Baldwin, T. (2021). *IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization*. http://arxiv.org/abs/2109.04607

Li, J., Cheng, J.-H., Shi, J.-Y., & Huang, F. (2012). Brief Introduction of Back Propagation (BP) Neural Network Algorithm and Its Improvement. *AISC*, *2*, 553–558. https://doi.org/https://doi.org/10.1007/978-3-642-30223-7_87

Lunardi, A., Viterbo, J., Boscarioli, C., Bernardini, F., & Maciel, C. (2016). Domain-tailored multiclass classification of user reviews based on binary splits. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9742*, 298–309. https://doi.org/10.1007/978-3-319-39910-2_28

Mardianti, S., Zidny Naf, M., Hidayatulloh, I., & Teknologi, F. (2018). Seminar Nasional Teknologi Informasi dan Multimedia. *UNIVERSITAS AMIKOM Yogyakarta*.

Mcilroy, S., Ali, N., Khalid, H., & Hassan, A. E. (2015). Analyzing and Automatically Labelling The Types of User Issues that are Raised in Mobile App Reviews. *Empirical Software Engineering*. https://doi.org/10.1007/s10664-015-9375-7

Mijwil, M. M., Alsaadi, A., Mijwel, M. M., Esen, A., & Shamil, A. (2019). *Overview of Neural Networks*. https://www.researchgate.net/publication/332655457

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

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). *Distributed Representations of Words and Phrases and their Compositionality*. http://arxiv.org/abs/1310.4546

Nasiri, D. F., & Budi, I. (2019, November 1). Aspect Category Detection on Indonesian E-commerce Mobile Application Review. *Proceedings of 2019 International Conference on Data and Software Engineering, ICoDSE 2019*. https://doi.org/10.1109/ICoDSE48700.2019.9092619

Noori, B. (2021). Classification of Customer Reviews Using Machine Learning Algorithms. *Applied Artificial Intelligence*, *35*(8), 567–588. https://doi.org/10.1080/08839514.2021.1922843

Power Reviews. (2023). *Survey: The Ever-Growing Power of Reviews (2023 Edition)*. https://www.powerreviews.com/power-of-reviews-2023/

Purnamasari, N. M. G. D., Ali Fauzi, M., Indriati, & Dewi, L. S. (2020). Cyberbullying identification in twitter using support vector machine and information gain based feature selection. *Indonesian Journal of Electrical Engineering and Computer Science*, *18*(3), 1494–1500. https://doi.org/10.11591/ijeecs.v18.i3.pp1494-1500

Rachmatullah, M. I. C., Santoso, J., & Surendro, K. (2020). A novel approach in determining neural networks architecture to classify data with large number of attributes. *IEEE Access*, *8*, 204728–204743. https://doi.org/10.1109/ACCESS.2020.3036853

Rayhan, F. M., Hadi Wijoyo, S., Hayuhardhika, W., & Putra, N. (2017). Analisis Sentimen Root Cause Analisis Kepuasan Pengguna Aplikasi Tokopedia pada Ulasan Menggunakan Metode Random Forest. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1*(1), 2548–2964. http://j-ptiik.ub.ac.id

Ririn, S., & Astanti, D. (2024). Supervised multilabel classification techniques for categorising customer requirements during the conceptual phase in the new product development. *Engineering Management in Production and Services*, *16*(1), 31–47. https://doi.org/10.2478/emj-2024-0003

Sajid, N. A., Rahman, A., Ahmad, M., Musleh, D., Basheer Ahmed, M. I., Alassaf, R., Chabani, S., Ahmed, M. S., Salam, A. A., & AlKhulaifi, D. (2023). Single vs. Multi-Label: The Issues, Challenges and Insights of Contemporary Classification Schemes. In *Applied Sciences (Switzerland)* (Vol. 13, Issue 11). MDPI. https://doi.org/10.3390/app13116804

Sarica, S., & Luo, J. (2020). Stopwords in Technical Language Processing. *ArXiv*. https://doi.org/10.1371/journal.pone.0254937

Sharma, S., Sharma, S., & Athaiya, A. (2020). Activation Functions in Neural Networks. *International Journal of Engineering Applied Sciences and Technology*, *4*, 310–316. https://doi.org/10.33564/IJEAST.2020.v04i12.054

Statista Research Department. (2024, August 17). *Number of users of e-commerce in Indonesia from 2020 to 2029*. Statista Research Department. https://www.statista.com/forecasts/251635/e-commerce-users-in-indonesia

Stephen J. Payne, & Andrew Howes. (2013). E-commerce. In *Adaptive Interaction*. https://doi.org/https://doi.org/10.1007/978-3-031-02199-2_7

Tanaka, H., Shinnou, H., Cao, R., Bai, J., & Ma, W. (2020). Document Classification by Word Embeddings of BERT. *Communications in Computer and Information Science*, *1215 CCIS*, 145–154. https://doi.org/10.1007/978-981-15-6168-9_13

Uysal, A. K., & Gunal, S. (2014). The impact of preprocessing on text classification. *Information Processing and Management*, *50*(1), 104–112. https://doi.org/10.1016/j.ipm.2013.08.006

Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). *IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding*. http://arxiv.org/abs/2009.05387

Yanti, D., Kristya Ningsih, N., Ony, J. G., & Suhalim, S. P. (2024). The Influence of Online Customer Reviews and Online Customer Ratings on Product Purchase Decisions on The Tokopedia Marketplace. In *Jurnal Pemasaran Kompetitif* (Vol. 07, Issue 2). http://www.openjournal.unpam.ac.id/index.php/JPK

Zhao, W. (2020). Classification of Customer Reviews on E-commerce Platforms Based on Naive Bayesian Algorithm and Support Vector Machine. *Journal of Physics: Conference Series*, *1678*(1). https://doi.org/10.1088/1742-6596/1678/1/012081