

**KLASIFIKASI ULASAN PRODUK *E-COMMERCE*
MENGUNAKAN *NEURAL NETWORK***

SKRIPSI

Oleh:

FAHRENDRA KHOIRUL IHTADA

NIM. 200605110124



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

**KLASIFIKASI ULASAN PRODUK *E-COMMERCE*
MENGUNAKAN *NEURAL NETWORK***

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
FAHRENDRA KHOIRUL IHTADA
NIM. 200605110124

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

HALAMAN PERSETUJUAN

**KLASIFIKASI ULASAN PRODUK *E-COMMERCE*
MENGUNAKAN *NEURAL NETWORK***

SKRIPSI

**Oleh:
FAHRENDRA KHOIRUL IHTADA
NIM. 200605110124**

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 3 Juni 2024

Pembimbing I,


Dr. Cahyo Crysdiyan, MCS
NIP. 19740424 200901 1 008

Pembimbing II,


Dr. Zainal Abidin, M.Kom
NIP. 19760613 200501 1 004

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



Dr. Fachrudin Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

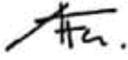
**KLASIFIKASI ULASAN PRODUK *E-COMMERCE*
MENGUNAKAN *NEURAL NETWORK***

SKRIPSI

Oleh:
FAHRENDRA KHOIRUL IHTADA
NIM. 200605110124

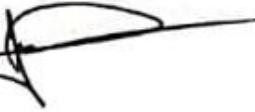
Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 7 Juni 2024

Susunan Dewan Penguji

Ketua Penguji	: <u>Dr. M. Amin Hariyadi, M.T</u> NIP. 19670118 200501 1 001	()
Anggota Penguji I	: <u>Fatchurrohman, M.Kom</u> NIP. 19700731 200501 1 002	()
Anggota Penguji II	: <u>Dr. Cahyo Crysdian, MCS</u> NIP. 19740424 200901 1 008	()
Anggota Penguji III	: <u>Dr. Zainal Abidin, M.Kom</u> NIP. 19760613 200501 1 004	()

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




Fatchur Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : FAHRENDRA KHOIRUL IHTADA
NIM : 200605110124
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : *KLASIFIKASI ULASAN PRODUK
E-COMMERCE MENGGUNAKAN NEURAL
NETWORK*

Menyatakan dengan sebenarnya bahwa Skripsi 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, 7 Juni 2024
Yang membuat pernyataan,



FAHRENDRA KHOIRUL IHTADA
NIM. 200605110124

MOTTO

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

“Step by step to the next step”

HALAMAN PERSEMBAHAN

Puji Syukur atas kehadiran Allah Subhanahu wa ta'ala, karena berkat rahmat dan petunjuk-Nya penulis dapat menyelesaikan skripsi ini. Shalawat serta salam kepada Rasulullah Shallallahu 'alaihi wasallam, yang telah membawa kita dari zaman jahiliyah menuju addinul Islam.

Penulis mempersembahkan karya ini kepada kedua orang tua, kerabat, teman, kakak tingkat, dan sahabat yang telah menemani perkuliahan daring maupun luring. Tanpa bantuan mereka penulis tidak akan bisa sampai di titik ini.

KATA PENGANTAR

Assalamualaikum wr wb.

Puji syukur penulis panjatkan kepada kehadiran Allah Swt yang telah melimpahkan nikmat serta karunia-Nya sehingga penulis mampu menyelesaikan penulisan Skripsi yang berjudul “Klasifikasi Ulasan Produk *E-Commerce* Menggunakan *Neural Network*” dengan baik dan tepat waktu.

Dalam penulisan skripsi ini, penulis menyadari banyak pihak yang terlibat baik dalam proses membimbing penulisan dan juga memberikan semangat dan dukungan moril atau materiil. 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. Fachrul Kurniawan, M.MT., IPM, selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang, serta Dosen Pembimbing II dalam skripsi ini dari awal hingga akhir.
4. Dr. Cahyo Crysdiyan, MCS selaku Dosen Pembimbing I yang telah memberikan banyak dukungan dan bimbingan dalam penulisan skripsi ini.
5. Dr. Zainal Abidin, M.Kom selaku Dosen Pembimbing II yang telah memberikan banyak dukungan dan bimbingan dalam penulisan skripsi ini.
6. Dr. M. Amin Hariyadi, M.T selaku Ketua Penguji yang telah memberikan banyak saran untuk menyelesaikan skripsi ini.

7. Fatchurrohman, M.Kom selaku dosen penguji II yang telah menguji serta memberikan masukan dalam penulisan skripsi ini.
8. Seluruh Dosen dan Jajaran Staf Program Studi Teknik Informatika yang telah memberikan banyak bantuan dalam skripsi ini.
9. Kedua orang tua, Bapak Harun Munif dan Ibu Lilis Sarmini serta Kakak Anindya dan Adik Aliya yang telah memberikan dukungan kepada penulis untuk menyelesaikan skripsi ini.
10. Teman-teman Angkatan 2020 Teknik Informatika “INTEGER” yang telah memberikan banyak bantuan baik material maupun dukungan intelektual, semangat, serta motivasi dalam menyelesaikan penyusunan skripsi ini.
11. Semua pihak yang tidak dapat disebutkan satu per satu, termasuk responden yang telah memberikan kontribusi, saran, dan dukungan dalam perjalanan penulisan skripsi ini.

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

Wassalamualaikum Warahmatullahi Wabarakatuh.

Malang, 7 Juni 2024

Penulis

DAFTAR ISI

HALAMAN PENGANTAR	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
ABSTRAK	xv
ABSTRACT	xvi
المخلص	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Pernyataan Masalah	5
1.3 Tujuan Penelitian	5
1.4 Batasan Masalah	5
1.5 Manfaat Penelitian	6
BAB II KAJIAN PUSTAKA	7
2.1 Kendala Banyaknya Ulasan <i>E-commerce</i>	7
2.2 Klasifikasi Ulasan	8
2.3 <i>Theoretical Framework</i>	12
BAB III DESAIN DAN IMPLEMENTASI	15
3.1 Akuisisi Data.....	15
3.2 Desain Sistem.....	15
3.3 <i>Data Cleaning</i>	16
3.3.1 <i>Case Folding</i>	16
3.3.2 <i>Punctuational Removal</i>	17
3.3.3 <i>Tokenization</i>	18
3.3.4 <i>Stopword Removal</i>	18
3.3.5 <i>Stemming</i>	19
3.4 <i>Term Frequency – Inverse Document Frequency (TF-IDF)</i>	19
3.5 <i>Neural Network</i>	21
3.6 Implementasi Sistem.....	25
3.6.1 <i>Proses Training</i>	25
BAB IV UJI COBA DAN PEMBAHASAN	31
4.1 Data Pengujian	31
4.2 Skenario Pengujian	32
4.3 Hasil Uji Coba.....	38
4.3.1 Model A1	40
4.3.2 Model A2	42
4.3.3 Model A3	44

4.3.4 Model B1	46
4.3.5 Model B2	48
4.3.6 Model B3	50
4.3.7 Model C1	52
4.3.8 Model C2	54
4.3.9 Model C3	56
4.4 Pembahasan.....	58
BAB V KESIMPULAN DAN SARAN	73
5.1 Kesimpulan	73
5.2 Saran	74
DAFTAR PUSTAKA	

DAFTAR TABEL

Tabel 4.1 Hasil survei <i>ground truth</i>	31
Tabel 4.2 Data ulasan produk <i>e-commerce</i>	32
Tabel 4.3 Skenario pengujian.....	34
Tabel 4.4 <i>Confusion matrix</i>	36
Tabel 4.5 <i>Confusion matrix</i> kelas produk	36
Tabel 4.6 <i>Confusion matrix</i> kelas layanan pelanggan.....	36
Tabel 4.7 <i>Confusion matrix</i> kelas pengiriman	37
Tabel 4.8 Hasil <i>preprocessing</i> data ulasan.....	38
Tabel 4.9 Hasil <i>preprocessing</i> TFIDF data ulasan	39
Tabel 4.10 Rasio data dan skenario pemberhentian proses <i>training</i>	39
Tabel 4.11 Hasil proses <i>training</i> skenario pengujian.....	59
Tabel 4.12 Hasil evaluasi skenario pengujian.....	61
Tabel 4.13 <i>Confusion matrix</i> model B3 kelas produk.....	61
Tabel 4.14 <i>Confusion matrix</i> model B3 kelas layanan pelanggan	62
Tabel 4.15 <i>Confusion matrix</i> model B3 kelas pengiriman.....	62
Tabel 4.16 Hasil <i>k-fold cross validation</i> model B3.....	66
Tabel 4.17 Skor TF-IDF tertinggi kelas produk	67
Tabel 4.18 Skor TF-IDF tertinggi kelas layanan pelanggan	68
Tabel 4.19 Ranking skor TF-IDF kelas pengiriman	69

DAFTAR GAMBAR

Gambar 1.1 Grafik proyeksi nilai ekonomi sektor <i>e-commerce</i> Indonesia.....	2
Gambar 2.1 <i>Theoretical framework</i>	13
Gambar 3.1 Desain sistem	16
Gambar 3.8 Topologi desain <i>neural network</i>	21
Gambar 3.9 <i>Feedforward Neural Network</i>	23
Gambar 3.10 <i>Backpropagation neural network</i>	27
Gambar 3.11 Flowchart proses <i>training</i>	30
Gambar 4.1 <i>Flowchart</i> proses <i>testing</i>	33
Gambar 4.2 Grafik <i>cost</i> proses <i>training</i> model A1	41
Gambar 4.3 <i>Confusion matrix</i> model A1	41
Gambar 4.4 Metrik performa model A1	42
Gambar 4.5 Grafik <i>cost</i> proses <i>training</i> model A2	43
Gambar 4.6 <i>Confusion matrix</i> model A2	43
Gambar 4.7 Metrik performa model A2	44
Gambar 4.8 Grafik <i>cost</i> proses <i>training</i> model A3	45
Gambar 4.9 <i>Confusion matrix</i> model A3	45
Gambar 4.10 Metrik performa model A3	46
Gambar 4.11 Grafik <i>cost</i> proses <i>training</i> model B1	47
Gambar 4.12 <i>Confusion matrix</i> model B1	47
Gambar 4.13 Metrik performa model B1	48
Gambar 4.14 Grafik <i>cost</i> proses <i>training</i> model B2	49
Gambar 4.15 <i>Confusion matrix</i> model B2	49
Gambar 4.16 Metrik performa model B2.....	50
Gambar 4.17 Grafik <i>cost</i> proses <i>training</i> model B3	51
Gambar 4.18 <i>Confusion matrix</i> model B3	51
Gambar 4.19 Metrik performa model B3.....	52
Gambar 4.20 Grafik <i>cost</i> proses <i>training</i> model C1	53
Gambar 4.21 <i>Confusion matrix</i> model C1	53
Gambar 4.22 Metrik performa model C1.....	54
Gambar 4.23 Grafik <i>cost</i> dan <i>accuracy</i> proses <i>training</i> model C2.....	55
Gambar 4.24 <i>Confusion matrix</i> model C2	55
Gambar 4.25 Metrik performa model C2.....	56
Gambar 4.26 Grafik <i>cost</i> dan <i>accuracy</i> proses <i>training</i> model C3.....	57
Gambar 4.27 <i>Confusion matrix</i> model C3	57
Gambar 4.28 Metrik performa model C3.....	58
Gambar 4.29 Hasil <i>k-fold cross validation</i> model B3 dengan nilai $k=4$	64
Gambar 4.30 Hasil <i>k-fold cross validation</i> model B3 dengan nilai $k=5$	65
Gambar 4.31 Hasil <i>k-fold cross validation</i> model B3 dengan nilai $k=10$	66
Gambar 4.32 <i>Wordcloud</i> kelas produk	67
Gambar 4.33 <i>Wordcloud</i> kelas layanan pelanggan	68

Gambar 4.34 *Wordcloud* kelas pengiriman..... 69

ABSTRAK

Ihtada, Fahrendra Khoirul. 2024. **Klasifikasi Ulasan Produk *E-commerce* Menggunakan *Neural network***. Skripsi. 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: *Klasifikasi, Neural Network, Ulasan Produk*

E-commerce telah menjadi bagian integral dalam cara masyarakat berbelanja, dengan meningkatnya ulasan pelanggan di berbagai platform. Ulasan-ulasan ini memberikan wawasan penting mengenai kualitas produk, layanan pelanggan, dan pengiriman. Namun, pemilahan ulasan secara manual memakan waktu dan rentan terhadap kesalahan. Penelitian ini bertujuan untuk mengklasifikasikan ulasan *e-commerce* ke dalam tiga kategori: produk, layanan pelanggan, dan pengiriman, menggunakan *neural network*. Dari sembilan skenario pengujian, model B3 dengan 50 *node* pada *hidden layer* pertama dan *learning rate* 0.1 memberikan performa terbaik dengan akurasi 65,85%, presisi 62,27%, *recall* 58,61%, dan *F1-score* 59,71%. Validasi menggunakan *k-fold cross validation* menunjukkan rata-rata akurasi 64,17% pada $k=10$. Analisis kata dengan TF-IDF mengidentifikasi kata dominan pada setiap kategori ulasan. Meskipun hasil menunjukkan potensi, model B3 belum mampu mengklasifikasikan ulasan dengan sempurna, disebabkan oleh dataset yang kurang banyak dan tidak seimbang, arsitektur model yang kurang kompleks, dan *preprocessing* TF-IDF yang kurang efektif. Temuan ini menekankan kebutuhan akan pendekatan lebih lanjut untuk meningkatkan akurasi klasifikasi ulasan *e-commerce*.

ABSTRACT

Ihtada, Fahrendra Khoirul. 2024. **Classification E-commerce Product Reviews Using Neural Networks**. Undergraduate 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.

E-commerce has become an integral part of how people shop, with the rise of customer reviews on various platforms. These reviews provide important insights into product quality, customer service, and delivery. However, manually sorting reviews is time-consuming and prone to errors. This study aims to classify e-commerce reviews into three categories: product, customer service, and delivery, using a neural network. From nine test scenarios, model B3 with 50 nodes in the first hidden layer and a learning rate of 0.1 provided the best performance with an accuracy of 65.85%, precision of 62.27%, recall of 58.61%, and f1-score of 59.71%. Validation using k-fold cross validation shows an average accuracy of 64.17% at k=10. Word analysis with TF-IDF identified dominant words in each review category. Although the results show potential, the B3 model is not yet able to classify reviews perfectly, due to the large and unbalanced dataset, less complex model architecture, and less effective TF-IDF preprocessing. These findings emphasize the need for further approaches to improve the classification accuracy of e-commerce reviews.

Keywords: *Classification, Neural Network, Product Review*

الملخص

الاهتدى، فهرندرا خير. 2024. التصنيف على أساس جوانب مراجعة المنتج في التجارة الإلكترونية باستخدام الشبكات العصبية. 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: التصنيف، الشبكة العصبية، جانب المراجعة.

أصبحت التجارة الإلكترونية جزءاً لا يتجزأ من الطريقة التي يتسوق بها الأشخاص، مع ارتفاع مراجعات العملاء عبر المنصات. توفر هذه المراجعات رؤى مهمة حول جودة المنتج وخدمة العملاء وتسليمه. ومع ذلك، فإن فرز المراجعات يدوياً يستغرق وقتاً طويلاً وعرضة للأخطاء. هدف البحث إلى تصنيف مراجعات التجارة الإلكترونية إلى ثلاث فئات: المنتجات وخدمات العملاء وتسليمها، باستخدام شبكة عصبية. من بين سيناريوهات الاختبار التسعة، يوفر نموذج ب3 مع 50 عقدة على الطبقة المخفية الأولى ومعدل تعلم 0.1 أفضل أداء بدقة 65.85% ودقة 62.27% واستدعاء 58.61% ودرجة ف1 بنسبة 59.71%. أظهر التحقق من الصحة باستخدام تقنية التحقق المتقاطع *K-Fold* متوسط دقة بنسبة 64.17% عند $k = 10$. حدد تحليل الكلمات مع *TF-IDF* الكلمة السائدة في كل فئة من فئات المراجعة. على الرغم من أن النتائج أظهرت إمكانات، إلا أن نموذج ب3 لم يتمكن من تصنيف المراجعات بشكل مثالي، بسبب مجموعة البيانات غير الكافية وغير المتوازنة، وبنية النموذج الأقل تعقيداً، والمعالجة المسبقة الأقل فعالية لـ *TF-IDF*. وأبرزت هذه النتائج الحاجة إلى مزيد من المدخل لتحسين دقة تصنيف مراجعات التجارة الإلكترونية.

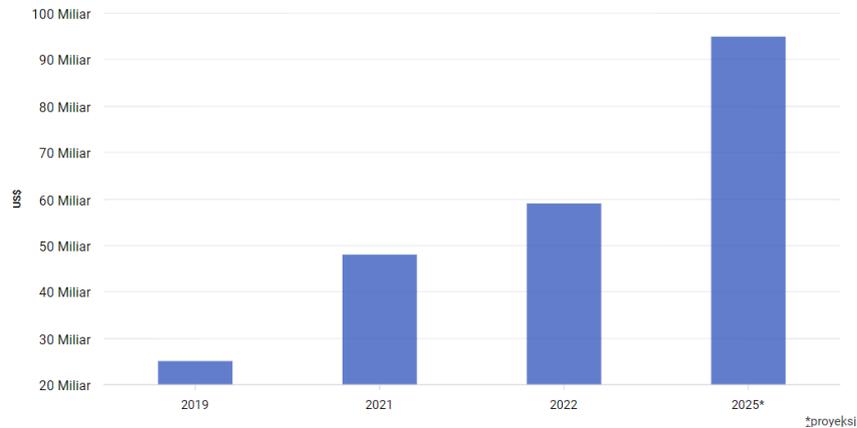
BAB I

PENDAHULUAN

1.1 Latar Belakang

Salah satu kegiatan yang paling populer di kalangan pengguna internet adalah berbelanja (O'Regan, 2018). Berbelanja secara daring di internet sering juga disebut dengan *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 telah berkembang pesat sejak beberapa tahun terakhir ini. Menurut data Badan Pusat Statistik (BPS), hasil survei menunjukkan peningkatan jumlah usaha *e-commerce* di Indonesia selama beberapa tahun terakhir. Pada tahun 2020, sekitar 25,25 persen usaha berbasis *e-commerce*, kemudian meningkat menjadi 32,23 persen pada tahun 2021, dan kembali stabil sekitar 24,10 persen pada tahun 2022 (BPS *e-commerce* 2022). Riset yang dilakukan oleh Google, Temasek, dan Bain & Company juga menunjukkan nilai ekonomi sektor *e-commerce* di Indonesia pada tahun 2019 sebesar US\$25 miliar, 2021 sebesar US\$48 miliar, 2022 mencapai US\$59 miliar, dan diproyeksikan akan terus tumbuh hingga US\$95 miliar di tahun 2025 (Annur, 2022).



Gambar 1.1 Grafik proyeksi nilai ekonomi sektor *e-commerce* Indonesia

E-commerce merubah bagaimana cara konsumen berbelanja di kalangan Masyarakat. Dengan adanya *e-commerce*, Masyarakat dapat berbelanja dan membeli produk yang mereka inginkan secara online. Hampir di setiap platform *e-commerce* saat ini sudah memiliki fitur ulasan atau ulasan untuk memfasilitasi konsumen agar dapat mengekspresikan pendapat, kekhawatiran, dan *feedback* mereka dengan lebih terbuka tentang produk *e-commerce*. Dengan tren yang terus berkembang pada penggunaan *e-commerce* di kalangan masyarakat, jumlah ulasan yang diberikan oleh pelanggan di berbagai platform *e-commerce* juga meningkat dengan tajam.

Pada saat seseorang ingin membeli barang, mereka seringkali mencari dan mengumpulkan informasi terlebih dahulu tentang barang yang mereka ingin beli. Salah satu sumber informasi yang bisa didapat yaitu melalui ulasan. Disana mereka dapat mengumpulkan informasi tentang barang tersebut dari pengalaman orang lain yang sudah membeli barang tersebut sebelumnya (Stephen J. Payne & Andrew Howes, 2013). Berdasarkan survei penelitian, sebanyak 93,5% responden mengatakan bahwa mereka selalu menggunakan fitur ulasan sebelum melakukan

transaksi (Mardianti *et al.*, 2018). Tentu hal ini jelas menjadi faktor yang sangat penting bagi pelaku bisnis.

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. Informasi-informasi ini menjadi pertimbangan bagi calon pelanggan dan juga memberikan *insight* kepada pelaku bisnis tentang aspek mana saja yang perlu ditingkatkan. Memanfaatkan informasi yang ada didalam ulasan dalam jumlah sangat banyak secara efisien memerlukan pendekatan sistematis untuk mengkategorikan ke dalam beberapa jenis. Kategori dapat dibagi menjadi ulasan yang berhubungan dengan produk, ulasan yang berhubungan dengan layanan pelanggan (*customer service*), dan ulasan yang berhubungan dengan pengiriman produk (*shipping/delivery*).

Proses pemilahan kategori ulasan ke dalam 3 aspek tersebut dengan cara manual akan memakan banyak waktu dan tenaga. Tidak hanya itu, proses dengan cara manual juga akan rentan terhadap subjektivitas dan kesalahan. Oleh karena itu, kebutuhan akan sistem klasifikasi yang akurat sangatlah dibutuhkan. Sistem seperti itu akan meringankan dan mempercepat proses klasifikasi. Sistem seperti itu juga dapat memberikan wawasan atau insight yang berharga bagi pelaku bisnis untuk menghasilkan keputusan yang tepat, menawarkan penawaran yang optimal, dan menyempurnakan strategi operasional bisnis.

Penelitian ini melakukan pengklasifikasian ulasan-ulasan kedalam 3 aspek ini dikarenakan masih belum banyak penelitian sebelumnya yang mengkaji aspek-

aspek ini secara terpisah dalam konteks ulasan ulasan *e-commerce*. Kebanyakan peneliti sebelumnya lebih fokus pada mengklasifikasikan sentimen positif dan negative dari ulasan-ulasan tersebut. Abien (2020) melakukan penelitian mengenai sentiment analisis pada ulasan *e-commerce* dengan menggunakan *Bidirectional Recurrent Neural Network*. Hasil uji didapatkan F1-score sebesar 93%, akurasi 92,8%, presisi 93%, dan *recall* 93% (Agarap, 2018). Qiang Ye, Ziqiong Zhang, dan Rob Law (2009) juga melakukan penelitian tentang sentiment analisis ulasan destinasi travel menggunakan *Naïve Bayes* dan SVM. Hasil *Naïve Bayes* didapatkan akurasi sebesar 80,71% dan SVM mendapatkan akurasi sebesar 85.14% (Ye *et al.*, 2009).

Dalam Al-Qur'an Surah Ar-Rum ayat 22 Allah SWT berfirman:

وَمِنْ آيَاتِهِ ۖ خَلْقُ السَّمٰوٰتِ وَالْاَرْضِ وَاٰخْتِلَافُ السِّيَاقِ وَالْوٰنِكُمْ اِنَّ فِيْ ذٰلِكَ لَاٰيٰتٍ لِّلْعٰلَمِيْنَ ۚ ۲۲

“Di antara tanda-tanda (kebesaran)-Nya ialah penciptaan langit dan bumi, perbedaan bahasa dan warna kulitmu. Sesungguhnya pada yang demikian itu benar-benar terdapat tanda-tanda (kebesaran Allah) bagi orang-orang yang berilmu.” (QS. Ar-Rum 30: 41).

Surah Ar-Rum ayat 22 menjelaskan analogi tentang tanda-tanda kekuasaan Allah yang tercermin dalam penciptaan langit dan bumi yang begitu besar, rinci, dan cermat. Analogi ini mencerminkan peranan Allah yang teliti dalam menciptakan tatanan alam semesta. Setelah menyebutkan kebesaran Allah melalui penciptaan langit dan bumi, ayat di atas menyatakan adanya keanekaragaman bahasa dan warna kulit. Sebagaimana langit dan bumi memiliki berbagai elemen yang saling terkait dan teratur, klasifikasi menggunakan *neural network* memungkinkan pemisahan dan pengelompokkan aspek-aspek ulasan produk

secara efisien. Pentingnya memahami perbedaan pandangan dalam ulasan produk, sebagaimana langit dan bumi menciptakan keanekaragaman, tercermin dalam kemampuan teknologi klasifikasi menggunakan *neural network* untuk mengidentifikasi, mengelompokkan, dan memberikan arti terhadap aspek-aspek yang beragam dalam ulasan konsumen.

1.2 Pernyataan Masalah

Seberapa baik performa metode *neural network* dalam melakukan klasifikasi aspek produk, layanan pelanggan, dan pengiriman terhadap ulasan *e-commerce*?

1.3 Tujuan Penelitian

Tujuan pada penelitian ini adalah untuk mengetahui bagaimana performa dari metode *neural network* dalam melakukan klasifikasi aspek ulasan *e-commerce*.

1.4 Batasan Masalah

- a. Data yang digunakan adalah komentar-komentar pelanggan platform *e-commerce* Tokopedia dari berbagai kategori produk. *Ground truth* dataset didapatkan dari survei yang melibatkan responden umum.
- b. Bahasa yang digunakan pada data yang dipakai di penelitian ini hanya terbatas pada Bahasa Indonesia.

1.5 Manfaat Penelitian

Hasil output penelitian ini di masa depan dapat dimanfaatkan oleh:

- a. 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.
- b. 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

KAJIAN PUSTAKA

2.1 Kendala Banyaknya Ulasan *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 tahun terakhir, *e-commerce* mengalami pertumbuhan yang sangat pesat. Hal tersebut juga meningkatnya pengguna pelanggan online. Sangat penting untuk menyadari bagaimana dampak yang dimiliki ulasan online terhadap jumlah pengguna *e-commerce* yang terus berkembang ini. Menurut [statista.com](https://www.statista.com), angka pengguna *e-commerce* di Indonesia diprediksi akan terus meningkat secara signifikan. Diperkirakan akan ada 48,2 juta pengguna baru (+24,53 persen) antara tahun 2023 dan 2027 di pasar *e-commerce* Indonesia. Setelah sembilan tahun sejak 2018, total pengguna *e-commerce* akan mencapai 244,67 juta pada 2027 (Statista Research Department, 2023).

Peningkatan jumlah pengguna ini menandakan juga bertambahnya ulasan online pada *e-commerce*. Banyaknya jumlah ulasan ini memiliki pengaruh yang besar terhadap keputusan membeli pelanggan dikemudian hari. Ulasan akan sangat membantu pelanggan yang ragu-ragu untuk memutuskan akan membeli produk atau tidak (Singh *et al.*, 2022). Ulasan menjadi informasi utama dan terpercaya bagi pelanggan. Tidak seperti deskripsi produk yang dapat dimanipulasi, ulasan berasal dari pengalaman-pengalaman yang didapatkan dari

pelanggan-pelanggan sebelumnya. Hal Ini menunjukkan ulasan sendiri bersifat nyata dan menggambarkan informasi sebenarnya dari produk maupun toko online itu sendiri. Oleh karena itu ulasan pelanggan sangatlah penting dalam meningkatkan kepercayaan dan reputasi dari pelaku bisnis di *e-commerce*.

2.2 Klasifikasi Ulasan

Seiring dengan meningkatnya jumlah pengguna *e-commerce*, ulasan pelanggan juga melonjak secara signifikan. Ulasan pelanggan memiliki peranan yang sangat penting bagi pelaku bisnis *e-commerce*. Pelaku bisnis perlu secara aktif mengelola ulasan-ulasan tersebut. Mengelola ulasan dengan jumlah yang melimpah akan memungkinkan pelaku bisnis untuk memahami harapan dan kebutuhan pelanggan dengan lebih mendalam (Noori, 2021). Salah satu caranya adalah dengan mengklasifikasikan ulasan-ulasan tersebut ke dalam tiga aspek utama yaitu produk, layanan pelanggan (*customer service*), dan pengiriman. Ketiga aspek ini sangat sesuai dengan komponen-komponen utama dalam *e-commerce*.

Kategori produk menggambarkan kepuasan pelanggan dengan kualitas, performa, dan kesesuaian produk dengan deskripsi yang diberikan. Kategori layanan pelanggan mencakup interaksi antara pelanggan dengan pelayanan yang diberikan, seperti keramahan, responsivitas, dan hal lainnya yang terkait dengan pelayanan pelanggan. Sementara itu, kategori pengiriman melibatkan aspek kecepatan pengiriman, kondisi barang saat diterima, dan ketepatan waktu pengiriman.

Dengan melakukan pendekatan ini, pelaku bisnis dapat menganalisis pola pelanggan, mengevaluasi kepuasan pelanggan, dan mengidentifikasi bidang mana yang perlu diperbaiki dan ditingkatkan. Pengklasifikasian ulasan pelanggan berdasarkan aspek-aspek seperti produk, layanan pelanggan, dan pengiriman akan sangat bermanfaat bagi pelaku bisnis dalam menganalisis umpan balik pelanggan. Namun, mengerjakan proses ini secara manual bukanlah opsi yang efisien, karena akan memakan waktu yang sangat lama. Tidak hanya itu, cara manual juga akan rentan dengan kesalahan subjektif. Oleh karena itu, sistem klasifikasi ulasan otomatis sangatlah penting dan menjadi teknologi kunci untuk mengelola data teks dengan efisien (Noori, 2021). Dengan pendekatan ini, pelaku bisnis dapat mengekstrak informasi dari ulasan dengan lebih cepat dan akurat. Hal ini berperan penting dalam meningkatkan kepercayaan dan reputasi dari pelaku bisnis di *e-commerce*.

Dalam konteks ini, teknik *natural language processing* (NLP) memiliki peran penting dalam menganalisis teks secara mendalam. NLP merupakan teknik pemrosesan komputasi bahasa natural seperti suara dan teks (Ramdhani *et al.*, 2020). Beberapa penelitian sebelumnya telah melakukan pengklasifikasian ulasan online. Afzaal *et al.*, (2019) melakukan penelitian aplikasi mobile pariwisata dengan focus pada identifikasi aspek dan analisis sentiment berbasis aspek bagi wisatawan. Data yang dipakai adalah data ulasan restoran dan hotel. Teknik seleksi fitur yang dipakai yaitu POS *Tagging* dan *N-grams*. Hasil penelitian mendapatkan dalam domain restoran model *Naïve Bayes Multinomial* (NBM) mendapatkan akurasi sebesar 88.08%. Model *Support Vector Machine* (SVM)

dan *Random Forest* (RF) memiliki akurasi berturut-turut sekitar 87.02% dan 87.42%. Sementara itu, pada domain hotel, model NBM memiliki akurasi tertinggi dengan nilai 90.53%. SVM mendapatkan 89.93% dan RF mendapatkan 87.42%. Hasil ini mencerminkan efektivitas pendekatan berbasis aspek yang diusulkan dalam mengklasifikasikan sentimen wisatawan terhadap restoran dan hotel dalam aplikasi *mobile* pariwisata ini.

Yu *et al.*, (2019) melakukan penelitian klasifikasi konten media sosial X yang berfokus pada *situational awareness* bencana alam. Data yang digunakan adalah data bencana Badar Sandy, Harvey, dan Irma. Penelitian ini mengklasifikasikan konten X ke dalam lima kelas yang berbeda, yaitu *caution and advice, casualties and damage, information sources, infrastructure and resources*, serta *donation and aid*. Seleksi fitur yang digunakan adalah *N-most common words* (2.000.000 kata paling umum pertama) dan teknik *embedding* menggunakan *word2vec embedding*. Hasil didapatkan bahwa metode *Convolutional Neural Network* (CNN) memiliki akurasi sebesar 81%, sedangkan metode lainnya yaitu SVM dan *Logistic Regression* (LR) berturut-turut hanya 63% hingga 72% dan 44% hingga 60%. Penelitian ini menunjukkan bahwa model CNN memiliki kemampuan yang bagus dalam mengklasifikasi data teks media sosial, khususnya pada konteks bencana alam.

Hassan & Shoaib (2020) melakukan penelitian klasifikasikan ulasan pelanggan dalam beberapa kategori rating. Dalam penelitian ini, dua dataset utama digunakan, yaitu IMDB dan ulasan hotel. Metode yang digunakan adalah *Recurrent Neural Network* (RNN) berbasis *Gated-Recurrent-Unit* (GRU) dengan

arsitektur *Deep Semantic Word Embeddings - GRNN* (DSWE-GRNN). Seleksi fitur dilakukan menggunakan dua teknik, yaitu *Continuous Bag of Words* (CBOW) dan *Skip-Gram* (SG), serta *embedding*. Hasil penelitian ini menunjukkan bahwa model DSWE-GRNN memperoleh hasil terbaik dengan akurasi sebesar 87.80% untuk dataset IMDB dan 81.32% untuk dataset ulasan hotel. Model ini mengungguli model-model lainnya, termasuk *Simple Neural Network* (WE-SimpleNN), CNN, *Long Short-Term Memory* (LSTM), dan CNN-LSTM. Penelitian ini menunjukkan keefektifan performa DSWE-GRNN dalam mengatasi kompleksitas klasifikasi multi-kategori dalam ulasan pelanggan.

Noori (2021) dalam penelitiannya melakukan klasifikasi ulasan pelanggan dari situs Tripadvisor.com. Penelitian ini menggunakan seleksi fitur teknik *N-Gram*, *Term Frequency - Inverse Document Frequency* (TF-IDF), *Principal Component Analysis* (PCA), dan *Gini Index*. Hasil penelitian menunjukkan bahwa pada dataset 3 dengan 1800 fitur, model *Naive Bayes* (NB) mendapatkan akurasi sebesar 98.9%, sementara *Decision Tree* (DT) mencapai 98%. Pada dataset 2 dengan 100 fitur, SVM mendapatkan akurasi terbaik dengan nilai sebesar 68.58%. Temuan ini menunjukkan bahwa metode klasifikasi NB dan DT sangat efektif ketika diterapkan pada dataset dengan fitur yang lebih besar, sementara SVM menjadi pilihan terbaik untuk dataset dengan fitur yang lebih terbatas.

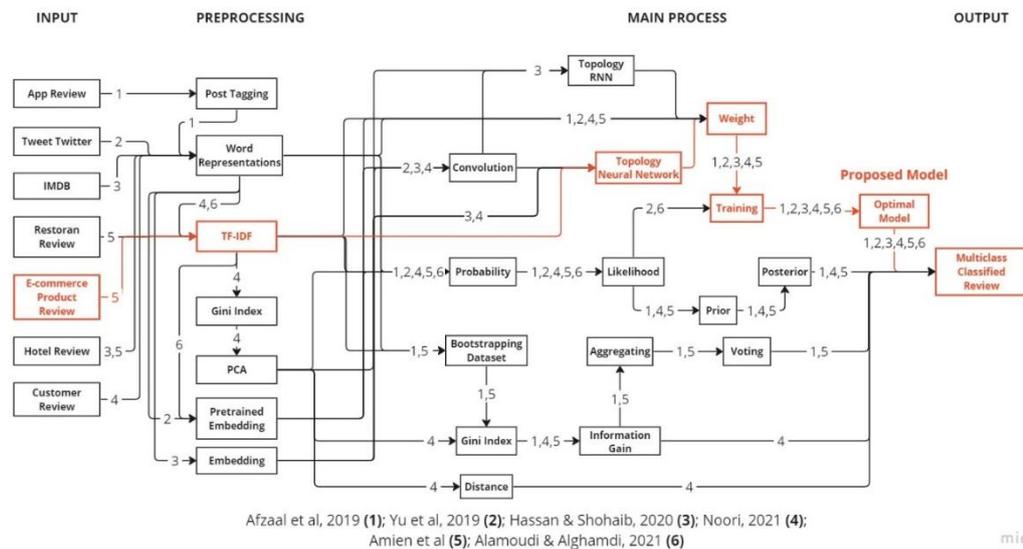
Amien *et al.*, (2021) melakukan analisis sentimen berbasis aspek dalam ulasan media online dalam tiga domain yang berbeda, yaitu restoran, hotel, dan *e-commerce*. Dalam penelitian ini, data ulasan dikelompokkan menjadi tiga kategori sentimen, yaitu positif, negatif, dan netral. TF-IDF dipakai sebagai metode

pembobotan. Metode *machine learning* dalam penelitian ini antara lain NB, SVM, dan RF. Pembagian data dibagi menjadi 80% data *training* dan 20% data *testing*. Penelitian ini didapatkan hasil bahwa model RF mendapatkan performa terbaik di seluruh domain, dengan nilai *precision* 79,3%, *recall* 91,3%, dan *F1-score* 84,3% pada domain restoran. Nilai *precision* 93,0%, *recall* 87,0%, dan *F1-score* 89,3% pada domain *e-commerce*. Nilai *precision* 91,8%, *recall* 81,2%, dan *F1-score* 85,7% pada domain hotel.

Selanjutnya Alamoudi & Alghamdi, (2021) pada penelitiannya melakukan klasifikasi sentimen dan analisis sentimen aspek apada ulasan restoran di Yelp dengan pendekatan *deep learning* dan *word embedding*. Seleksi fitur yang digunakan yaitu *Bag of Words* (BOW), TF-IDF, dan GloVe *embedding*. Hasil penelitian ini mendapatkan model LR mencapai akurasi sebesar 82.815%, NB sebesar 81.112%, dan CNN sebesar 88.295%. Hasil tersebut merupakan klasifikasi 3 kelas dari salah satu skenario yang dilakukan.

2.3 Theoretical Framework

Dalam membuat penelitian tentang klasifikasi ulasan produk pada *e-commerce* menggunakan *neural network* perlu mengacu pada penelitian-penelitian terdahulu yang membentuk sebuah kerangka teori seperti yang terlihat pada Gambar 2.1.



Gambar 2.1 *Theoretical framework*

Blok berwarna merah merepresentasikan alur yang digunakan pada penelitian ini. Secara khusus, data-data yang digunakan sebagai *input* penelitian-penelitian terkait mencakup beragam sumber antara lain ulasan aplikasi, *tweet*, ulasan film (IMDB), ulasan restoran, ulasan produk *e-commerce*, ulasan hotel, dan juga ulasan pelanggan. Dari semua penelitian tersebut memiliki pendekatan yang berbeda pada proses *preprocessing* dan juga metode utama atau *main proses*. Proses *preprocessing* dalam penelitian-penelitian terkait melibatkan langkah-langkah seperti *post-tagging*, *word representation*, TF-IDF, *gini index*, PCA, *word embedding*, dan *embedding* lainnya. Sementara itu, metode utama atau main proses yang diterapkan mencakup penggunaan berbagai model seperti *neural network*, *Recurrent Neural network* (RNN), *Long Short-Term Memory* (LSTM), *Decision Tree*, *Random Forest*, *Naive Bayes*, *K-Nearest Neighbors* (KNN), dan *Support Vector Machine* (SVM). Meskipun terdapat perbedaan dalam pendekatan

dan metode yang digunakan, namun semua penelitian tersebut memiliki tujuan akhir yang serupa, yaitu melakukan klasifikasi ulasan secara *multiclass*.

Pada penelitian ini menggunakan metode *neural network* untuk klasifikasi berdasarkan aspek ulasan produk pada *e-commerce*. Pendekatan *preprocessing* yang digunakan adalah TF-IDF sebagai ekstraksi fitur. Pemilihan alur ini didasarkan pada kompleksitas yang lebih sederhana dibandingkan dengan penelitian – penelitian sebelumnya. Penggunaan metode *neural network* dipilih karena kemampuannya yang terbukti dalam mengatasi tugas klasifikasi yang kompleks. Dengan menggunakan pendekatan ini, diharapkan proses klasifikasi ulasan produk pada *e-commerce* dapat menjadi lebih akurat dan efisien.

BAB III

DESAIN DAN IMPLEMENTASI

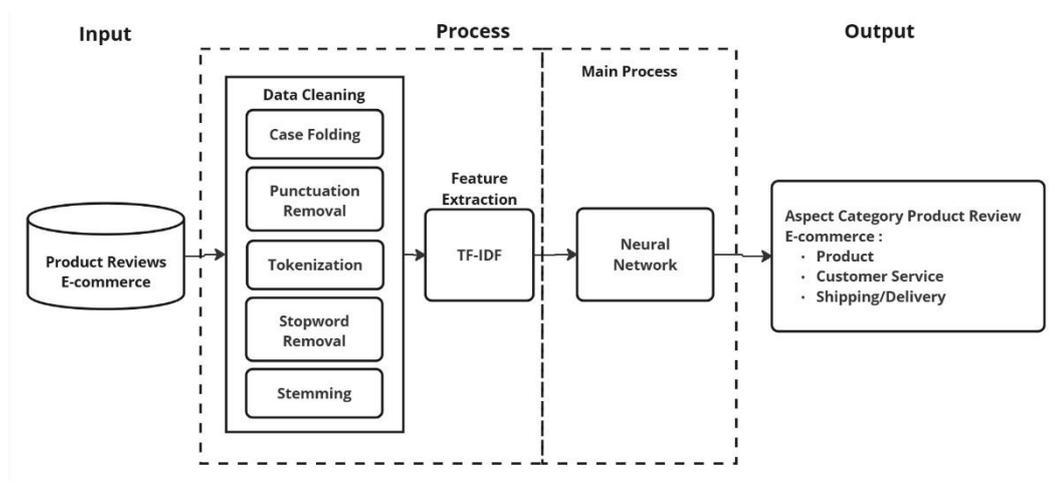
3.1 Akuisisi Data

Pada penelitian ini data yang digunakan diambil dari Kaggle. Data berupa ulasan produk Tokopedia dari berbagai kategori produk, mulai dari fashion, elektronik, olahraga, pertukangan, handphone, dan lain lain. Data yang diambil masih berupa data mentah tanpa label *output* kategori aspek produk, layanan pelanggan, dan pengiriman. Oleh karena itu, perlu dilakukan pelabelan terlebih dahulu pada data yang diambil.

Pelabelan dilakukan dengan melakukan survei kepada orang-orang di setiap data ulasan. Dengan survei ini, setiap data ulasan dapat di kategorikan ke dalam satu tiga kategori yang telah ditetapkan yaitu aspek produk, pelayanan pelanggan, dan pengiriman. Dalam penelitian ini akan digunakan 1024 data ulasan yang telah dilabeli. Data tersebut akan menjadi dasar untuk melatih dan menguji model klasifikasi yang akan dikembangkan pada penelitian ini.

3.2 Desain Sistem

Desain sistem penelitian ini diilustrasikan dalam Gambar 3.1 yang terdiri dari beberapa tahapan, yaitu *input* ulasan produk *e-commerce*, *text preprocessing*, TF-IDF, *neural network*, dan *output* berupa ulasan produk *e-commerce* terklasifikasi.



Gambar 3.1 Desain sistem

3.3 Data Cleaning

Data cleaning tahapan yang penting untuk dilakukan. *Data cleaning* mengubah data mentah menjadi data yang bersih dengan memeriksa nilai yang hilang, data *noisy*, dan juga ketidakkonsistenan. Hal ini dilakukan agar membantu meningkatkan akurasi model klasifikasi (Uysal & Gunal, 2014).

Pada tahap *text preprocessing*, data ulasan produk *e-commerce* yang sudah dimasukkan kedalam *repository* kemudian diekstrak teksnya untuk melewati beberapa proses. Prosesnya yaitu *case folding*, *tokenization*, *stopword removal*, dan yang terakhir adalah *stemming*.

3.3.1 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.

3.3.2 Punctuational 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.

3.3.3 *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*"].

3.3.4 *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".

3.3.5 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". Dari proses ini membantu mengelompokkan kata-kata yang memiliki dasar kata yang sama sehingga pemrosesan teks dapat lebih efektif.

3.4 *Term Frequency – Inverse Document Frequency (TF-IDF)*

Dalam penelitian ini, *Term Frequency – Inverse Document Frequency (TF-IDF)* merupakan tahapan yang sangat penting. Setelah melalui *text preprocessing*, TF-IDF berperan mengubah teks menjadi vektor numerik yang merepresentasikan tingkat kepentingan kata dalam kumpulan data. TF-IDF salah satu algoritma yang sering digunakan dalam penelitian *text mining* (Hakim *et al.*, 2014). TF-IDF mengukur seberapa sering kemunculan suatu kata dalam sebuah dokumen. TF-IDF memiliki konsep jika suatu kata muncul di satu atau sedikit dokumen, maka kata tersebut penting dan ini seharusnya memiliki nilai TF-IDF yang tinggi. Jika kata sering muncul di semua atau kebanyakan dokumen, maka kata tersebut dianggap tidak penting dan memiliki nilai TF-IDF yang rendah (Dalaorao *et al.*, 2019).

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:

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

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

IDF menilai seberapa penting suatu kata di semua dokumen dan dihitung dengan Persamaan 3.2:

$$IDF_{(t)} = \log \frac{|D|}{\sum d_t \in D} \quad (3.2)$$

Keterangan:

$|D|$ = total semua dokumen
 d_t = dokumen d yang mengandung kata t
 $\sum d_t \in D$ = jumlah dokumen d yang mengandung kata t

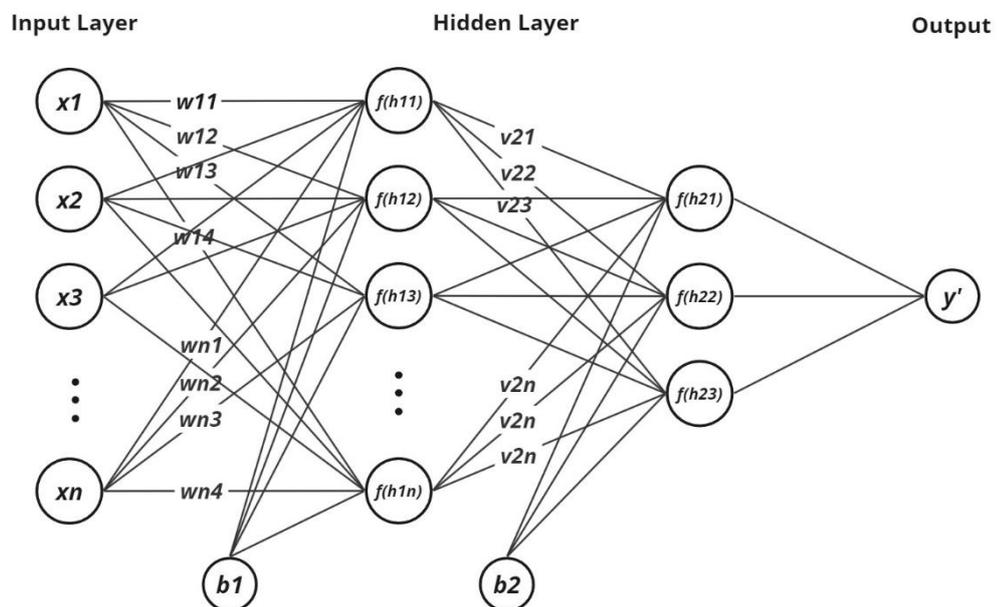
TF-IDF dihitung dengan mengalikan nilai TF dan IDF seperti pada Persamaan 3.3:

$$TFIDF_{(t,d)} = TF_{(t,d)} \times IDF_{(t)} \quad (3.3)$$

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

3.5 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*. Desain topologi *neural network* yang diusulkan dapat dilihat pada Gambar 3.8.



Gambar 3.2 Topologi desain *neural network*

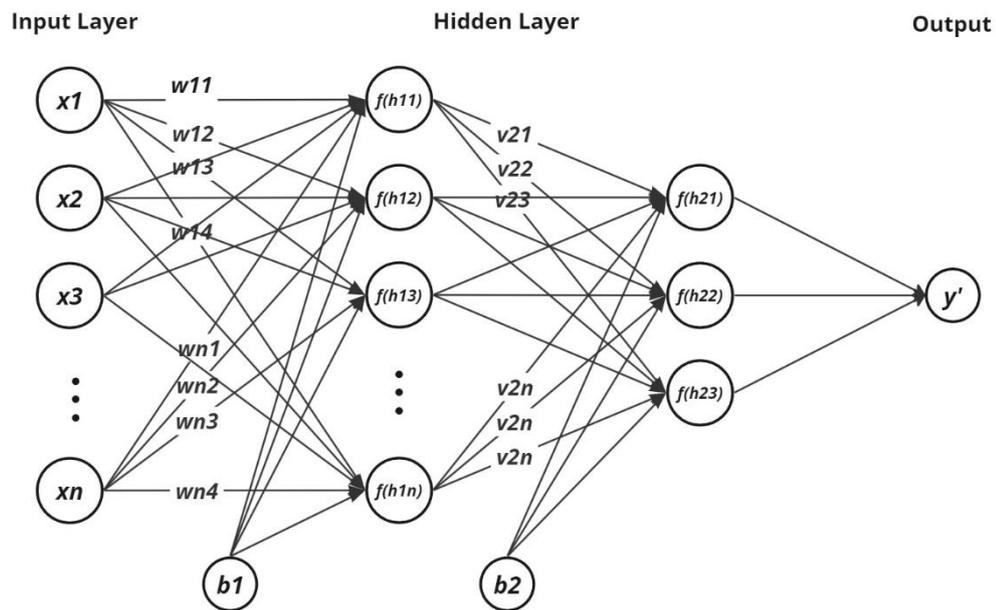
Pada *neural network* terdapat beberapa bagian didalamnya. *Input layer* digunakan untuk menerima informasi dari nilai *input*. *Hidden layer* berlokasi di antara *input layer* dan *output layer* yang memiliki fungsi memetakan pola informasi internal. *Output layer* merupakan hasil dari informasi yang 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* berikutnya (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.

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

- a. *Input layer* : Panjang menyesuaikan *output* vektor TF-IDF.
- b. *Hidden layer 1* : jumlah *node* disesuaikan dengan skenario pengujian.
- c. *Hidden layer 2* : menggunakan 3 *node* karena menyesuaikan dengan total *output* kelas yang akan diprediksi.
- d. *Output layer* : menggunakan 1 *node* karena *output* dari sistem ini adalah salah satu dari kelas yang prediksi.

Dalam *neural network* proses menghitung *input* hingga ke *layer output* dikenal sebagai *feedforward*. Alur proses *feedforward* ditunjukkan pada Gambar 3.9. 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*.



Gambar 3.3 Feedforward neural network

Berikut adalah langkah proses yang terjadi selama *feedforward*:

1. Pertama dilakukan perhitungan pada *hidden layer 1*.

$$h_{1i} = b_{1i} + \sum_j w_{ji}x_i \quad (3.4)$$

Keterangan:

h_{1i} adalah *node hidden layer* pertama indeks ke i .

b_{1i} adalah nilai bias pada *hidden layer* pertama indeks ke i .

$\sum_j w_{ji}x_i$ adalah total dari perkalian bobot (*weight*) *hidden layer* pertama 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.

$$a_i = f(h_i) \quad (3.5)$$

$$a_{1i} = \max(0, a_i) \quad (3.6)$$

Keterangan:

a_{1i} adalah nilai aktivasi dari nilai *hidden layer* pertama 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.

$$h_{2i} = b_{2i} + \sum_j v_{ji} h_{1i} \quad (3.7)$$

Keterangan:

h_{2i} adalah *node hidden layer* kedua indeks ke i .

b_{2i} adalah nilai bias pada *hidden layer* kedua indeks ke i

$\sum_j v_{ji} h_{1i}$ adalah total dari perkalian bobot (*weight*) *hidden layer* kedua dan *hidden layer* pertama.

4. Hasil *hidden layer* 2 dilakukan proses aktivasi menggunakan fungsi *softmax*. Fungsi *softmax* digunakan karena *output* yang diinginkan berupa *categorical*.

$$S(h_{2i}) = \frac{\exp(h_i)}{\sum_{j=1}^3 \exp(h_j)} \quad (3.8)$$

$$y' = S(h_{2i}) \quad (3.9)$$

Keterangan:

S adalah fungsi aktivasi *softmax*.

y' adalah hasil *output* dari model *neural network*.

5. Hasil proses fungsi aktivasi *softmax* merupakan hasil *output*. Langkah selanjutnya adalah menghitung nilai *error* menggunakan *cross-entropy loss*.

$$E = - \sum_{j=1}^3 y \log (y') \quad (3.10)$$

Keterangan:

y adalah *output* target *ground truth*.

y' adalah *output* prediksi dari model.

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

3.6 Implementasi Sistem

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). Proses *training* ini akan menghasilkan nilai *error* yang selanjutnya digunakan untuk melakukan *update* dan optimasi nilai bobot yang terdapat di dalam model *neural network*. Setelah dilakukan *training*, model diharapkan dapat mempelajari data yang ada. Pada tahap selanjutnya, model *neural network* akan melalui proses pengujian (*testing*) untuk mengukur kemampuan dari model yang telah dibuat.

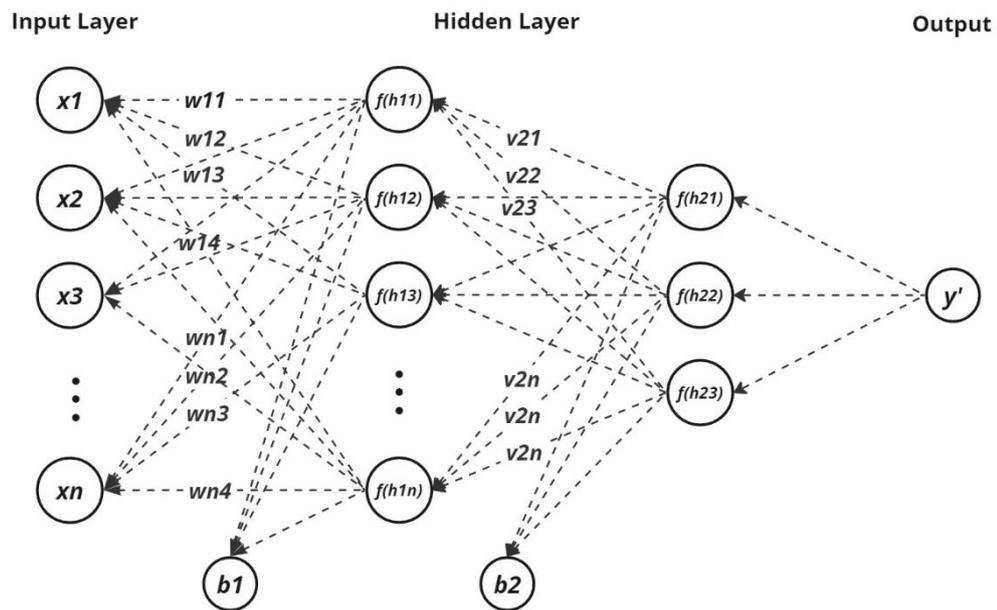
3.6.1 Proses *Training*

Tahapan pertama yang dilakukan pada model adalah *training*. Pada tahap ini, model *neural network* dilatih menggunakan data *training*. 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*, dimana data tersebut berbentuk vektor TF-IDF. Nilai dari vektor TF-IDF dimasukkan kedalam *input layer*. Kemudian, menjalani *training* (*feedforward* dan *backpropagation*) sebanyak jumlah *epoch* yang telah ditentukan. Alur dari proses *backpropagation* ditunjukkan pada Gambar 3.10.



Gambar 3.4 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 *hidden layer* 2.

$$\delta_2 = (y - h_2) \quad (3.11)$$

$$\frac{dL}{dV} = \delta_2 (h_1)^T \quad (3.12)$$

$$\frac{dL}{db_2} = \delta_2 \quad (3.13)$$

Keterangan:

δ_2 adalah *error* pada *hidden layer* kedua.

y adalah target *output*.

$\frac{dL}{dV}$ adalah turunan bobot pada *hidden layer* kedua.

$\frac{dL}{db_2}$ adalah turunan bias pada *hidden layer* kedua.

$(h_1)^T$ adalah nilai *transpose hidden layer* pertama.

$\sigma'(h_2)$ adalah turunan dari fungsi aktivasi pada *output layer*.

Error atau kesalahan pada *hidden layer* kedua dihitung dan diteruskan ke *layer* sebelumnya untuk menghitung kesalahan bobot dan bias pada *hidden layer* kedua.

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

$$\delta_1 = V^T * \delta_2 * \sigma'(a_1) \quad (3.14)$$

$$\frac{dL}{dW} = \delta_1 x \quad (3.15)$$

$$\frac{dL}{db_1} = \delta_1 \quad (3.16)$$

Keterangan:

δ_1 adalah *error* pada *hidden layer* pertama.

$\frac{dL}{dW}$ adalah turunan bobot pada *hidden layer* pertama.

$\frac{dL}{db_1}$ adalah turunan bias pada *hidden layer* pertama.

V^T adalah nilai *transpose* matriks bobot antara *hidden layer* pertama dan *hidden layer* kedua.

$\sigma'()$ adalah turunan dari fungsi aktivasi ReLU pada *hidden layer* pertama.

x adalah nilai *input*.

Error dari *hidden layer* kedua digunakan untuk mencari *error* pada *hidden layer* pertama. Selanjutnya hasil dari *error hidden layer* pertama digunakan untuk menghitung turunan dari bobot dan bias pada *hidden layer*

pertama. Dengan begitu penyesuaian yang diperlukan untuk bobot dan bias *hidden layer* pertama berdasarkan apa yang terjadi pada *hidden layer* kedua.

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

$$V^t = V^{t-1} - \alpha \frac{dL}{dV} \quad (3.17)$$

$$b_2^t = b_2^{t-1} - \alpha \frac{dL}{db_2} \quad (3.18)$$

$$W^t = W^{t-1} - \alpha \frac{dL}{dW} \quad (3.19)$$

$$b_1^t = b_1^{t-1} - \alpha \frac{dL}{db_1} \quad (3.20)$$

Keterangan:

α adalah laju pembelajaran (*learning rate*).

$\frac{dL}{dV}$ adalah turunan bobot pada *hidden layer* kedua.

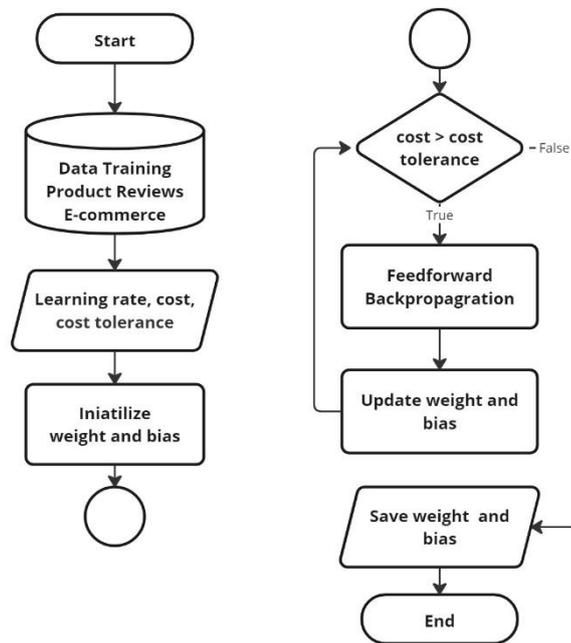
$\frac{dL}{db_2}$ adalah turunan bias pada *hidden layer* kedua.

$\frac{dL}{dW}$ adalah turunan bobot pada *hidden layer* pertama.

$\frac{dL}{db_1}$ adalah turunan bias pada *hidden layer* pertama.

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 3.5 Flowchart proses training

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Data Pengujian

Data yang digunakan dalam penelitian ini didapatkan dari platform Kaggle. Namun, data tersebut masih belum memiliki label atau *ground truth*. *Ground truth* data ini diperoleh dari survei yang melibatkan responden umum. Rincian jumlah ulasan produk e-commerce berdasarkan hasil survei kepada responden umum ditunjukkan pada Tabel 4.1.

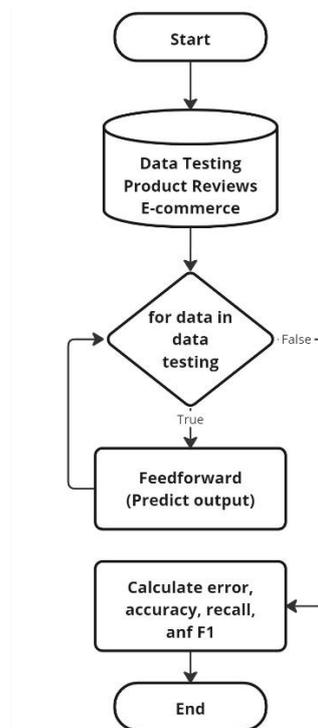
Tabel 4.1 Hasil survei *ground truth*

Label Ground Truth	Jumlah
Produk	523
Layanan Pelanggan	329
Pengiriman	172
Total	1024

Total responden: 81

Masing-masing ulasan dijadikan sebagai pertanyaan dalam survei dan dinilai berdasarkan sistem voting, Dimana nilai terbesar akan dijadikan *ground truth* yang paling relevan. Total responden dalam survei adalah 81 responden. Contoh *ground truth* ditampilkan pada Tabel 4.2.

melihat bagaimana performa kinerja dari model *neural network* yang dibuat untuk klasifikasi aspek ulasan produk *e-commerce*.



Gambar 4.1 *Flowchart* proses *testing*

Pada penelitian ini dilakukan pengujian terhadap model klasifikasi aspek ulasan produk *e-commerce* dengan mengukur akurasi, presisi, *recall*, dan *F1-score*. Data akan dibagi menggunakan rasio data *training* dan data *testing* dengan pembagian rasio data 8:2, yaitu 819 data untuk *training* dan 204 data untuk *testing*. Untuk menemukan parameter model yang optimal, penelitian ini melakukan beberapa skenario terhadap dua parameter penting, yaitu:

- a. Jumlah *node hidden layer* pertama: Parameter ini menentukan kompleksitas model dan kemampuannya dalam mempelajari pola dalam data.
- b. *Learning rate*: Parameter ini mengendalikan kecepatan model dalam mempelajari dan memperbarui bobot dan biasanya dalam proses *training*.

Pada penelitian ini digunakan beberapa skenario pengujian pada model. Masing masing model memiliki variasi yang berbeda pada jumlah *node hidden layer* pertama dan *learning rate*. Keseluruhan skenario pengujian pada model disajikan pada Tabel 4.3.

Tabel 4.3 Skenario pengujian

Model	Jumlah Node Hidden Layer Pertama	Learning Rate
A1	25	0.01
A2	25	0.05
A3	25	0.1
B1	50	0.01
B2	50	0.05
B3	50	0.1
C1	100	0.01
C2	100	0.05
C3	100	0.1

Variasi jumlah *node hidden layer* pertama ditampilkan dalam tiga pilihan, yaitu 25, 50, dan 100. Jumlah *node input* dalam penelitian ini adalah 1085, yang telah disesuaikan dengan panjang vektor dari proses TF-IDF. Jumlah kelas *output* adalah 3 kelas. Berdasarkan penelitian Rachmatullah *et al.*,(2020), rumus untuk menentukan jumlah *node hidden layer* dengan *input* 1085 dan *output* 3 kelas menghasilkan 57 *node hidden layer* pertama. Sedangkan berdasarkan rumus pada penelitian Yotov *et al.*,(2023), dengan *input* 1085, jumlah *node hidden layer* pertama adalah 46. Pada penelitian ini diambil nilai di antara kedua hasil tersebut, yaitu 50. Variasi 25 dan 100 dipilih untuk melihat bagaimana model berperilaku terhadap jumlah *node hidden layer* pertama yang berbeda.

Pada *learning rate* terdapat tiga variasi yang akan digunakan, yaitu 0.01, 0.05, dan 0.1. Penelitian yang dilakukan oleh Suhendra & Saputra, (2020) menunjukkan bahwa *learning rate* sebesar 0.1 memberikan proses *training* yang stabil dan nilai *error* yang kecil. Sebagai tambahan, *error* dapat dicapai lebih kecil lagi dengan nilai *learning rate* 0.01. Akan tetapi, nilai *learning rate* 0.01 memerlukan proses komputasi yang lebih lama daripada nilai *learning rate* 0.1. Pada penelitian nilai *learning rate* 0.05 akan digunakan sebagai pembandingan tambahan untuk melihat bagaimana model berperilaku pada proses *training*.

Dari 9 model skenario pengujian yang diajukan, dilakukan proses *training* dan *testing*. Setelah mendapatkan nilai akurasi tertinggi, Langkah selanjutnya adalah menghitung *confusion matrix*. *Confusion matrix* membantu untuk memberikan gambaran performa klasifikasi dari model (Bujang *et al.*, 2021). Hasil dari *confusion matrix* terdiri dari empat kategori. *True Positive* (TP) adalah Ketika data ulasan yang berhasil diprediksi benar sebagai kelas yang positif. Ketika model salah memprediksi kelas positif, maka akan menghasilkan *False Positive* (FP). *True Negative* (TN) adalah Ketika model benar memprediksi kelas yang negatif. Ketika model salah memprediksi kelas yang negatif, maka akan menghasilkan *False Negative* (FN). Pada kasus permasalahan multiclass atau lebih dari dua kelas kategori, kelas positif adalah label dimana perhitungan itu dilakukan, sedangkan kelas negatif adalah kelas lainnya (Ali *et al.*, 2022).

Tabel 3.2 merepresentasikan *confusion matrix* digunakan untuk mengukur performa model klasifikasi aspek ulasan produk *e-commerce* dimana A, B, C merepresentasikan kelas aspek ulasan produk *e-commerce*. A untuk aspek produk,

B untuk aspek Layanan Pelanggan, dan C untuk aspek Pengiriman. *Confussion matrix* untuk penelitian ini ditampilkan pada Tabel 4.4.

Tabel 4.4 *Confussion matrix*

Kelas Aktual	Kelas Prediksi		
	Produk (A)	Layanan Pelanggan (B)	Pengiriman (C)
Produk (A)	AA	AB	AC
Layanan Pelanggan (B)	BA	BB	BC
Pengiriman (C)	CA	CB	CC

Dalam penelitian ini, hasilnya terdiri dari tiga kelas *output*. *Confussion matrix* untuk setiap kelas memiliki empat kategori: TP, TN, FP, dan FN. Kategori-kategori ini digunakan untuk menunjukkan berapa banyak kasus *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) dalam setiap kelas. *Confussion matrix* masing-masing kelas ditampilkan pada Tabel 4.5, Tabel 4.6, Tabel 4.7.

Tabel 4.5 *Confussion matrix* kelas produk

Kelas Aktual	Kelas Prediksi	
	Produk (A)	Bukan Produk
Produk (A)	<i>True Positive</i> (AA)	<i>False Negative</i> (AB + AC)
Bukan Produk	<i>False Positive</i> (BA + CA)	<i>True Negative</i> (BB + BC + CB + CC)

Tabel 4.6 *Confussion matrix* kelas layanan pelanggan

Kelas Aktual	Kelas Prediksi	
	Layanan Pelanggan (B)	Bukan Layanan Pelanggan
Layanan Pelanggan (B)	<i>True Positive</i> (BB)	<i>False Negative</i> (BA + BC)
Bukan Layanan Pelanggan	<i>False Positive</i> (AB + CB)	<i>True Negative</i> (AA + AC + CA + CC)

Tabel 4.7 *Confusion matrix* kelas pengiriman

Kelas Aktual	Kelas Prediksi	
	Pengiriman (C)	Bukan Pengiriman
Pengiriman (C)	<i>True Positive</i> (CC)	<i>False Negative</i> (CA + CB)
Bukan Pengiriman	<i>False Positive</i> (AC + BC)	<i>True Negative</i> (AA+ AB+ BA+ BB)

Berikut adalah penjelasan mengenai empat komponen utama dari *confusion matrix*:

1. *True Positive* (TP) : Ulasan dengan hasil pengklasifikasian kategori *i*, diprediksi benar oleh sistem bahwa ulasan masuk kategori *i*.
2. *True Negative* (TN) : Ulasan dengan hasil pengklasifikasian bukan kategori *i*, diprediksi benar oleh sistem bahwa ulasan tidak masuk kategori *i*.
3. *False Positive* (FP) : Ulasan dengan hasil pengklasifikasian bukan kategori *i*, diprediksi salah oleh sistem bahwa ulasan masuk kategori *i*.
4. *False Negative* (FN) : Ulasan dengan hasil pengklasifikasian kategori *i*, diprediksi salah oleh sistem bahwa ulasan tidak masuk kategori *i*.

Dari tabel *confusion matrix* dapat digunakan untuk mengukur nilai akurasi, presisi, *recall*, dan *F1-score* dari model dengan menggunakan persamaan berikut:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4.21)$$

$$Precision = \frac{TP}{TP + FP} \quad (4.22)$$

$$Recall = \frac{TP}{TP + FN} \quad (4.23)$$

$$F1 - Score = 2 \frac{Recall \times Precision}{Recall + Precision} \quad (4.24)$$

Accuracy adalah nilai persentase prediksi benar terhadap semua data. *Precision* adalah rasio prediksi benar kelas positif terhadap total semua kelas positif. *Recall* adalah proporsi kelas positif yang berhasil diprediksi dengan benar. *F1-score* adalah rata-rata bobot dari presisi dan *recall*. Nilai *F1-score* yang lebih tinggi menunjukkan bahwa performa model yang lebih baik (Ali *et al.*, 2022).

4.3 Hasil Uji Coba

Sebelum data ulasan dimasukkan ke dalam *neural network*, masing masing data ulasan perlu dilakukan *preprocessing* terlebih dahulu melalui *case folding*, *punctuation removal*, *tokenization*, *stopword removal*, *stemming*, dan dilanjutkan pembobotan TFIDF. Hasil dari proses *preprocessing* data ulasan ditampilkan pada Tabel 4.8.

Tabel 4.8 Hasil *preprocessing* data ulasan

No	Sebelum Preprocessing	Setelah Preprocessing (tanpa TFIDF)
1	barang sudah saya terima thanks	barang terima thanks
2	Fast ResponseFriendly Recommended Seller	fast responsefriendly recommended seller
3	Seller fast respon Produk sudah diterima buat stok. Thanks	seller fast respon produk terima stok thanks
...
1024	Barang bagus harga maknyuuzzz Top markotop hahahaa	barang bagus harga maknyuuzzz top markotop hahahaa

Setelah data melewati proses *stemming*, dilakukan proses TF-IDF pada data *training* untuk semua data. Didapatkan model TF-IDF yang kemudian digunakan untuk mengubah data ulasan *testing* menjadi bentuk vektor TF-IDF. Hasil dari proses *preprocessing* TFIDF data ulasan ditampilkan pada Tabel 4.9.

Tabel 4.9 Hasil *preprocessing* TFIDF data ulasan

No	Teks Ulasan	Vektor TFIDF
1	barang terima thanks	[0.16663067, 0.34765416, 0.4372591, ..., 0, 0, 0]
2	fast responsefriendly recommended seller	[0, 0, 0, ..., 0, 0, 0]
3	seller fast respon produk terima stok thanks	[0, 0, 0, ..., 0, 0, 0]
...	...	
1024	barang bagus harga maknyuuzzz top markotop hahahaa	[0, 0, 0, ..., 0, 0, 0]

Vektor TF-IDF terdiri dari angka-angka yang merepresentasikan bobot setiap kata dalam ulasan berdasarkan frekuensi kemunculan dan kepentingannya dalam keseluruhan kumpulan data. Panjang vektor dari vektor TFIDF ini adalah 1085. Vektor TFIDF inilah yang akan dimasukkan kedalam *input layer neural network*. Banyaknya *node input layer neural network* akan diatur sesuai dengan panjang vektor TFIDF.

Sesuai dengan skenario pengujian, penelitian ini dilakukan 9 kali uji coba dengan scenario jumlah *node hidden layer* pertama dan *learning rate* yang berbeda. Skenario pemberhentian proses *training* dipilih menggunakan perubahan nilai *cost* yang telah ditentukan. Nilai rasio data dan perubahan *cost* yang dipakai pada penelitian ini ditampilkan pada Tabel 4.10.

Tabel 4.10 Rasio data dan skenario pemberhentian proses *training*

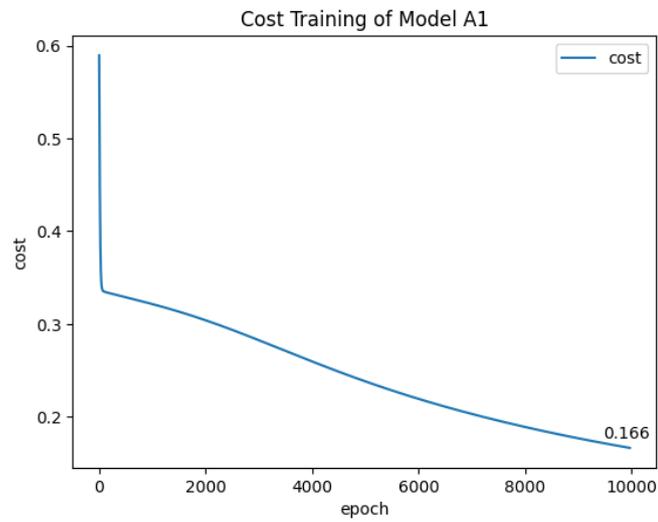
Variabel	Nilai
Rasio data <i>training</i> dan <i>testing</i>	8:2 (819 data <i>training</i> dan 204 data <i>testing</i>)
Skenario pemberhentian (selisih <i>cost</i> setiap <i>epoch</i>)	0,00001

Rasio pembagian dataset yang digunakan untuk semua skenario adalah sama, yaitu 80:20, yaitu 819 data *training* dan 205 data *testing*. Proses *training* dilakukan hingga grafik *cost* terjadi konvergen atau sudah tidak berubah lagi

dengan waktu yang cukup lama. Hal ini dapat dilakukan dengan menentukan nilai selisih *cost* untuk skenario pemberhentian pada proses *training*. Berdasarkan pada penelitian yang dilakukan oleh Prechelt, (1998) disebutkan bahwa tidak ada ketentuan pasti dalam pemilihan nilai selisih *cost* untuk skenario pemberhentian. Penelitian tersebut menetapkan nilai selisih *cost* yang dapat digunakan adalah di bawah 0,1. Pada penelitian ini nilai selisih *cost* yang dipilih adalah 0,00001. Skenario pemberhentian ini memiliki arti jika selisih nilai *cost* pada *epoch* terakhir dengan *epoch* sebelumnya tidak lebih kecil dari 0,00001 maka proses *training* akan terus berlanjut, sedangkan jika selisih *cost* lebih kecil dari 0,00001 maka akan proses *training* akan diberhentikan. Tujuan dari skenario ini adalah untuk menemukan parameter terbaik yang dapat memberikan proses *training* paling efisien dan hasil paling optimal untuk model klasifikasi. Masing-masing model klasifikasi yang dihasilkan dievaluasi performanya dengan menggunakan parameter akurasi, presisi, *recall*, dan F1-score. Untuk mendapatkan nilai-nilai tersebut, kelas aktual dibandingkan dengan kelas prediksi sehingga dapat terlihat berapa data ulasan produk yang diprediksi dengan benar dan tidak.

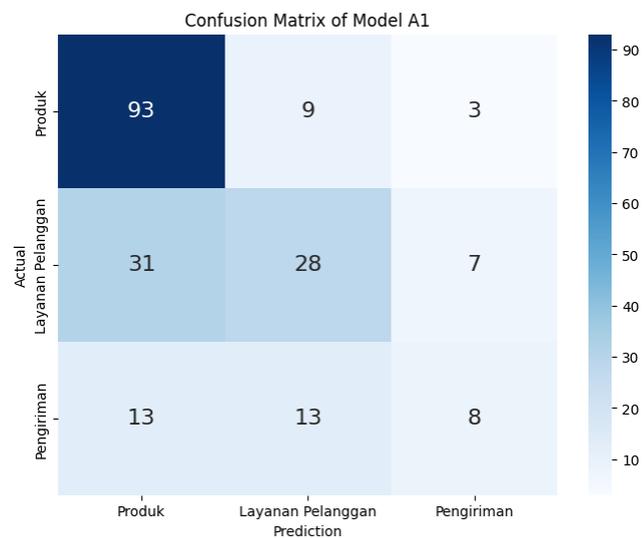
4.3.1 Model A1

Uji coba model A1 menggunakan jumlah *node hidden layer* pertama sebanyak 25 *node* dan *learning rate* sebesar 0.01. Proses *training* model A1 yang menggunakan data *training* ditunjukkan pada Gambar 4.2. Dari proses *training* model A1 memakan waktu sekitar 388,33 detik atau 6,47 menit dengan jumlah *epoch* sebanyak 9963 *epoch*. Dari proses *training* model A1 juga didapatkan nilai *cost* sebesar 0,166438.



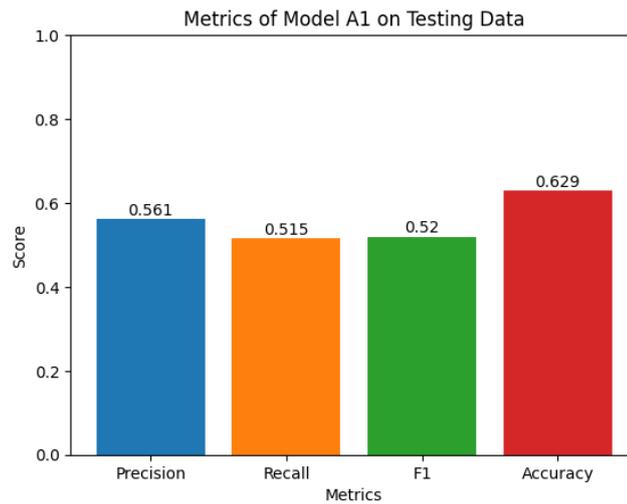
Gambar 4.2 Grafik *cost* proses *training* model A1

Hasil *confusion matrix* model A1 pada data *testing* dapat dilihat pada Gambar 4.3. Hasil tersebut menunjukkan model A1 berhasil memprediksi kelas dengan benar sebanyak 129 data dan gagal memprediksi kelas dengan benar sebanyak 76 data.



Gambar 4.3 *Confusion matrix* model A1

Berdasarkan Persamaan 4.1, 4.2, 4.3, dan 4.4, dapat diperoleh performa model dalam hal akurasi, presisi, *recall*, dan *F1-score*. Nilai-nilai performa dari model A1 ini kemudian ditampilkan secara visual pada Gambar 4.4.



Gambar 4.4 Metrik performa model A1

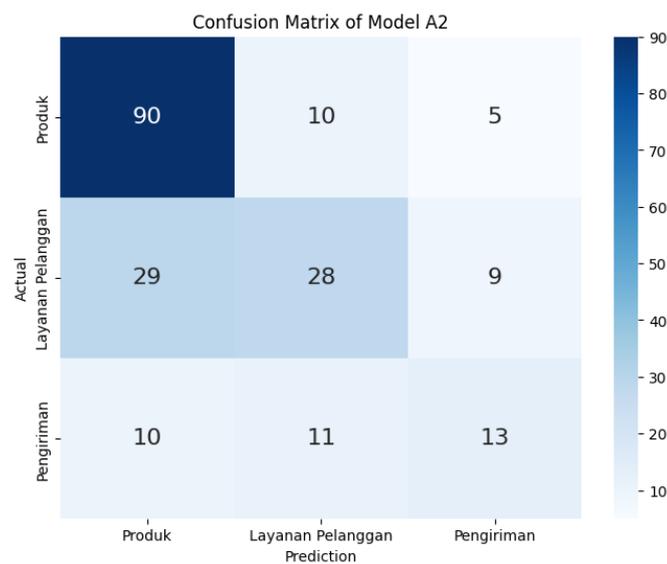
4.3.2 Model A2

Uji coba model A2 menggunakan jumlah *node hidden layer* pertama sebanyak 25 *node* dan *learning rate* sebesar 0.05. Proses training model A2 yang menggunakan data *training* ditunjukkan pada Gambar 4.5. Dari proses training model A2 memakan waktu sekitar 283,68 detik atau 4,73 menit dengan jumlah epoch sebanyak 4987 epoch. Dari proses training model A2 juga didapatkan nilai *cost* sebesar 0,098499.



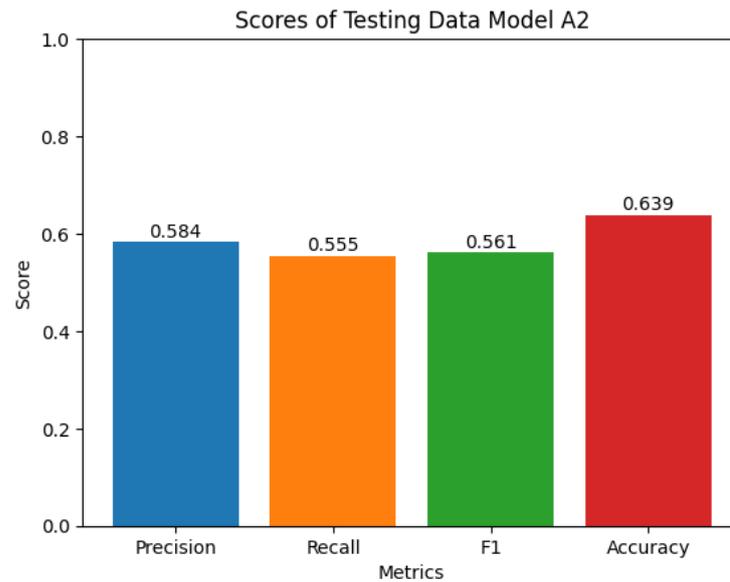
Gambar 4.5 Grafik *cost* proses *training* model A2

Hasil *confusion matrix* pada data *testing* dapat dilihat pada Gambar 4.6. Hasil tersebut menunjukkan model berhasil memprediksi kelas dengan benar sebanyak 131 data dan gagal memprediksi kelas dengan benar sebanyak 74 data.



Gambar 4.6 *Confusion matrix* model A2

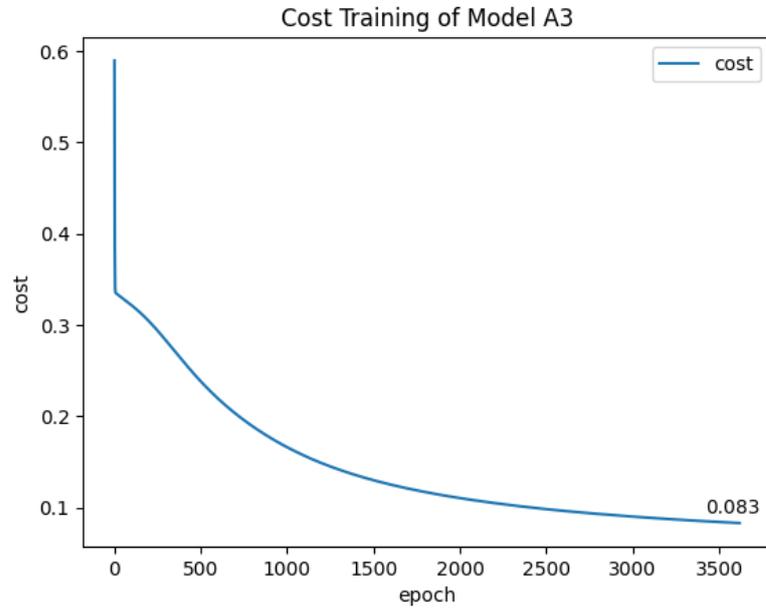
Berdasarkan Persamaan 4.1, 4.2, 4.3, dan 4.4, dapat diperoleh performa model dalam hal akurasi, presisi, *recall*, dan *F1-score*. Nilai-nilai performa dari model A2 ini kemudian ditampilkan secara visual pada Gambar 4.7.



Gambar 4.7 Metrik performa model A2

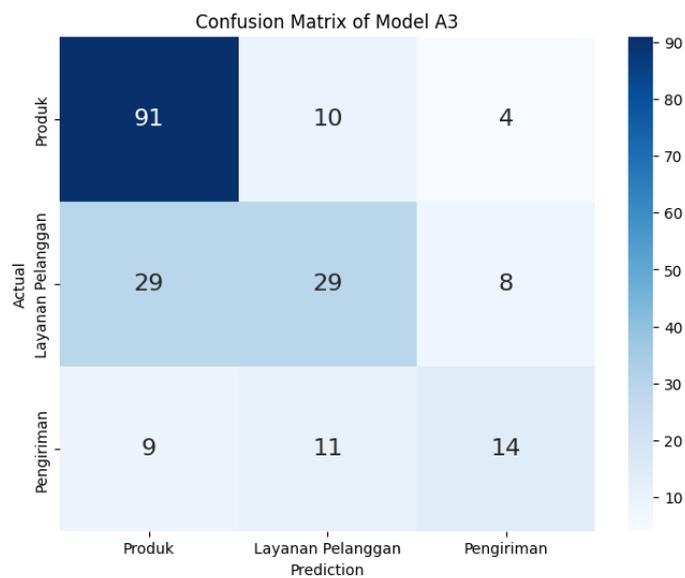
4.3.3 Model A3

Uji coba model A3 menggunakan jumlah *node hidden layer* pertama sebanyak 25 *node* dan *learning rate* sebesar 0.1. Proses *training* model A3 yang menggunakan data *training* dapat ditunjukkan pada Gambar 4.8. Dari proses *training* model A3 memakan waktu sekitar 214,31 detik atau 3,57 menit dengan jumlah *epoch* sebanyak 3618 *epoch*. Dari proses *training* model A3 juga didapatkan nilai *cost* sebesar 0,083149.



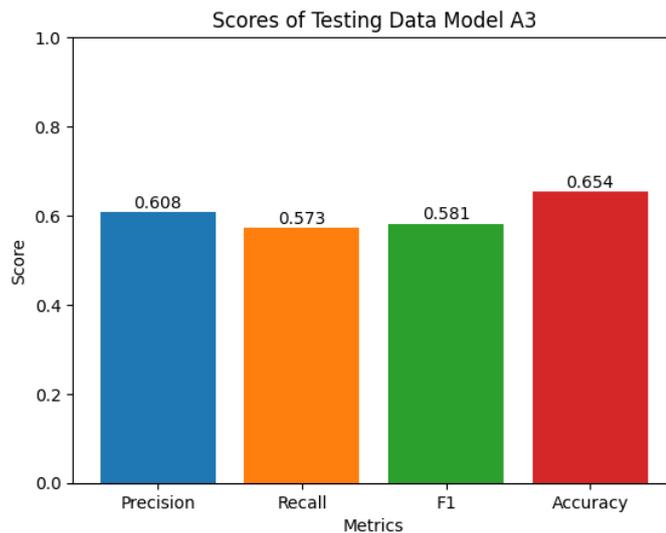
Gambar 4.8 Grafik *cost* proses *training* model A3

Hasil *confusion matrix* pada data *testing* dapat dilihat pada Gambar 4.9. Hasil tersebut menunjukkan model berhasil memprediksi kelas dengan benar sebanyak 134 data dan gagal memprediksi kelas dengan benar sebanyak 71 data.



Gambar 4.9 *Confusion matrix* model A3

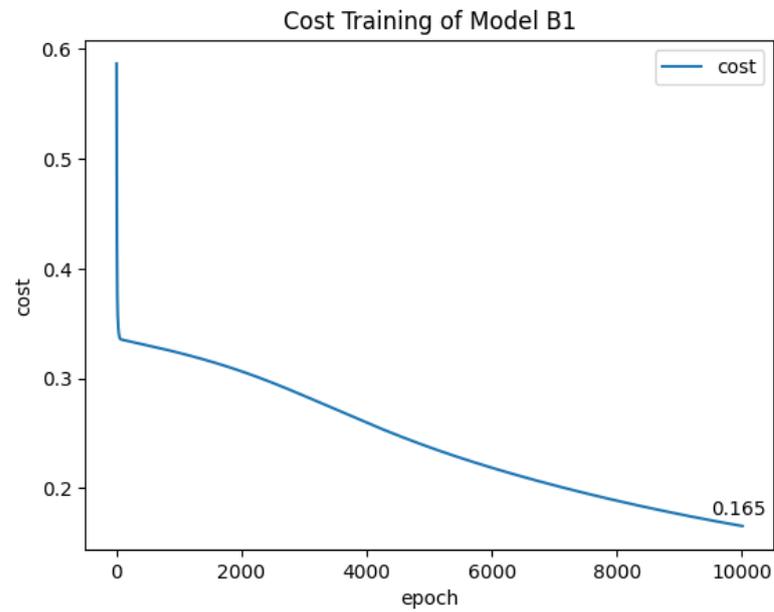
Berdasarkan Persamaan 4.1, 4.2, 4.3, dan 4.4, dapat diperoleh performa model dalam hal akurasi, presisi, *recall*, dan *F1-score*. Nilai-nilai performa dari model A3 ini kemudian ditampilkan secara visual pada Gambar 4.10.



Gambar 4.10 Metrik performa model A3

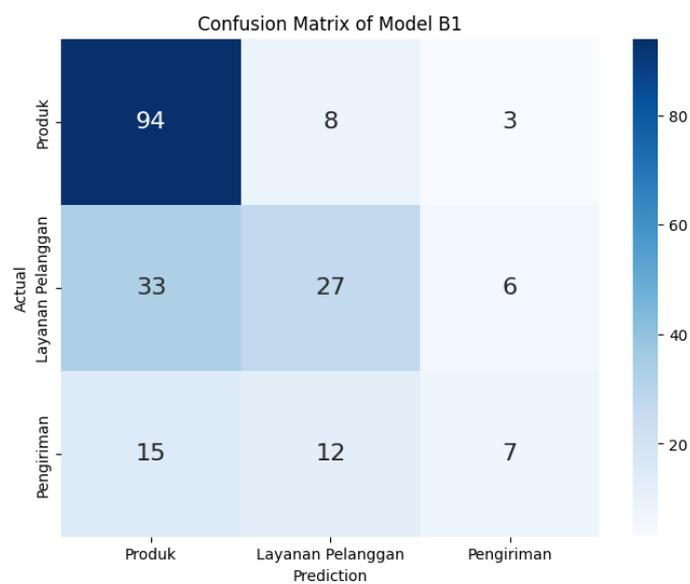
4.3.4 Model B1

Uji coba model B1 menggunakan jumlah node hidden layer pertama sebanyak 50 node dan learning rate sebesar 0.01. Proses *training* model B1 yang menggunakan data *training* ditunjukkan pada Gambar 4.11. Dari proses *training* model B1 memakan waktu sekitar 817,38 detik atau 13,62 menit dengan jumlah *epoch* sebanyak 10009 *epoch*. Dari proses *training* model B1 juga didapatkan nilai *cost* sebesar 0,165462.



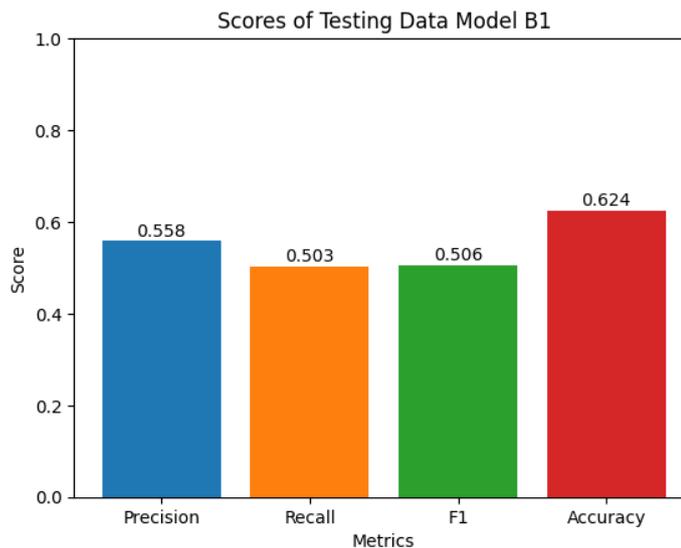
Gambar 4.11 Grafik *cost* proses *training* model B1

Hasil *confusion matrix* pada data *testing* dapat dilihat pada Gambar 4.12. Hasil tersebut menunjukkan model berhasil memprediksi kelas dengan benar sebanyak 128 data dan gagal memprediksi kelas dengan benar sebanyak 77 data.



Gambar 4.12 *Confusion matrix* model B1

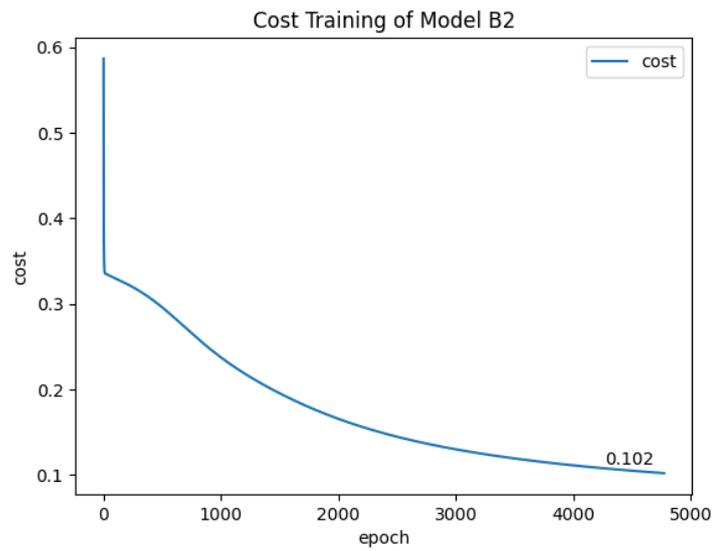
Berdasarkan Persamaan 4.1, 4.2, 4.3, dan 4.4, dapat diperoleh performa model dalam hal akurasi, presisi, *recall*, dan *F1-score*. Nilai-nilai performa dari model B1 ini kemudian ditampilkan secara visual pada Gambar 4.13.



Gambar 4.13 Metrik performa model B1

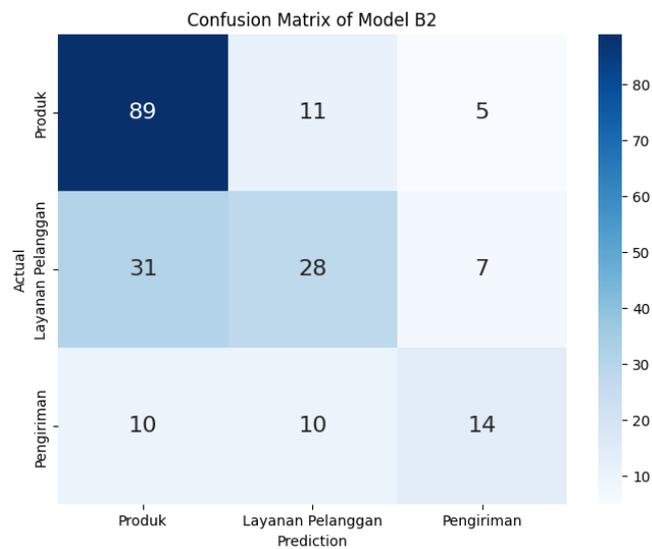
4.3.5 Model B2

Uji coba model B2 menggunakan jumlah *node hidden layer* pertama sebanyak 50 *node* dan *learning rate* sebesar 0.05. Proses *training* model B2 yang menggunakan data *training* ditunjukkan pada Gambar 4.14. Dari proses *training* model B2 memakan waktu sekitar 428,86 detik atau 7,15 menit dengan jumlah *epoch* sebanyak 4775 *epoch*. Dari proses *training* model B2 juga didapatkan nilai *cost* sebesar 0,102014.



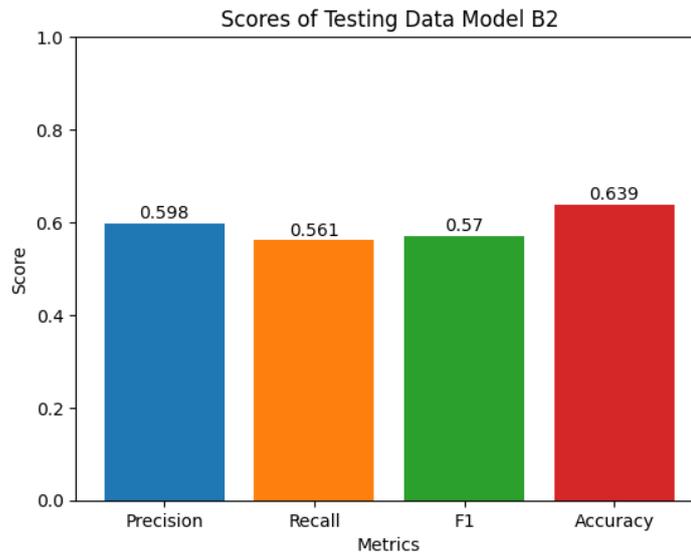
Gambar 4.14 Grafik *cost* proses *training* model B2

Hasil *confusion matrix* pada data *testing* dapat dilihat pada Gambar 4.15. Hasil tersebut menunjukkan model berhasil memprediksi kelas dengan benar sebanyak 131 data dan gagal memprediksi kelas dengan benar sebanyak 74 data.



Gambar 4.15 *Confusion matrix* model B2

Berdasarkan Persamaan 4.1, 4.2, 4.3, dan 4.4, dapat diperoleh performa model dalam hal akurasi, presisi, *recall*, dan *F1-score*. Nilai-nilai performa dari model B2 ini kemudian ditampilkan secara visual pada Gambar 4.16.



Gambar 4.16 Metrik performa model B2

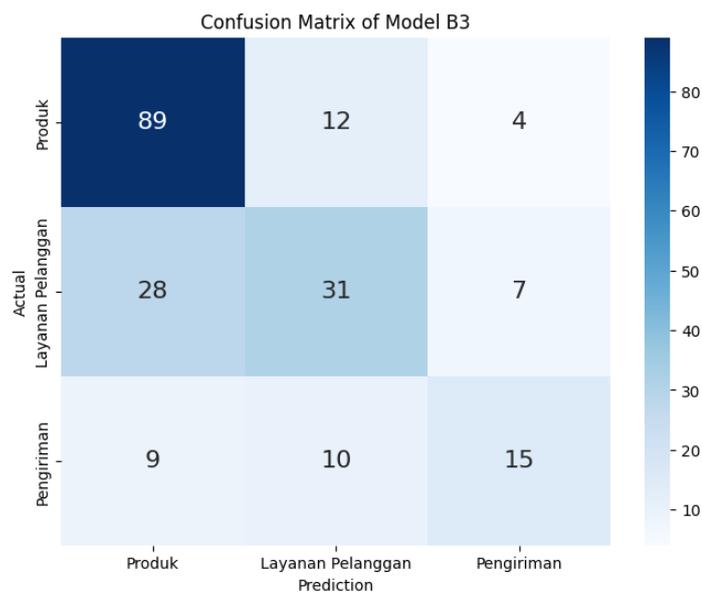
4.3.6 Model B3

Uji coba model B3 menggunakan jumlah *node hidden layer* pertama sebanyak 50 *node* dan *learning rate* sebesar 0.1. Proses *training* model B3 yang menggunakan data *training* ditunjukkan pada Gambar 4.17. Dari proses *training* model B3 memakan waktu sekitar 283,35 detik atau 4,72 menit dengan jumlah *epoch* sebanyak 3379 *epoch*. Dari proses *training* model B3 juga didapatkan nilai *cost* sebesar 0,088147.



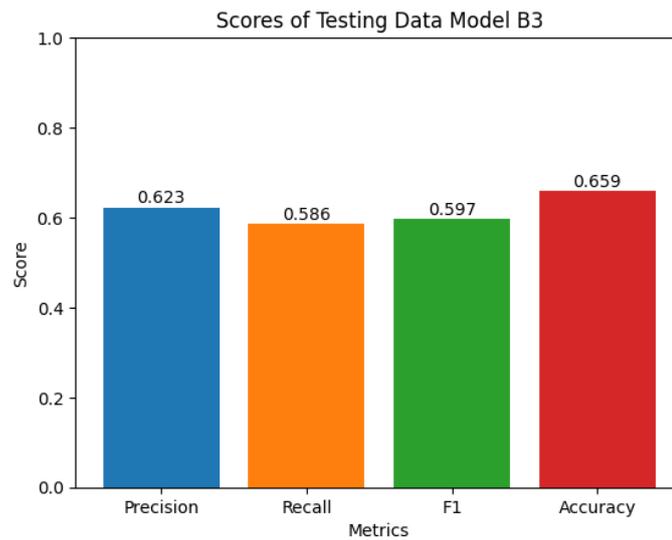
Gambar 4.17 Grafik *cost* proses *training* model B3

Hasil *confusion matrix* pada data *testing* dapat dilihat pada Gambar 4.18. Hasil tersebut menunjukkan model berhasil memprediksi kelas dengan benar sebanyak 135 data dan gagal memprediksi kelas dengan benar sebanyak 70 data.



Gambar 4.18 *Confusion matrix* model B3

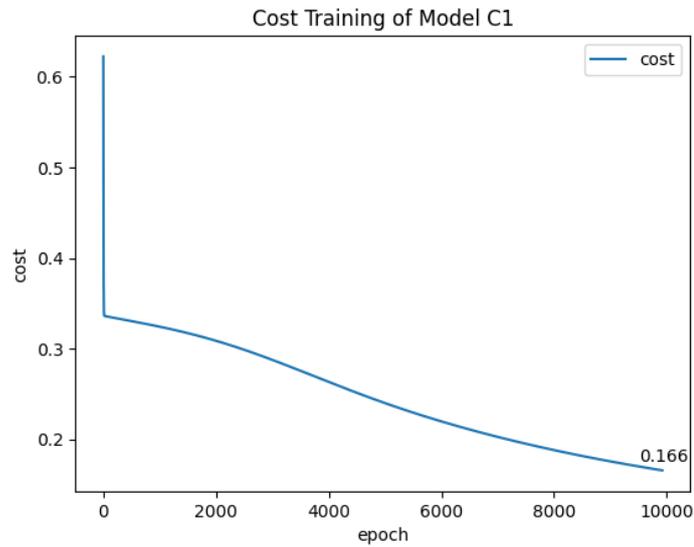
Berdasarkan Persamaan 4.1, 4.2, 4.3, dan 4.4, dapat diperoleh performa model dalam hal akurasi, presisi, *recall*, dan *F1-score*. Nilai-nilai performa dari model B3 ini kemudian ditampilkan secara visual pada Gambar 4.19.



Gambar 4.19 Metrik performa model B3

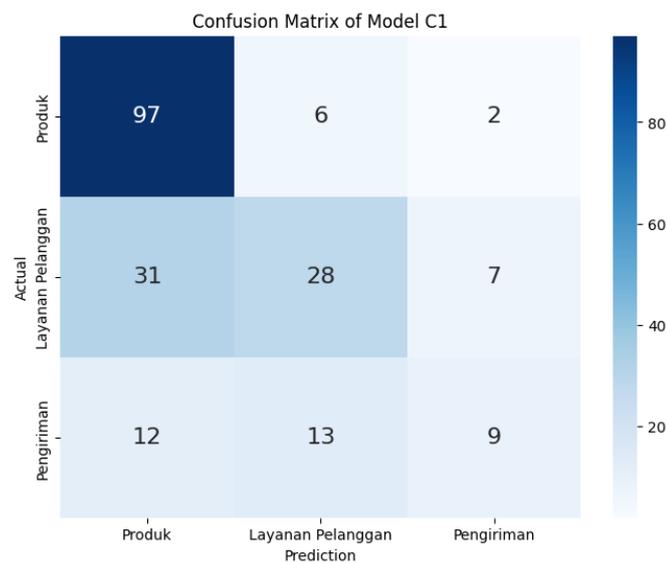
4.3.7 Model C1

Uji coba model C1 menggunakan jumlah *node hidden layer* pertama sebanyak 100 *node* dan *learning rate* sebesar 0.01. Proses *training* model C1 yang menggunakan data *training* ditunjukkan pada Gambar 4.20. Dari proses *training* model C1 memakan waktu sekitar 1575,18 detik atau 26,25 menit dengan jumlah *epoch* sebanyak 9922 *epoch*. Dari proses *training* model C1 juga didapatkan nilai *cost* sebesar 0,165794.



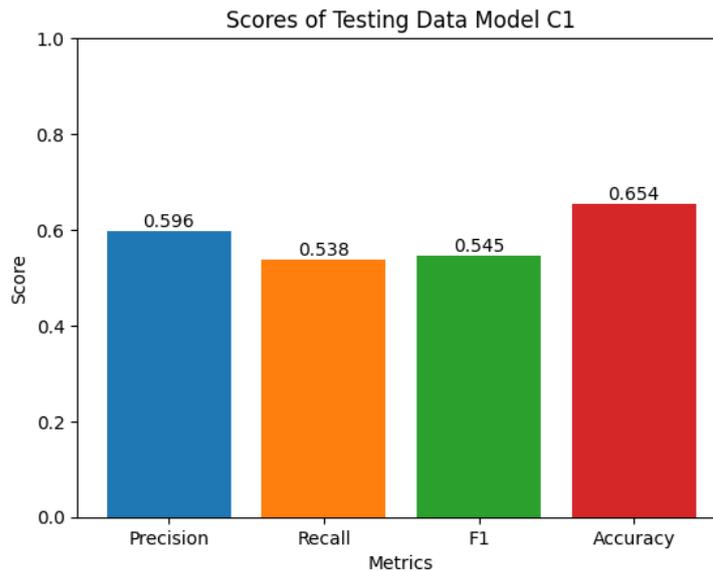
Gambar 4.20 Grafik *cost* proses *training* model C1

Hasil *confusion matrix* pada data *testing* dapat dilihat pada Gambar 4.21. Hasil tersebut menunjukkan model berhasil memprediksi kelas dengan benar sebanyak 134 data dan gagal memprediksi kelas dengan benar sebanyak 71 data.



Gambar 4.21 *Confusion matrix* model C1

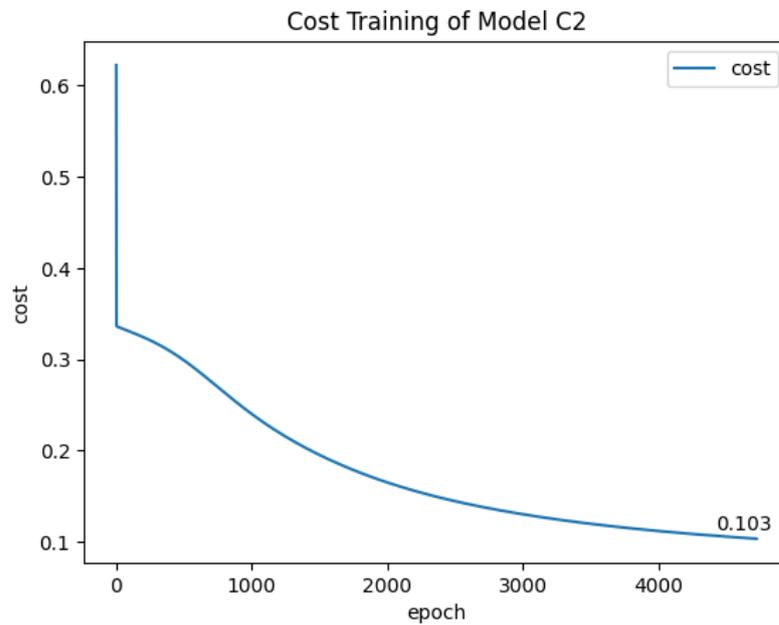
Berdasarkan Persamaan 4.1, 4.2, 4.3, dan 4.4, dapat diperoleh performa model dalam hal akurasi, presisi, *recall*, dan *F1-score*. Nilai-nilai performa dari model C1 ini kemudian ditampilkan secara visual pada Gambar 4.22.



Gambar 4.22 Metrik performa model C1

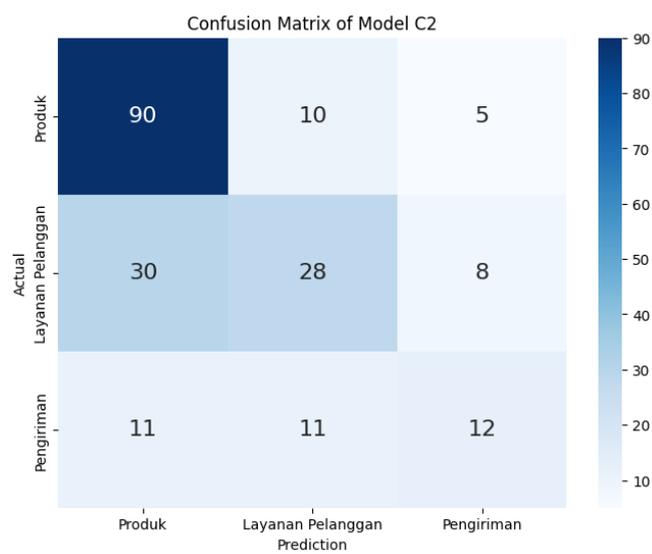
4.3.8 Model C2

Uji coba model C2 menggunakan jumlah *node hidden layer* pertama sebanyak 100 *node* dan *learning rate* sebesar 0.05. Proses *training* model C2 yang menggunakan data *training* ditunjukkan pada Gambar 4.23. Dari proses *training* model C2 memakan waktu sekitar 772,01 detik atau 12,87 menit dengan jumlah *epoch* sebanyak 4723 *epoch*. Dari proses *training* model C2 juga didapatkan nilai *cost* sebesar 0,103425.



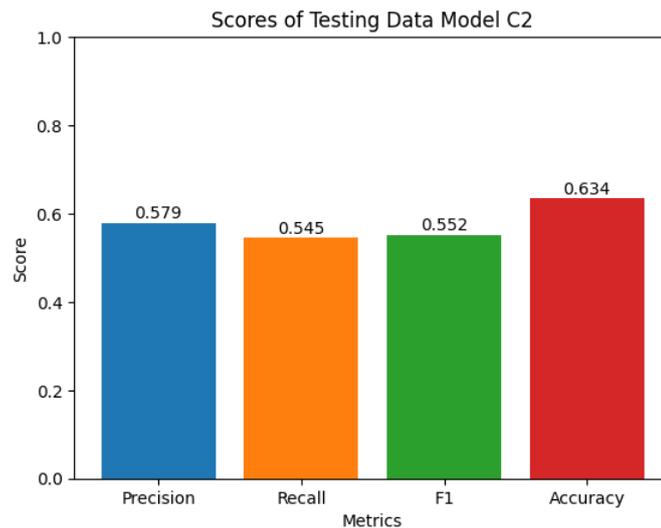
Gambar 4.23 Grafik *cost* dan *accuracy* proses *training* model C2

Hasil *confusion matrix* pada data *testing* dapat dilihat pada Gambar 4.24. Hasil tersebut menunjukkan model berhasil memprediksi kelas dengan benar sebanyak 130 data dan gagal memprediksi kelas dengan benar sebanyak 75 data.



Gambar 4.24 *Confusion matrix* model C2

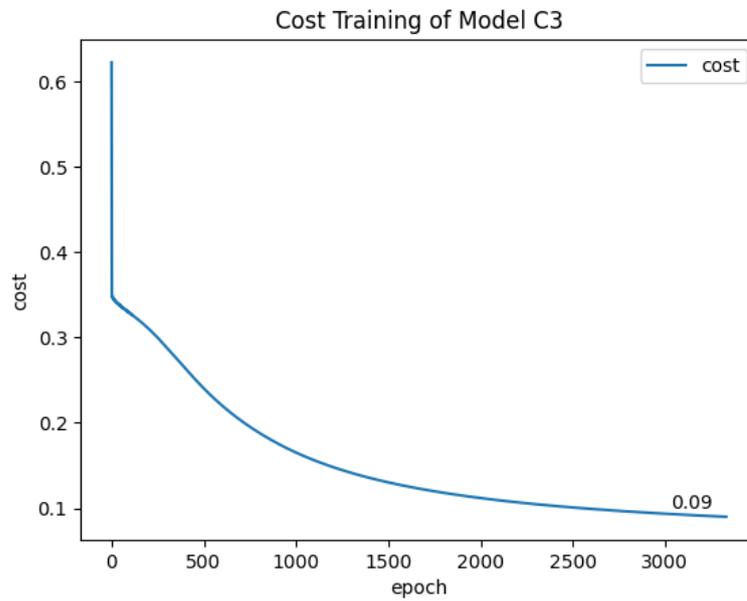
Berdasarkan Persamaan 4.1, 4.2, 4.3, dan 4.4, dapat diperoleh performa model dalam hal akurasi, presisi, *recall*, dan *F1-score*. Nilai-nilai performa dari model C2 ini kemudian ditampilkan secara visual pada Gambar 4.25.



Gambar 4.25 Metrik performa model C2

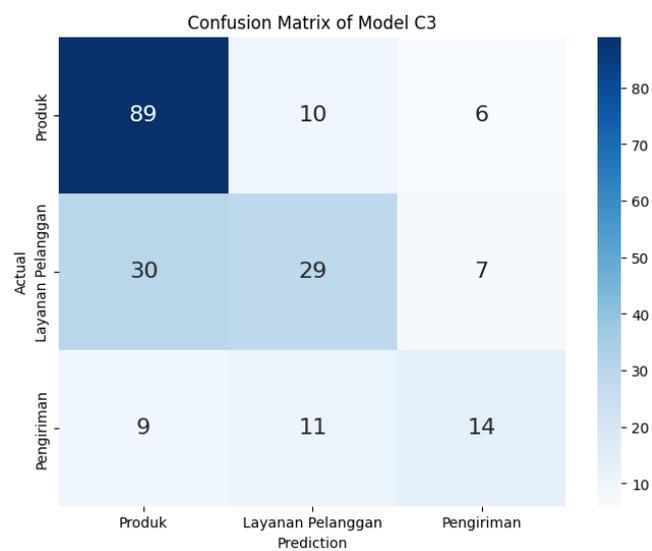
4.3.9 Model C3

Uji coba model C3 menggunakan jumlah *node hidden layer* pertama sebanyak 100 *node* dan *learning rate* sebesar 0.1. Proses *training* model C3 yang menggunakan data *training* ditunjukkan pada Gambar 4.26. Dari proses *training* model C3 memakan waktu sekitar 1167,37 detik atau 19,46 menit dengan jumlah *epoch* sebanyak 3332 *epoch*. Dari proses *training* model C3 juga didapatkan nilai *cost* sebesar 0,090012.



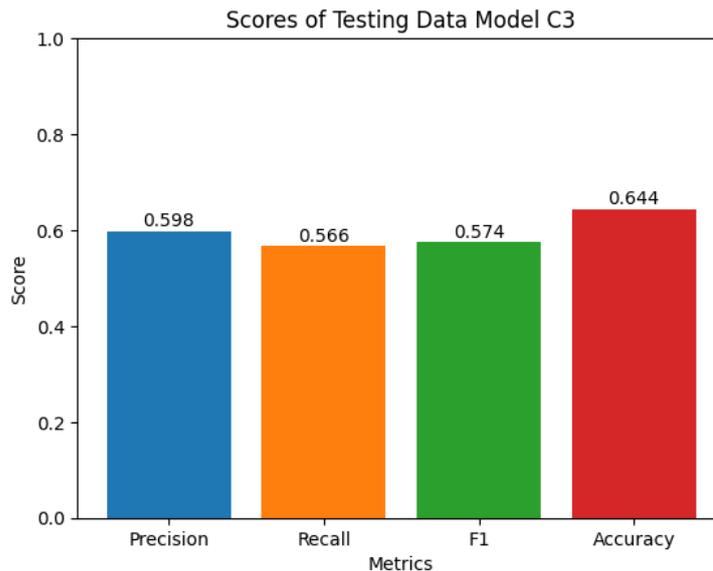
Gambar 4.26 Grafik *cost* dan *accuracy* proses *training* model C3

Hasil *confusion matrix* pada data *testing* dapat dilihat pada Gambar 4.27. Hasil tersebut menunjukkan model berhasil memprediksi kelas dengan benar sebanyak 132 data dan gagal memprediksi kelas dengan benar sebanyak 73 data.



Gambar 4.27 *Confusion matrix* model C3

Berdasarkan Persamaan 4.1, 4.2, 4.3, dan 4.4, dapat diperoleh performa model dalam hal akurasi, presisi, *recall*, dan *F1-score*. Nilai-nilai performa dari model C3 ini kemudian ditampilkan secara visual pada Gambar 4.28.



Gambar 4.28 Metrik performa model C3

4.4 Pembahasan

Berdasarkan hasil skenario pengujian yang telah dilakukan, didapatkan hasil proses *training* masing-masing model. Hasil proses *training* berisi nilai-nilai performa model, diantaranya *cost*, akurasi, *epoch*, dan durasi waktu proses *training*. Nilai *cost* mencerminkan seberapa baik model dapat menyesuaikan diri dengan data pelatihan, Akurasi mencerminkan seberapa baik model dapat memprediksi kelas dengan benar. Informasi jumlah *epoch* dan waktu *training* penting untuk mengevaluasi efisiensi dan efektivitas proses *training* model tersebut.

Dalam penelitian ini menggunakan skenario pemberhentian *training* dengan memantau perubahan *cost*. Proses *training* akan diberhentikan ketika nilai *cost*

tidak mengalami perubahan yang signifikan antara *epoch* sekarang dan *epoch* sebelumnya. Hal ini menandakan bahwa model sudah konvergen jika tidak ada perubahan yang signifikan dalam nilai *cost*. Nilai perubahan yang digunakan dalam penelitian ini yaitu 0,00001, artinya jika *epoch* sekarang memiliki selisih kurang dari 0,00001 dengan *epoch* sebelumnya, maka proses *training* akan dihentikan. Hasil proses *training* dari masing-masing model ditampilkan dalam Tabel 4.11.

Tabel 4.11 Hasil proses *training* skenario pengujian

Model	Training			
	Cost	Epoch	Waktu (detik)	Waktu (menit)
A1	0,166438	9963	388,33	6,47
A2	0,098499	4987	283,68	4,73
A3	0,083149	3618	214,31	3,57
B1	0,165462	10009	817,38	13,62
B2	0,102014	4775	428,86	7,15
B3	0,088147	3379	283,35	4,72
C1	0,165794	9922	1575,18	26,25
C2	0,103425	4723	772,01	12,87
C3	0,090012	3332	1167,37	19,46

Model-model dibedakan berdasarkan jumlah *node* pada *hidden layer* pertama dan besar *learning rate*. Kode huruf pada nama model menunjukkan jumlah *node hidden layer* pertama, A untuk 25, B untuk 50, dan C untuk 100. Sedangkan angka pada nama model menunjukkan besar *learning rate* pada model. Angka 1 untuk 0,01, angka 2 untuk 0,05, dan angka 3 untuk 0,001.

Terlihat perbedaan waktu yang signifikan antara masing-masing model pada Tabel 4.11. Model dengan *learning rate* rendah (model dengan kode angka 1) cenderung lebih lama daripada model dengan *learning rate* tinggi (model dengan

kode angka 3) karena *learning rate* mempengaruhi kecepatan proses *training* pada model.

Perbedaan waktu yang signifikan juga terlihat dari jumlah *node hidden layer* pertama. Model dengan jumlah *node hidden layer* pertama yang lebih tinggi cenderung memerlukan waktu yang lebih lama daripada model dengan jumlah *node hidden layer* pertama yang lebih rendah. Jumlah *node* layer *neural network* menunjukkan kompleksitas dari model tersebut. Dari analisis ini, semakin kompleks model semakin lama waktu yang diperlukan untuk proses *training*.

Namun, kompleksitas model tidak selalu menghasilkan performa yang lebih baik. Model terbaik dalam proses *training* adalah model B3, yang memiliki jumlah *node hidden layer* pertama paling kecil, yaitu 25, dan memiliki *learning rate* yang terbesar sebesar 0,1. Model A3 merupakan model paling efisien dibandingkan dengan model-model lainnya. Model A3 mendapatkan nilai *cost* terkecil, yaitu 0,083149 untuk *cost*. Model A3 juga memerlukan waktu paling singkat dibandingkan dengan model-model lainnya, yaitu hanya sekitar 214,13 detik atau 3,57 menit.

Setelah proses *training*, model-model skenario pengujian dilakukan evaluasi pada data *testing*. Pada tahapan ini, performa model diukur dengan menggunakan data *testing* yang belum pernah dilihat sebelumnya oleh model selama proses *training*. Evaluasi dilakukan untuk memahami seberapa baik model dalam mempelajari pola dari data *training* dan memprediksinya ke data *testing*. Ini memungkinkan evaluasi performa model secara objektif dan memastikan bahwa

model dapat digunakan dengan baik dalam mengklasifikasikan data baru. Hasil evaluasi dari masing-masing model ditampilkan dalam Tabel 4.12.

Tabel 4.12 Hasil evaluasi skenario pengujian

Model	Presisi	Recall	F1-Score	Akurasi
A1	0,56109	0,51508	0,51968	0,62927
A2	0,58353	0,55458	0,56080	0,63902
A3	0,60796	0,57261	0,58148	0,65366
B1	0,55798	0,50340	0,50635	0,62439
B2	0,59819	0,561209	0,57036	0,63902
B3	0,62272	0,58616	0,59719	0,65854
C1	0,59620	0,53758	0,54452	0,65366
C2	0,579486	0,54477	0,55215	0,63415
C3	0,59794	0,56626	0,57432	0,64390

Proses pengujian ini dilakukan pada data *testing*. Terlihat bahwa semua model memiliki hasil masing masing metrik yang tidak beda jauh satu sama lain. Semakin mendekati nilai 1,0 pada metrik presisi, *recall*, *F1-score*, dan akurasi, maka performa semakin bagus. Hasil menunjukkan metrik presisi mencapai nilai pada rentang 0,55 - 0,62. Pada metrik *recall* nilai didapatkan nilai yang cukup rendah pada rentang 0,51 - 0,58. Begitu juga pada metrik *F1-score*, didapatkan nilai pada rentang 0,51 - 0,59. Sedangkan untuk metrik akurasi didapatkan nilai yang sedikit besar daripada metrik lainnya, yaitu nilai pada rentang 0,62 – 0,65.

Tabel 4.13 *Confusion matrix* model B3 kelas produk

Kelas Aktual	Kelas Prediksi	
	Produk (A)	Bukan Produk (B & C)
Produk (A)	89	16
Bukan Produk (B & C)	37	63

Tabel 4.14 *Confusion matrix* model B3 kelas layanan pelanggan

Kelas Aktual	Kelas Prediksi	
	Layanan Pelanggan (B)	Bukan Layanan Pelanggan (A & C)
Layanan Pelanggan (B)	31	35
Bukan Layanan Pelanggan (A & C)	22	117

Tabel 4.15 *Confusion matrix* model B3 kelas pengiriman

Kelas Aktual	Kelas Prediksi	
	Pengiriman C)	Bukan Pengiriman (A & B)
Pengiriman (C)	15	19
Bukan Pengiriman (A & B)	11	160

Dari semua model skenario pengujian, model B3 merupakan model yang terbaik pada proses pengujian. Model B3 memiliki nilai presisi, *recall*, *F1-score*, dan akurasi tertinggi dibanding model lainnya. Nilai presisi pada model B3 adalah 0,62272 atau 62,27%. Nilai ini didapatkan dari *confusion matrix* pada Gambar 4.17 dimana jumlah FP kelas produk sebanyak 37 data dengan 89 data TP kelas produk, FP kelas layanan pelanggan sebanyak 22 data dengan 31 TP kelas layanan pelanggan, dan FP kelas pengiriman sebanyak 11 data dengan 15 data TP kelas pengiriman. Dari hasil tersebut menunjukkan bahwa model B3 masih memprediksi FP yang banyak.

Pada metrik *recall*, model B3 mendapatkan nilai 0,58616 atau 58,61%. Berdasarkan *confusion matrix* pada Gambar 4.17 dimana FN kelas produk sebanyak 16 data dengan 89 data TP kelas produk, FN kelas layanan pelanggan sebanyak 35 data dengan 31 TP kelas layanan pelanggan, dan FN kelas pengiriman sebanyak 19 data dengan 15 data TP kelas pengiriman. Dari hasil

tersebut dapat dilihat bahwa FN kelas layanan produk dan kelas pengiriman memiliki jumlah lebih banyak daripada TP kelas layanan produk dan kelas pengiriman. Ini menunjukkan model cenderung untuk memprediksi benar kelas produk dibanding kelas lainnya.

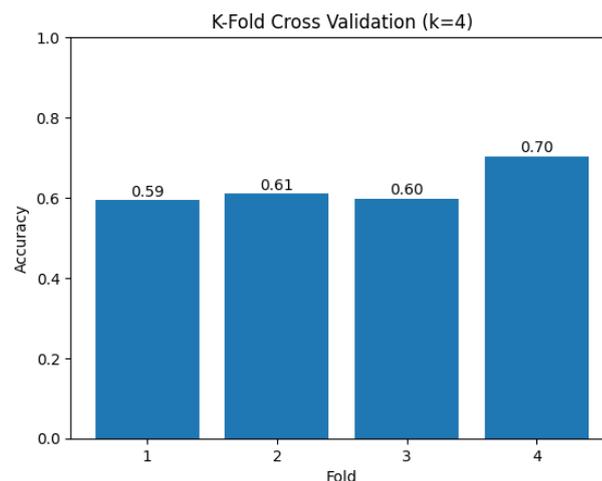
Nilai metrik akurasi model B3 adalah 0,65853 menunjukkan bahwa model B3 mampu memprediksi 65,85% data keseluruhan dengan benar. Nilai ini masih tergolong rendah. Model B3 mampu memprediksi kelas dengan benar pada sebagian data. Rendahnya nilai akurasi menunjukkan bahwa model B3 belum mampu memprediksi data secara optimal. Nilai akurasi yang rendah juga berdampak pada banyak data yang diprediksi salah oleh model, walaupun jumlah secara keseluruhan lebih sedikit daripada data yang diprediksi benar oleh model.

Berdasarkan analisis metrik-metrik diatas, menunjukkan bahwa performa model B3 masih terbilang kurang baik. Nilai yang didapatkan pada metrik presisi, *recall*, *F1-score* dan akurasi merupakan nilai yang rendah. Nilai performa yang rendah tentu tidak praktis dan tidak efisien jika diimplementasikan pada kasus dunia nyata karena masih memiliki probabilitas eror yang sangat besar.

Pada penelitian ini menerapkan *k-fold cross validation* untuk mengevaluasi kinerja model dan menguji performanya pada semua distribusi data. Ini dilakukan untuk melihat apakah model tidak mengalami penurunan performa pada *input* baru dari distribusi data yang sama. Dataset dibagi menjadi *k* bagian atau yang disebut *fold*. Model dilatih dan dievaluasi sebanyak *k* iterasi menggunakan *fold* yang berbeda pada setiap iterasi. Metrik performa dihitung pada setiap iterasi kemudian dihitung rata-ratanya untuk mengstimasi performa general model.

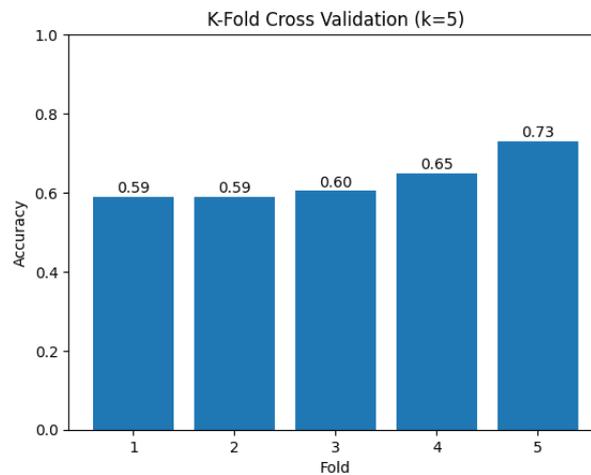
Penelitian ini menggunakan 3 skenario *k-fold cross validation* dengan masing masing nilai *k-fold*, yaitu 4, 5, dan 10. Berdasarkan penelitian yang dilakukan oleh Berrar, (2018) Nilai *k-fold* 10 memberikan keseimbangan yang baik antara bias dan varians, menjadikannya pilihan yang sering digunakan dalam praktik evaluasi model. Nilai $k=10$ menghasilkan kesalahan prediksi yang hamper tidak bias diabndingkan dengan nilai k yang lebih kecil atau lebih besar. Nilai $k=4$ dan $k=5$ juga digunakan untuk melihat bagaimana model berperilaku dengan nilai k yang berbeda. Model yang akan digunakan pada *k-fold cross validation* adalah model B3.

Hasil *k-fold cross validation* model B3 dengan nilai $k=4$ ditunjukkan pada Gambar 4.29 dengan capaian rata-rata akurasi sebesar 0.6259 atau 62,59%. Nilai akurasi terendah didapatkan pada *fold* ke-1 sebesar 0,59 atau 59% dan nilai tertinggi akurasi pada *fold* ke-4 sebesar 0,70 atau 70%. Nilai standar deviasi yang didapatkan adalah 0,0449.



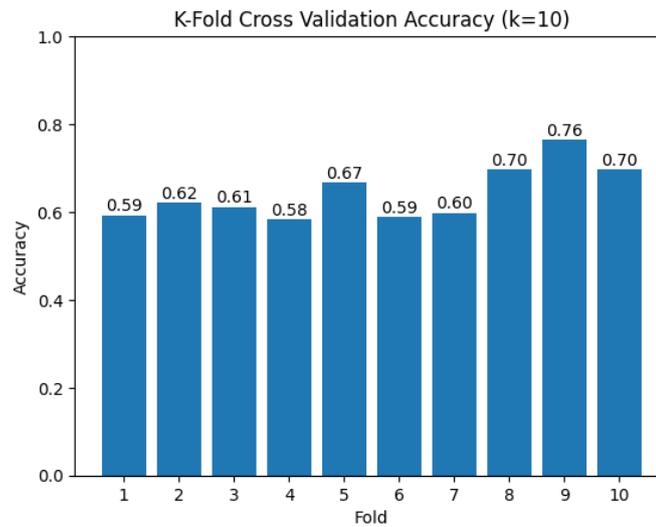
Gambar 4.29 Hasil *k-fold cross validation* model B3 dengan nilai $k=4$

Hasil *k-fold cross validation* model B3 dengan nilai $k=5$ ditunjukkan pada Gambar 4.30 dengan capaian rata-rata akurasi sebesar 0.6329 atau 63,29%. Nilai akurasi terendah didapatkan pada *fold* ke-1 dan ke- dengan nilai serupa, yaitu sebesar 0,59 atau 59% dan nilai tertinggi akurasi pada *fold* ke-5 sebesar 0,73 atau 73%. Nilai standar deviasi yang didapatkan adalah 0,0532.



Gambar 4.30 Hasil *k-fold cross validation* model B3 dengan nilai $k=5$

Hasil *k-fold cross validation* model B3 dengan nilai $k=10$ ditunjukkan pada Gambar 4.31 dengan capaian rata-rata akurasi sebesar 0.6417 atau 64,17%. Nilai akurasi terendah didapatkan pada *fold* ke-4, yaitu sebesar 0,58 atau 58% dan nilai tertinggi akurasi pada *fold* ke-9 sebesar 0,76 atau 76%. Nilai standar deviasi yang didapatkan adalah 0,0580.



Gambar 4.31 Hasil *k-fold cross validation* model B3 dengan nilai $k=10$

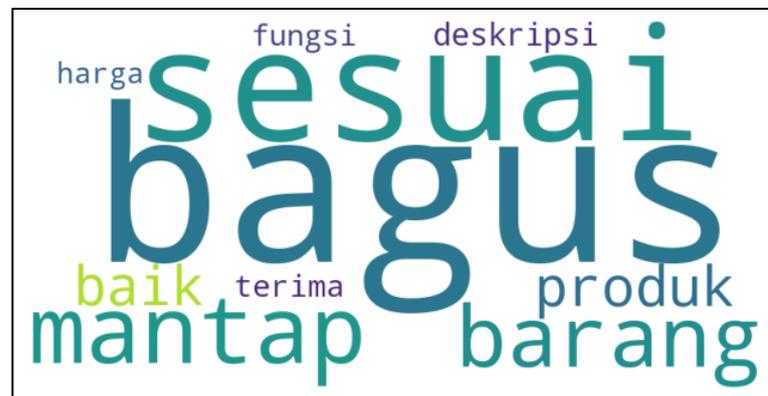
Tabel 4.16 Hasil *k-fold cross validation* model B3

Nilai K	Standar Deviasi	Rata-rata Akurasi
4	0.0449	62,59%.
5	0.0532	63,29%
10	0,0580	64,17%

Proses *preprocessing* pada penelitian ini menggunakan TF-IDF. Nilai dari TF-IDF ini dapat digunakan untuk menganalisis kata. Nilai TF-IDF dapat digunakan untuk mengidentifikasi kata-kata yang paling mendominasi atau mewakili dalam sebuah kelas ulasan. Analisis kata ini diterapkan untuk memahami preferensi dan penilaian pelanggan terhadap kelas produk, layanan pelanggan, dan juga pengiriman. Kata-kata paling mendominasi pada kelas produk ditampilkan pada Tabel 4.17 dan divisualisasikan pada Gambar 4.32.

Tabel 4.17 Skor TF-IDF tertinggi kelas produk

No	Kata	Skor TF-IDF
1	Bagus	0,08822
2	Sesuai	0,08082
3	Mantap	0,05956
4	Barang	0,05645
5	Produk	0,05169
6	Baik	0,05042
7	Deskripsi	0.04703
8	Terima	0.04359
9	Fungsi	0.04112
10	Harga	0.03931



Gambar 4.32 Wordcloud kelas produk

Kata-kata seperti “bagus”, “sesuai”, ”mantap” merupakan 3 kata yang paling mendominasi atau paling relevan dengan kelas produk. Kata-kata tersebut memiliki nilai TF-IDF yang cukup tinggi pada kelas produk. Hal ini menunjukkan bahwa aspek-aspek seperti kualitas barang, kesesuaian deskripsi, dan juga kepuasan pelanggan memiliki pengaruh yang cukup besar dalam penilaian produk oleh pelanggan.

Selanjutnya, kata-kata paling mendominasi pada kelas layanan pelanggan ditampilkan pada Tabel 4.18 dan divisualisasikan pada Gambar 4.33.

Tabel 4.18 Skor TF-IDF tertinggi kelas layanan pelanggan

No	Kata	Skor TF-IDF
1	Respon	0,08489
2	Cepat	0,08005
3	Barang	0,07574
4	Seller	0,07318
5	Fast	0,06868
6	Rapi	0,06190
7	Gan	0,05862
8	Recommended	0,05629
9	Kirim	0,05115
10	Mantap	0,04891



Gambar 4.33 Wordcloud kelas layanan pelanggan

Kata-kata seperti “cepat”, “respon”, ”barang” merupakan 3 kata yang paling mendominasi atau paling relevan dengan kelas layanan pelanggan. Kata-kata tersebut memiliki nilai TF-IDF yang cukup tinggi pada kelas layanan pelanggan. Hal ini menunjukkan bahwa aspek-aspek seperti bagaimana respon penjual, kecepatan penjual dalam melayani pelanggan, dan juga kualitas barang yang diterima pelanggan memiliki pengaruh yang cukup besar dalam penilaian pelayanan pelanggan.

Selanjutnya, kata-kata paling mendominasi pada kelas pengiriman ditampilkan pada Tabel 4.19 dan divisualisasikan pada Gambar 4.34.

Tabel 4.19 Ranking skor TF-IDF kelas pengiriman

No	Kata	Skor TF-IDF
1	Cepat	0,15582
2	Kirim	0,11844
3	Terima	0,08296
4	Selamat	0,07486
5	Barang	0,07431
6	Kasih	0,07010
7	Aman	0,05379
8	Coba	0,05348
9	Sampai	0,04849
10	Fungsi	0,04672



Gambar 4.34 Wordcloud kelas pengiriman

Kata-kata seperti “cepat”, “kirim”, ”terima” merupakan 3 kata yang paling mendominasi atau paling relevan dengan kelas pengiriman. Kata-kata tersebut memiliki nilai TF-IDF yang cukup tinggi pada kelas pengiriman. Hal ini menunjukkan bahwa aspek-aspek seperti kecepatan dan ketepatan pengiriman sangat diperhatikan oleh pelanggan dalam menilai aspek pengiriman. Adanya kata “selamat” dan “aman” juga menunjukkan bahwa keselamatan dan keamanan pengiriman barang menjadi perhatian pelanggan yang tidak kalah penting.

Hasil analisis kata menggunakan nilai TF-IDF memberikan pemahaman yang lebih mendalam terhadap aspek-aspek yang sering diperhatikan oleh

pelanggan. Dengan analisis ini, penjual atau pelaku bisnis *e-commerce* dapat memanfaatkannya untuk mengidentifikasi aspek apa yang perlu ditingkatkan atau dievaluasi dalam toko online mereka sesuai dengan *feedback* pelanggan melalui ulasan yang diberikan.

Dalam menentukan aspek ulasan produk dari pengguna, tentunya terdapat indikator-indikator yang menandakan atau menjadi karakteristik dari aspek-aspek tersebut. Aspek produk memiliki karakteristik ulasan yang menggambarkan kepuasan pelanggan terhadap kualitas, performa, dan kesesuaian produk dengan deskripsi yang diberikan. Aspek layanan pelanggan memiliki karakteristik ulasan yang mencakup interaksi antara pelayanan dengan pelayanan yang diberikan, seperti keramahan, responsivitas, dan hal lainnya yang terkait dengan pelayanan pelanggan. Sementara itu, Aspek pengiriman memiliki karakteristik ulasan yang melibatkan kecepatan pengiriman, kondisi barang saat diterima, dan ketepatan waktu pengiriman. Sejalan dengan hal ini, Al-Qur'an menerangkan tentang karakteristik-karakteristik yang mendefinisikan sesuatu, seperti bagaimana karakteristik dari orang-orang mukmin. Seperti yang tertuang dalam QS. Al-Anfal ayat 2:

إِنَّمَا الْمُؤْمِنُونَ الَّذِينَ إِذَا ذُكِرَ اللَّهُ وَجِلَّتْ قُلُوبُهُمْ وَإِذَا تُلِيَتْ عَلَيْهِمْ آيَاتُهُ زَادَتْهُمْ إِيمَانًا وَعَلَىٰ رَبِّهِمْ يَتَوَكَّلُونَ ﴿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'an, dan (3) hanya bertaqwa kepada Allah Swt dan tidak berserah diri kepada yang lain.

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:

ذٰلِكَ الْكِتٰبُ لَا رَيْبَ فِيْهِ هُدًى لِّلْمُتَّقِيْنَ ﴿٢﴾ الَّذِيْنَ يُؤْمِنُوْنَ بِالْغَيْبِ وَيُقِيْمُوْنَ الصَّلٰوةَ وَمِمَّا رَزَقْنٰهُمْ يُنْفِقُوْنَ ﴿٣﴾ وَالَّذِيْنَ يُؤْمِنُوْنَ بِمَا اُنزِلَ اِلَيْكَ وَمَا اُنزِلَ مِنْ قَبْلِكَ وَبِالْآخِرَةِ هُمْ يُوقِنُوْنَ ﴿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 orang-orang yang beriman kepada perkara gaib, yaitu hal-hal yang tidak bisa ditangkap oleh panca indera dan tersembunyi, seperti hari akhir. Orang-orang yang mendirikan sholat, yaitu menunaikannya sesuai syariat, meliputi syarat, rukun, wajib, dan sunnahnya. Dan orang-orang yang gemar menginfakkan sebagian rezeki yang diterima oleh Allah Swt. Kemudian, orang-orang yang beriman kepada wahyu yang Allah turunkan, serta mereka yang beriman kepada adanya akhirat.

اِنَّا كُلَّ شَيْءٍ خَلَقْنٰهُ بِقَدَرٍ ﴿٤٩﴾

“Sesungguhnya Kami menciptakan segala sesuatu sesuai dengan ukuran.” (QS. Al-Qamar 54:49)

Menurut Tafsir Ibnu Katsir, Allah Swt. telah menentukan kadar atau ukuran dari masing-masing makhluk-Nya dan memberi petunjuk kepada semua makhluk-Nya. Hal ini mencakup makhluk dan seluruh alam semesta. Sejalan dengan ini, masing-masing kelas yang diobservasi pada penelitian ini juga memiliki sesuatu yang menjadi tolar ukur, seperti karakteristik atau ciri khas yang mendefinisikan kelas tersebut.

Sejalan dengan konsep pada QS. Al-Anfal ayat 2 dan QS. Al-Baqarah ayat 2-4 ini, orang-orang yang beriman dan orang-orang yang bertaqwa memiliki ciri-cirinya masing-masing. Begitu juga dengan aspek 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 V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Klasifikasi aspek ulasan produk pada *e-commerce* menggunakan *neural network* melakukan 9 skenario pengujian. Variabel bebas yang digunakan adalah parameter jumlah *node hidden layer* pertama dan *learning rate* pada model *neural network*. Model dievaluasi menggunakan *confusion matrix* untuk memberikan gambaran performa model dalam mengklasifikasikan data, kemudian model dengan performa terbaik divalidasi menggunakan *k-fold cross validation* untuk memvalidasi bagaimana berperilaku terhadap data baru.

Model B3 dengan jumlah *node hidden layer* pertama sebanyak 50 *node* dan *learning rate* sebesar 0.1, memberikan performa terbaik dibandingkan dengan model lainnya. Model B3 mencapai presisi 62,27%, *recall* 58,61%, *F1-score* 59,71% dan akurasi 65,85%. Validasi model B3 menggunakan *k-fold cross validation* mendapatkan hasil rata-rata akurasi yang tidak jauh berbeda. Hasil terbaik didapatkan pada $k=10$ dengan nilai rata-rata akurasi 64,17%.

Berdasarkan analisis kata menggunakan TF-IDF, didapatkan bahwa pada kelas produk, tiga kata yang dominan, yaitu “bagus”, “sesuai”, dan “mantap”. Pada kelas layanan pelanggan, tiga kata yang dominan, yaitu “respon”, “cepat”, dan “barang”. Dan yang terakhir pada kelas pengiriman, tiga kata yang dominan, yaitu “cepat”, “kirim”, “terima”. Kata-kata tersebut yang sering digunakan oleh pelanggan dalam memberikan ulasan terhadap penilaian produk,

layanan pelanggan, dan pengiriman toko online pada *e-commerce*.

Berdasarkan Hasil *confusion matrix* dan *k-fold cross validation* pada model B3 menunjukkan bahwa model B3 masih belum bisa mengklasifikasikan data ulasan produk *e-commerce* dengan sempurna. Nilai akurasi ini masih belum mencapai hasil yang diharapkan. Hal ini terjadi dikarenakan oleh beberapa faktor, diantaranya adalah dataset yang digunakan kurang banyak dan tidak seimbang untuk tidak kelasnya, arsitektur yang diajukan pada penelitian ini kurang mampu menangkap pola yang kompleks dari data, dan penggunaan *preprocessing* TF-IDF yang kurang efektif untuk kasus klasifikasi aspek ulasan produk pada *e-commerce* ini.

5.2 Saran

Penelitian klasifikasi aspek ulasan produk pada *e-commerce* menggunakan *neural network* masih memiliki potensi untuk peningkatan lebih lanjut, Terdapat beberapa saran yang dapat dipertimbangkan untuk diterapkan pada penelitian yang akan mendatang:

1. Melakukan pengumpulan data yang lebih banyak dan menggunakan dataset yang seimbang untuk setiap kelasnya. Data yang kurang banyak dan tidak seimbang akan membuat model cenderung hanya mempelajari pola pada kelas yang dominan, hal ini akan membuat model terjadi bias yang besar dan berdampak performa yang buruk pada model. Dengan seimbang nya jumlah data yang disetiap kelas, tentu akan mengurangi bias pada model.
2. Melakukan eksperimen arsitektur *neural network* yang lebih kompleks, contohnya dengan menambah *hidden layer* dan *node* pada *neural network*.

Hal ini dilakukan dengan harapan model dapat mempelajari pola kompleks dari data.

3. Melakukan metode *preprocessing* lain yang lebih bagus untuk kasus klasifikasi teks selain TF-IDF.

DAFTAR PUSTAKA

- Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Umar, A. M., Linus, O. U., Arshad, H., Kazaura, 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>
- Agarap, A. F. (2018). Statistical Analysis on E-Commerce Reviews, with Sentiment Classification using Bidirectional Recurrent Neural Network (RNN). *ArXiv E-Prints*. <https://doi.org/10.48550/arXiv.1805.03687>
- Ali, K., Shaikh, Z. A., Khan, A. A., & Laghari, A. A. (2022). Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer. *Neuroscience Informatics*, 2(4), 100034. <https://doi.org/10.1016/j.neuri.2021.100034>
- 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>
- Berrar, D. (2018). Cross-validation. In *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* (Vols. 1–3, pp. 542–545). Elsevier. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>
- Bujang, S. D. A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H., & Ghani, N. A. M. (2021). Multiclass Prediction Model for Student Grade Prediction Using Machine Learning. *IEEE Access*, 9, 95608–95621. <https://doi.org/10.1109/ACCESS.2021.3093563>
- Dalaorao, G. A., Sison, A. M., Aguinaldo, E., & Medina, R. P. (2019). Integrating Collocation as TF-IDF Enhancement to Improve Classification Accuracy. *IEEE*.
- 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)*.
- 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>

- 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
- Mardianti, S., Zidny Naf, M., & Hidayatulloh, I. (2018). Ekstraksi TF-IDF N-Gram dari Komentar Pelanggan Produk Smartphone pada Website E-Commerce. *Seminar Nasional Teknologi Informasi Dan Multimedia*.
- Mijwil, M. M., Alsaadi, A., Mijwel, M. M., Esen, A., & Shamil, A. (2019). *Overview of Neural Networks*. <https://www.researchgate.net/publication/332655457>
- Prechelt, L. (1998). *Early Stopping - But When?* (pp. 55–69). https://doi.org/10.1007/3-540-49430-8_3
- 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>
- 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>
- 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
- Suhendra, C. D., & Saputra, A. C. (2020). Penentuan Parameter Learning Rate Selama Pembelajaran Jaringan Syaraf Tiruan Backpropagation Menggunakan Algoritma Genetika. *Jurnal Teknologi Informasi*, 14(2). <https://doi.org/https://doi.org/10.47111/JTI>
- 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>
- Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications*, 36(3 PART 2), 6527–6535. <https://doi.org/10.1016/j.eswa.2008.07.035>

Yotov, K., Hadzhikolev, E., Hadzhikoleva, S., & Cheresarov, S. (2023). Finding the Optimal Topology of an Approximating Neural Network. *Mathematics*, *11*(1). <https://doi.org/10.3390/math11010217>