

**KLASIFIKASI KOMENTAR MENGANDUNG *BODY SHAMING*
PADA MEDIA SOSIAL DENGAN MENGGUNAKAN
ALGORITMA *CONVOLUTIONAL
NEURAL NETWORK***

SKRIPSI

MONICA JULIANA EIRENE TAMPUBOLON

201402081



**PROGRAM STUDI S1 TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2025**

**KLASIFIKASI KOMENTAR MENGANDUNG *BODY SHAMING*
PADA MEDIA SOSIAL DENGAN MENGGUNAKAN
ALGORITMA *CONVOLUTIONAL
NEURAL NETWORK***

SKRIPSI

Diajukan untuk melengkapi tugas dan memenuhi syarat meperoleh ijazah
Sarjana Teknologi Informasi

**MONICA JULIANA EIRENE TAMPUBOLON
201402081**



**PROGRAM STUDI S1 TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2025**

PERSETUJUAN

Judul : Klasifikasi Komentar Mengandung Body Shaming pada
 Media Sosial dengan Menggunakan Algoritma
 Convolutional Neural Network

Kategori : Skripsi

Nama : Monica Juliana Eirene Tampubolon

Nomor Induk Mahasiswa : 201402081

Program Studi : Sarjana (S1) Teknologi Informasi

Fakultas : Ilmu Komputer Dan Teknologi Informasi Universitas
 Sumatera Utara

Medan, 10 Januari 2025

Komisi Pembimbing :

Pembimbing 2



Annisa Fadhillah Pulungan S.Kom., M.Kom.,
NIP. 199308092020012001

Pembimbing 1



Sarah Purnamawati, S.T., M.Sc.,
NIP. 198302262010122003

Diketahui/disetujui oleh

Program Studi S1 Teknologi Informasi

Ketua,



Dedy Arisandi S.T., M.Kom.

NIP. 197908312009121002

PERNYATAAN

KLASIFIKASI KOMENTAR MENGANDUNG *BODY SHAMING* PADA
MEDIA SOSIAL DENGAN MENGGUNAKAN ALGORITMA
CONVOLUTIONAL NEURAL NETWORK

SKRIPSI

Saya mengakui bahwa skripsi ini adalah hasil karya saya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing telah disebutkan sumbernya.

Medan, 10 Januari 2025

Monica Juliana Eirene Tampubolon

201402081

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa untuk segala rahmat, karunia, dan berkat-Nya yang telah menuntun penulis dalam merampungkan tugas akhir berupa skripsi yang berjudul “Klasifikasi Komentar Mengandung *Body Shaming* Pada Media Sosial Dengan Menggunakan Algoritma *Convolutional Neural Network*” sebagai salah satu syarat dalam meraih sebuah gelar Sarjana Komputer pada program studi Teknologi Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara.

Selain itu, penulis juga ingin menyampaikan rasa terima kasih sebesar-besarnya kepada berbagai pihak yang terlibat yang telah memberikan dukungan dan bimbingan kepada penulis selama menghadapi kendala dalam proses penulisan skripsi. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya dengan ketulusan hati kepada:

1. Keluarga penulis Papa, Mami, Dior, Lexy, Nico, Oscanic dan seluruh keluarga besar yang telah memberikan dukungan, doa, dan semangat tiada henti kepada penulis hingga akhirnya penulis dapat merampungkan skripsi ini.
2. Ibu Sarah Purnamawati, S.T., M.Sc., selaku Dosen Pembimbing I yang telah meluangkan banyak waktu dan pemikiran untuk memberikan saran dan kritik kepada penulis dalam proses menyelesaikan skripsi ini.
3. Ibu Annisa Fadhillah Pulungan S.Kom., M.Kom., selaku Dosen Pembimbing II yang telah bersedia memberikan waktu dan pemikiran berupa kritik dan saran kepada penulis untuk dapat menyelesaikan skripsi ini.
4. Bapak Dedy Arisandi ST., M.Kom., selaku Ketua Program Studi S1 Teknologi Informasi Universitas Sumatera Utara dan Bapak Ivan Jaya S.Si., M.Kom., selaku Sekretaris Program Studi S1 Teknologi Informasi Universitas Sumatera Utara.
5. Bapak dan Ibu dosen pengajar di Program Studi Teknologi Informasi, Universitas Sumatera Utara, yang telah memberikan pengetahuan yang luas selama kelas perkuliahan dan kegiatan akademik lainnya.

6. Bapak dan Ibu *staff* pegawai Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara, yang telah memberikan bantuan dan waktunya dalam keperluan administrasi selama masa perkuliahan.
7. Teman-teman seperjuangan penulis Fildzah Zata, Retno Wulan Sari, Syavira Nur Khairani, Pretty Ohara, Jane Anabel, Yeftha El Imani, Vicky Natanael, dan Arief Fadhlal yang telah menemani penulis mulai dari awal perkuliahan sebagai mahasiswa baru hingga masa penyusunan tugas akhir.
8. Teman-teman magang penulis, Christine Amanda, Stephani Uli, Bobby Adithya, dan Deby Enina yang telah berbagi keceriaan dengan penulis selama masa magang di PT Telkom Indonesia dan memberikan dukungan serta masukan kepada penulis selama penggeraan skripsi.
9. Teman-teman seperjuangan Teknologi Informasi Angkatan 2020 yang telah menemani dan membantu penulis selama perkuliahan yang tidak dapat disebutkan satu persatu.
10. Sahabat masa SMA penulis yang telah menemani, memberikan masukan, dan berbagi kebahagiaan kepada penulis dari waktu yang lama hingga proses menyelesaikan skripsi ini terwujud.
11. Dan terakhir kepada diri penulis, Monica Juliana Eirene Tampubolon. Terima kasih telah selalu berusaha yang terbaik, bertahan dalam segala kesulitan sejauh ini, dan tidak menyerah meski dihadapkan dengan kesulitan dalam masa penyusunan skripsi ini. Terima kasih telah bekerja keras dan menjaga semangat dengan baik.

Semoga Tuhan Yang Maha Esa memberkati kepada seluruh pihak terkait yang telah membantu penulis dalam penyelesaian tugas akhir ini. Serta, penulis berharap tugas akhir dapat bermanfaat terhadap semua orang yang membacanya.

Medan, 10 Januari 2025

Penulis,

Monica Juliana Eirene Tampubolon

ABSTRAK

Penggunaan media sosial yang dapat dengan mudah diakses tanpa batas, dapat berkontribusi menjadi sarana perundungan secara online yaitu *body shaming*. Dalam media sosial, tindakan ini dapat dijumpai pada kolom komentar, cuitan, ataupun dapat berupa konten. *Body shaming* pada media sosial dapat berupa kritikan negatif terhadap bentuk tubuh secara keseluruhan kepada orang lain. Hal ini memberikan kekhawatiran dan perasaan terganggu kepada para pengguna media sosial yang menjadi korban. Banyaknya jumlah komentar *body shaming* yang dapat dijumpai , menyebabkan kesulitan untuk melakukan klasifikasi komentar *body shaming* secara manual yang dilakukan oleh tenaga ahli. Oleh karena itu, dengan berkembangnya kemajuan teknologi informasi sekarang ini dapat dirancang sebuah sistem yang dapat mengklasifikasikan komentar yang mengandung unsur *body shaming* secara lebih efektif. Penelitian ini bertujuan untuk mengklasifikasikan komentar mengandung *body shaming* pada media sosial dengan menerapkan algoritma *Convolutional Neural Network* dan *IndoBERT* sebagai *word embedding*. Data yang digunakan merupakan 3500 komentar berbahasa Indonesia hasil *crawled* pada aplikasi media sosial *Instagram* dan *TikTok* yang mencakup komentar *body shaming* dan komentar yang tidak mencakup *body shaming*. Pada hasil evaluasi kinerja model melalui *confusion matrix* diperoleh akurasi sebesar 89%. Berdasarkan hasil evaluasi tersebut menunjukkan bahwa sistem yang dikembangkan mampu melakukan klasifikasi komentar *body shaming* dengan tingkat keakuratan yang memadai.

Kata Kunci : Klasifikasi *body shaming*, *IndoBERT Embedding*, *Convolutional Neural Network*

***CLASSIFICATION OF COMMENTS CONTAINING BODY SHAMING
ON SOCIAL MEDIA BY USING THE CONVOLUTIONAL
NEURAL NETWORK ALGORITHM***

ABSTRACT

The use of social media, which can be easily accessed without limits, can contribute to online bullying, namely body shaming. In social media, this action can be found in comments, tweets, or contents. Body shaming on social media can be in the form of negative criticism of the overall body shape to others. This causes concern and annoyance to the victimized social media users. The large number of body shaming comments that can be found makes it difficult to classify body shaming comments manually by professional. Therefore, with the development of information technology today, a system can be designed that can classify comments containing body shaming elements more effectively. This research aims to classify comments containing body shaming on social media by applying the Convolutional Neural Network algorithm and IndoBERT as word embedding. The data used are 3500 Indonesian comments crawled on Instagram and TikTok social media applications which include body shaming comments and comments that do not include body shaming. The results of the model performance evaluation through confusion matrix obtained an accuracy of 89%. Based on the evaluation results, it shows that the developed system is able to classify body shaming comments with sufficient accuracy.

Keywords: Body shaming classification, IndoBERT Embedding, Convolutional Neural Network

DAFTAR ISI

| | |
|---|------|
| PERSETUJUAN | iii |
| PERNYATAAN | iv |
| UCAPAN TERIMA KASIH | v |
| ABSTRAK | vii |
| <i>ABSTRACT</i> | viii |
| DAFTAR ISI | ix |
| DAFTAR TABEL | xi |
| DAFTAR GAMBAR | xii |
| DAFTAR PSEUDOCODE | xiii |
| BAB 1 PENDAHULUAN | 2 |
| 1.1 Latar Belakang | 2 |
| 1.2 Rumusan Masalah | 4 |
| 1.3 Tujuan Penelitian | 4 |
| 1.4 Batasan Penelitian | 4 |
| 1.5 Manfaat Penelitian | 5 |
| 1.6 Metodologi Penelitian | 5 |
| 1.7 Sistematika Penulisan | 6 |
| BAB 2 LANDASAN TEORI | 8 |
| 2.1 <i>Body Shaming</i> | 8 |
| 2.2 <i>Text Preprocessing</i> | 10 |
| 2.3 <i>Word Embedding</i> | 10 |
| 2.4 <i>Convolutional Neural Network</i> | 11 |
| 2.5 <i>Confusion Matrix</i> | 14 |
| 2.5.1 <i>Accuracy</i> | 15 |
| 2.5.2 <i>Precision</i> | 15 |
| 2.5.3 <i>Recall</i> | 15 |
| 2.5.4 <i>F-1 Score</i> | 16 |
| 2.6 Penelitian Terdahulu | 16 |
| 2.7 Perbedaan penelitian | 21 |

| | |
|---|----|
| BAB 3 ANALISIS DAN PERANCANGAN SISTEM | 22 |
| 3.1 Data yang Digunakan | 22 |
| 3.2 Analisis Sistem | 26 |
| 3.2.1 <i>Preprocessing</i> | 27 |
| 3.2.2 <i>Word Embedding</i> | 34 |
| 3.2.3 <i>Convolutional Neural Network</i> | 35 |
| 3.2.4 Output | 40 |
| 3.3 Perancangan Sistem | 40 |
| 3.3.1 Rancangan Tampilan Halaman Beranda | 40 |
| 3.3.2 Rancangan Tampilan Halaman <i>Training</i> dan <i>Testing</i> | 42 |
| 3.3.3 Rancangan Tampilan Halaman <i>Input User</i> | 43 |
| 3.4 Metode Evaluasi | 44 |
| BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM | 46 |
| 4.1 Implementasi Sistem | 46 |
| 4.1.1 Spesifikasi Perangkat Keras dan Perangkat Lunak | 46 |
| 4.1.2 Penerapan Rancangan Tampilan Antarmuka Pengguna | 47 |
| 4.2 Pelatihan Model | 52 |
| 4.3 Hasil Pengujian Sistem | 55 |
| 4.4 Pengukuran Kinerja Model | 58 |
| BAB 5 KESIMPULAN DAN SARAN | 63 |
| 5.1 Kesimpulan | 63 |
| 5.2 Saran | 64 |
| DAFTAR PUSTAKA | 65 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 2. 1 Penerapan confusion matrix | 15 |
| Tabel 2. 2 Penelitian Terdahulu | 18 |
| Tabel 3. 1 Dataset | 23 |
| Tabel 3. 2 Komentar Body Shaming Bentuk Badan | 24 |
| Tabel 3. 3 Komentar Body Shaming Bentuk Wajah | 25 |
| Tabel 3. 4 Komentar Body Shaming Warna Kulit | 25 |
| Tabel 3. 5 Pembagian Dataset | 26 |
| Tabel 3. 6 Implementasi Cleaning | 29 |
| Tabel 3. 7 Implementasi Case Folding | 30 |
| Tabel 3. 8 Implementasi Punctuation Removal | 30 |
| Tabel 3. 9 Implementasi Normalization | 31 |
| Tabel 3. 10 Implementasi Stopword Removal | 32 |
| Tabel 3. 11 Implementasi Tokenization | 34 |
| Tabel 3. 12 Pengubahan Token Menjadi Token id | 34 |
| Tabel 3. 13 Perhitungan Confusion Matrix | 45 |
| Tabel 4. 1 Hasil Akurasi Pengujian Parameter | 52 |
| Tabel 4. 2 Kombinasi Parameter Terbaik | 54 |
| Tabel 4. 3 Hasil Pengujian Sistem Pada Data Testing | 55 |
| Tabel 4. 4 Keterangan Confusion Matrix Label Body Shaming Bentuk Badan | 59 |
| Tabel 4. 5 Keterangan Confusion Matrix Label Body Shaming Bentuk Wajah | 59 |
| Tabel 4. 6 Keterangan Confusion Matrix Label Body Shaming Warna Kulit | 59 |
| Tabel 4. 7 Keterangan Confusion Matrix Label Non-Body Shaming | 60 |
| Tabel 4. 8 Hasil Perhitungan Kinerja Model | 61 |

DAFTAR GAMBAR

| | |
|---|----|
| Gambar 2. 1 Contoh Komentar Body Shaming Bentuk Badan | 9 |
| Gambar 2. 2 Contoh Komentar Body Shaming Bentuk Wajah | 9 |
| Gambar 2. 3 Contoh Komentar Body Shaming Warna Kulit | 10 |
| Gambar 2. 4 Arsitektur CNN (Kim, 2014) | 12 |
| Gambar 3. 1 Arsitektur Umum | 28 |
| Gambar 3. 2 Diagram Alur Pelatihan Model | 37 |
| Gambar 3. 3 Diagram Alur Aplikasi pada Menu Training dan Testing | 38 |
| Gambar 3. 4 Diagram Alur Aplikasi pada Menu Input User | 38 |
| Gambar 3. 5 Halaman Beranda | 41 |
| Gambar 3. 6 Halaman Training dan Testing | 43 |
| Gambar 3. 7 Halaman Input User | 44 |
| Gambar 4. 1 Tampilan Halaman Beranda | 47 |
| Gambar 4. 2 Tampilan Halaman Training dan Testing | 48 |
| Gambar 4. 3 Tampilan Halaman Hasil Training | 49 |
| Gambar 4. 4 Tampilan Halaman Hasil Testing | 49 |
| Gambar 4. 5 Tampilan Halaman Model Evaluation | 50 |
| Gambar 4. 6 Tampilan Halaman Confusion Matrix | 51 |
| Gambar 4. 7 Tampilan Halaman Input User | 52 |
| Gambar 4. 8 Confusion Matrix Klasifikasi Body Shaming | 58 |

DAFTAR PSEUDOCODE

| | |
|---|----|
| Pseudocode 3. 1 Cleaning | 29 |
| Pseudocode 3. 2 Stopword Removal | 32 |
| Pseudocode 3. 3 Tokenization | 33 |
| Pseudocode 3. 4 Model CNN | 39 |

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pada zaman *digital* saat ini, penggunaan media sosial di masyarakat merupakan sebuah kebutuhan yang menjadi bagian dari kehidupan sehari-hari. Akses media sosial yang mudah, dapat di lakukan dimana saja, dan kapan saja membuat masyarakat menjadikan media sosial sebagai media online untuk berkomunikasi, berbagi pendapat, dan berbagi informasi. Namun, seiring dengan adanya dampak positif penggunaan media sosial ada juga dampak negatif yang didapat. Salah satunya, pengguna media sosial dapat menjadikan media sosial sebagai sarana perundungan berbasis online. Dalam hal ini, tindakan *body shaming* merupakan salah satu bentuk perundungan yang kerap terjadi di masyarakat. *Body shaming* adalah istilah yang mengacu pada tindakan mengkritik atau mengomentari secara negatif terhadap fisik atau tubuh orang lain (Erdianti et al., 2022). Objek *body shaming* dapat terdiri dari bentuk tubuh dan berat badan seseorang, wajah atau rupa, dan terhadap warna kulit (Pratiwi, 2021; Fauzia & Rahmiaji, 2019). Tindakan *body shaming* sering sekali masih diremehkan dan diwajarkan di masyarakat. Padahal hal itu dapat mempengaruhi rasa percaya diri korban yang dapat memungkinkan mempengaruhi tingkat rasa rendah diri dan memicu masalah gangguan psikologis serta tekanan mental pada korban(Melizza et al., 2023).

Sebuah survei yang diselenggarakan oleh Yahoo! Health pada tahun 2016 menunjukkan bahwa wanita lebih sering mendapatkan perilaku *body shaming* daripada pria. Terdapat 64% remaja laki-laki dan 94% remaja putri pernah mengalami tindakan *body shaming*, menurut survei yang dilakukan terhadap 2000 orang berusia 13 hingga 64 tahun. Kemudian, Markas Besar Polisi Republik Indonesia (Mabespolri) mengungkapkan bahwa terdapat 966 kasus *body shaming* di tahun 2018, dengan 374

kasus terselesaikan yang mana diantaranya tindakan tersebut terjadi di media sosial. Jumlah ini meningkat dari 206 kasus yang terjadi pada tahun 2015. Berdasarkan jumlah kasus ini *body shaming* menjadi salah satu isu sosial yang meresahkan di masyarakat yang tidak hanya dapat terjadi di dunia nyata, tindakan *body shaming* dapat juga terjadi melalui media sosial. Penggunaan media sosial yang dapat dengan mudah dijangkau oleh khalayak ramai dapat menjadi sarana yang umum digunakan untuk melakukan tindakan *body shaming*. Banyak dari pengguna media sosial tersebut tidak memahami bahwa tulisan komentar yang dibuat dapat merupakan bentuk tindakan *body shaming*. Karena itu, diperlukan metode untuk memudahkan pengklasifikasian komentar-komentar tersebut. Melakukan proses klasifikasi terhadap komentar *body shaming* penting dilakukan karena bertujuan membantu dalam melindungi para pengguna media sosial dari dampak buruk serangan komentar *body shaming* dan membantu para pengguna media sosial mengetahui informasi terkait pernyataan komentar *body shaming* seperti apa yang mereka dapatkan. Dengan melalui *machine learning*, diharapkan dapat membantu mempermudah dan mempercepat proses pengklasifikasian komentar-komentar tersebut yang mana sebelumnya melakukan pengklasifikasian secara manual yang membutuhkan waktu lama dikarenakan penilaian yang dilakukan secara individual tidak efisien untuk jumlah komentar yang banyak.

Berdasarkan beberapa penelitian sebelumnya yang telah dilakukan proses pengklasifikasian teks, diantaranya adalah penelitian berjudul “*Toxic Speech Classification via Deep Learning using Combined Features from BERT & FastText Embedding*”(Asmi P & M S Sanaj, 2021). Penelitian ini melakukan pendekstian teks berbahasa inggris yang mengandung ujaran *toxic* berdasarkan 25296 cuitan komentar pada twitter dengan 2 percobaan. Pertama, menggunakan klasifikasi 2 kelas dengan pembagian ujaran kebencian dan ujaran yang menyenggung. Dan kedua, menggunakan klasifikasi *multi-class* ujaran kebencian, ujaran yang menyenggung, dan tidak termasuk keduanya. Tingkat akurasi *BERT* dan *CNN* berdasarkan *macro average F1 measure* pada pembagian 2 kelas sebesar 92,59% dan pada pembagian 3 kelas sebesar 72,59%.

Pada penelitian lainnya yang yang berjudul “*Aspect Based Sentiment Analysis Market Place Product Reviews Using BERT, LSTM, and CNN*”(Syaiful Imron et al., 2023) melakukan klasifikasi *sentiment analysis* terhadap review produk pada *marketplace* dengan menggunakan IndoBERT sebagai *word embedding* dan

menggunakan algoritma LSTM dan CNN untuk mengklasifikasi kata. Hasilnya menunjukkan BERT dan algoritma CNN memiliki akurasi sebesar 90,04%. Dengan penelitian lain yang berjudul “*Text Classification Using IndoBert Fine-Tuning Modeling with Convolutional Neural Network and Bi-LSTM*”(Zevana & Riana, 2024). Pada penelitian ini peneliti melakukan pengklasifikasian teks pendapat masyarakat terhadap layanan ekspedisi dengan 4 faktor. Berdasarkan penelitian ini IndoBERT digunakan untuk tokenisasi dan *modelling*, serta CNN digunakan sebagai algoritma classifier menghasilkan akurasi sebesar 83% dengan *text preprocessing*. Penelitian yang lain yang berjudul “*A BERT-CNN Based Approach on Movie Review Sentiment Analysis*” (Zhang, 2023)melakukan klasifikasi terhadap sentimen analisis terhadap *review film* dengan menggunakan dataset sebanyak 50.000 data, berdasarkan penelitian ini penggunaan *pre-trained* model BERT yaitu *best-base-uncased* dan algoritma CNN memberikan hasil akurasi sebesar 94,4%.

Penelitian sebelumnya yang telah dilakukan berjudul “Klasifikasi Komentar Body Shaming Beauty Vlogger Pada Youtube Menggunakan Metode BM25 dan K-Nearest Neighbor”(Aditya Prana & Pandu Adikara, 2019), penelitian ini melakukan pengklasifikasian komentar mengandung body shaming dan tidak body shaming pada komentar *beauty vlogger* di media sosial Youtube dengan menggunakan metode BM25 untuk mendapatkan skor kata. Kemudian menggunakan algoritma *K-Nearest Neighbor* utnuk mengklasifikasi komentar dan pada pengujian data menghasilkan akurasi sebesar 86,67%. Penelitian lainnya yang berjudul “Sentiment Analysis of the Body-Shaming Beauty Vlog Comments”(Jaman et al., 2020). Penelitian ini mengidentifikasi dan mengklasifikasikan komentar *body shaming* dengan menggunakan 33,044 komentar yang 986 diantaranya merupakan komentar *body shaming* dan sisanya bukan komentar *body shaming*. Pada tahap persiapan data, peneliti menggunakan teknik *text mining* dengan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM), kemudian pengklasifikasian komentar menggunakan metode *Naïve Bayes Classifier* yang menghasilkan akurasi tertinggi sebesar 98,48% dengan perbandingan data latih dengan data uji 90:10. Kemudian pada penelitian lain yang berjudul “Identifikasi Komentar Body Shaming Pada Media Sosial Menggunakan Term Frequency Chi-square dan Random Forest Classifier”(Pardede, 2022). Pada penelitian ini, penulis melakukan pengidentifikasi komentar mengandung *body shaming* pada media sosial menggunakan fitur *Term Frequency Chi-Square* yang digunakan untuk memberikan

pembobotan pada kata yang ada, selanjutnya akan diidentifikasi dengan menggunakan *Random Forest Classifier*. Dengan menggunakan 4300 komentar, pengidentifikasiannya menghasilkan akurasi sebesar 96,16%.

Berdasarkan latar belakang dan penelitian terdahulu di atas, penulis mengajukan penerapan algoritma *Convolutional Neural Network* dan pengekstrasiannya menggunakan *IndoBERT Embedding* untuk mengklasifikasi komentar yang mengandung *body shaming* pada sosial media. Penelitian ini diberi judul “**KLASIFIKASI KOMENTAR MENGANDUNG BODY SHAMING PADA MEDIA SOSIAL DENGAN MENGGUNAKAN ALGORITMA CONVOLUTIONAL NEURAL NETWORK**”.

1.2 Rumusan Masalah

Banyaknya jumlah komentar *body shaming* yang ada di media sosial dapat disebabkan oleh rendahnya kesadaran masyarakat terhadap dampak merugikan dari komentar-komentar yang disampaikan kepada individu lain. Padahal komentar-komentar tersebut dapat mengakibatkan munculnya rasa rendah diri dan gangguan mental yang akan dialami oleh korban. Melalui sistem yang dirancang ini, diharapkan mampu memberikan kemudahan dalam pengolahan komentar dengan cepat dan teliti. Dengan alasan tersebut, diperlukan sebuah sistem untuk dapat mengklasifikasi tindakan *body shaming* pada komentar di media sosial untuk memberikan informasi kepada masyarakat tentang bagian-bagian dari *body shaming*.

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah mengklasifikasikan komentar *body shaming* pada media sosial dengan menggunakan metode *IndoBERT embedding* dan algoritma *Convolutional Neural Network*.

1.4 Batasan Penelitian

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan berupa teks yang dimuat dalam file berekstensi .xlsx.
2. Dataset berupa komentar hasil *crawled* dari media sosial Instagram dan TikTok
3. Dataset yang digunakan merupakan komentar Bahasa Indonesia.

4. Klasifikasi komentar berfokus terhadap kata dan tidak menggunakan angka, gambar, ataupun emoji.
5. Klasifikasi komentar *body shaming* dibagi kedalam tiga kelas yaitu berdasarkan bentuk badan, bentuk wajah, dan warna kulit.
6. Aplikasi tidak dapat melakukan pengklasifikasian multikelas, hanya dapat mengklasifikasikan komentar ke dalam satu kelas saja.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Membantu menghindari pengguna media sosial untuk memberikan komentar mengandung *body shaming* yang menggunakan Bahasa Indonesia.
2. Mengetahui kinerja dari metode *IndoBERT Embedding* dan algoritma *Convolutional Neural Network* dalam pengklasifikasian komentar mengandung *body shaming* di media sosial.

1.6 Metodologi Penelitian

Tahapan-Tahapan yang akan dilakukan pada penelitian ini adalah:

1. Studi Literatur

Tahap ini mencakup pengumpulan data dan referensi yang berhubungan dengan *text processing*, metode *IndoBERT Embedding*, algoritma *Convolutional Neural Network*, dan pemahaman tentang *body shaming* dari berbagai sumber seperti jurnal, skripsi, buku, artikel, dan berbagai sumber bacaan lainnya.

2. Analisis Permasalahan

Pada tahap ini, melakukan analisis referensi dan data yang telah dikumpulkan terlebih dahulu untuk mendapatkan pemahaman tentang metode *IndoBERT Embedding* dan *Convolutional Neural Network* yang akan digunakan dalam penelitian untuk klasifikasi pernyataan mengandung *body shaming* berbahasa Indonesia.

3. Perancangan Sistem

Berdasarkan analisis yang telah dilakukan di tahap sebelumnya, dilakukan perancangan sistem mencakup perancangan arsitektur, penentuan data *training, validation* dan *testing*.

4. Implementasi

Tahap ini merupakan proses mengimplementasikan rancangan sistem yang telah dibuat pada tahap sebelumnya sehingga akan menghasilkan sistem yang sesuai dengan tujuan penelitian.

5. Pengujian Sistem

Pada tahap ini dilakukan pengujian terhadap performa sistem yang telah dikembangkan untuk memastikan bahwa sistem yang telah dibuat dapat digunakan dan untuk memperoleh nilai akurasi dari penggunaan metode *IndoBERT Embedding* dan *Convolutional Neural Network* dalam pengklasifikasian komentar mengandung *body shaming* berbahasa Indonesia.

6. Penyusunan Laporan

Pada tahap ini dilakukan penyusunan laporan yang memiliki tujuan untuk pencatatan terhadap tahapan dan proses penelitian yang digunakan dalam sehingga mencapai hasil akhir.

1.7 Sistematika Penulisan

Sistematika penulisan pada penelitian ini disusun berdasarkan lima bagian utama, yaitu:

Bab 1: Pendahuluan

Tahapan awal yaitu Pendahuluan, berisi pemberian penjelasan tentang latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metodologi penelitian, dan sistematika penulisan penelitian.

Bab 2: Landasan Teori

Bab ini berisi pembahasan terhadap landasan teori-teori yang pada penelitian yang bertujuan untuk membantu dalam memahami dan memecahkan permasalahan yang ada. Pada bagian landasan teori ini berisi teori terhadap komentar yang mengandung unsur *Body Shaming* serta memberikan penjelasan tentang metode yang digunakan dalam penelitian yaitu *IndoBERT Embedding* dan *Convolutional Neural Network*. Selain itu, pada bab ini juga menjelaskan penelitian sebelumnya sebagai referensi yang akan digunakan pada penelitian ini dan pembeda penelitian ini terhadap penelitian sebelumnya.

Bab 3: Analisis dan Perancangan Sistem

Pada bab ini memberikan penjelasan tentang peninjauan dan perancangan sistem pengklasifikasian komentar mengandung *body shaming* pada media sosial dengan menggunakan *IndoBERT embedding* dan *Convolutional Neural Network*. Tak hanya itu, pada bab ini juga berisi alur dari arsitektur yang digunakan untuk mengembangkan sistem.

Bab 4: Implementasi dan Pengujian Sistem

Bab ini berisikan penjelasan terhadap proses implementasi yang dilakukan terkait perancangan sistem yang telah ditetapkan. Kemudian sistem yang telah dikembangkan akan diuji dalam melakukan klasifikasi terhadap komentar *body shaming* sebagai cara untuk evaluasi dalam penelitian ini.

Bab 5: Kesimpulan dan Saran

Bagian ini berisi kesimpulan terhadap hasil penelitian yang telah dilakukan beserta saran dari penulis yang dapat ditingkatkan pada penelitian berikutnya dikarenakan setiap penelitian memiliki kekurangan yang perlu untuk dilakukan perbaikan untuk kedepannya.

BAB 2

LANDASAN TEORI

2.1 *Body Shaming*

Body Shaming adalah ungkapan yang mengacu pada tindakan mengkritisi atau mengomentari dengan pandangan negatif terhadap fisik atau penampilan individu lain (Erdianti et al., 2022). *Body shaming* termasuk bentuk perundungan yang pada umumnya dilakukan secara verbal atau tanpa adanya sentuhan fisik. Dalam kehidupan sehari-hari *body shaming* masih kerap kali dianggap sebagai candaan walaupun tindakan ini dapat memberikan dampak serius terhadap psikologis korbannya. Media sosial seperti Instagram dan Tiktok dapat menjadi *platform* digital utama, perilaku ini dapat terjadi dikarenakan sifatnya yang berbasis visual dan adanya interaksi publik (Fitria & Febrianti, 2020). Menurut penelitian (Sri & Astuti, 2019) tindakan *body shaming* di media sosial dapat berupa komentar yang berisi kritikan yang mencela terhadap suatu bagian tubuh, memberikan komentar puji dengan berkedok penghinaan fisik, dan memberikan saran dengan mengatur yang pantas ataupun tidak pantas dipakai. Objek *body shaming* dapat terdiri dari bentuk tubuh dan berat badan seseorang, wajah atau rupa, dan terhadap warna kulit (Pratiwi, 2021; Fauzia & Rahmiaji, 2019). Adapun penjelasan terkait kategori klasifikasi yang akan saya lakukan adalah (Araaf et al., 2023):

1. Bentuk badan

Pada bagian ini dapat dikategorikan dengan individu-individu yang memiliki bentuk badan gemuk atau *plus size*, termasuk juga dengan orang-orang dengan memiliki bentuk badan kurus ataupun terlalu kurus, selain itu bagian tubuh lainnya juga dapat menjadi objek *body shaming* seperti lengan, pinggul, paha, betis, dan bagian tubuh lainnya

2. Bentuk wajah

Dalam kategori ini, orang-orang yang memiliki bentuk wajah yang tidak sesuai dengan *standard* kecantikan pada masyarakat menjadi objek tindakan *body shaming*. Bagian wajah yang dapat menjadi objek *body shaming* adalah hidung, bibir, pipi, alis mata, mata, gigi, hingga tekstur pada wajah.

3. Warna kulit

Pada kategori ini merupakan tindakan *body shaming* berupa diskriminasi terhadap warna kulit yang berbeda, dapat berupa disebabkan oleh warna kulit yang berbeda seperti warna kulit yang gelap ataupun warna kulit yang terlalu putih atau pucat.

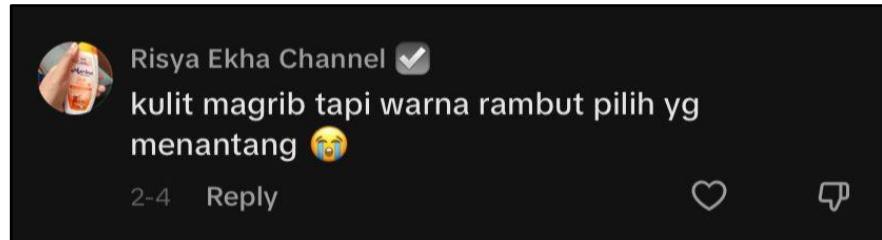
Tindakan *body shaming* merupakan salah satu dari sekian banyak permasalahan sosial yang masih sering terjadi di masyarakat secara tanpa disadari. Bahkan pada khidupan sehari-hari perilaku *body shaming* masih banyak dianggap sebagai bahan bercandaan. Padahal dapat memungkinkan mempengaruhi tingkat rasa rendah diri, kecemasan hingga dapat memicu masalah gangguan psikologis serta tekanan mental pada korban(Melizza et al., 2023). Berikut merupakan contoh komentar mengandung *body shaming* pada aplikasi TikTok dan Instagram.



Gambar 2. 1 Contoh Komentar *Body Shaming* Bentuk Badan



Gambar 2. 2 Contoh Komentar *Body Shaming* Bentuk Wajah



Gambar 2. 3 Contoh Komentar *Body Shaming* Warna Kulit

2.2 *Text Preprocessing*

Text preprocessing merupakan berbagai tahapan yang dilakukan dalam penelitian yang bertujuan untuk menghasilkan representasi data yang lebih terstruktur, teratur, dan sesuai sehingga dapat mempermudah dalam memahami proses analisis data pada penelitian. Dalam penelitian ini tahapan *text preprocessing* diperlukan dikarenakan pada data yang akan di ambil pada media sosial dapat banyak mengandung emoji, angka, dan karakter khusus yang dapat mempengaruhi kinerja algoritma saat mengidentifikasi teks. Adapun tahapan *text preprocessing* yang digunakan pada penelitian yaitu *data cleaning*, *case folding*, *punctuation removal*, *normalization*, *stopword removal*, *stemming* dan *tokenization*.

2.3 *Word Embedding*

Word embedding atau representasi kata adalah teknik yang dipakai untuk menyajikan teks ke dalam bentuk vektor. Teknik ini memberikan pengaruh dalam peningkatan kinerja proses *text classification*. Terdapat tiga kategori yang umum digunakan dalam representasi kata , yaitu *Traditional word embedding*, *Static word embedding*, dan *Contextualized word* (Birunda & Devi, 2021). *Contextualized word embedding* merupakan salah satu kategori *word embedding* yang yang bekerja dengan mengubah rangkaian huruf ke dalam bentuk vektor dengan memperhatikan konteks yang ada sehingga dapat menghasilkan hasil representasi yang lebih kaya dan informatif jika dibandingkan dengan menggunakan bentuk *word embedding* lainnya.

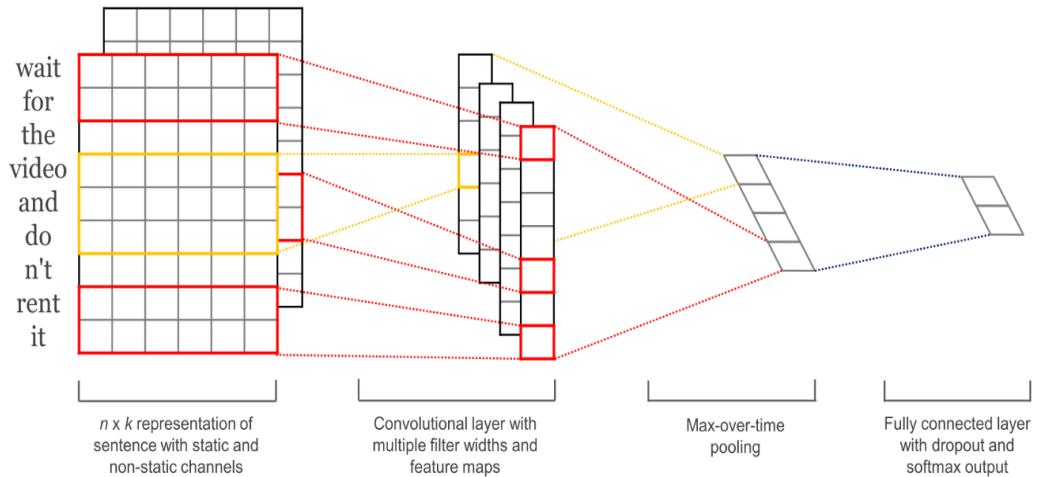
Salah satu jenis *contextualized word embedding* yang dapat melakukan representasi kata dengan baik adalah BERT (*Bidirectional Encoder Representation Transformers*). Dalam hal ini BERT dapat digunakan dalam mengatasi permasalahan

kata bermakna ganda seperti kata “*bank*” yang dapat diartikan dalam konteks berhubungan dengan keuangan ataupun berhubungan dengan sungai seperti “*river bank*” yang berarti tepi sungai dalam Bahasa Indonesia. BERT merupakan *contextualized word embedding* yang bekerja dengan pendekatan menggunakan model *transformers* dan mekanisme *self-attention* yang mana mekanisme ini bekerja dengan memperhatikan interaksi antar kata dan menentukan kata yang perlu lebih diperhatikan sehingga dapat mengidentifikasi hubungan antar kata. Pada BERT terdapat metode tokenisasi khusus yang digunakan yaitu *WordPiece tokenizer*, pada *WordPiece tokenizer text processing* dilakukan dengan memisahkan teks menjadi ke dalam subword units. Selain itu, BERT menggunakan *special token* yaitu [CLS] yang berada pada permulaan kalimat dan [SEP] yang berada di penutup kalimat dalam pernyataan yang diinput. Untuk membantu ekstraksi fitur dalam menghasilkan analisis dan mempelajari hubungan kata berdasarkan konteks diperlukan adanya *pre-trained model*, salah satu pre-trained dari BERT adalah IndoBERT. IndoBERT telah dilatih dengan menggunakan jutaan kata yang diambil dari Wikipedia Indonesia sebanyak 74 juta, artikel berita seperti Kompas, Tempo, dan Liputan 6 sebanyak 55 juta, dan juga *Indonesian Web Corpus* sebanyak 90 juta kata (Nabiilah et al., 2024).

2.4 Convolutional Neural Network

Convolutional Neural Network (CNN) merupakan *multi-layer* jaringan saraf tiruan yang memiliki kemampuan untuk mengekstraksi fitur-fitur kompleks dalam data. CNN merupakan algoritma *deep learning* yang awalnya dimanfaatkan untuk melakukan penganalisisan citra digital, akan tetapi seiring dengan berkembangnya penelitian, seperti yang dilakukan oleh (Kim, 2014) menunjukkan bahwa algoritma CNN dapat digunakan untuk melakukan *text processing*. Dibandingkan dengan menggunakan *traditional model*, *deep learning model* dapat melakukan klasifikasi teks tanpa perlu menentukan aturan ataupun *features* secara manual karena menggunakan pendekatan berbasis data yang secara otomatis dapat menyediakan representasi secara semantik untuk *text mining*. Keunggulan CNN dari model *deep learning* lainnya adalah dapat dengan baik menemukan pola pada kalimat tanpa tergantung pada letaknya pada suatu kalimat (Li et al., 2022). Terdapat serangkaian lapisan yang perlu dilakukan dalam algoritma CNN termasuk di dalamnya *Convolution layer*, *Pooling layer*, dan *Fully Connected layer*. Selain itu terdapat teknik regularisasi seperti *dropout* yang digunakan

agar mencegah terjadinya overfitting pada model dan dalam mengklasifikasikan kelas diperlukan adanya fungsi aktivasi *softmax* pada model. Berikut merupakan arsitektur algoritma *Convolutional Neural Network*.



Gambar 2. 4 Arsitektur CNN (Kim, 2014)

Pada arsitektur *Convolutional Neural Network* diatas, dapat dilihat panjang kalimat input yang diwakilkan dengan n dan k yang merupakan perwakilan untuk ukuran dimensi *word embedding* akan menjadi representasi matriks input untuk model CNN. Berikutnya, matriks input ini yang akan digunakan pada pemrosesan *convolutional layer*, berlanjut pada lapisan berikutnya yaitu *pooling layer*, dan kemudian sampai pada *fully connected layer* untuk melakukan klasifikasi berdasarkan kelas yang telah ditentukan.

1) Convolutional Layer

Lapisan ini merupakan lapisan awal pada algoritma CNN yang bertugas untuk mengekstraksi fitur penting yang terdapat pada matriks yang merupakan bentuk input pada CNN. Di lapisan ini pada matriks akan dilakukan penerapan *filter* dan *kernel* yang merupakan pendekripsi fitur guna mempelajari pola penting pada data input, yang kemudian hasil dari lapisan ini akan menjadi *features map*. Berikut merupakan rumus untuk mendapatkan *output* dalam lapisan konvolusi.

$$Z[i] = \sum_{j=0}^{k-1} (X[i+j] \cdot W[j]) + b \quad \dots \quad (2.1)$$

Keterangan:

X = input data

W = bobot filter

b = bias

k = ukuran kernel

j = indeks elemen filter

i = posisi

2) *Pooling Layer*

Lapisan ini bertujuan untuk mengurangi dimensi dari *features map* dengan hanya mengambil fitur-fitur terpenting yang ada. Pada umumnya *pooling layer* mengambil nilai maksimum yang terdapat pada setiap jendela *filter* untuk menentukan fitur yang paling menonjol. Selain itu, proses ini juga membantu menjaga informasi yang penting dan mengabaikan detail yang tidak terkait. Hasil dari *pooling layer* berupa peta fitur yang berisi perwakilan fitur-fitur terpenting dari teks tersebut. Berikut merupakan rumus untuk hasil lapisan maksimum *pooling*.

$$P[i] = \max\{Z[i], Z[i+1], \dots, Z[i+l-1]\} \dots \quad (2.2)$$

Keterangan:

l = ukuran jendela *pooling*

3) *Fully Connected Layer*

Lapisan *fully connected* merupakan lapisan akhir yang ada di model *Convolutional Neural Network*, pada lapisan ini setiap jaringan saraf akan terhubung dengan jaringan saraf hasil dari lapisan sebelumnya. Lapisan ini bekerja dengan melakukan transformasi linier pada vektor input dengan menggunakan matriks bobot. Lapisan ini diperlukan karena bertujuan untuk menghasilkan prediksi kelas pada klasifikasi teks. Berikut merupakan rumus umum yang digunakan pada lapisan *fully connected layer*.

$$y = Wx + b \dots \dots \dots \quad (2.3)$$

Keterangan:

x = vektor input lapisan sebelumnya

W = bobot matriks

y = vektor hasil

b = bias

2.5 Confusion Matrix

Confusion matrix adalah sistem yang dilakukan dengan tujuan mengkaji kinerja sebuah model yang terkait dalam pengklasifikasian data. Dengan membandingkan hasil prediksi dan label asli, *confusion matrix* menyajikan informasi tentang prediksi yang dibuat oleh model. Hasil perhitungan dari *confusion matrix* dapat ditentukan menjadi empat bagian, yaitu:

- a. *True Negative* (TN) adalah model memproyeksikan hasil data bernilai negatif tetapi data sebenarnya bernilai negatif.
 - b. *True Positive* (TP) adalah model memproyeksikan hasil data bernilai positif tetapi data sebenarnya bernilai positif.
 - c. *False Negative* (FN) adalah model memproyeksikan hasil data bernilai negatif tetapi data sebenarnya bernilai positif.
 - d. *False Positive* (FP) adalah model memproyekasikan hasil data bernilai positif, tetapi hasil data sebenarnya bernilai negatif.

Adapun di tabel 2.1 berikut menyajikan bentuk penerapan dari *confusion matrix* tersebut.

Tabel 2. 1 Penerapan *confusion matrix*

| | | Aktual | |
|----------|---------|---------|---------|
| | | Negatif | Positif |
| Prediksi | Negatif | TN | FN |
| | Positif | FP | TP |

Terdapat beberapa perhitungan metrik secara umum yang digunakan untuk performansi, yaitu:

2.5.1 Accuracy

Accuracy merupakan persentase perhitungan persentase prediksi yang benar dari semua prediksi yang telah dibuat oleh model. Semakin banyak nilai data yang diprediksi dengan tepat, maka semakin tinggi pula nilai akurasi dalam suatu penelitian.

$$Accuracy = \frac{TP+TN}{Total\ Data} \times 100\% \quad \dots \dots \dots \quad (2.3)$$

2.5.2 Precision

Precision merupakan persentase perhitungan prediksi positif yang stabil dan tepat di antara semua prediksi positif. Hal ini dilakukan untuk memahami seberapa akurat model dalam menentukan kelas positif.

$$Precision = \frac{TP}{TP+FP} \times 100\% \dots \quad (2.4)$$

2.5.3 Recall

Recall atau sensitivitas menunjukkan seberapa baik model dapat menemukan kelas positif dengan menghitung persentase sampel positif yang dapat berhasil dideteksi oleh model dari seluruh sampel positif.

$$Recall = \frac{TP}{TP+FN} \times 100\% \dots \dots \dots \quad (2.5)$$

2.5.4 F-1 Score

F-1 Score merupakan metrik yang mengkombinasikan dua parameter yaitu, *precision* dan *recall* menjadi sebuah nilai tunggal, ini bermanfaat dilakukan jika jumlah kelas positif dan negatif tidak seimbang. Dengan kata lain, *F-1 Score* dapat memberikan keseimbangan antara *precision* dan *recall*. Sehingga dengan ini dapat menunjukkan kemampuan model dalam mencapai akurasi yang seimbang dalam menemukan data positif yang tepat dan membuat prediksi yang benar.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \dots \quad (2.6)$$

2.6 Penelitian Terdahulu

Berdasarkan beberapa penelitian sebelumnya yang telah dilakukan proses pengklasifikasian teks, diantaranya adalah penelitian berjudul “*Toxic Speech Classification via Deep Learning using Combined Features from BERT & FastText Embedding*”(Asmi P & M S Sanaj, 2021). Penelitian ini melakukan pendekripsi teks berbahasa inggris yang mengandung ujaran kebencian berdasarkan 25296 cuitan komentar pada Twitter dengan 2 percobaan. Pertama, menggunakan klasifikasi 2 kelas dengan pembagian ujaran kebencian dan ujaran yang menyindir. Dan kedua, menggunakan klasifikasi *multi-class* ujaran kebencian, ujaran yang menyindir, dan tidak termasuk keduanya. Setiap percobaan yang dibuktikan melalui penelitian ini melakukan pengklasifikasian berdasarkan algoritma serta *embedding* berbeda. Diantaranya menggunakan gabungan *fastText* embedding dengan CNN dan Bi-LSTM, Bert Embedding dengan CNN dan Bi-LSTM, dan menggabungkan *TF-IDF*, *Bert Embedding*, *fastText Embedding* dengan CNN dan Bi-LSTM. Tingkat akurasi BERT dan CNN yang terlah dihitung pada pembagian 2 kelas sebesar 92,59% dan pada pembagian 3 kelas sebesar 72,59%.

Pada penelitian lainnya yang berjudul “*Aspect Based Sentiment Analysis Market Place Product Reviews Using BERT, LSTM, and CNN*” (Syaiful Imron et al., 2023) melakukan *multi-class* klasifikasi sentimen analisis terhadap review produk pada marketplace berdasarkan 6 aspek yaitu kualitas, pengiriman, akurasi, pengemasan, harga, dan pelayanan. Kemudian mempertegas setiap aspek menjadi 1 sentimen yaitu

berupa positif atau negatif. Dengan menggunakan IndoBERT sebagai *word embedding* dan algoritma LSTM untuk klasifikasi aspek dan CNN untuk mengklasifikasi sentiment. Hasilnya menunjukkan BERT dan algoritma CNN memiliki akurasi sebesar 90,04%.

Penelitian yang lain yang berjudul “*A BERT-CNN Based Approach on Movie Review Sentiment Analysis*” (Zhang, 2023) melakukan klasifikasi terhadap sentiment analisis terhadap *review film* dengan menggunakan dataset yang diambil dari IMDb sebanyak 50.000 data, berdasarkan penelitian ini penggunaan *pre-trained* model BERT yaitu *best-base-uncased* dan algoritma CNN memberikan hasil akurasi sebesar 94,4%.

Dengan penelitian lain yang berjudul “*Text Classification Using IndoBert Fine-Tuning Modeling with Convolutional Neural Network and Bi-LSTM*”(Zevana & Riana, 2024). Pada penelitian ini peneliti melakukan pengklasifikasian teks pendapat masyarakat terhadap layanan ekspedisi dengan 4 faktor yaitu harga, waktu, pengembalian dan lainnya pada media sosial Twitter. Berdasarkan penelitian ini IndoBERT digunakan untuk tokenisasi dan *modelling*, serta CNN digunakan sebagai algoritma pengklasifikasi menghasilkan akurasi sebesar 83% dengan *text preprocessing*.

Penelitian sebelumnya yang telah dilakukan berjudul “*Klasifikasi Komentar Body Shaming Beauty Vlogger Pada Youtube Menggunakan Metode BM25 dan K-Nearest Neighbor*”(Aditya Prana & Pandu Adikara, 2019), pada penelitian ini menggunakan 600 kata mengandung dengan 300 komentar *body shaming* dan 300 komentar tidak *body shaming* yang diambil dari komentar pada konten *beauty vlogger* pada media sosial Youtube. Penelitian ini menggunakan metode BM25 untuk mendapatkan skor kata. Kemudian menggunakan algoritma *K-Nearest Neighbor* utnuk mengklasifikasi komentar dan pada pengujian data menghasilkan akurasi sebesar 86,67%.

Penelitian lainnya yang berjudul “*Sentiment Analysis of the Body-Shaming Beauty Vlog Comments*”(Jaman et al., 2020). Penelitian ini mengidentifikasi dan mengklasifikasikan komentar body shaming dengan menggunakan 33,044 komentar yang 986 diantaranya merupakan komentar body shaming dan sisanya bukan komentar body shaming. Pada tahap persiapan data, peneliti menggunakan teknik text mining

dengan metode *Cross-Industry Standard Process for Data Mining (CRISP-DM)*, kemudian pengklasifikasian komentar menggunakan metode *Naïve Bayes Classifier* yang menghasilkan akurasi tertinggi sebesar 98,48% dengan perbandingan data latih dengan data uji 90:10.

Kemudian, pada penelitian lain yang berjudul “*Identifikasi Komentar Body Shaming Pada Media Sosial Menggunakan Term Frequency Chi-square dan Random Forest Classifier*”(Pardede, 2022). Pada penelitian ini, penulis melakukan pengidentifikasi komentar mengandung *body shaming* pada media sosial menggunakan fitur *Term Frequency Chi-Square* yang digunakan untuk memberikan pembobotan pada kata yang ada, selanjutnya akan diidentifikasi dengan menggunakan *Random Forest Classifier*. Dengan menggunakan 4300 data komentar, pengidentifikasi komentar menghasilkan akurasi sebesar 96,16%.

Tabel 2. 2 Penelitian Terdahulu

| No | Penulis | Tahun | Metode | Keterangan |
|----|-------------------------------|-------|------------------------------|--|
| 1 | Asmi P dan Sanaj M S | 2021 | BERT, FastText, CNN, Bi-LSTM | Penelitian ini melakukan pendekripsi teks berbahasa Inggris yang mengandung ujaran <i>toxic</i> berdasarkan 25296 cuitan komentar pada Twitter dengan dua percobaan. Pertama, menggunakan klasifikasi dua kelas. Dan kedua, menggunakan klasifikasi <i>multi-class</i> . Tingkat akurasi <i>BERT</i> dan <i>CNN</i> berdasarkan <i>macro average F1 measure</i> pada pembagian 2 kelas sebesar 92,59% dan pada pembagian 3 kelas sebesar 72,59%. |
| 2 | Syaiful Imron, Esther Irawati | 2023 | IndoBERT, CNN, LSTM | Melakukan klasifikasi <i>sentiment analysis</i> terhadap review produk pada <i>marketplace</i> dengan menggunakan IndoBERT sebagai <i>word embedding</i> |

| | | | | | |
|----------|---|------|------------------------------------|---|---|
| | | | | Setiawan, Joan Santoso, dan Mauridhi Hery Purnomo | dan menggunakan algoritma LSTM dan CNN untuk mengklasifikasi kata. Hasilnya menunjukkan BERT dan algoritma CNN memiliki akurasi sebesar 90,04%. |
| 3 | Alda Zevana Putri Widodo dan Dwiza Riana | 2023 | IndoBERT, CNN | Pada penelitian ini peneliti melakukan pengklasifikasian teks pendapat masyarakat terhadap layanan ekspedisi dengan 4 faktor. Berdasarkan penelitian ini IndoBert digunakan untuk tokenisasi dan <i>modelling</i> , serta CNN digunakan sebagai algoritma <i>classifier</i> menghasilkan akurasi sebesar 83%. | |
| 4 | Bowen Zhang | 2023 | BERT dan CNN | Pengklasifikasian terhadap sentimen analisis terhadap <i>review film</i> dengan menggunakan dataset yang diambil dari IMDb sebanyak 50.000 data, berdasarkan penelitian ini penggunaan <i>pre-trained</i> model BERT yaitu <i>best-base-uncased</i> dan algoritma CNN memberikan hasil akurasi sebesar 94,4%. | |
| 5 | Pengkuh Aditya Prana, Indriati, dan Putra Pandu Adikara | 2019 | BM25 dan <i>K-Nearest Neighbor</i> | Penelitian ini mengerjakan pengklasifikasian komentar mengandung <i>body shaming</i> dan tidak <i>body shaming</i> pada komentar <i>beauty vlogger</i> di media sosial Youtube dengan menggunakan metode BM25 untuk mendapatkan skor kata. | |

| | | | | | |
|---|---|------|---|---|--|
| | | | | | Kemudian, menerapkan algoritma <i>K-Nearest Neighbor</i> dalam klasifikasi komentar dan pada pengujian data menghasilkan akurasi sebesar 86,67%. |
| 6 | Jajam Haerul Jaman, Hannie, dan Martina Sari Simatupang | 2020 | <i>Naïve Bayes Classifier</i> | Penelitian ini mengidentifikasi dan mengklasifikasikan komentar <i>body shaming</i> dengan menggunakan 33,044 komentar. Pada tahap persiapan data, peneliti memanfaatkan pendekatan <i>text mining</i> dengan teknik <i>Cross-Industry Standard Process for Data Mining</i> (CRISP-DM) kemudian, klasifikasi komentar mempergunakan <i>Naïve Bayes Classifier</i> dan menghasilkan akurasi tertinggi sebesar 98,48% dengan perbandingan data pelatihan dengan data pengujian 90:10. | |
| 7 | Michellle Christine Natalia Pardede | 2022 | <i>Term Frequency Chi-Square, Random Forest</i> | Pengidentifikasi komentar mengandung <i>body shaming</i> pada media sosial menggunakan fitur <i>Term Frequency Chi-Square</i> yang digunakan untuk memberikan pembobotan pada kata yang ada, selanjutnya akan diidentifikasi dengan menggunakan <i>Random Forest Classifier</i> . Dengan menggunakan 4300 komentar, pengidentifikasi komentar menghasilkan akurasi sebesar 96,16%. | |

2.7 Perbedaan penelitian

Penelitian yang akan dilaksanakan adalah klasifikasi *body shaming* menjadi 3 kelas yaitu berdasarkan bentuk badan, bentuk wajah, dan warna kulit. Penelitian terdahulu yang dikerjakan oleh Prana et al (2019) adalah mengidentifikasi komentar *body shaming* pada *beauty vlogger* dengan mengoperasikan teknik BM25 serta *K-Nearest Neighbor*. Hannie & Simatupang (2020) melakukan penelitian mengidentifikasi *body shaming* dengan menggunakan metode *Naïve Bayes*. Pardede (2022) melakukan penelitian dengan mengidentifikasi *body shaming* dengan mengoperasikan teknik *Term Frequency Chi-Square* dan *Random Forest*. Adapun penelitian yang akan dilaksanakan ini dengan memanfaatkan metode *IndoBERT Embedding* sebagai *word embedding* untuk menghasilkan representasi vektor untuk meningkatkan akurasi dan proses klasifikasi menggunakan algoritma *Convolutional Neural Network*.

BAB 3

ANALISIS DAN PERANCANGAN SISTEM

3.1 Data yang Digunakan

Data yang diterapkan melalui penelitian ini mengandung komentar yang dibuat selama proses pengumpulan informasi, kemudian komentar yang diperoleh akan dipilah secara manual untuk dikategorikan dataset yang dapat mewakili objek yang diinginkan sekaligus mengurangi kemungkinan data menganggu (*noise*) yang dapat mempengaruhi proses dari pengujian dan pelatihan. Proses pengumpulan komentar-komentar tersebut didapatkan melalui hasil *crawling* dari kolom komentar dari beberapa akun di Instagram dan Tiktok. Total data yang didapat setelah dilakukan pemilahan mencapai 3500 data yang kemudian disimpan pada bentuk .xlsx.

Pada data tercantum komentar dan kelas, yang digunakan untuk menentukan apakah komentar tersebut mengandung unsur *body shaming* berdasarkan bentuk badan, *body shaming* berdasarkan bentuk wajah, *body shaming* berdasarkan warna kulit, ataupun tidak mengandung *body shaming*. Masing-masing komentar akan diberikan label kelas berupa angka, dimana komentar yang berunsur *body shaming* terhadap bentuk badan akan diberi kelas 1, *body shaming* terhadap bentuk wajah diberi kelas 2, *body shaming* terhadap warna kulit akan diberi kelas 3, dan Adapun komentar yang bukan merupakan *body shaming* akan diberi kelas 0. Berdasarkan penelitian yang dilakukan oleh (Sri & Astuti, 2019) terhadap gambaran komentar *body shaming* pada media sosial yang dapat diterima oleh korban *body shaming*, penulis mendapatkan referensi dalam penetapan kategori *body shaming* pada dataset dan juga penulis melakukan diskusi kepada ahli bahasa terhadap penetapan label kelas yang penulis lakukan pada dataset yang ada.

Tabel 3. 1 Dataset

| No | Komentar | Label |
|------|--|-------|
| 1. | bisa nahan rasa sakit kata2nya tapi ngk bisa nahan air mata ðŸ¥ºðŸ¥º | 0 |
| 2. | muka kamu tuh kaya imut lucu bgt gitu lo kaya yg gemes Ã°Å,Â¤Â¤Ã°Å,Â¤Â»Ã°Å,ËœÂ½ | 0 |
| 3. | ga ada salahnya mba kalo mulai coba diet dari sekarang, biar bisa tetap sehat badannya | 0 |
| 4. | Yg baju merah udah tonggos komuk aneh buruk rupa kali kau .. minimal ngaca gigi tonggos lu ga enak dilihat | 2 |
| 5. | Udah kurus kering Ã°Å,â€”Ã¢ minta tobrut? Dikelep di tengah gunung tinggal nama | 1 |
| 6. | Ninra yg mukanya kyk centong penyok itu kan?pdhl si ninra itu buka tarot ama jasa buka aura. Lha yg menyekutukan Tuhan dsini siapa cb? Wkwkwk kocak bgt anjir | 2 |
| 7. | Ayok kalian para orang2 gendut, silahkan marah di komen, marah juga membakar kalori lho, yuk dipersilahkan Ã°Å,â„¢Â¤Ã°Å,Â¤Ã» | 1 |
| 8. | Editannya cakep cerah, begitu ketemu langsung sm orang nya kusam | 3 |
| 9. | bjir dah kulit gelap ini perkara bentuk mata aja dah merasa cindo | 3 |
| 10. | Beda bgt ya bibir item karena rokok sama item karena emang keturunan. Klo ngerokok gini itemnya kek abu'?, butek wkwkðŸ¤£ | 3 |
| . | . | . |
| . | . | . |
| . | . | . |
| . | . | . |
| . | . | . |
| 3495 | Njir!! Masih Mending Pentil Gua Warna Nya Pink Peach Gtu Ko Dia Malah Item Anjir??????? | 3 |
| 3496 | Ihh lehernya kok gelap banget sih jorok banget | 3 |
| 3497 | Setidaknya punya mata teh yang bener atuh, jangan sebelah ke kanan sebelah lagi ke kiri... Biar apa? Wkwk | 2 |

| | | |
|------|---|---|
| 3498 | rejeki tinggi tapi ga halal itu bukan rejeki ya ky tapi itu maruk, yee gue kirim susu peninggi badan baru tau rasa lu | 1 |
| 3499 | OWHHH TOLOL EMANG, FISIK LO EMANG PENDEK APALAGI AKALNYA | 1 |
| 3500 | Padahal komennya disini pada baik lho cuma elu nya aja mbak yang nolak kebenaran yang terjadi sama elu | 0 |

Berikut merupakan beberapa contoh kalimat pada komentar terhadap masing-masing kelas klasifikasi *body shaming*.

Tabel 3. 2 Komentar Body Shaming Bentuk Badan

| No. | Komentar |
|-----|--|
| 1. | "kirain kalo jadi dokter tuh badannya terawat bentukannya eh ternyata aku salah 😂" |
| 2. | "Udah kurus kering ðŸ—_ minta tobrut? Dikelep di tengah gunung tinggal nama" |
| 3. | "Keren kah begitu? Berhijab tapi mamerin tetek sama ventilnya yang tepos itu? Mereka ga sadar kah kalau yang mereka lakuin malah ngejatuhin harga diri mereka sendiri? Jijik banget sat" |
| 4. | "Yaallah rose gw tinggal tulang sama kulit doang 🤣🤣🤣🤣" |
| 5. | "Ngehina? Lah fakta boss orang gendut wkwkwk. Masa dia gendut tapi dikatain gendut ga terima" |

Tabel 3.2 berisi contoh-contoh komentar yang mengandung *body shaming* terhadap bentuk badan, yang dapat dilihat berisi kata-kata yang mengandung unsur penghinaan terhadap anggota badan.

Tabel 3. 3 Komentar Body Shaming Bentuk Wajah

| No. | Komentar |
|-----|--|
| 1. | "Blm kena bom muka udah pada ancur gitu 😂 minimal klo muka kyk sampah dasar jelek. Hati dan etika baik lah 😥 " |
| 2. | "beberapa orang tambah manis pake lesung pipi tp bbrapa orang juga gak pantes contohnya kamuðŸ~,ðŸ~," |
| 3. | "besok besok mending perbaikin bagian muka yang lain aja dari pada itu pipi jadi makin jelek kan lu" |
| 4. | "200 itu udah cakep kocak, kan emang dasarnya muka lu banyak bopeng, trus lu berharap jadi mulus? diampelas dulu harusnya, ngeselin bgt, kalo ga terima ya harusnya meap sendiri aja deh, ribet" |
| 5. | "mata udh 4 masih aja byk laga nya, minimal ituu gigi rada di bagusin dlu biar di liatnya gee cakepan dikit gigi pada maju kaya mo nempel aspal 🤡 " |

Tabel 3.3 berisi komentar-komentar yang mengandung body shaming terhadap bentuk wajah, yang dapat dilihat dari beberapa contoh tersebut yang menyebutkan unsur menghina dan ejekan terhadap bentuk wajah.

Tabel 3. 4 Komentar Body Shaming Warna Kulit

| No. | Komentar |
|-----|---|
| 1. | "najiss sadar diri woi mana ada kulit putih kek gini,,, lu tuh emang item maghrib GOBLOK" |
| 2. | "Bibir item, kulit item, pake baju coklat, setelan di hujat" |
| 3. | "Njir!! Masih Mending Pentil Gua Warna Nya Pink Peach Gtu Ko Dia Malah Item Anjir?????? " |
| 4. | "Kalau kalian sering ketemu bule, kulit pucat mereka itu gak secakep di layar hp kalian. Kek, biasa aja gitu. Sampe bikin aku mikir kalau tan skin itu lebih manis" |
| 5. | " ngerokok 10 batang sehari kali ya makanya moncong dia bisa biru " |

Tabel 3.4 berisi beberapa komentar mengandung *body shaming* terhadap warna kulit, seperti yang dapat diperhatikan terdapat kata-kata yang mengandung unsur penghinaan dan ejekan terhadap warna kulit tertentu.

Berikutnya koleksi data diambil melalui dengan *crawled* akan dipecah menjadi dua bagian yaitu data pelatihan dan data pengujian. Proporsi yang dipakai saat memecah dataset adalah dengan 80:20, sebanyak 80% dari dataset akan dipergunakan untuk data pelatihan, dan sisa sebanyak 20% akan dipergunakan sebagai data uji. Jumlah keseluruhan data pelatihan juga akan dipecah dengan menggunakan perbandingan 90:10, dimana 90% akan dipakai sebagai data latih utama dan 10% yang merupakan sisanya akan dipakai sebagai data validasi. Untuk menunjukkan detail terkait pemecahan dataset dapat diperhatikan melalui tabel 3.5.

Tabel 3. 5 Pembagian Dataset

| Data Training | | Data Testing |
|----------------------------|----------------------|---------------------|
| Data Training Utama | Data Validasi | |
| 2450 | 350 | 700 |

3.2 Analisis Sistem

Pada langkah ini penulis menerapkan beberapa rangkaian penggerjaan yang perlu dilakukan yaitu, dimulai dengan pengumpulan data, dimana komentar yang dipakai merupakan data yang diperoleh melalui kolom komentar aplikasi Instagram dan Tiktok dengan menggunakan proses *crawling* dan disimpan dengan format .xlsx. Lalu, data kemudian akan dipecah menjadi tiga bagian yaitu data latih, data validasi, dan data uji. Penggunaan data latih memiliki tujuan yakni melatih model pembelajaran mesin untuk mempelajari kerangka berpikir dari data yang diberikan. Kemudian, selama proses pelatihan ini model akan melakukan evaluasi dengan menggunakan data validasi. Setelah itu, akan dilakukan pengujian terhadap data model hasil pelatihan dengan menggunakan data uji.

Berikutnya merupakan tahapan-tahapan pada teks *preprocessing*, pada serangkaian tahapan ini terdapat langkah-langkah yang perlu dilakukan diantaranya adalah *cleaning* (pembersihan), *case folding* (penyeragaman huruf), *punctuation*

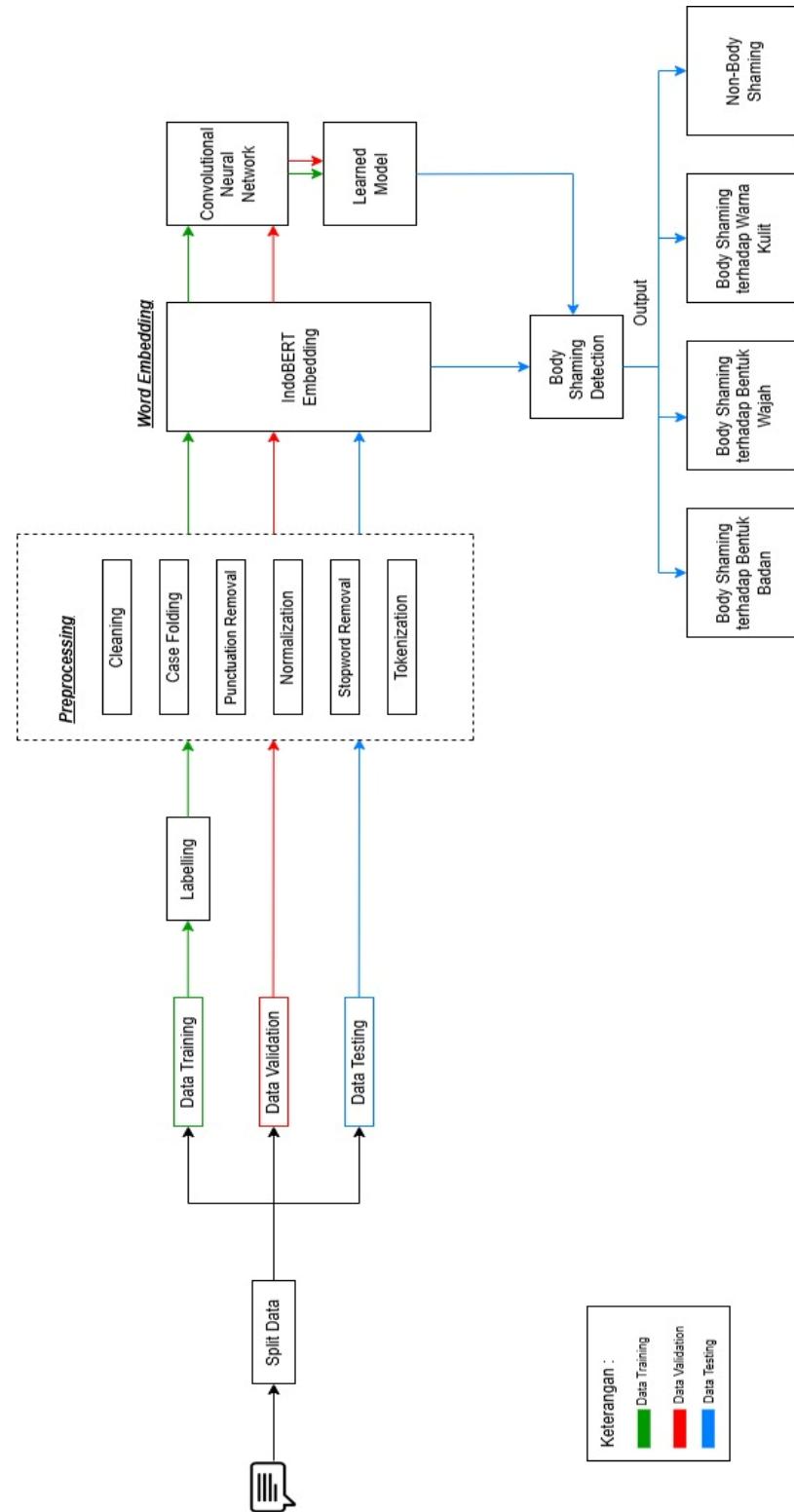
removal (penghapusan tanda baca ataupun simbol khusus), *normalization* (normalisasi terhadap kata singkatan dan *typo*), *stopword removal* (penghapusan kata yang tidak memberikan pengaruh jika kata tersebut dihapus), dan *tokenization*. Setelah menyelesaikan tahapan teks *preprocessing* data hasil dari langkah *tokenization* tersebut yang berupa token akan diproses oleh *IndoBERT embedding* untuk mendapatkan konversi *embedding* dalam bentuk vektor numerikal yang memiliki dimensi yang sama dari kata yang ada di dalam data. Langkah terakhir, hasil *embedding* yang dilakukan sebelumnya oleh *IndoBERT embedding* tersebut akan di *training* dengan data *train* dan bersamaan dengan itu melakukan evaluasi dengan data *validation* menggunakan algoritma *Convolutional Neural Network* dan pengujian model dilakukan pada proses *testing* dengan hasil berupa klasifikasi empat kelas komentar *body shaming* yaitu komentar *body shaming* berdasarkan bentuk badan, *body shaming* berdasarkan bentuk wajah, *body shaming* berdasarkan warna kulit, dan *non-body shaming*. Pada gambar 3.1 menujukkan gambaran arsitektur umum yang digunakan dalam mengembangkan penelitian klasifikasi *body shaming* ini.

3.2.1 *Preprocessing*

Preprocessing merupakan langkah yang dilakukan guna mendapatkan hasil data yang terskruktur dengan baik dan dapat dipahami oleh model dengan benar sehingga dapat menghasilkan tingkat keakuratan dengan tepat dan baik. Tahapan pada yang dilakukan dalam proses ini adalah yaitu *data cleaning*, *case folding*, *punctuation removal*, *normalization*, *stopword removal*, dan *tokenization*.

a. Cleaning

Cleaning adalah tahapan pertama dan fundamental yang dilakukan pada *preprocessing* yang bertujuan untuk membersihkan data sehingga menghasilkan data yang andal untuk digunakan pada pengklasifikasian komentar. Bentuk pengimplementasiannya dapat dilihat dalam *pseudocode* 3.1.



Gambar 3. 1 Arsitektur Umum

```

FUNCTION text_preprocessing (text):
    #Menghapus URL dari teks
    url_pattern = compile_url_pattern()
    text = replace_with_empty(text, url_pattern)
    # Menghapus tag HTML dari teks
    html_pattern = compile_html_pattern()
    text = replace_with_empty(text, html_pattern)
    #Mengatasi kata dengan pengulangan huruf berlebih
    text = handle_word_elongation(text)
    #Menghapus spasi berlebih dari teks
    text = remove_extra_spaces(text)
    RETURN text
END FUNCTION

```

Pseudocode 3. 1 Cleaning

Berikut adalah hasil dari langkah *preprocessing* data *cleaning* pada dataset yang dimuat pada tabel 3.6.

Tabel 3. 6 Implementasi *Cleaning*

| Sebelum Implementasi Cleaning | Setelah Implementasi Cleaning |
|--|--|
| Kuat banget kak Å°Å,ËœÅ- aku dikatain jelek ama 1 org aja udh buat aku gk mau lagi up diri sendiri di sosmed, cukup lama bgt buat gk peduli sama perkataan org wkwkÅ°Å,ËœÅ | Kuat banget kak aku dikatain jelek ama org aja udh buat aku gk mau lagi up diri sendiri di sosmed, cukup lama bgt buat gk peduli sama perkataan org wkwk |

b. *Case Folding*

Case folding ialah tahapan dalam *preprocessing* yang dijalankan bertujuan dalam penyeragaman huruf yang pada data terdapat perbedaan bentuk huruf yaitu huruf besar dan huruf kecil. Dengan *case folding* bentuk huruf akan diseragamkan menjadi bentuk huruf kecil. Berikut tersaji di tabel 3.7 yang merupakan hasil dari *case folding*.

Tabel 3. 7 Implementasi Case Folding

| Sebelum Implementasi Case Folding | Setelah Implementasi Case Folding |
|--|--|
| Kuat banget kak aku dikatain jelek ama org aja udh buat aku gk mau lagi up diri sendiri di sosmed, cukup lama bgt buat gk peduli sama perkataan org wkwk | kuat banget kak aku dikatain jelek ama org aja udh buat aku gk mau lagi up diri sendiri di sosmed, cukup lama bgt buat gk peduli sama perkataan org wkwk |

c. Punctuation Removal

Punctuation removal adalah tahapan pada *preprocessing* yang memiliki tujuan terhadap penghapusan karakter yang tidak berkaitan pada proses klasifikasi yang dapat berupa tanda baca dan simbol khusus, hal ini dilakukan agar pemrosesan hanya terfokus pada teks saja dan tidak memperhatikan karakter lainnya. Dengan dilakukannya *punctuation removal* diharapkan dapat mempermudah proses klasifikasi. Berikut merupakan implementasi *punctuation removal* yang terdapat pada tabel 3.8.

Tabel 3. 8 Implementasi Punctuation Removal

| Sebelum Implementasi Punctuation Removal | Setelah Implementasi Punctuation Removal |
|--|---|
| kuat banget kak aku dikatain jelek ama org aja udh buat aku gk mau lagi up diri sendiri di sosmed, cukup lama bgt buat gk peduli sama perkataan org wkwk | kuat banget kak aku dikatain jelek ama org aja udh buat aku gk mau lagi up diri sendiri di sosmed cukup lama bgt buat gk peduli sama perkataan org wkwk |

d. Normalization

Normalization merupakan tahapan yang terdapat pada *preprocessing* yang berguna dalam menormalisasi kata yang terdapat pada dataset. Kata yang dinormalisasi dapat berupa bentuk singkatan ataupun *typo* pada penulisan yang biasanya dapat dijumpai pada penulisan di kolom komentar sosial media. Dengan dilakukannya normalisasi, maka akan didapatkan bentuk kata yang lebih

terstruktur sehingga dapat memudahkan dalam proses klasifikasi. Tabel 3.9 dibawah menunjukkan implementasi normalisasi.

Tabel 3. 9 Implementasi *Normalization*

| Sebelum Implementasi Normalization | Setelah implementasi Normalization |
|---|---|
| kuat banget kak aku dikatain jelek ama org aja udh buat aku gk mau lagi up diri sendiri di sosmed cukup lama bgt buat gk peduli sama perkataan org wkwk | kuat banget kak aku dikatai jelek sama orang saja sudah buat aku enggak mau up diri sendiri sosmed cukup lama banget buat enggak peduli sama perkataan orang wkwk |

e. *Stopword removal*

Stopword removal yaitu tahapan pada preprocessing yang bertujuan dalam penghapusan kata yang sering ditemukan pada kolom komentar akan tetapi keberadaannya tidak memberikan pengaruh terhadap proses klasifikasi jika dihapuskan. Hal ini dikarenakan jika kata tersebut dihapus, maka tidak akan memberikan perubahan makna keseluruhan secara signifikan pada kalimat komentar. Berikut merupakan pseudocode yang digunakan dalam menghapus stopwords. Dalam proses ini penulis menambahkan beberapa kata yang perlu dihapus seperti 'anjir', 'bjir', 'wkwk', 'anj', 'hahaha', 'plis', 'woi', 'wkwkwk', 'hehe', 'asu', 'buset', dan 'anjay'. Pada *pseudocode* 3.2 yang tertera dibawah ini dapat dilihat proses penghapusan *stopwords* pada kalimat komentar dan hasilnya secara berurut di tabel 3.10.

```
Function RemoveStopwords (text):
```

```
    more_stopword = ['wkwkwk', 'wkwk', 'bjir', 'wkwk', 'hehe', 'anjay',
    'anjir', 'anj', 'plis', 'woi', 'asu', 'buset', 'hahaha' ]

    default_stopword = StopwordRemoverFactory.get_stop_words()

    combined_stopword = combined(default_stopword, more_stopword)

    dictionary = ArrayDictionary(combined_stopword)

    cleaned_text = stopword_remover.remove(text)

    RETURN cleaned_text
```

```
END FUNCTION
```

Pseudocode 3. 2 Stopword Removal

Tabel 3. 10 Implementasi *Stopword Removal*

| Sebelum Implementasi <i>Stopword Removal</i> | Setelah Implementasi <i>Stopword Removal</i> |
|---|--|
| kuat banget kak aku dikatai jelek sama orang saja sudah buat aku enggak mau up diri sendiri sosmed cukup lama banget buat enggak peduli sama perkataan orang wkwk | kuat banget kak aku dikatai jelek sama orang saja sudah buat aku enggak mau up diri sendiri sosmed cukup lama banget buat enggak peduli sama perkataan orang |

f. *Tokenization*

Tahap tokenisasi, tahapan ini merupakan proses pembagian teks menjadi bentuk unit lebih kecil yang disebut token. Dalam tahap tokenisasi di penelitian ini, penulis menggunakan indobert-base-p2, ini merupakan bentuk dari model BERT yang telah dilatih terlebih dahulu dengan menggunakan korpus Bahasa Indonesia. Kemudian, *IndoBERT* akan memberi token khusus pada setiap kalimat yang hendak ditokenisasi. Token khusus tersebut yaitu [CLS] atau *classification* yang letaknya berada di awal kalimat sebagai penanda, dan [SEP] atau *separator* sebagai penghubung dan berada di akhir kalimat. Dalam penelitian ini tokenisasi dataset dilakukan dengan tahap yang pertama yaitu dengan memastikan token khusus telah ditambahkan pada teks. Kedua, menentukan panjang maksimum kata inputan dimana pada penelitian ini panjang maksimum ditetapkan menjadi 128.

Ketiga, menambahkan *padding* pada token yang bertujuan untuk menyamaratakan panjang teks inputan pada teks yang panjang kalimatnya tidak mencapai panjang maksimum. Keempat, memotong kata pada teks yang memiliki panjang yang melebihi batas maksimum yang telah ditentukan sebelumnya. Kelima, memastikan untuk tidak memberikan informasi mengembalikan token id dikarenakan penelitian ini merupakan klasifikasi teks dengan satu segmen. Keenam, memastikan menambahkan *attention mask* yang berguna untuk menandai kata yang akan digunakan pada klasifikasi. Berikut dapat diperhatikan *pseudocode* 3.3 sebagai gambaran pada proses ini.

```

FUNCTION tokenize_text(text, tokenizer, max_len):
    #Menambahkan token khusus
    add_special_token = True
    #Menentukan panjang maksimum tokenisasi
    maximal_length = max_len
    #Menambahkan padding sesuai panjang maximal length
    padding='max_length'
    #Memotong kata jika kalimat melebihi maximal length
    truncation=True
    #Menentukan tidak mengembalikan token type IDs
    return_token_type_ids=False
    #Menambahkan attention mask untuk menandai token yang digunakan
    return_attention_mask=True
END FUNCTION

```

Pseudocode 3. 3 Tokenization

Pseudocode diatas merupakan kode program yang digunakan untuk mendapatkan hasil tokenisasi yang dipakai untuk memecah teks menjadi bentuk token kata yang selanjutnya akan diubah kedalam nilai numerikal yang akan menjadi representasi dari token kata tersebut. Dibawah ini tertera hasil dari implementasi dari tokenisasi ke dalam bentuk sub-kata pada tabel 3.11 dan implementasi kedalam bentuk numerikal pada tabel 3.12.

Tabel 3. 11 Implementasi *Tokenization*

| Sebelum Implementasi <i>Tokenization</i> | Setelah Implementasi <i>Tokenization</i> |
|--|--|
| kuat banget kak aku dikatai jelek sama orang saja sudah buat aku enggak mau up diri sendiri sosmed cukup lama banget buat enggak peduli sama perkataan orang | ['[CLS]', 'kuat', 'banget', 'kak', 'aku', 'dika', '##ta', '##i', 'jelek', 'sama', 'orang', 'saja', 'sudah', 'buat', 'aku', 'enggak', 'mau', 'up', 'diri', 'sendiri', 'sos', '##med', 'cukup', 'lama', 'banget', 'buat', 'enggak', 'peduli', 'sama', 'perkataan', 'orang', '[SEP]'] |

Tabel 3. 12 Pengubahan Token Menjadi Token id

| Sebelum Diubah Menjadi Index | Setelah Diubah Menjadi Token Id |
|--|---|
| ['[CLS]', 'kuat', 'banget', 'kak', 'aku', 'dika', '##ta', '##i', 'jelek', 'sama', 'orang', 'saja', 'sudah', 'buat', 'aku', 'enggak', 'mau', 'up', 'diri', 'sendiri', 'sos', '##med', 'cukup', 'lama', 'banget', 'buat', 'enggak', 'peduli', 'sama', 'perkataan', 'orang', '[SEP]'] | [2, 1541, 2174, 1844, 304, 28416, 155, 30356, 8302, 500, 232, 377, 259, 968, 304, 6878, 422, 912, 745, 536, 13616, 11633, 724, 985, 2174, 968, 6878, 4625, 500, 7784, 232, 3] |

3.2.2 *Word Embedding*

Pada tahapan ini, data yang sudah diekstraksi selama langkah *preprocessing* nantinya diubah menjadi vektor, pada tahap ini *IndoBERT Embedding* digunakan sebagai *word embedding*. *IndoBERT Embedding* merupakan *contextualized word embedding* yang mana dapat dengan baik menangkap konteks teks dan hubungan antara kata-kata sehingga dapat mengidentifikasi pola yang rumit dan makna ganda dalam suatu kata, hal ini dapat dilakukan dikarenakan *IndoBERT Embedding* merupakan *word embedding* dengan pendekatan *bidirectional* dan menggunakan arsitektur *transformers* dalam melakukan representasi vektor. *IndoBERT* memiliki fitur berbentuk *bidirectional* ini

yang dapat membantu *IndoBERT* dalam memahami teks dari dua arah yang berbeda baik dari sisi kiri ke kanan ataupun dari sisi kanan ke kiri, sehingga dapat mempermudah dalam mengkaji konteks yang ada dalam kalimat. Representasi input dari *embedding* terdiri dari tiga komponen yaitu *token embedding*, *segment embedding*, dan *position embedding* (Devlin et al., n.d.). Pada tahap *token embedding*, inputan teks yang berupa kata tersebut akan ditokenisasi ke dalam bentuk vektor numerik. Kemudian, akan dilanjut pada tahap *segment embedding* yang merupakan tahapan untuk menandai segmen pada inputan teks dan pada penelitian saya yang merupakan klasifikasi teks tunggal maka segmen id akan diberi nilai 0 pada semua token. Setelah itu, tahap *position embedding* merupakan tahapan memberikan index posisi pada urutan token.

Penelitian ini menggunakan *indobert-base-p2* sebagai *word embedding* yang didapat melalui pustaka *huggingface*. Hasil dari embedding *IndoBERT* dinamakan *last hidden state* yang mana lapisan ini nantinya yang akan digunakan berikutnya dalam tugas pengklasifikasian pada model, pada lapisan ini tensor akhir yang dihasilkan memiliki tiga dimensi yaitu ukuran *batch*, panjang urutan token pada input teks, dan *hidden size*. Dimana *hidden size* merupakan dimensi yang digunakan untuk merepresentasikan setiap token, untuk *IndoBERT* sendiri terdapat 768 dimensi yang nantinya akan diterapkan pada setiap token.

3.2.3 Convolutional Neural Network

Di bagian ini, dalam penelitian dilakukan pengembangan model yang akan memanfaatkan algoritma *Convolutional Neural Network* (CNN). CNN yakni algoritma dengan berbasis arsitektur jaringan saraf yang menghubungkan setiap neuron atau saraf yang dihasilkan dari setiap lapisan yang mana hasilnya akan digunakan dalam operasi pengenalan pola dan untuk pengambilan keputusan dalam pembelajaran mesin.

Pada penelitian ini terdapat beberapa lapisan yang digunakan untuk pemrosesan algoritma *Convolutional Neural Network*. Berikut merupakan penjelasan terhadap lapisan-lapisan pada tersebut.

1. Lapisan *Convolutional*

Lapisan ini merupakan lapisan yang digunakan untuk menangkap pola penting seperti urutan kata dalam teks input hasil dari *word embedding*. Dimana, hal ini dilakukan dengan adanya penerapan kernel dan juga filter pada inputan yang

mempengaruhi hasil lapisan ini pada model dalam memperhatikan urutan kata yang akan diproses. Pada lapisan ini juga ditambahkan aktivasi *ReLU* (*Rectified Linear Unit*) yang berguna untuk mengatasi permasalahan non-linier pada data sehingga algoritma dapat lebih baik dalam memahami pola yang lebih kompleks.

2. Lapisan *Pooling*

Lapisan pooling adalah lapisan yang diaplikasikan setelah melalui lapisan konvolusi, lapisan ini digunakan untuk menyederhanakan ukuran data dengan hanya mengambil nilai tertinggi yang paling signifikan dari hasil lapisan konvolusi.

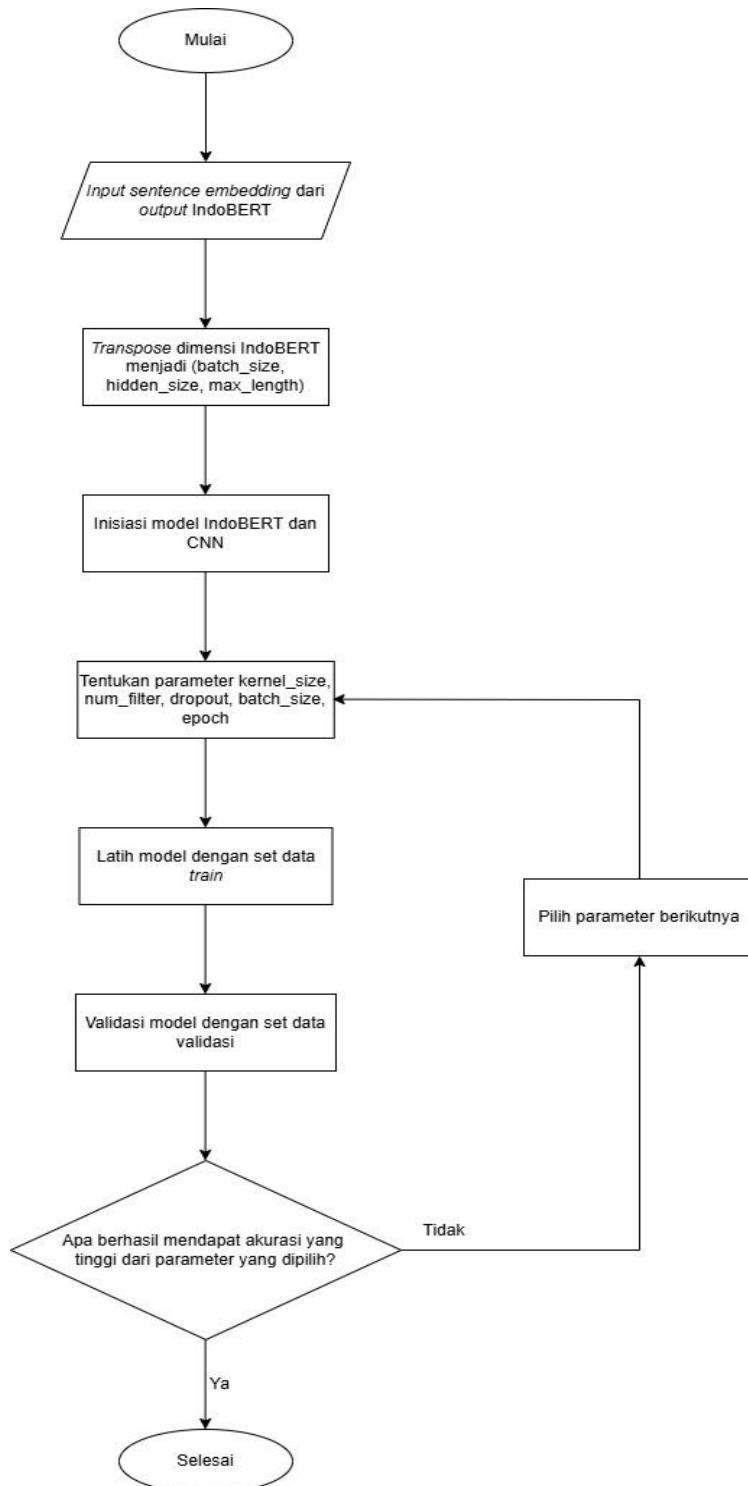
3. Lapisan *Dropout*

Lapisan dropout merupakan teknik regularisasi dimana sebagian jaringan saraf kemudian ditetapkan dengan random dan tidak akan dipergunakan pada saat langkah *training*, hal tersebut dilakukan agar jaringan saraf tidak tergantung pada subset tertentu. Selain itu, *dropout* dapat membantu dalam mencegah permasalahan *overfitting*, dikarenakan pada lapisan *dropout* akan terjadi proses *drop* yang mematikan *neuron* secara acak agar model menjadi tidak tergantung pada saraf tertentu dan dengan ini model dapat bekerja lebih baik pada data yang baru.

4. Lapisan *Fully Connected*

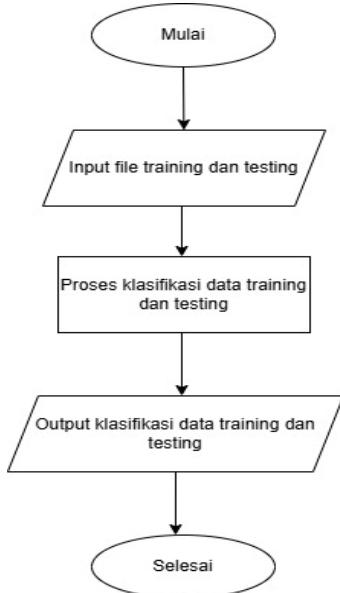
Lapisan *fully connected* merupakan lapisan untuk menggabungkan semua jaringan saraf yang beraktivitas pada lapisan sebelumnya dengan jaringan saraf pada lapisan selanjutnya. Hasil dari lapisan ini nantinya berupa logit yang merupakan nilai mentah yang nantinya akan digunakan sebagai inputan untuk fungsi lain yang diterapkan pada lapisan *fully connected*. Logit akan diproses oleh fungsi *softmax* berperan untuk mendapatkan probabilitas dari model untuk setiap kategori kelas. Setelah mendapatkan nilai probabilitas dari *softmax*, selanjutnya diterapkan fungsi *argmax* yang berperan untuk memilih nilai probabilitas tertinggi dan menentukan kelas prediksi. Pada lapisan *fully connected* ini juga ditambahkan fungsi *CrossEntropyLoss* yang merupakan fungsi *loss* dimana fungsi ini digunakan untuk meminimalisir *error* oleh model terhadap data label sebenarnya dengan estimasi perkiraan kelas yang diprediksi oleh model.

Gambar 3.2 yang diperlihatkan berikut merupakan gambar alur pelatihan yang digunakan dalam pengembangan model.



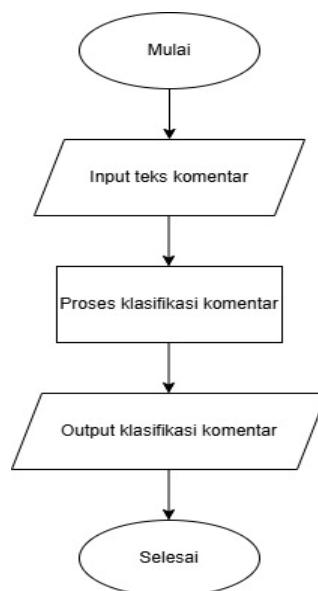
Gambar 3.2 Diagram Alur Pelatihan Model

Berikut pada gambar 3.3 merupakan gambar diagram alur aplikasi pada halaman menu *training* dan *testing*. Berdasarkan gambar, proses *training* dan *testing* dilakukan dengan input file *training* dan *testing*, kemudian sistem akan melakukan proses klasifikasi pada data *training* dan *testing*, lalu di dapat output yaitu hasil klasifikasi data *training* dan *testing*.



Gambar 3. 3 Diagram Alur Aplikasi pada Menu *Training* dan *Testing*

Berikut pada gambar 3.4 merupakan diagram alur aplikasi pada halaman menu *input user*. Berdasarkan gambar proses dimulai dengan input teks komentar, kemudian klasifikasi komentar, dan hasil dari klasifikasi komentar.



Gambar 3. 4 Diagram Alur Aplikasi pada Menu *Input User*

Berikut merupakan penerapan model *Convolutional Neural Network* dan *IndoBERT embedding* yang dapat dilihat pada *pseudocode* 3.4 dibawah ini.

```

FUNCTION BertWithCNN(model_name, num_labels, kernel_size, num_filters,
dropout):
    LOAD pretrained BERT model for embeddings
    INITIALIZE CNN layer with kernel_size and num_filters
    INITIALIZE Dropout layer with rate `dropout`
    INITIALIZE Fully Connected layer with output `num_labels`
    END FUNCTION

FUNCTION forward(input_ids, attention_mask):
    EMBEDDING = BERT(input_ids, attention_mask) → last_hidden_state
    EMBEDDING = TRANSPOSE dimensions to (batch_size, hidden_size,
seq_len)
    FEATURES = APPLY CNN → ReLU → MaxPooling
    APPLY Dropout
    LOGITS = Fully Connected Layer(FEATURES)
    LOSS = CrossEntropyLoss(LOGITS, labels)
    RETURN LOGITS

END FUNCTION
#Melakukan prediksi
PROBABILITIES = APPLY softmax(LOGITS ALONG dim=1)
PREDICTED_LABELS = FIND argmax(PROPBABILITY ALONG axis=1)

```

Pseudocode 3. 4 Model CNN

Penjelasan lebih rinci terhadap proses operasional pelatihan model akan diuraikan sebagai berikut.

1. Tensor akhir yang dihasilkan oleh *IndoBERT embedding* memiliki bentuk dimensi (batch_size, max_length, hidden_size) dimana bentuk perlu diubah posisinya agar sesuai untuk bentuk input yang dapat diolah oleh model *Convolutional Neural Network* yaitu dengan bentuk (batch_size, hidden_size, max_length).
2. Setelahnya, dilakukan inisiasi model *Convolutional Neural Network* yang dimulai dari lapisan konvolusi yang selama lapisan ini diterapkan aktivasi ReLU, lapisan pooling, lapisan dropout, dan terakhir lapisan *fully connected*.
3. Menentukan parameter terbaik untuk diterapkan di kernel_size, num_filter, dropout, batch_size, dan epoch. Tahapan ini dilakukan untuk mengukur skor akurasi, presisi, *f1-score*, dan *recall* terbaik yang dapat dihasilkan oleh model.

4. Melakukan pelatihan terhadap model dengan menggunakan parameter yang telah ditetapkan sebelumnya pada set data *training* untuk menjadi basis pembelajaran terhadap pola data sehingga dapat digunakan untuk proses prediksi pada data yang baru.
5. Melakukan evaluasi model dengan menggunakan set data validasi yang bertujuan untuk mengevaluasi kinerja dari model yang diaplikasikan terhadap data baru yang belum pernah ada dalam pelatihan sehingga dapat diketahui kemampuan model dalam generalisasi data dengan baik.
6. Model ini kemudian dapat digunakan dalam pengambilan keputusan dalam tugas klasifikasi pada komentar mengandung *body shaming* terhadap bentuk badan, bentuk wajah, warna kulit, dan non-*body shaming*.

3.2.4 Output

Hasil dari proses-proses yang telah diterapkan diatas adalah pembagian komentar *body shaming* menjadi empat kategori, diantaranya komentar *body shaming* berdasarkan bentuk badan, komentar *body shaming* berdasarkan bentuk wajah, komentar *body shaming* berdasarkan warna kulit, dan komentar mengandung non-*body shaming*.

3.3 Perancangan Sistem

Tahapan perancangan sistem merupakan tahapan menguraikan secara terperinci mengenai konsep antarmuka yang akan diimplementasikan pada sistem untuk mengakomodasi kebutuhan pengguna. Perancangan sistem ini akan ditampilkan dengan hasil akhir berupa sebuah platform yang berbasis *web*. Proses pada sistem ini berfokus pada klasifikasi komentar mengandung *body shaming* pada media sosial dengan menggunakan algoritma CNN. Perancangan sistem ini dimaksudkan untuk memvisualisasikan sistem serta operasinya secara menyeluruh yang dapat digunakan sebagai panduan yang dapat bermanfaat bagi pengguna.

3.3.1 Rancangan Tampilan Halaman Beranda

Halaman beranda ialah halaman depan yang akan ditampilkan kepada *users* ketika telah mengakses laman web. Rincian halaman beranda meliputi beragam detail keterangan yang mencakup judul penelitian, identitas penulis, pilihan tombol menu yang akan menuntun pengguna ke halaman-halaman berikutnya yang disediakan oleh sistem, dan informasi mengenai *body shaming* beserta jenis-jenis *body shaming*. Rancangan tampilan halaman beranda akan dimuat dalam gambar 3.5.



Gambar 3. 5 Halaman Beranda

Keterangan:

- Label 1 pada gambar merupakan tombol yang menunjukkan tombol home, tombol ini merupakan tombol yang mengarah ke halaman beranda.
- Label 2 pada gambar merupakan tombol yang menujukkan tombol *training* dan *testing*, tombol ini merupakan tombol yang mengarah ke halaman *training* dan *testing*.
- Label 3 pada gambar merupakan tombol yang menunjukkan tombol *input user*, tombol ini merupakan tombol yang mengarah ke halaman *input user*.

3.3.2 Rancangan Tampilan Halaman *Training* dan *Testing*

Halaman ini ialah halaman operasi pelatihan dan pengujian terhadap model, proses pelatihan dan pengujian tersebut dapat dilakukan dengan mengunggah file dataset melalui format .xlsx. Untuk memulai proses pelatihan pengguna dapat mengklik tombol "Choose file" sebagai langkah awal untuk mengunggah *file* pelatihan. Setelah berhasil, maka akan didapat hasil proses pelatihan yaitu tampilan data sebelum dan setelah proses *preprocessing* serta berlangsungnya proses pelatihan. Selanjutnya, untuk proses pengujian pengguna dapat menguji model dengan mengklik tombol "Choose file" terlebih dahulu untuk mengunggah *file* pengujian. Sesudah sukses dilakukan, akan didapat hasil proses pengujian berupa tampilan data awal dan setelah preprocessing serta hasil prediksi. Untuk mempermudah pengguna hasil dari proses pelatihan dan pengujian akan dimuat dalam bentuk tabel sehingga dapat membantu dalam membandingkan hasil dari setiap operasi sistem pada model ini. Selain itu, dalam halaman ini juga akan ditampilkan hasil *confusion matrix* dari proses pengujian yang meliputi akurasi, *precision*, *recall*, *f1-score*. Melalui gambar 3.6 berikut dapat diamati desain tampilan halaman *training* dan *testing*.

Keterangan:

- Label 1 merupakan tombol yang digunakan untuk memilih dokumen *training* yang ingin diunggah untuk tahap pelatihan.
- Label 2 merupakan tombol yang digunakan untuk memilih dokumen *testing* yang ingin diunggah untuk tahap pengujian.
- Label 3 merupakan tombol yang digunakan untuk memulai proses *training-testing* pada *file* yang telah diunggah.
- Label 4 merupakan tabel hasil dari file data *training* yang diunggah.
- Label 5 merupakan tabel hasil dari file data *testing* yang diunggah.
- Label 6 akan menampilkan hasil akurasi yang dimuat dalam bentuk *confusion matrix*.

TUGAS AKHIR

Home Training & Testing User Input

Training dan Testing Data

File Training
1 Choose File

File Testing
2 Choose File

Analysis
3

Training File 4

| Teks Awal | Teks Setelah Preprocessing | Label |
|-----------|----------------------------|-------|
| | | |

Testing File 5

| Teks Awal | Teks Setelah Preprocessing | Label | Hasil Prediksi |
|-----------|----------------------------|-------|----------------|
| | | | |

Akurasi 6

| Confussion Matrix | | | |
|-------------------|--|--|--|
| | | | |

FOOTER

Gambar 3. 6 Halaman *Training dan Testing*

3.3.3 Rancangan Tampilan Halaman *Input User*

Halaman ini ialah halaman yang dapat dimanfaatkan oleh *users* untuk melakukan proses klasifikasi *body shaming* dengan bentuk data tunggal (*single*). Pada halaman ini

pengguna hanya perlu menginput satu data komentar Bahasa Indonesia. Setelah menerima input komentar tersebut, sistem akan mulai mengklasifikasikan komentar. Melalui gambar 3.7 ditampilkan detail rancangan halaman *Input User*.

| Teks | Hasil Preprocessing | Hasil Prediksi |
|------|---------------------|----------------|
| | | |

Gambar 3. 7 Halaman *Input User*

Keterangan:

- Label 1 merupakan wadah yang disediakan untuk menginput teks komentar yang ingin diklasifikasikan oleh sistem.
- Label 2 adalah tombol yang dipakai untuk mengawali proses klasifikasi pada data komentar.
- Label 3 merupakan tabel hasil prediksi data tunggal yang di input.

3.4 Metode Evaluasi

Metode penilaian dalam evaluasi model yang dipakai adalah menggunakan Matriks Kebingungan (*Confusion Matrix*). Pengevaluasian pada penelitian ini perlu dilakukan dengan tujuan untuk menilai seberapa baik kemampuan model dalam melakukan klasifikasi komentar berisi *body shaming* dan mengelompokkannya berdasarkan jenis *body shaming* tersebut. Evaluasi tersebut diukur melalui perhitungan dengan membandingkan hasil prediksi yang dilakukan oleh model dengan label data asli. Melalui *confusion matrix*, akan didapat rincian kemampuan model dalam mengukur tingkat kesalahan yang dilakukan. Dalam *confusion matrix* terdapat empat parameter

yang digunakan sebagai standar penilaian yaitu *accuracy*, *recall*, *precision*, dan *F-1 Score*. Perhitungan dengan *confusion matriks* dapat diperhatikan dalam tabel 3.13 yang disajikan sebagai berikut.

Tabel 3. 13 Perhitungan *Confusion Matrix*

| Klasifikasi | Nilai Prediksi Positif | Nilai Prediksi Negatif |
|-----------------------------|-------------------------------|-------------------------------|
| Nilai Aktual Positif | TP (True Positive) | FN (False Negative) |
| Nilai Aktual Negatif | FP (False Negative) | TN (True Negative) |

Keterangan:

TP = Kuantitas data yang bernilai positif pada hasil prediksi oleh model dengan kelas aktual yang bernilai positif.

FN = Kuantitas data yang bernilai negatif pada hasil prediksi oleh model dengan kelas aktual yang bernilai positif.

FP = Kuantitas data yang bernilai positif pada hasil prediksi oleh model dengan kelas aktual yang bernilai negatif.

TN = Kuantitas data yang bernilai negatif pada hasil prediksi oleh model dengan kelas aktual yang bernilai negatif.

BAB 4

IMPLEMENTASI DAN PENGUJIAN SISTEM

4.1 Implementasi Sistem

Terkait dengan membangun sistem untuk klasifikasi komentar mengandung *body shaming* pada media sosial menggunakan algoritma *Convolutional Neural Network*, diperlukan dukungan dari perangkat lunak (*software*) dan perangkat keras (*hardware*) sebagai komponen penunjang dan komponen tersebut diantaranya, yaitu:

4.1.1 Spesifikasi Perangkat Keras dan Perangkat Lunak

Spesifikasi teknis *hardware* yang dimanfaatkan dalam mengembangkan sistem pada penelitian yang dilaksanakan, yakni:

1. Laptop Acer Swift SF314-41
2. Processor AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx 2.10 GHz
3. Daya tampung memori 8 GB RAM
4. Daya tamping memori SSD 512 GB

Adapun spesifikasi dari *software* yang dipakai dalam membentuk sistem pada penelitian yang dilaksanakan diuraikan, yaitu:

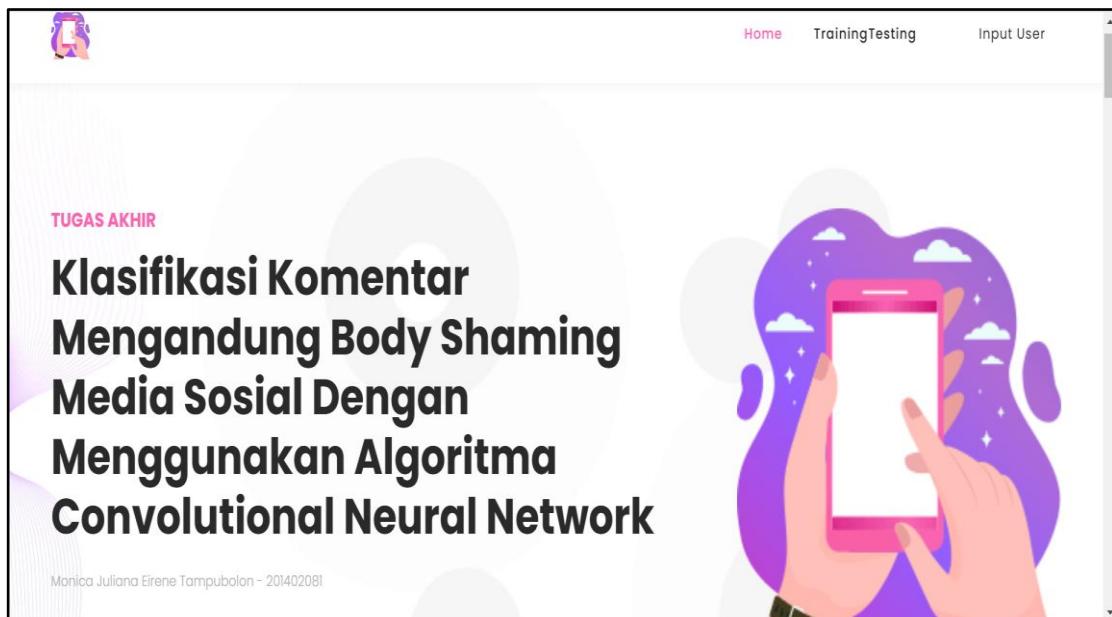
1. OS : *Windows 11 Home Single Language*
2. Bahasa Pemrograman *Python* versi 3.9.13
3. Perpustakaan (*library*) dalam permrograman *python*, mencakup:
 - *torch* versi 2.5.1
 - *transformers* versi 4.46.2
 - *indonlp* versi 0.2.0
 - *sastrawi* versi 1.0.1
4. IDE : *Microsoft Visual Studio Code*

4.1.2 Penerapan Rancangan Tampilan Antarmuka Pengguna

Penerapan rancangan tampilan antarmuka sistem berbasis web yang sudah diterangkan pada bab sebelumnya adalah berikut:

1. Tampilan Halaman Beranda

Laman beranda adalah laman tampilan awal yang dijangkau oleh pengguna ketika mengawali menjalankan sistem. Dalam halaman ini, disertakan informasi umum tentang penelitian, yang diantaranya berupa judul penelitian, identitas penulis, dan pengertian tentang *body shaming* berserta jenisnya. Selain hal itu, di halaman ini telah disediakan tombol navigasi yang dapat menuntun para pengguna untuk dapat mengunjungi menu lainnya yang disediakan oleh sistem. Tampilan halaman beranda yang telah dirancang dapat diperhatikan dalam gambar 4.1.

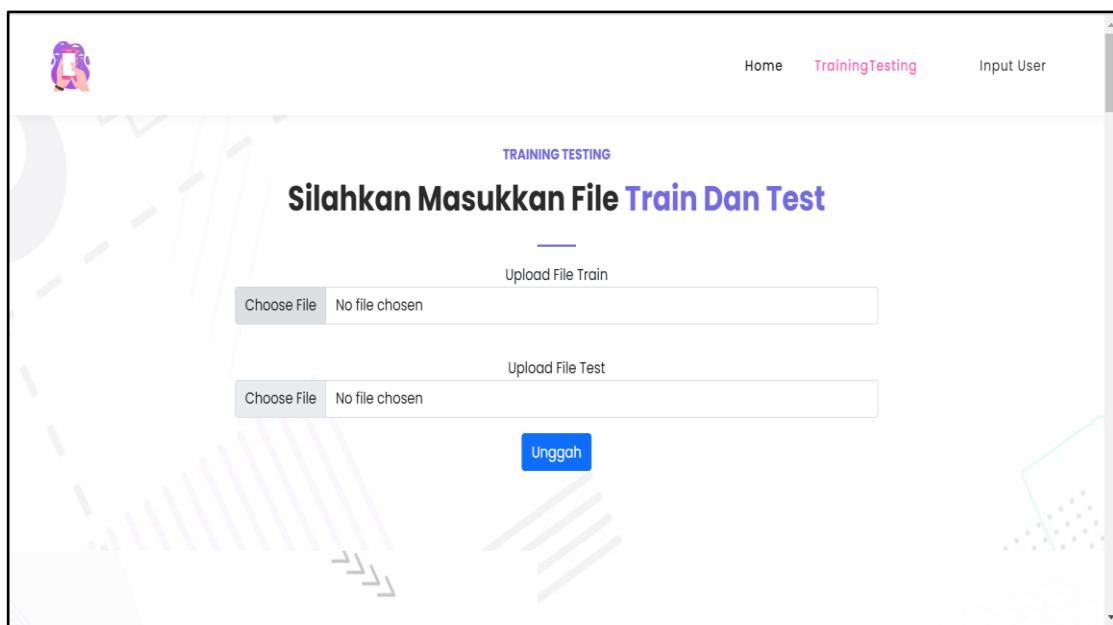


Gambar 4. 1 Tampilan Halaman Beranda

2. Tampilan Halaman *Training* dan *Testing*

Halaman *training* dan *testing* merupakan bagian yang memfasilitasi pengguna untuk dapat mengunggah *file training* dan *file testing* yang akan dimanfaatkan untuk menjalankan klasifikasi *body shaming*. Bentuk dokumen yang dapat diunggah pada bagian *training* dan *testing* berupa dokumen

dengan format xlsx. Selanjutnya setelah mengunggah dokumen, maka proses *training* dan *testing* baru akan dapat berjalan. Setelah proses selesai dilakukan, berikutnya akan didapatkan hasil dari proses *training* dan *testing* beserta *classification report* dan *confusion matrix*. Pada hasil proses training akan ditampilkan tabel yang berisi teks sebelum di teks *preprocessing*, teks sesudah *preprocessing*, label, dan hasil klasifikasi. Kemudian, untuk hasil proses *testing* akan ditampilkan tabel yang berisi teks keseluruhan, label, dan hasil prediksi kelas. Selain itu, sistem juga memberikan *pagination* dan *filter* yang dapat mempermudah pengguna mencari hasil prediksi. Tampilan halaman *training* dan *testing* dapat diamati pada gambar 4.2. Pada gambar 4.3 menunjukkan tampilan halaman hasil data pelatihan, dan pada gambar 4.4 menunjukkan tampilan halaman hasil data uji. Gambar 4.5 merupakan tampilan *classification report* dan gambar 4.6 merupakan gambar tampilan *confusion matrix* hasil data *testing*.



Gambar 4. 2 Tampilan Halaman *Training* dan *Testing*

Berikut pada gambar 4.3 merupakan tampilan hasil file *.xlsx training* pada halaman *training testing* yang menampilkan teks komentar asli, hasil teks setelah *preprocessing*, dan label yang diberikan.

| Hasil Training | | | |
|-----------------------|--|---|-------|
| Show 5 entries | Search: | | |
| No | Teks Awal | Teks Setelah Preprocessing | Label |
| 1 | MUNGKIN KONTOLNYA ARHAN KECIL, KALO KONTOLNYA SALIN GEDE JADI ZIZAH LEBIH PUAS SECARA DIA ARABKAN ASUMSI KU GITU | mungkin kontolnya arhan kecil kalau kontolnya salin besar jadi zizah lebih puas arabkan asumsi ku | 1 |
| 2 | Kamu terlalu tinggi, jadi kelihatan aneh di antara teman-temanmu. | kamu terlalu tinggi jadi kelihatan aneh teman temanmu | 1 |
| 3 | AKu pake sering pake masker putih telur dulu di muka menurutku ga ada efeknya sama sekali sih | aku pakai sering pakai masker putih telur dulu muka menurutku enggak efeknya sama sekali sih | 0 |
| 4 | Berhijab tapi pake baju masih kependekan gitu mending ga usah pake aja sekalian | berhijab pakai baju kependekan mending enggak usah pakai sekalian | 0 |
| 5 | maaf aja ya kak tapi ga ada cantiknya menurutku kulit seputat ini | maaf kak enggak cantiknya menurutku kulit seputat ini | 3 |

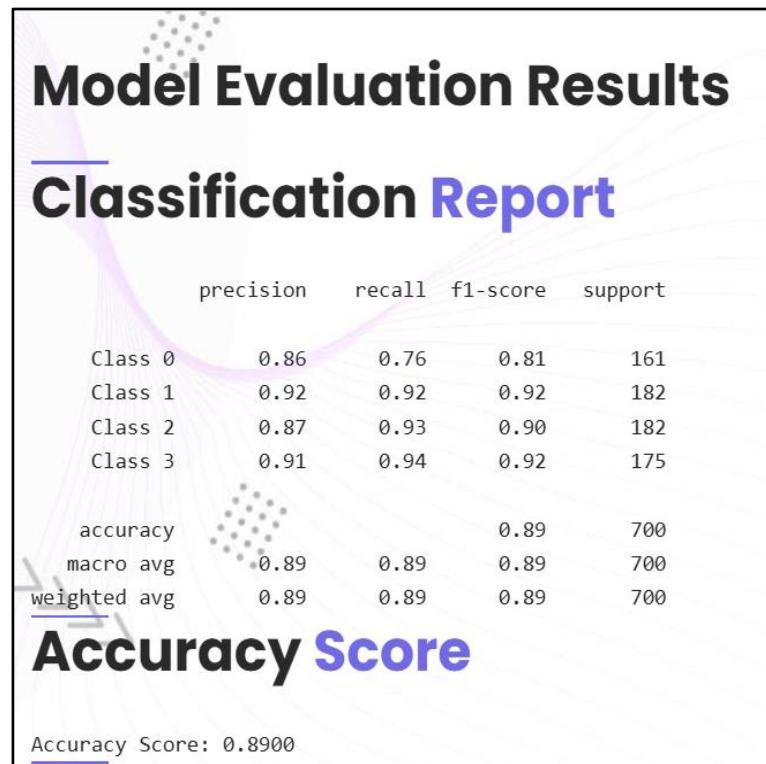
Gambar 4. 3 Tampilan Halaman Hasil *Training*

Berikut pada gambar 4.4 merupakan tampilan hasil file *.xlsx testing* pada halaman *training testing* yang menampilkan teks komentar asli, hasil teks setelah melalui *preprocessing*, label sebenarnya, dan hasil prediksi oleh sistem.

| Hasil Testing | | | |
|----------------------|--|--|----------------------------|
| Show 5 entries | Search: | | |
| No | Teks Awal | Teks Setelah Preprocessing | True Label Predicted Label |
| 1 | Semangat terus yaa, konsisten terus. smpe jadi versi terbaik dr diri kamu | semangat terus konsisten terus jadi versi terbaik diri kamu | 0 0 |
| 2 | lucu sekali alisnya cocok deh sama kamu | lucu sekali alisnya cocok deh sama kamu | 0 0 |
| 3 | yg model kek bapak itu Banyak banget kita jumpain | model kayak bapak banyak banget jumpain | 0 1 |
| 4 | tapi memang bener loh kamu itu pundak kecil kepala besar | memang benar loh kamu pundak kecil kepala besar | 1 1 |
| 5 | Ya orang Indonesia kan bogel2 karena gizi buruk. Paling bogel sedunia malah wkwkwk | orang indonesia kan bogel gizi buruk paling bogel sedunia malah wkwkwk | 1 1 |

Gambar 4. 4 Tampilan Halaman Hasil *Testing*

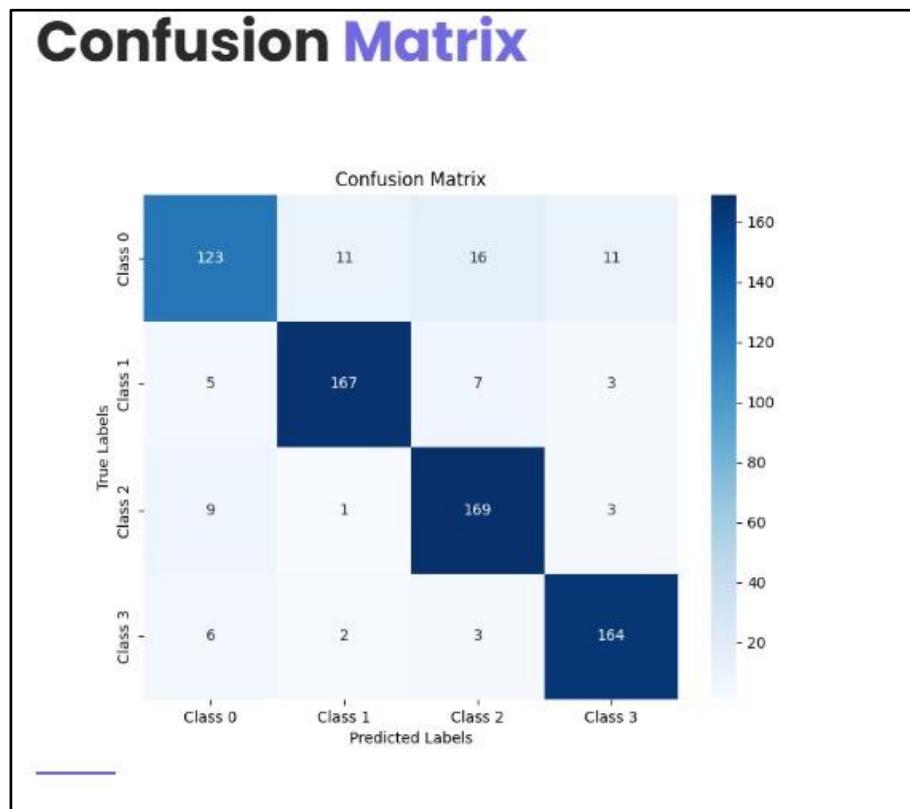
Berikut pada gambar 4.5 merupakan tampilan hasil model evaluasi yang memperlihatkan nilai *precision*, *recall*, *f-1 score*, dan *support* pada setiap kelas. Dimana, *class 0* untuk komentar non-*body shaming*, *class 1* untuk komentar *body shaming* bentuk badan, *class 2* untuk komentar *body shaming* berdasarkan bentuk wajah, dan *class 3* untuk komentar *body shaming* warna kulit.



Gambar 4.5 Tampilan Halaman *Model Evaluation*

Berikut pada gambar 4.6 merupakan tampilan hasil model evaluasi yang memperlihatkan *class 0* untuk komentar non-*body shaming*, *class 1* untuk komentar *body shaming* bentuk badan, *class 2* untuk komentar *body shaming* berdasarkan bentuk wajah, dan *class 3* untuk komentar *body shaming* warna kulit. Pada gambar ini menyatakan *true labels* sebagai label yang telah diberikan sebelumnya, dan *predicted labels* merupakan label hasil sistem. Tertera pada kelas 0 yang dilabel asli terdapat 123 komentar yang diprediksi sistem dengan benar di kelas 0, 11 komentar diprediksi sebagai kelas 1, 16 komentar diprediksi sebagai kelas 2, dan 11 komentar diprediksi sebagai kelas 3. Pada kelas 1 pada kelas 1 yang dilabel asli terdapat 167 komentar yang diprediksi sistem dengan benar di kelas 1, 5 komentar diprediksi sebagai kelas 0, 7 komentar diprediksi sebagai kelas 2, dan 3 komentar diprediksi

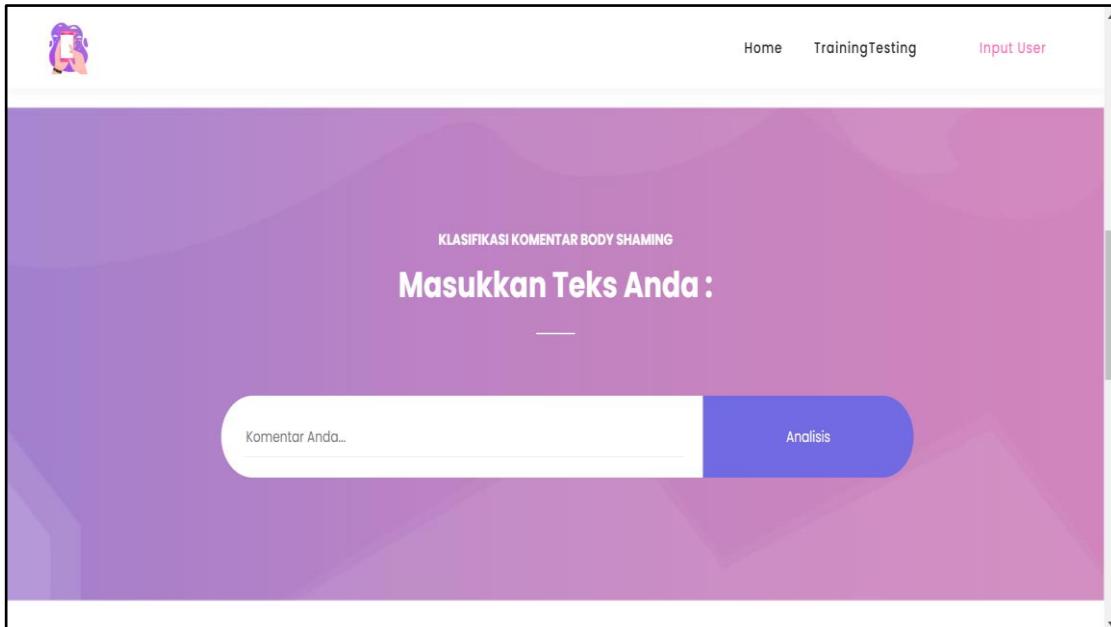
sebagai kelas 3. Lalu, pada kelas 2 yang dilabel asli terdapat 169 komentar yang diprediksi sistem dengan benar di kelas 2, 9 komentar diprediksi sebagai kelas 0, 1 komentar diprediksi sebagai kelas 1, dan 3 komentar diprediksi kelas 3. Kemudian, pada kelas 3 yang dilabel asli terdapat 164 komentar yang diprediksi sistem dengan benar di kelas 3, 6 komentar diprediksi sebagai kelas 0, 2 komentar diprediksi sebagai kelas 1, dan 3 komentar diprediksi sebagai kelas 2.



Gambar 4. 6 Tampilan Halaman *Confusion Matrix*

3. Tampilan Halaman *Input User*

Laman *input user* adalah halaman yang memfasilitasi pengguna untuk dapat menginput teks komentar yang ingin diklasifikasi. Melalui *textbox* yang telah disediakan, pengguna dapat memasukkan teks komentar yang berencana diklasifikasi. Setelah menginput teks, pengguna kemudian akan mendapatkan hasil klasifikasi dari komentar tersebut berupa tabel yang berisi teks komentar asli, teks setelah *preprocessing* dan hasil prediksi oleh sistem. Adalpun gambar 4.7 adalah tampilan halaman *input user* yang telah dikembangkan.



Gambar 4. 7 Tampilan Halaman *Input User*

4.2 Pelatihan Model

Penelitian yang dilakukan ini menerapkan algoritma *Convolutional Neural Network* untuk mengoptimalkan kinerja dalam melakukan klasifikasi. Dalam melakukan pelatihan model, diperlukan berbagai eksperimen terhadap parameter guna mendapatkan hasil yang terbaik. Variasi jenis parameter yang digunakan untuk mendapatkan hasil terbaik pada penelitian ini diantaranya adalah dengan *filter_size*, *kernel_size*, *batch_size*, *dropout*, dan *epoch*. Dengan melakukan hyperparameter tuning pada parameter tersebut, akan didapatkan kombinasi parameter yang dapat memberikan performa terbaik. Hasil akurasi dari setiap pengujian parameter akan disajikan pada tabel 4.1.

Tabel 4. 1 Hasil Akurasi Pengujian Parameter

| No. | Filter_Size | Kernel_Size | Batch_Size | Dropout | Epoch | Akurasi |
|-----|-------------|-------------|------------|---------|-------|---------|
| 1. | 100 | 3 | 16 | 0,3 | 5 | 87,00 |
| 2. | 100 | 3 | 16 | 0,5 | 5 | 87,43 |
| 3. | 100 | 3 | 32 | 0,3 | 5 | 84,00 |

| | | | | | | |
|------------|-----|---|----|-----|----|-------|
| 4. | 100 | 3 | 32 | 0,5 | 5 | 84,86 |
| 5. | 100 | 5 | 16 | 0,3 | 5 | 87,70 |
| 6. | 100 | 5 | 16 | 0,5 | 5 | 87,88 |
| 7. | 100 | 5 | 32 | 0,3 | 5 | 86,90 |
| 8. | 100 | 5 | 32 | 0,5 | 5 | 87,71 |
| 9. | 200 | 3 | 16 | 0,3 | 5 | 88,10 |
| 10. | 200 | 3 | 16 | 0,5 | 5 | 87,57 |
| 11. | 200 | 3 | 32 | 0,3 | 5 | 86,29 |
| 12. | 200 | 3 | 32 | 0,5 | 5 | 85,86 |
| 13. | 200 | 5 | 16 | 0,3 | 5 | 88,29 |
| 14. | 200 | 5 | 16 | 0,5 | 5 | 88,42 |
| 15. | 200 | 5 | 32 | 0,3 | 5 | 87,29 |
| 16. | 200 | 5 | 32 | 0,5 | 5 | 88,20 |
| 17. | 100 | 3 | 16 | 0,3 | 10 | 88,43 |
| 18. | 100 | 3 | 16 | 0,5 | 10 | 88,14 |
| 19. | 100 | 3 | 32 | 0,3 | 10 | 88,57 |
| 20. | 100 | 3 | 32 | 0,5 | 10 | 89,00 |
| 21. | 100 | 5 | 16 | 0,3 | 10 | 88,47 |
| 22. | 100 | 5 | 16 | 0,5 | 10 | 88,66 |
| 23. | 100 | 5 | 32 | 0,3 | 10 | 88,10 |
| 24. | 100 | 5 | 32 | 0,5 | 10 | 88,16 |
| 25. | 200 | 3 | 16 | 0,3 | 10 | 88,58 |
| 26. | 200 | 3 | 16 | 0,5 | 10 | 88,72 |

| | | | | | | |
|------------|-----|---|----|-----|----|-------|
| 27. | 200 | 3 | 32 | 0,3 | 10 | 87,50 |
| 28. | 200 | 3 | 32 | 0,5 | 10 | 85,71 |
| 29. | 200 | 5 | 16 | 0,3 | 10 | 88,60 |
| 30. | 200 | 5 | 16 | 0,5 | 10 | 88,76 |
| 31. | 200 | 5 | 32 | 0,3 | 10 | 87,86 |
| 32. | 200 | 5 | 32 | 0,5 | 10 | 87,71 |

Akurasi yang ditampilkan pada tabel 4.1 yang berada diatas merupakan rata-rata pada beberapa kombinasi parameter yang diuji pada penelitian ini, dari hasil ini didapatkan perolehan kesimpulan bahwa kombinasi yang optimal pada *Convolutional Neural Network* adalah dengan menggunakan ukuran filter senilai 100, ukuran kernel senilai 3, ukuran batch senilai 32, dropout senilai 0,5, dan epoch yang digunakan senilai 10. Kombinasi dari setiap parameter ini memberikan hasil akurasi yaitu mencapai 89,00% pada klasifikasi *body shaming*.

Tabel 4.2 berikut menampilkan kombinasi parameter yang dengan hasil akurasi terbaik pada gabungan *IndoBERT word embedding* dengan algoritma *Convolutional Neural Network*.

Tabel 4. 2 Kombinasi Parameter Terbaik

| Parameter | Nilai |
|-------------|-------|
| Filter Size | 100 |
| Kernel Size | 3 |
| Batch Size | 32 |
| Dropout | 0,5 |
| Epoch | 10 |

4.3 Hasil Pengujian Sistem

Pada bagian ini, dijalankan tahapan menguji kinerja dari model yang sudah dibuat dengan menggunakan data pelatihan dan data uji untuk mengevaluasi kemampuan model *Convolutional Neural Network* dalam melakukan klasifikasi *body shaming*. Pengujian model ini menggunakan sebanyak 700 data komentar, dimana data ini akan melalui tahapan *pre-processing* dan *word embedding* IndoBERT yang juga diterapkan pada data pelatihan, sebelum akhirnya akan diuji menggunakan model *Convolutional Neural Network* dalam melakukan prediksi. Pengujian kinerja tersebut akan ditampilkan pada tabel 4.3.

Tabel 4. 3 Hasil Pengujian Sistem Pada Data *Testing*

| No | Komentar | Label | Hasil Prediksi |
|----|---|-------|----------------|
| 1. | Semangat terus yaa, konsisten terus. smpe jadi versi terbaik dr diri kamu | 0 | 0 |
| 2. | Ya orang Indonesia kan bogel2 karena gizi buruk. Paling bogel sedunia malah wkwkwk | 1 | 1 |
| 3. | salah warna foundation 🎨 jadi abu abu serem gtu mukanya | 3 | 3 |
| 4. | Knp mesti gigit bibir sih 😔, | 0 | 2 |
| 5. | makanya punya daging bg, bukan kulit doang | 1 | 3 |
| 6. | Percaya diri penting tapi sadar diri jauh lebih penting, udah tau badannya begitu mbok ya gausah diumbar kak, lebih tertutup lebih baik kok | 1 | 0 |
| 7. | Kayaknya abis itu, idung deh yg perlu dipermak, kan duitnya masih bnyk yg nganggur biar sekalian dah, supaya bagus dikit hidungnya | 2 | 2 |
| 8. | Pernah dikatain item sama temen gua jawab aja "oh waktu pembagian warna di pabrik kebagian dark brown soalnya warna creamnya udah abis gatau dah diambil siapa" 😔—; | 0 | 0 |

| | | | |
|------|--|---|---|
| 9. | bersiin dulu itu banyak bat daki prindapan 😂😂 | 3 | 3 |
| 10. | Orang orang enak ya pulang ke rumah matahari masih keliatan lah aku pulang langitnya dah gelap banget | 0 | 3 |
| 11. | jadi teringat sama mayang jilbab abu abu | 0 | 3 |
| . | . | . | . |
| . | . | . | . |
| . | . | . | . |
| 692. | Lihat aslinya bongsor banget beda di instagram liatnya imut begitu jauh sama real nya | 1 | 1 |
| 693. | Penting untuk memahami bahwa obesitas adalah masalah kesehatan, dan saya senang melihat kamu mengambil langkah-langkah positif untuk mengatasinya. | 0 | 1 |
| 694. | kuning langsat tuh kulit nya klo diliat agak ke kuning2an gitu ga si?soalnya tgn gw kek gitu suka berubah ² warnanya kdang cerah kadang gelap trgntung pencahayaan | 0 | 3 |
| 695. | bjir dah kulit gelap ini perkara bentuk mata aja dah merasa cindo | 3 | 3 |
| 696. | makan bang tuh ngaca tinggal kulit aja lo | 1 | 3 |
| 697. | Kaget liat hidungnya sekarang ga natural hasil operasinya jadi ga cocok | 2 | 2 |
| 698. | dipermak habis pun itu keseluruhan paras dia tetap ga tertolong | 2 | 2 |
| 699. | kasian itu kursinya ga kuat ditimpa obesitas kek dia | 1 | 1 |
| 700. | Ayok kalian para orang2 gendut, silahkan marah di komen, marah juga membakar kalori lho, yuk dipersilahkan » | 1 | 1 |

Saat tahapan uji model, terjadi beberapa ketidakakuratan sistem dalam melakukan klasifikasi yang dapat disebabkan oleh beberapa kondisi. Keterbatasan variasi data komentar turut menjadi penyebab pengaruh pada ketidakakuratan model dalam melakukan klasifikasi komentar *body shaming* dengan baik. Selain adanya keterbatasan data, penulis melakukan analisis terhadap beberapa kemungkinan yang menyebabkan ketidaktepatan sistem dalam memprediksi hasil, yaitu:

1. Pada suatu komentar ditemukan kata yang cenderung berorientasi pada kelas tertentu, sehingga memungkinkan adanya kecenderungan salah dalam memprediksi kelas.

Misal:

- makan bang tuh ngaca tinggal kulit aja lo
- makanya punya daging bg, bukan kulit doang

Dalam komentar ini sistem memprediksi bahwa komentar tersebut masuk menjadi bagian kelas *body shaming* berdasarkan warna kulit, dimana label sebenarnya dari data tersebut adalah *body shaming* berdasarkan bentuk tubuh. Hal ini, dapat terjadi dikarenakan kata “kulit” umum digunakan dalam *body shaming* berdasarkan warna kulit.

2. Kemungkinan penyebab salah prediksi oleh model berikutnya adalah terdapat komentar-komentar dengan kata ambigu yang membuat model cenderung melakukan kesalahan dalam memprediksi.

Misal:

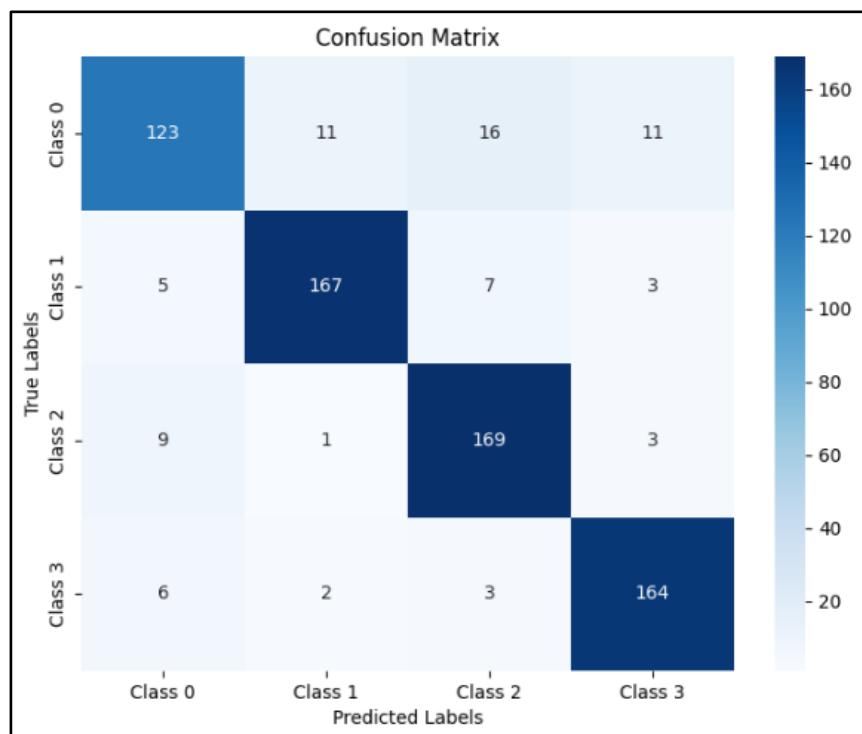
- Percaya diri penting tapi sadar diri jauh lebih penting, udah tau badannya begitu mbok ya gausah diumbat kak, lebih tertutup lebih baik kok

Komentar tersebut secara pelabelan dianggap merupakan komentar yang mengandung *body shaming* terhadap bentuk badan. Akan tetapi, penggunaan kata-kata yang tidak secara langsung merujuk pada *body shaming* seperti “badan begitu” dan “gausah diumbat” menyebabkan model melakukan kesalahan dalam memprediksi. Dalam contoh ini,

model memprediksi komentar tersebut merupakan komentar non-*body shaming*.

4.4 Pengukuran Kinerja Model

Pengukuran kinerja model merupakan proses yang dilakukan untuk meninjau kinerja perpaduan *IndoBERT word embedding* dan *Convolutional Neural Network* dalam menjalankan pengklasifikasian komentar yang berhubungan dengan *body shaming*. Dalam ruang lingkup penelitian ini, penulis menggunakan metode *Confusion Matrix* dalam peninjauan performa. Penggunaan metode dilakukan untuk memberikan ilustrasi komparasi terhadap data yang berhasil diklasifikasi dengan tepat dan data yang diklasifikasi secara tidak tepat. Hasil tinjauan tersebut meliputi nilai *precision*, *recall*, *f1-score*, dan nilai *accuracy*. Perbandingan tersebut divisualisasikan dengan format *heatmap* yang tertera pada gambar 4.8.



Gambar 4. 8 Confusion Matrix Klasifikasi *Body Shaming*

Pada *Confusion Matrix* yang telah disajikan tersebut, perhitungan nilai data yang dapat diklasifikasi dengan tepat dan tidak tepat dapat diketahui. Akumulasi perhitungan nilai data tersebut akan dijabarkan pada tabel 4.4, tabel 4.5, tabel 4.6, dan tabel 4.7.

Tabel 4.4 merupakan tabel keterangan data yang memuat komentar mengandung *body shaming* terhadap bentuk badan (*Class 1*).

Tabel 4. 4 Keterangan *Confusion Matrix Label Body Shaming* Bentuk Badan

| No | Data Body Shaming Bentuk Badan | Jumlah |
|----|--------------------------------|--------|
| 1. | <i>True Positive Comments</i> | 167 |
| 2. | <i>True Negative Comments</i> | 504 |
| 3. | <i>False Positive Comments</i> | 14 |
| 4. | <i>False Negative Comments</i> | 15 |

Tabel 4.5 berikut merupakan tabel keterangan data yang memuat komentar mengandung *body shaming* terhadap bentuk wajah (*Class 2*).

Tabel 4. 5 Keterangan *Confusion Matrix Label Body Shaming* Bentuk Wajah

| No | Data Body Shaming Bentuk Wajah | Jumlah |
|----|--------------------------------|--------|
| 1. | <i>True Positive Comments</i> | 169 |
| 2. | <i>True Negative Comments</i> | 492 |
| 3. | <i>False Positive Comments</i> | 26 |
| 4. | <i>False Negative Comments</i> | 13 |

Tabel 4.6 merupakan tabel keterangan data yang memuat komentar mengandung *body shaming* terhadap warna kulit (*Class 3*).

Tabel 4. 6 Keterangan *Confusion Matrix Label Body Shaming* Warna Kulit

| No | Data Body Shaming Warna Kulit | Jumlah |
|----|--------------------------------|--------|
| 1. | <i>True Positive Comments</i> | 164 |
| 2. | <i>True Negative Comments</i> | 508 |
| 3. | <i>False Positive Comments</i> | 17 |

| | | |
|-----------|--------------------------------|----|
| 4. | <i>False Negative Comments</i> | 11 |
|-----------|--------------------------------|----|

Tabel 4.7 merupakan tabel keterangan data yang memuat komentar yang tidak mengandung *body shaming* (Class 0).

Tabel 4.7 Keterangan *Confusion Matrix Label Non-Body Shaming*

| No | Data Tidak Mengandung <i>Body Shaming</i> | Jumlah |
|-----------|--|---------------|
| 1. | <i>True Positive Comments</i> | 123 |
| 2. | <i>True Negative Comments</i> | 519 |
| 3. | <i>False Positive Comments</i> | 20 |
| 4. | <i>False Negative Comments</i> | 38 |

Dengan demikian, berdasarkan data yang telah tersaji di atas akan digunakan untuk menentukan nilai *Precision*, *F1-Score*, *Recall*, dan *Accuracy*. Berikut merupakan hasil perhitungan tersebut:

a. *Body Shaming* Bentuk Badan

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% = \frac{167}{167 + 14} \times 100\% = 92,2\%$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% = \frac{167}{167 + 15} \times 100\% = 91,7\%$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% = 2 \times \frac{0,892 \times 0,917}{0,892 + 0,917} \times 100\% = 90,4\%$$

b. *Body Shaming* Bentuk Wajah

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% = \frac{169}{169 + 26} \times 100\% = 86,7\%$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% = \frac{169}{169 + 13} \times 100\% = 92,8\%$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% = 2 \times \frac{0,866 \times 0,928}{0,866 + 0,928} \times 100\% = 89,6\%$$

c. *Body Shaming* Warna Kulit

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% = \frac{164}{164 + 17} \times 100\% = 90,6\%$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% = \frac{164}{164 + 11} \times 100\% = 93,7\%$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% = 2 \times \frac{0,906 \times 0,937}{0,906 + 0,937} \times 100\% = 92,1\%$$

d. Tidak Mengandung *Body Shaming*

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% = \frac{123}{123 + 20} \times 100\% = 86\%$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% = \frac{123}{123 + 38} \times 100\% = 76,4\%$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% = 2 \times \frac{0,860 \times 0,763}{0,860 + 0,763} \times 100\% = 80,9\%$$

Akurasi keseluruhan data:

$$\begin{aligned} \text{Accuracy} &= \frac{TP}{\text{Total Data}} \times 100\% = \frac{623}{700} \times 100\% \\ &= 89\% \end{aligned}$$

Hasil perhitungan data komentar pada klasifikasi *body shaming* dapat diperhatikan selanjutnya melalui tabel 4.8.

Tabel 4. 8 Hasil Perhitungan Kinerja Model

| Jenis Klasifikasi | Precision | Recall | F1-Score |
|---|-----------|--------|----------|
| <i>Body Shaming</i> Terhadap Bentuk Badan | 92,2% | 91,7% | 90,4% |
| <i>Body Shaming</i> Terhadap Bentuk Wajah | 86,7% | 92,8% | 89,6% |
| <i>Body Shaming</i> Terhadap Warna Kulit | 90,6% | 93,7% | 92,1% |
| Tidak Mengandung <i>Body Shaming</i> | 86% | 76,4% | 80,9% |
| Rata-Rata | 88,8% | 88,7% | 88,3% |

Berdasarkan hasil pengujian pada penelitian ini, penggunaan metode *IndoBERT word embedding* dan *Convolutional Neural Network* dalam mengklasifikasikan komentar mengandung *body shaming* pada media sosial memperlihatkan kinerja yang memadai. Berdasarkan perhitungan rata-rata, model mampu mencapai rata-rata *precision* senilai 88,8%, *recall* senilai 88,7%, *f1-score* senilai 88,3%, dan nilai akurasi sebesar 89%. Berdasarkan hal tersebut, dapat disimpulkan bahwa model berhasil mencapai tujuannya dalam mengklasifikasikan komentar mengandung *body shaming* dengan baik.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dalam kajian yang sudah dilaksanakan, diperoleh ditarik sejumlah kesimpulan terkait implementasi gabungan *IndoBERT word embedding* dan algoritma *Convolutional Neural Network* dalam mengklasifikasikan komentar mengandung *body shaming* pada media sosial, yaitu:

1. Penggunaan *IndoBERT word embedding* dan *Convolutional Neural Network* terbukti efektif dalam melakukan klasifikasi komentar yang mengandung *body shaming* pada media sosial. Penelitian ini menggunakan sejumlah 3.500 data komentar yang diperoleh dari media sosial, dengan perbandingan 80:20 dimana 80% data digunakan sebagai data latih yang kemudian 10% diantaranya akan dimanfaatkan sebagai data validasi, serta selebihnya sebanyak 20% data komentar akan diolah menjadi data pengujian. Dari riset ini diperoleh hasil akurasi sebesar 89% yang mengacu pada penilaian performa model. Dengan rata-rata dari setiap kelas untuk *precision* senilai 88,8%, *recall* senilai 88,7%, dan *f1-score* senilai 88,3%.
2. Variasi nilai parameter dalam mengklasifikasikan komentar mengandung *body shaming* pada media sosial memberikan pengaruh terhadap performa model dalam melakukan tugasnya. Adapun variasi nilai yang memberikan hasil paling optimal dalam algoritma *Convolutional Neural Network* adalah dengan ukuran *filter* sebesar 100, ukuran *kernel* sebesar 3, ukuran *batch* sebesar 32, *dropout* sebesar 0,5, dan *epoch* sebanyak 10.
3. Terdapat komentar yang mengandung kata-kata yang bersifat ambigu serta kata-kata yang mengarah pada konteks suatu jenis klasifikasi tertentu yang mempengaruhi performa model dalam melakukan tugas pengklasifikasian pada penelitian ini.

5.2 Saran

Saran yang dapat disampaikan oleh penulis terhadap pengembangan sistem pada riset berikutnya meliputi:

1. Diharapkan dapat megoptimalisasi pengumpulan data komentar yang lebih banyak dan bervariasi untuk memberikan kemajuan positif pada kinerja model. Upaya ini diharapkan dapat memberikan tingkat akurasi yang lebih unggul dibandingkan pada hasil penelitian penulis.
2. Memperkaya daftar jumlah *stopword* dan penambahan kata yang perlu diperbaiki pada normalisasi, dengan meningkatkan jumlahnya dapat menyajikan pemahaman yang lebih mendalam terkait data oleh model dan memperkecil dampak kurang berhubungan dengan klasifikasi yang dilakukan oleh model.
3. Diharapkan pada penelitian berikutnya untuk menerapkan metode klasifikasi lainnya untuk dapat digunakan sebagai pembanding dan divalidasi hasilnya dengan metode *IndoBERT embedding* dan *Convolutional Neural Network*.
4. Disarankan pada penelitian selanjutnya untuk memberikan perhatian khusus pada komentar yang memiliki kemungkinan tergolong ke dalam lebih dari satu kelas, sehingga dapat menangani klasifikasi secara lebih baik.

DAFTAR PUSTAKA

- Aditya Prana, P., & Pandu Adikara, P. (2019). *Klasifikasi Komentar Body Shaming Beauty Vlogger Pada Youtube Menggunakan Metode BM25 dan K-Nearest Neighbor* (Vol. 3, Issue 8). <http://j-ptiik.ub.ac.id>
- Araaf, K. L., Nur, M., & Nurdin, H. (2023). *Hubungan Antara Body Shaming Dengan Body Image Pada Remaja Akhir Perempuan Di Kota Makassar.* <https://doi.org/https://doi.org/10.52472/jci.v6i1.355>
- Asmi P, & M S Sanaj. (2021). *Toxic Speech Classification via Deep Learning using Combined Features from BERT & FastText Embedding.* www.ijert.org
- Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (n.d.). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.* <https://github.com/tensorflow/tensor2tensor>
- Erdianti, R. N., Samira, E., & Al Fatih, S. (2022). Body Shaming Act as A Crime in Indonesian Criminal Law Wasis. *Justitia Jurnal Hukum*, 16(1), 155–172. <https://doi.org/10.36501/justitia.v1i2>
- Fajariani Fauzia, T., & Ratri Rahmiaji, L. (2019). *Memahami Pengalaman Body Shaming Pada Remaja Perempuan.*
- Fitria, K., & Febrianti, Y. (2020). The Interpretation And Attitude Of Body Shaming Behavior On Social Media (A Digital Ethnography Study On Instagram). *Diakom : Jurnal Media Dan Komunikasi*, 3(1), 12–25. <https://doi.org/10.17933/diakom.v3i1.78>
- Jaman, J. H., Hannie, & Simatupang, M. S. (2020). Sentiment Analysis of the Body-Shaming Beauty Vlog Comments. *Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar; MSCEIS 2019.* <https://doi.org/10.4108/eai.12-10-2019.2296530>
- Kim, Y. (2014). *Convolutional Neural Networks for Sentence Classification.* <http://nlp.stanford.edu/sentiment/>

- Li, Q., Peng, H., Li, J., Xia, C., Yang, R., Sun, L., Yu, P. S., & He, L. (2022). A Survey on Text Classification: From Traditional to Deep Learning. In *ACM Transactions on Intelligent Systems and Technology* (Vol. 13, Issue 2). Association for Computing Machinery. <https://doi.org/10.1145/3495162>
- Melizza, N., Rahmawati, A., Ari Arfianto, M., Wahyu Mashfufa, E., Dwi Kurnia, A., & Lailaltul Masruroh, N. (2023). The Relationship Between Body Shaming and Self-esteem in Students. *KnE Medicine*. <https://doi.org/10.18502/kme.v3i2.13084>
- Nabiilah, G. Z., Alam, I. N., Purwanto, E. S., & Hidayat, M. F. (2024). Indonesian multilabel classification using IndoBERT embedding and MBERT classification. *International Journal of Electrical and Computer Engineering*, 14(1), 1071–1078. <https://doi.org/10.11591/ijece.v14i1.pp1071-1078>
- Pardede, M. C. N. (2022). *Identifikasi Komentar Body Shaming Pada Media Sosial Menggunakan Term Frequency Chi-Square Dan Random Forest Classifier*.
- Pratiwi, E. I. (2021). Law Enforcement Efforts against the Crime of Body Shaming Through Mediation. *Pancasila and Law Review*, 1(2), 101. <https://doi.org/10.25041/plr.v1i2.2127>
- Selva Birunda, S., & Kanniga Devi, R. (2021). A Review on Word Embedding Techniques for Text Classification. In J. S. Raj, A. M. Iliyasu, R. Bestak, & Z. A. Baig (Eds.), *Innovative Data Communication Technologies and Application* (pp. 267–281). Springer Singapore.
- Sri, Y., & Astuti, W. (2019). Body Shaming di Dunia Maya: Studi Netnografi pada Akun Youtube Rahmawati Kekeyi Putri Cantika Body Shaming in The Virtual: Netnographic Study on Youtube Accounts Rahmawati Kekeyi Putri Cantika. In *UNIVERSITAS* (Issue 1). Yenny, Body Shaming.
- Syaiful Imron, Setiawan, E. I., Joan Santoso, & Mauridhi Hery Purnomo. (2023). Aspect Based Sentiment Analysis Marketplace Product Reviews Using BERT, LSTM, and CNN. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 7(3), 586–591. <https://doi.org/10.29207/resti.v7i3.4751>

- Zevana, A., & Riana, D. (2024). Text Classification Using Indobert Fine-Tuning Modeling With Convolutional Neural Network And Bi-Lstm. *Jurnal Teknik Informatika (Jutif)*, 4(6), 1605–1610.
<https://doi.org/10.52436/1.jutif.2023.4.6.1650>
- Zhang, B. (2023). A BERT-CNN Based Approach on Movie Review Sentiment Analysis. *SHS Web of Conferences*, 163, 04007.
<https://doi.org/10.1051/shsconf/202316304007>