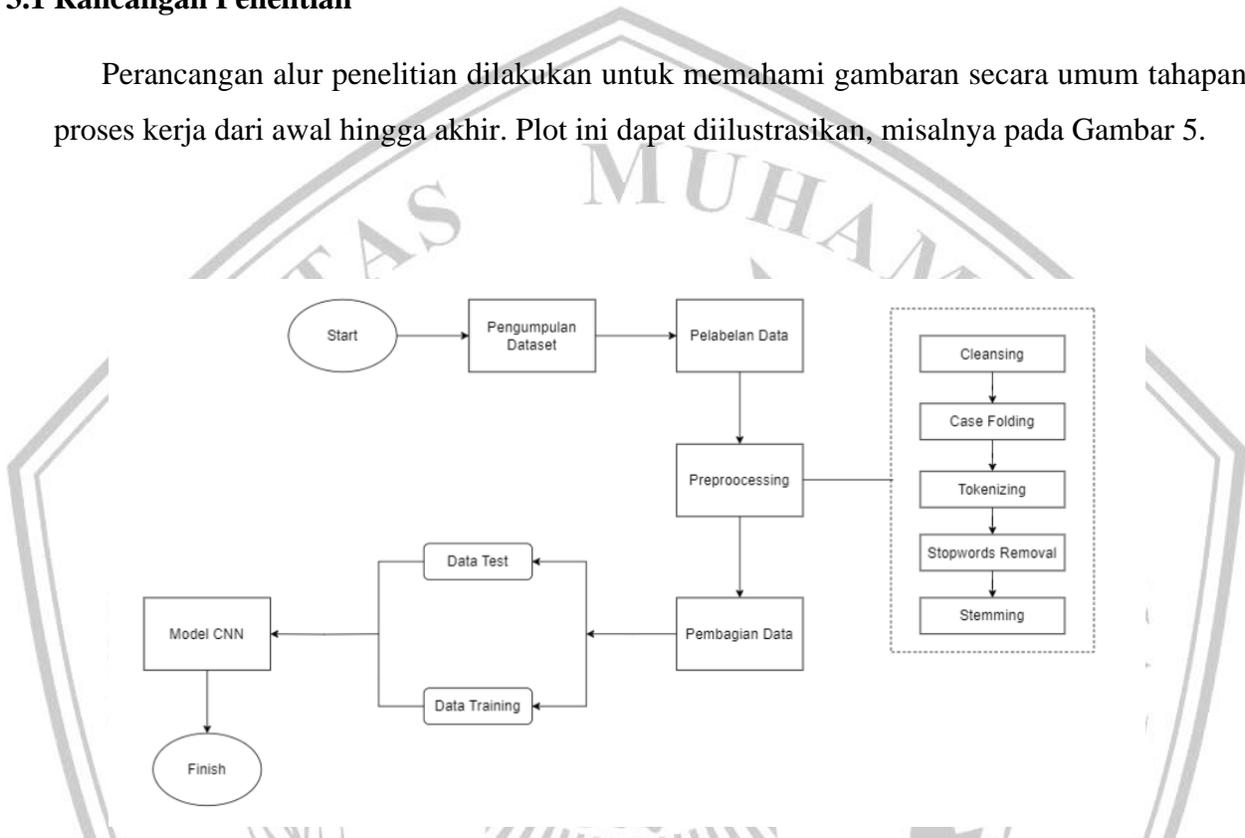


BAB III

METODOLOGI PENELITIAN

3.1 Rancangan Penelitian

Perancangan alur penelitian dilakukan untuk memahami gambaran secara umum tahapan proses kerja dari awal hingga akhir. Plot ini dapat diilustrasikan, misalnya pada Gambar 5.



Gambar 3.1 Rancangan Penelitian

Yang perlu dilakukan pertama kali adalah yaitu membentuk kumpulan data dengan mengumpulkan data komentar tentang pariwisata Labuan Bajo di X. *Scraping* teknik digunakan dalam tahapan data pengumpulan ini. Setelah data terkumpul maka dilanjutkan ke tahap pelabelan data. Pada titik ini, setiap ulasan akan memiliki label sentiment sesuai dengan apa pun dideskripsikan pada ulasan komentar tersebut. Label sentiment dapat berupa positif, negatif dan netral yang dilakukan secara manual [21]. Langkah selanjutnya adalah preprocessing, dimana pada tahap ini data akan mengalami pembersihan. Tujuan dari tahap ini untuk menghilangkan komponen-komponen data yang dianalisis tidak mempunyai ciri-ciri yang penting, sehingga data menjadi lebih konsisten agar data sesuai dengan formatnya yang diperlukan untuk pengolahan pada tahap selanjutnya. Hal yang

dilakukan seperti *cleaning*, *case folding*, *stopwords removal*, *tokenizing*, dan *stemming* [22].

Setelah data telah dibersihkan maka akan dipisahkan menjadi data latihan dan data uji. Fungsi dari data latihan adalah sebagai input dan pelatihan model. Dengan 90% sebagai data latihan dan 10% sebagai data test. Pada tahap berikutnya yaitu implementasi metode CNN. Model digambar dan analisa agar menghasilkan ketelitian yang tinggi. Setelah model selesai, maka akan dilanjutkan ke tahap akhir pada pengujian dan evaluasi merupakan dua komponen penelitian ini. Dalam pengujian ini tahap evaluasi data akan dicoba dalam model yang selanjutnya akan menghasilkan akurasi. Dengan melihat akurasi tersebut bisa mengidentifikasi bagaimana kinerjanya metode CNN. Selain itu juga dimanfaatkan *confusion matrix* untuk melihat berapa banyak titik data yang berhasil diklasifikasikan oleh model.

3.2 Pengumpulan Dataset

Informasi yang digunakan dalam penelitian ini adalah ulasan komentar berbahasa Indonesia yang diambil pada aplikasi X tentang pariwisata Labuan Bajo. Untuk mempermudah proses pencarian komentar tersebut memanfaatkan fitur pencarian komentar di aplikasi X. Komentar yang digunakan hanya membahas tentang Labuan Bajo dan Komodo. Dataset yang berhasil dikumpulkan sebanyak 2985 dan rentang waktu data yang diambil adalah dari tanggal 22 Desember 2023 – 22 Januari 2024. Pada periode ini mencakup liburan akhir tahun, yang dimana biasanya meningkatkan aktifitas wisata yang signifikan [23]. Selain itu periode Desember hingga Januari di kenal sebagai *high season* di Labuan Bajo karena cuaca yang mendukung dan daya tarik liburan tahun baru. Pengumpulan data pada waktu ini dapat menunjukkan pengelolaan destinasi selama puncak kunjungan wisatawan serta membantu memetakan perbedaan perilaku antara domestik dan internasional. Data ini juga berguna untuk merancang strategi pariwisata kedepan, seperti mempersiapkan kapasitas penginapan dan transportasi sesuai lonjakan permintaan.

Setelah itu untuk mengubah data menjadi file csv teknik *scraping*. Proses mengekstraksi informasi dari halaman web dikenal sebagai *scraping*. Setiap komentar informasi ditangkap dan diubah menjadi dokumen. Setelah pengumpulan data, semua data akan disusun menjadi satu dokumen [22]. Setelah data terkumpul maka dilanjutkan ke

tahap pelabelan data. Pada titik ini, setiap ulasan akan memiliki label sentiment sesuai dengan apa pun dideskripsikan pada ulasan komentar tersebut. Label sentiment dapat berupa positif, negatif dan netral yang dilakukan secara manual. Pelabelan dilakukan oleh 3 *reviewer* yang memiliki latar belakang ahli Bahasa dalam menentukan label. Teknik yang digunakan untuk tes ini memerlukan metode khusus, seperti metode pemungutan suara mayoritas. Di antara berbagai metode tersebut yang paling banyak digunakan adalah metode pemungutan suara mayoritas. Metode ini menggunakan konsep keputusan hasil pemungutan suara yang diperoleh dari jumlah suara yang dikeluarkan pada setiap suara yang ada [24]. Label sentiment dapat dilihat pada Tabel 2.

Tabel 3. 1 Label Sentimen

full_text	username	lang	sentimen
Mau aku ajak ke Labuan Bajo. 🤔	renjanaka	in	netral
pengen ke labuan bajo tp kenapa tiketnya ga masuk akal ya najis	shkittens	in	negatif
Gili Lawa Darat yang sunsetnya selalu memukau	novemlawalata	in	positif
Pengen ke sumba, labuan bajo, raja ampat, tp beneran tiket ke indonesia timur tuh mahal2 :(Padahal alam indonesia timur tuh bagus2	lilacnmoon	in	negatif
Emang Surganya Dunia Ga Si Labuan Bajo tuh 😊😊😊	huntzter	in	positif
Woi ikutan dong ke labuan bajo 🙌🙌🙌🙌	rainbowgurl_	in	netral
Masya Allah, potret romantis Gibran dan Selvi Ananda di Labuan Bajo sangat menginspirasi	ini_baik97426	in	positif

3.3 Preprocessing

Data yang diambil sebelumnya merupakan data yang tidak beraturan dan tidak kompleks. Di tahap inilah dilakukan pemrosesan data sebelum digunakan untuk melatih model. Langkah preprocessing akan membuat komponen-komponen dalam data dianggap kurang penting dan membuat data menjadi lebih konsisten. Tujuannya adalah untuk

memastikan bahwa data dapat diformat dengan tepat untuk dianalisis tahap sebagai berikut preprocessing memiliki banyak jenis yang berbeda. Hal ini kemudian sesuai dengan kebutuhan dataset yang digunakan. Setiap penelitian penelitian mempunyai persiapan yang berbeda-beda namun dalam penelitian ini, *preprocessing* yang sedang dilakukan *cleaning*, *case folding*, *stopwords removal*, *tokenizing*, dan *stemming*. Proses hal ini harus dilakukan secara metodis agar dapat memberikan hasil yang sesuai maksimal [16].

1.3.1. Cleaning

Cleaning atau membersihkan merupakan proses pembersihan informasi dari tweet yang memuat nama pengguna atau username, angka, RT, hastag, dan Uniform Resource Locator (URL), symbol, white space, ruang kosong, dan emoji [22]. Hasil *cleansing* dapat dilihat pada Tabel 3.

Tabel 3. 2 Cleaning

Sebelum	Sesudah
Mengunjungi Pantai Pink di Labuan Bajo, pemandangan yang langka dan memukau! 📍🌸🌟 #Seger8ener #Seruuubos	Mengunjungi Pantai Pink di Labuan Bajo pemandangan yang langka dan memukau
Pemerintah harus cepat menanganinya agar pariwisata di Labuan bajo segera bangkit lagi.	Pemerintah harus cepat menanganinya agar pariwisata di Labuan Bajo segera bangkit lagi
@aLineasa Bagus mbak kalau suka air. Disini pantainya mantap, bisa ke komodo juga. Seafoodnya segar semua. Wah, psikolog ternyata. Buka praktik sendiri?	Bagus mbak kalau suka air Disini pantainya mantap bisa ke komodo juga Seafoodnya segar semua Wah psikolog ternyata buka praktik sendiri
indahny Pantai Pink di Pulau Komodo. Saya sangat terpesona dengan keindahan alam Indonesia yang luar biasa! 😊❤️	Indahnya Pantai Pink di Pulau Komodo saya sangat terpesona dengan keindahan alam Indonesia yang luar biasa

1.3.2. Case folding

Proses mengubah setiap karakter huruf dalam suatu kalimat menjadi huruf kecil atau besar yang disebut sebagai case folding yang dilakukan pada penelitian ini adalah dengan memperkecil seluruh dataset. Huruf kapital biasanya ditemukan di setiap awal kalimat seperti “Udara di tempat ini sangat sejuk”. Dengan

menggunakan case lipat, kalimat tersebut diubah menjadi “Udara di tempat ini sangat dingin”. Tujuan dari case ini untuk memastikan bahwa kata “Udara” mempunyai dua bentuk yang berbeda, yaitu “Udara” dan “udara” namun hanya satu bentuk kecil [16] . Hasil *case folding* dapat dilihat pada Tabel 4.

Tabel 3. 3 Case Folding

Sebelum	Sesudah
Mengunjungi Pantai Pink di Labuan Bajo pemandangan yang Langka dan memukau	mengunjungi pantai pink di labuan bajo pemandangan yang langka dan memukau
Pemerintah harus cepat menanganinya agar pariwisata di Labuan Bajo segera bangkit lagi	pemerintah harus cepat menanganinya agar pariwisata di labuan bajo segera bangkit lagi
Bagus mbak kalau suka air Disini pantainya mantap bisa ke komodo juga Seafoodnya segar semua Wah psikolog ternyata buka praktik sendiri	bagus mbak kalau suka air disini pantainya mantap bisa ke komodo juga seafoodnya segar semua wah psikolog ternyata buka praktek sendiri
Indahnya Pantai Pink di Pulau Komodo saya sangat terpesona dengan keindahan alam Indonesia yang luar biasa	indahya pantai pink di pulau komodo saya sangat terpesona dengan keindahan alam indonesia yang luar biasa

1.3.3. Stopwords Removal

Stopword removal atau *filtering* merupakan proses menciptakan kata-kata penting dan yang kurang penting. Eliminasi stopword adalah metode untuk menciptakan kata-kata yang tidak penting. Tujuan dari stopword removal untuk menghilangkan kata-kata yang sering muncul namun tidak berkontribusi pada proses analisis data [17]. Tujuan dari penghapusan stopword adalah untuk mengurangi dimensi data dan mempercepat pemrosesan komputer. Contoh kata non esensial dalam bahasa indonesia antara lain “dan”, “yang”, “di”, dan “ke”. Hasil *stopwords* penghapusan dapat dilihat pada Tabel 5.

Tabel 3.4 *Stopword Removal*

Sebelum	Sesudah
mengunjungi pantai pink di labuan bajo pemandangan yang langkah dan memukau	mengunjungi pantai pink labuan bajo pemandangan langkah memukau
pemerintah harus cepat menanganinya agar pariwisata di labuan bajo segera bangkit lagi	pemerintah cepat menangani pariwisata labuan bajo segera bangkit
bagus mbak kalau air disini pantainya mantap bisa ke komodo juga seafoodnya segar semua wah psikolog ternyata buka praktek sendiri	bagus air pantai mantap bisa komodo juga seafood segar semua wah psikolog ternyata buka praktek sendiri
indahny pantai pink di pulau komodo saya sangat terpesona dengan keindahan alam indonesia yang luar biasa	indah pantai pink pulau komodo terpesona keindahan alam indonesia luar biasa

1.3.4. Tokenizing

Tokenisasi merupakan proses prosedur untuk mengubah dokumen menjadi kata atau bingkai symbol atau elemen lain dengan makna. Optimalisasi token dapat dilakukan dengan menghilangkan karakter ilegal pada dokumen, seperti tanda baca Symbol, HTML, angka, dan penyebutan. Pertimbangkan karakter ilegal itu dihilangkan %, &, >, (, {,), 1-9, dan seterusnya @fadlizon, https://twitter.com/last_inks/status/1738315990851637670 [25]. Hasil *tokenizing* dapat dilihat pada Tabel 6.

Tabel 3.5 *Tokenizing*

Sebelum	Sesudah
mengunjungi pantai pink labuan bajo pemandangan langkah memukau	['mengunjungi', 'pantai', 'pink', 'labuan', 'bajo', 'pemandangan', 'langkah', 'memukau']
pemerintah cepat menangani pariwisata labuan bajo segera bangkit	['pemerintah', 'cepat', 'menangani', 'pariwisata', 'labuan', 'bajo', 'segera', 'bangkit']

bagus air pantai mantap bisa komodo juga seafood segar semua wah psikolog ternyata buka praktek sendiri	['bagus', 'air', 'pantai', 'mantap', 'bisa', 'komodo', 'seafood', 'segar', 'semua', 'wah', 'psikolog', 'ternyata', 'buka', 'praktek', 'sendiri']
indah pantai pink pulau komodo terpesona keindahan alam indonesia luar biasa	['indah', 'pantai', 'pink', 'pulau', 'komodo', 'terpesona', 'keindahan', 'alam', 'indonesia', 'luar', 'biasa']

1.3.5. Stemming

Stemming adalah proses mengubah suatu kata menjadi bentuk dasar. Proses stemming dilakukan dengan menganalisis imbuhan, baik awalan maupun akhiran suatu kata tertentu untuk memperoleh dasarnya. Biasanya stemming digunakan pada teks berbahasa indonesia dengan menggunakan stemmer perpustakaan sastra yang dikembangkan berdasarkan algoritma Nazief-Adriani [26]. *Stemmer* Sastrawi merupakan salah satu stemmer library yang menetapkan algoritma Nazief-Adriani, dilanjutkan dengan algoritma confix stripping (CS) dan terakhir algoritma enhanced confix stripping (ESC). Akhirnya dikembangkan lebih lanjut sebagai algoritma modified ECS berdasarkan morfologi Bahasa Indonesia. Hasil *stemming* dapat dilihat pada Tabel 7.

Tabel 3. 6 Stemming

Sebelum	Sesudah
['mengunjungi', 'pantai', 'pink', 'labuan', 'bajo', 'pemandangan', 'langkah', 'memukau']	['kunjung', 'pantai', 'pink', 'labuan', 'bajo', 'pandang', 'langkah', 'memukau']
['pemerintah', 'cepat', 'menangani', 'pariwisata', 'labuan', 'bajo', 'segera', 'bangkit']	['pemerintah', 'cepat', 'tangani', 'pariwisata', 'labuan', 'bajo', 'segera', 'bangkit']
['bagus', 'air', 'pantai', 'mantap', 'bisa', 'komodo', 'seafood', 'segar', 'semua', 'wah', 'psikolog', 'ternyata', 'buka', 'praktek', 'sendiri']	['bagus', 'air', 'pantai', 'mantap', 'bisa', 'komodo', 'seafood', 'segar', 'semua', 'wah', 'psikolog', 'nyata', 'buka', 'praktek', 'diri']
['indah', 'pantai', 'pink', 'pulau', 'komodo', 'terpesona', 'keindahan', 'alam', 'indonesia', 'luar', 'biasa']	['indah', 'pantai', 'pink', 'pulau', 'komodo', 'pesona', 'indah', 'alam', 'indonesia', 'luar', 'biasa']

3.4 Implementasi Model CNN

Data yang sudah disediakan akan dipisahkan menjadi data latihan, data validasi, dan data uji. Data latihan digunakan untuk mengajar model tentang analisis sentiment, sedangkan validasi data digunakan untuk evaluasi dan menyesuaikan model selama pelatihan, dan mencegah *overfitting*. Data uji digunakan setelah pelatihan untuk memberikan evaluasi akhir kinerja model pada data baru. Pembagian data ini umumnya adalah 90% untuk data latihan dan 10% untuk data uji. Pembagian ini memungkinkan pelatihan dan evaluasi yang komprehensif, meningkatkan keandalan, mencegah *overfitting*, dan meningkatkan akurasi analisis.

Dalam Model *Convolutional Neural Network* (CNN) ada banyak lapisan (CNN). Informasi pertama kali memasuki lapisan penyematan, yang menganalisis teks menggunakan bobot acak dan mengajarkannya masukkan ke dalam dataset latih. Selanjutnya, data menuju ke *convolutional layer* dimana dilakukan ekstraksi fitur. Proses ini melibatkan *pooling layer* yang membantu mengurangi representasi masukan kalimat parameter, dan komputasi dalam jaringan. Lapisan terakhir adalah *dense layer* atau *fully connected layer*, dimana klasifikasi dilakukan terhadap fitur yang terekstrasi pada lapisan sebelumnya [9].

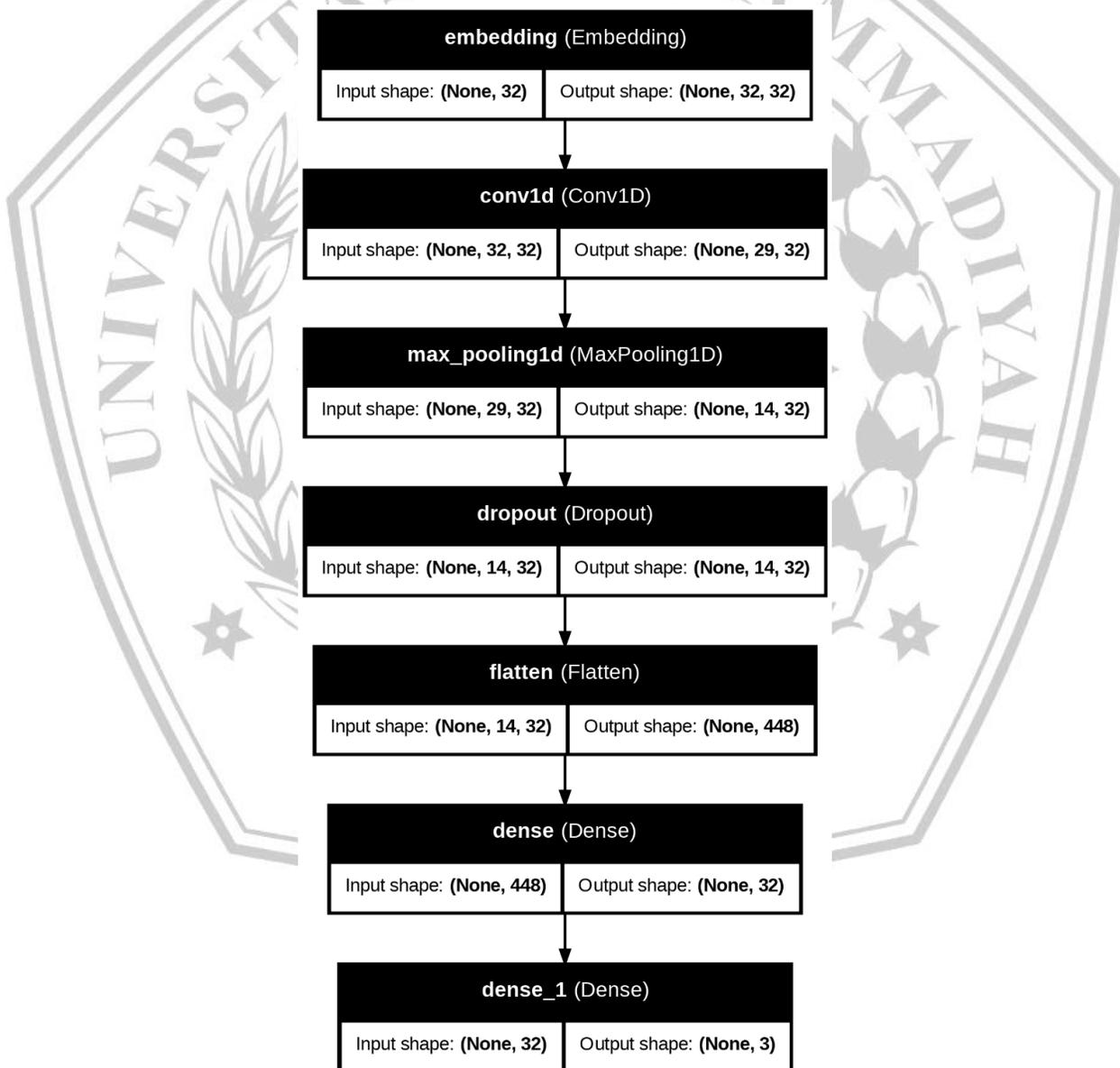
Pemilihan model adalah hal yang penting karena bisa mempengaruhi hasil keluaran. Penelitian ini memilih konvolusi 1D karena jenis konvolusi ini sesuai dengan data teks. Pengecualian adalah pengumpulan maks global 1D juga digunakan untuk menyajikan hasil itu optimal. Untuk menentukan hasil terbaik, diperlukan pemilihan parameter yang tepat. Parameter ini mencerminkan nilai kernel, ukuran filter, *hidden dimention*s, dan dropout. Untuk menentukan parameter yang sesuai, pada penelitian ini meneliti beberapa model itu dapat dilihat pada Tabel 8.

Tabel 3. 7 Model untuk Klasifikasi Sentimen

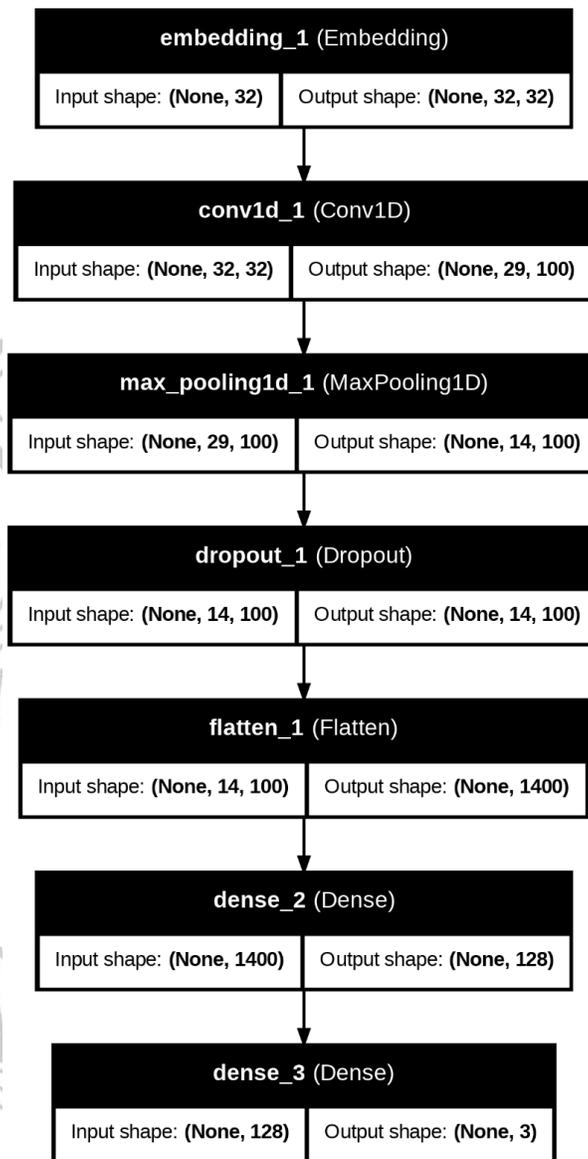
Model	Metode	Epoch	Kernel	Filter	<i>Hidden-dimention</i> s	Dropout
Model 1	CNN	20	4	32	32	0.5
Model 2		20	4	100	128	0.25

Model 3		20	5	32	32	0.5
Model 4		20	9	200	128	0.5

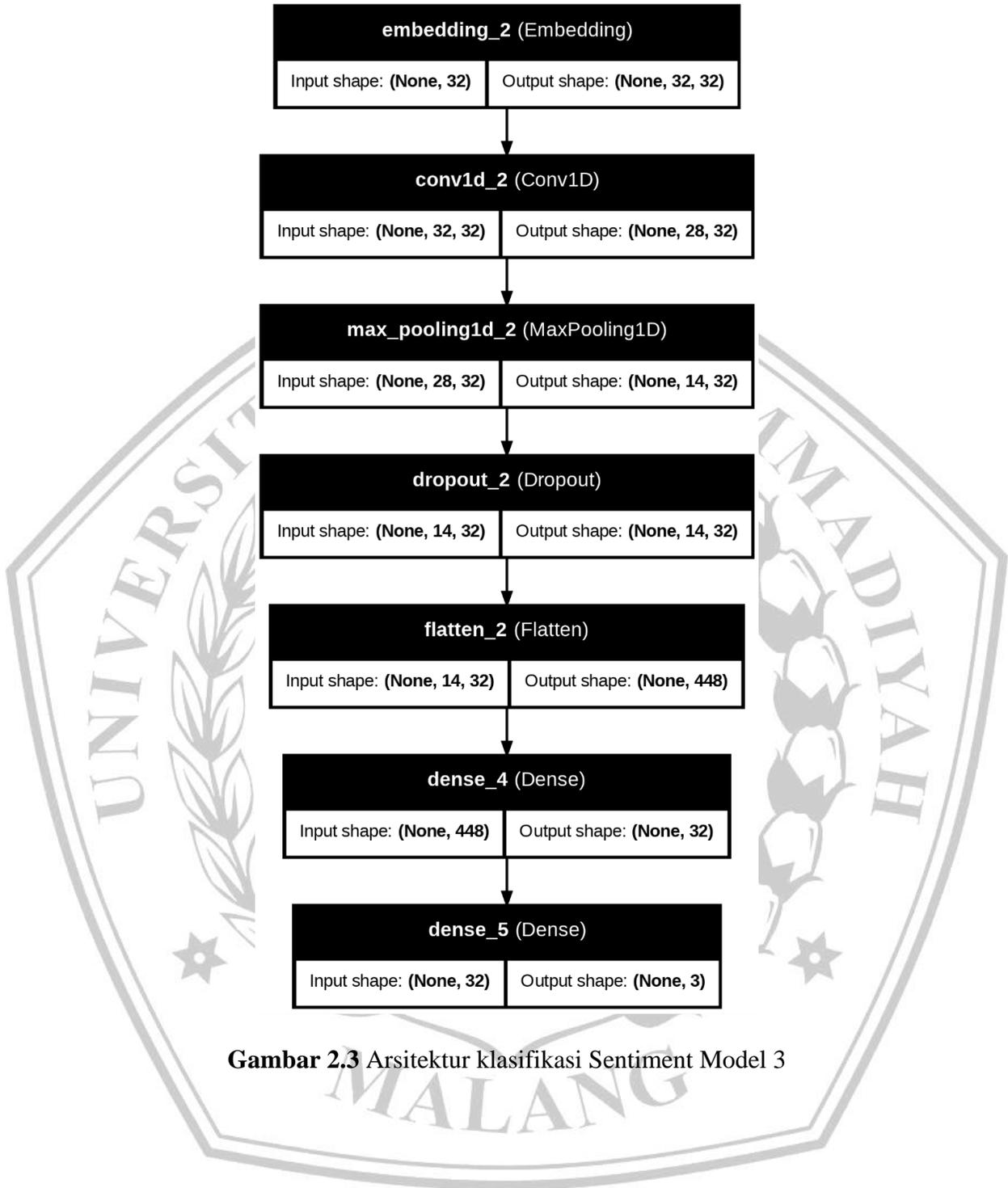
Pengujian yang berbeda ini membantu menentukan kapan model mulai *overfitting* atau mencapai performa terbaiknya, memungkinkan penyesuaian lebih lanjut untuk tingkat akurasi dan keandalan model. Visualisasi artistik berasal dari CNN hal ini digunakan agar dapat melihat lebih jelas bagaimana struktur model yang dimaksud dibangun. Klasifikasi arsitektur sentimen hal ini dapat dilihat pada gambar 6, 7, 8, dan 9 secara berurutan.



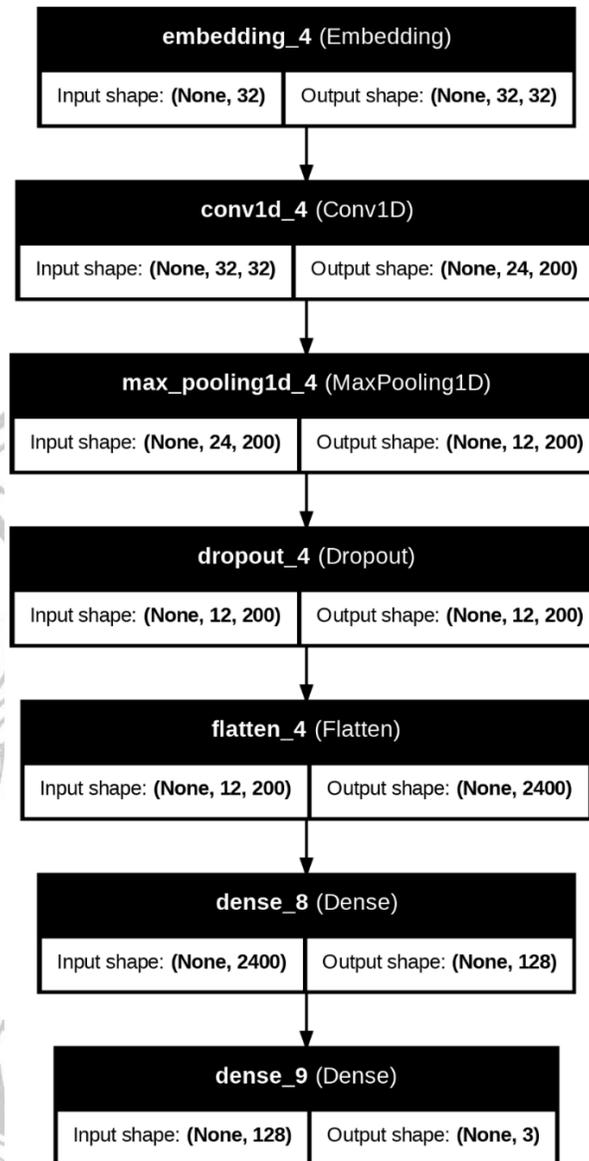
Gambar 2.1 Arsitektur klasifikasi Sentiment Model 1



Gambar 2.2 Arsitektur klasifikasi Sentiment Model 2



Gambar 2.3 Arsitektur klasifikasi Sentiment Model 3



Gambar 2. 4 Arsitektur klasifikasi Sentiment Model 4

3.5 Pengujian & Evaluasi

Setelah model dikembangkan, langkah selanjutnya adalah pengujian. Model yang digunakan dalam klasifikasi sentiment terdiri dari empat variasi, masing masing dengan parameter yang berbeda. Model 1 menggunakan kernel berukuran 4, 32 filter, 32 hidden-dimensions dan dropout sebesar 0.5, dilatih selama 20 epoch. Model 2 memiliki 100 filter, 128 hidden-dimensions, dan dropout 0.25, dilatih selama 20 epoch. Model 3 dengan kernel berukuran 5, menggunakan 32 filter dan hidden-dimensions yang sama dengan model 1 dan dilatih selama 20 epoch. Model 4 menggunakan kernel berukuran 9, 200 filter, 128 hidden-dimensions, dan dropout sebesar 0.5, dilatih selama 20 epoch. Evaluasi model dilakukan dengan menggunakan metrix seperti akurasi,

F1-Score, presisi, dan recall. Cofusion matrix digunakan untuk menghitung nilai-nilai tersebut. Evaluasi bertujuan untuk melihat beberapa model bagus dapat memprediksi sentiment dari data yang belum diperiksa sebelumnya, dan juga untuk membandingkan performa keempat model tersebut dalam klasifikasi sentiment.

