

BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini merupakan penjelasan hasil dan analisis dari penelitian pada bab sebelumnya. Bab hasil dan pembahasan mencakup implementasi metode CNN pada studi kasus judul penelitian ini. Berikut tahapan hasil penelitian ini.

4.1 Hasil Pengumpulan dan Preprocessing Data

Data yang dikumpulkan dengan menggunakan teknik web screepping di ambil dari Google Play Store sejumlah 2.260 data. Dataset tersebut tersimpan dalam format csv yang masih belum di lakukan proses pembersihan data. Setelah data dilakukan proses pembersihan atau proses preprocessing terkumpul 2.042 data dalam sentimen negatif dan 218 data sentimen positif. Dapat diketahui bahwa dataset tersebut belum seimbang maka dilakukan proses Smote agar dataset tersebut seimbang dengan jumlah masing-masing yaitu 2.042 dataset bersentimen positif dan negatif.

Data sebelum proses preprocessing	Data setelah proses preproceasing	Sentiment
Puas dengan aplikasi ini, karena banyak fitur menarik...Khususnya fitur musiknya pun sering diperbaharui	puas dengan aplikasi ini karena banyak fitur menarik khususnya fitur musiknya pun sering diperbaharui	1
Musiknya juga pas sudah dinaikin besoknya dibuat musik tidak tersedia.. kecewa banget??..	musiknya juga pas sudah dinaikin besoknya dibuat musik tidak tersedia kecewa banget	0
ketika upload menggunakan "fitur musik" tidak ada suara, update tidak merata di semua akun	ketika upload menggunakan fitur musik tidak ada suara update tidak merata di semua akun	0

Tabel 3 Hasil Preprocessing data

4.2 Hasil Evaluasi pengujian

Pada penelitian ini menunjukkan hasil dan evaluasi dari penggunaan metode CNN (Convolutional Neural Network), dengan melakukan 2 percobaan serta masing-masing memiliki 2 skenario model pengujian. Hasil penelitian tersebut berupa nilai akurasi dan perbandingan plot loss dan akurasi dalam 2 skenario yang di ujikan, untuk penggunaan optimizer percobaan pertama menggunakan Adam dan percobaan kedua model 2 RMSprop. Dengan menganalisis 2 percobaan akan dilihat hasil yang terbaik penggunaan metode CNN ini. Berikut perbandingan hasil dan hyperparameter dari 2 percobaan.

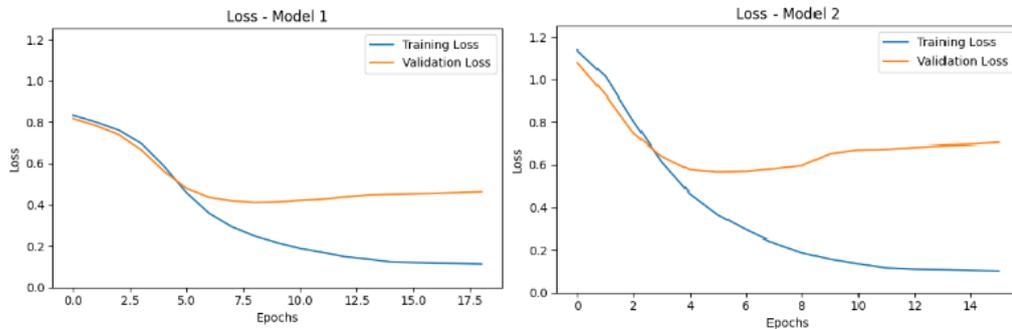
Percobaan	Skenario Model	Akurasi test	Loss	Filters	Random state (k-fold)	Epoch dan batch size
1	1	93%	24%	32,16,8	32	50 dan 20
	2	92%	34%	256,128,32		
2	1	97%	20%	64,32,16	42	80 dan 32
	2	95%	21%	128,64,32		

Tabel 4 Hasil perbandingan 2 skenario

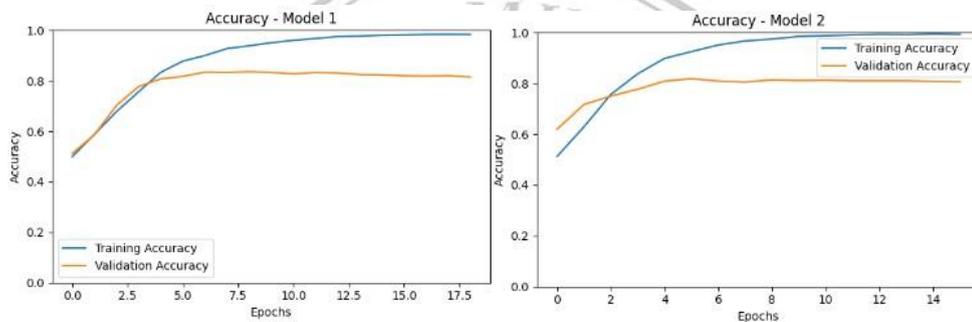
4.2.1 Percobaan pertama

Pada percobaan pertama memiliki 2 skenario model layer, layer yang digunakan yaitu Embedding, Conv1D, MaxPooling1D, GlobalMaxPooling1D, Dense, dan Dropout. Perbedaan 2 skenario model dalam percobaan ini pada layer Conv1D dengan filters model pertama menggunakan 32,16,8 dan 256,128,32 untuk model kedua. Penggunaan optimizer kedua model yaitu adam dengan jumlah learning rate 0,0001. Sebelum data melakukan pemodelan, data dilatih yang telah ditentukan oleh jumlah epoch sebanyak 50 dan 20 untuk ukuran

batch yang digunakan. Berikut hasil pelatihan data dilihat dari plot loss dan plot akurasi.



Gambar 4 Perbandingan Plot loss percobaan 1



Gambar 5 perbandingan plot akurasi percobaan 1

Pada gambar 4 dan 5 diatas, menjelaskan bahwa nilai plot loss akan berkurang seiring bertambahnya jumlah epoch, sedangkan nilai akurasi akan semakin meningkat. Perbandingan pada percobaan 1 pada plot loss maupun plot akurasi kedua model gambar 4 dan 5 diatas yang sudah memiliki nilai sumbu y yang sama pada masing-masing plot. Pada saat proses pelatihan rata-rata waktu eksekusi yang digunakan untuk percobaan 1 model 1 yaitu 44,17 *seconds*/detik dan untuk percobaan 1 model 2 yaitu 78,48 *seconds*/detik. Nilai loss validation model pertama dan kedua pada percobaan pertama ini memiliki nilai 24% dan 34%. Jadi pada percobaan pertama ini yang paling cepat adalah penggunaan model 1.

Maka dari itu, dapat dikatakan performa model yang digunakan stabil. Plot yang ditunjukkan pada gambar 3 dan 4 baik dan stabil, selain dipengaruhi oleh kualitas data yang bebas dari noise dan outliers,

juga dilakukan tahap early stopping dan learning rate scheduler. Pada tahap early stopping dilakukan untuk menghentikan pelatihan saat sudah tidak ada peningkatan jumlah epoch yang digunakan pada tahap ini sejumlah 10 mencegah terjadinya overfitting data pelatihan. Pada tahap learning rate scheduler dilakukan untuk membantu atau mengarahkan model agar lebih stabil jika memecapai patience sebesar 10 dalam plot yang ditunjukkan. Learning rate scheduler yang digunakan adalah ReduceLRonPlateau, maka jika epoch 1-5 nilai val_loss membaik jadi learning rate akan tetap, jika epoch 6-10 val_loss berhenti maka pada 1 epoch setelah nya learning rate dikurangi menjadi 20% dari nilai sebelumnya, begitupun juga untuk selanjutnya dan nilai learning rate tidak akan dikurangi lebih rendah dari nilai 0,00001. Setelah dilakukan pelatihan dan pemodelan data maka untuk mengetahui hasil tersebut ditunjukkan dengan classification report untuk percobaan pertama skenario model 1 dan 2 sebagai berikut.

	precision	recall	F1-score	support
0	0,93	0,94	0,94	430
1	0,93	0,92	0,93	387
Akurasi			0,93	817

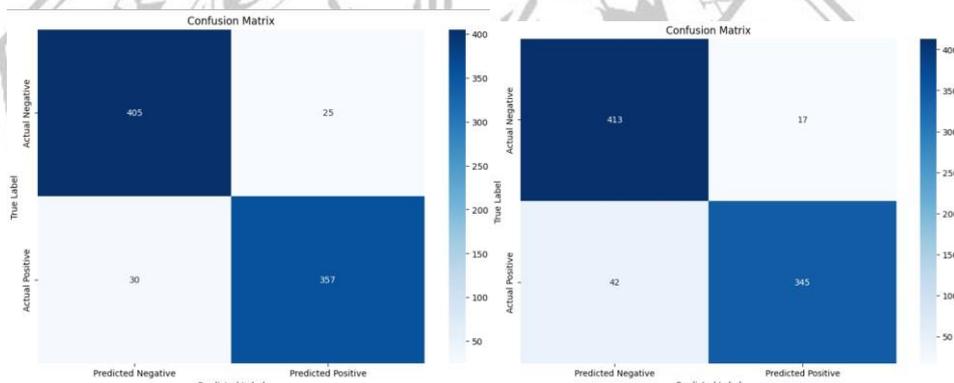
Tabel 5 Classification Report percobaan 1 model 1

	precision	recall	F1-score	support
0	0,91	0,96	0,93	430
1	0,95	0,89	0,92	387
Akurasi			0,92	817

Tabel 6 Classification Report percobaan 1 model 2

Pada tabel 5 dan 6 yang disajikan diatas, percobaan kedua memiliki nilai akurasi yang sama sebesar 93%. Walaupun memiliki akurasi yang sama kedua model tersebut memiliki perbedaan pada hasil precision,recall,f1-score, serta nilai loss memiliki nilai 24% dan 34%.

Hasil tersebut dikatakan baik karena dipengaruhi oleh kebersihan data sebelum dataset diolah masuk pada tahap preprocessing, dataset terlebih dahulu dibersihkan secara manual, misalnya kata yang mengandung angka atau kata yang tidak penting. Setelah dilakukan tahap pembersihan secara manual, maka dataset dapat diolah pada tahap preprocessing. Selain itu, juga dilakukan tahap k-fold cross validation memiliki nilai k sebanyak 5 untuk membantu hasil pengolahan data yang pasti. Data juga dilakukan teknik smote untuk mengatasi tidak seimbangan kelas dalam dataset, sebelum dataset seimbang berjumlah 2.042 negatif dan 218 positif dan setelah dilakukan smote masing-masing dataset berjumlah 2.042 data. Pada percobaan pertama ini selain menampilkan classification report juga menunjukkan visualisasi confusion matrix sebagai berikut.



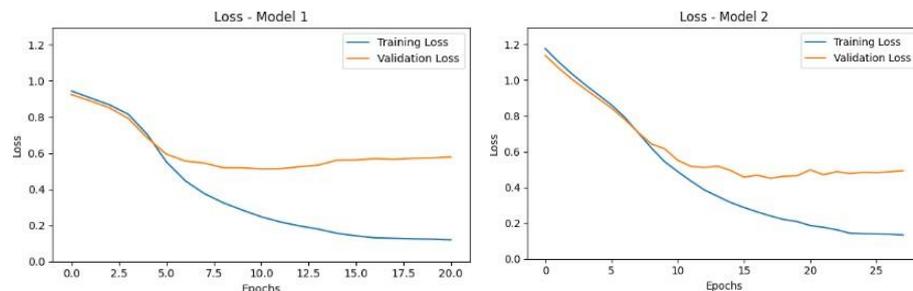
Gambar 6 confusion matrix percobaan 1

Pada gambar 6. Diatas menunjukkan visualisasi percobaan pertama dalam bentuk confusion matrix. visualisasi tersebut ditunjukkan untuk mengetahui kinerja model dalam melakukan klasifikasi prediksi pada masing-masing kelas yang jalankan dalam percobaan pertama ini. Selain itu, juga untuk mengetahui pola kesalahan klasifikasi yang dibuat oleh model pertama dan kedua pada masing-masing kelas. Pada gambar 6 yang ditunjukkan bahwa model 1 memiliki nilai True Negative (TN) sejumlah 405, artinya model memprediksi 405 ulasan negatif dengan benar. Nilai False Positives (FP) berjumlah 25, artinya model salah memprediksi 25 ulasan negatif sebagai positif. Nilai False

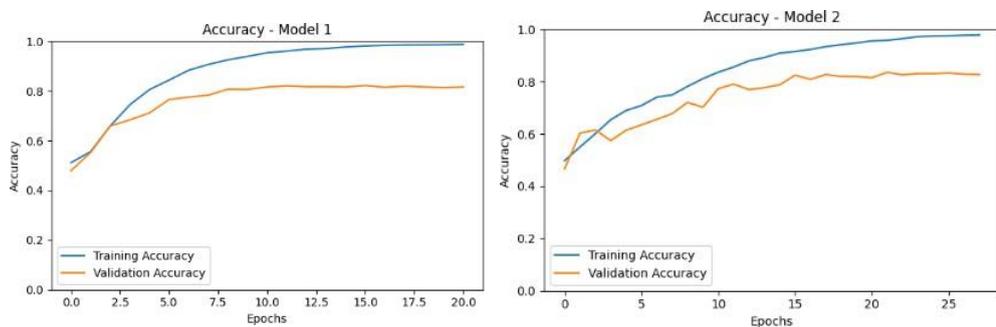
Negatives (FN) sejumlah 30, artinya model salah memprediksi sejumlah 30 ulasan positif sebagai negatif. Nilai True Positives (TP) sejumlah 357, artinya model memprediksi 357 ulasan dengan benar. Sedangkan, confusion matrix pada model kedua memiliki nilai True Negative (TN) sejumlah 413, artinya model memprediksi 413 ulasan negatif dengan benar. Nilai False Positives (FP) dengan nilai 17, artinya model salah memprediksi 17 ulasan negatif sebagai positif. Nilai False Negatives (FN) memiliki nilai 42, Artinya model salah memprediksi 42 ulasan positif sebagai negatif. Nilai True Positives memiliki nilai 345, artinya model memprediksi 345 ulasan positif dengan benar.

4.2.2 Percobaan kedua

Pada percobaan kedua yang memiliki 2 model yang akan dilakukan percobaan, layer layer yang digunakan sama dengan percobaan pertama. Akan tetapi, pada percobaan ini memiliki perbedaan hyperparameter yang digunakan, antara lain: filters pada model pertama 64,32,16 dan 128,64,32 untuk model kedua, jumlah epoch yang digunakan 80, 32 untuk ukuran batch, dan 42 untuk penggunaan random state k-fold cross validation. Penelitian percobaan kedua ini menggunakan 2 optimizer yang berbeda pada setiap modelnya, untuk model pertama menggunakan Adam sedangkan model kedua menggunakan RMSprop. Sebelum melakukan percobaan model ini, data dilakukan pelatihan dengan menghasilkan yang ditunjukkan oleh plot loss dan plot akurasi sebagai berikut.



Gambar 7 Perbandingan plot loss percobaan 2



Gambar 8 perbandingan plot akurasi percobaan 2

Pada gambar 7 dan 8 diatas dilakukan pelatihan data pada percobaan model pertama dan kedua yang memiliki sumbu y yang sama pada masing-masing model. Pada pelatihan gambar 7 dan 8 memerlukan waktu eksekusi untuk percobaan kedua model pertama yaitu 39,42 *seconds/detik* sedangkan percobaan kedua model kedua yaitu 66,79 *seconds/detik*. Dilihat dari waktu eksekusi tersebut model kedua percobaan ini lebih lama dibandingkan dengan model 2. Nilai loss validation model 1 dan 2 memiliki 20% dan 21%. Jika dilihat dari proses pelatihan percobaan kedua, maka percobaan kedua model pertama yang sudah mencukupi.

Nilai yang dihasilkan oleh plot percobaan kedua ini dapat dikatakan sudah mencukupi, dipengaruhi oleh peneliti yang melakukan tahap pembersihan data secara manual kembali. Peneliti juga melakukan beberapa uji coba angka untuk hyperparameter khususnya pada random state yang dimiliki oleh k-fold, jumlah epoch, dan ukuran batch karena hal itu juga mempengaruhi tingkat kinerja model yang dilakukan. Pada percobaan ini juga melakukan tahap seperti percobaan 2 yaitu melakukan early stopping dan learning rate scheduler yang sama, karena hal ini bisa melakukan monitoring untuk setiap langkah yang percobaan oleh peningkatan epoch. Setelah langkah tersebut maka dilakukan evaluasi yang ditunjukkan pada classification report percobaan kedua untuk model pertama dan kedua sebagai berikut.

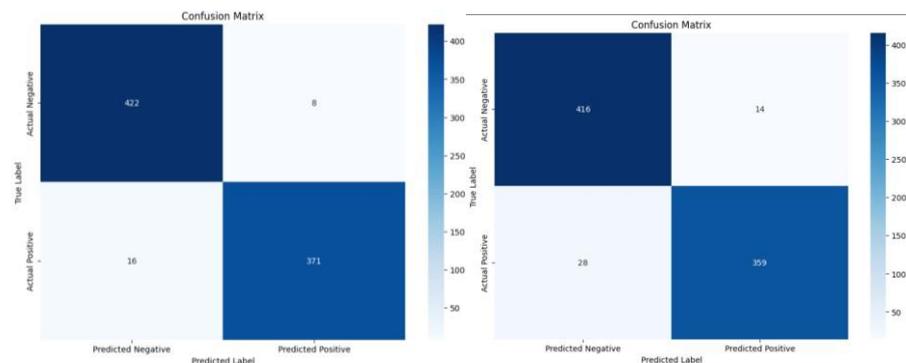
	precision	recall	F1-score	support
0	0,96	0,98	0,97	430
1	0,98	0,96	0,97	387
Akurasi			0,97	817

Tabel 7 Classification Report percobaan 2 model 1

	precision	recall	F1-score	support
0	0,94	0,97	0,95	430
1	0,96	0,93	0,94	387
Akurasi			0,95	817

Tabel 8 Classification Report percobaan 2 model 2

Pada tabel 7 dan 8 diatas menunjukkan bahwa hasil nilai akurasi percobaan kedua ini lebih baik dibandingkan dengan percobaan sebelumnya. Percobaan kedua ini memiliki nilai akurasi model pertama sejumlah 97% dan nilai loss sejumlah 20%. Model kedua memiliki jumlah akurasi 95% dan nilai loss sejumlah 21%. Selisih akurasi dari kedua model percobaan kedua ini adalah 2%. Dari proses pengolahan data yang dihasilkan percobaan kedua model pertama lebih baik atau mencukupi dibanding dengan percobaan pertama, dilihat dari hasil akurasi dan lamanya waktu eksekusi percobaan kedua model pertama yang paling cepat. Dapat di ketahui bahwa jika nilai akurasi semakin meningkat artinya performa yang dimiliki semakin baik, sedangkan untuk nilai loss semakin nilainya turun maka peforma semakin baik. Performa percobaan kedua dikatakan lebih baik dari percobaan sebelumnya karena penggunaan nilai hyperparameter pada random state pada k-fold sejumlah 42 , nilai epoch sebanyak 80,dan 32 untuk ukuran batch yang digunakan.



Gambar 9 Confusion matrix percobaan 2

Pada gambar 9. Yang di tunjukkan visualisasi confusion matrix pada model pertama tersebut memiliki nilai TN (True Negative) sejumlah 422, yang artinya jumlah sampel negatif yang diprediksi dengan benar sebagai negatif. Nilai FP (False Positive) sejumlah 8, yang artinya jumlah sampel negatif yang diprediksi salah sebagai positif. Nilai FN (False Negative) sejumlah 16, yang artinya jumlah 16 sampel positif yang diprediksi salah sebagai negatif. Nilai TP (True Positive) sejumlah 371, yang artinya jumlah 371 sampel positif yang diprediksi dengan benar.

Pada model kedua yang ditunjukkan pada gambar 8 confusion matrix diatas memiliki nilai TN sejumlah 416 sampel negatif yang diprediksi benar. Nilai FP sejumlah 14 sampel negatif yang diprediksi salah sebagai positif. Nilai FN sejumlah 28 sampel positif yang diprediksi sebagai negatif. Nilai TP yang berjumlah 359 sampel positif yang diprediksi dengan benar.