IMPLEMENTASI PERBANDINGAN MODEL ALGORITMA *DEEP LEARNING* BILSTM, CNN, DAN GRU UNTUK ANALISIS SENTIMEN KECANDUAN MEDIA SOSIAL

Unique Scovi Egnel, Sawali Wahyu
Teknik Informatika, Universitas Esa Unggul
Kampus Bekasi Jl. Harapan Indah Boulevard No. 2,
Pusaka Rakyat, Kec. Tarumajaya, Bekasi, Jawa Barat
uniquescoviegnel86@gmail.com

ABSTRAK

Ruang digital telah berkembang menjadi media utama bagi masyarakat untuk menyampaikan opini, sikap, dan respons terhadap berbagai fenomena sosial, termasuk kecenderungan kecanduan media sosial. Aktivitas tersebut menghasilkan data teks dalam jumlah besar bersifat tidak terstruktur, sehingga memerlukan pendekatan komputasional untuk dianalisis secara objektif dan berbasis data. Penelitian bertujuan untuk menganalisis sentimen kecanduan media sosial serta membandingkan kinerja tiga model deep learning, yaitu Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (BiLSTM), dan Gated Recurrent Unit (GRU). Data yang digunakan terdiri atas 12.730 cuitan berbahasa Indonesia yang dikumpulkan dari platform X (Twitter) selama periode 1 Januari hingga 31 Desember 2024. Proses pelabelan data dilakukan menggunakan pendekatan weak supervision berbasis aturan heuristik guna mendukung pengolahan data teks berskala besar. Tahapan pra-pemrosesan meliputi lowercase, pembersihan teks, normalisasi, tokenisasi, serta padding dan sequencing. Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model CNN memperoleh akurasi tertinggi sebesar 90,86%, diikuti oleh GRU sebesar 86,95% dan BiLSTM sebesar 86,26%. Distribusi sentimen terrdiri atas 28,1% sentimen positif, 33,5% sentimen netral, dan 38,4% sentimen negatif. Temuan ini menunjukkan bahwa CNN memiliki performa paling optimal dan stabil, serta menegaskan efektivitas pendekatan deep learning berbasis weak supervision dalam analisis sentimen media sosial berbahasa Indonesia.

Kata kunci : *Analisis Sentimen, Kecanduan Media Sosial, Deep Learning, Weak Supervision, X (Twitter)*

1. PENDAHULUAN

Ruang digital saat ini tidak hanya berfungsi sebagai media pertukaran informasi, tetapi juga berkembang menjadi wadah utama bagi masyarakat dalam menyampaikan pandangan, sikap, serta respons terhadap berbagai peristiwa. Melalui media sosial, pengguna dapat mengekspresikan pengalaman dan opini mereka secara terbuka dalam bentuk teks, sehingga membentuk kumpulan data yang merepresentasikan kecenderungan sikap serta pola respons pengguna di lingkungan digital. Di sisi lain, intensitas penggunaan media sosial yang tinggi berpotensi menimbulkan kecanduan, yang berdampak pada menurunnya produktivitas, gangguan kesehatan mental, serta berkurangnya kualitas interaksi sosial. Beberapa penelitian terdahulu mengaitkan kecanduan media sosial dengan kondisi psikologis tertentu dan menunjukkan bahwa fenomena tersebut dapat dianalisis menggunakan pendekatan komputasional, misalnya melalui penerapan algoritma *machine learning* berbasis data survei untuk mengidentifikasi tingkat ketergantungan pengguna[1].

Namun demikian, pendekatan berbasis survei umumnya belum memanfaatkan data teks yang dihasilkan langsung dari media sosial sebagai sumber utama analisis sentimen. Platform X (Twitter) merupakan salah satu media sosial yang banyak dimanfaatkan sebagai sarana penyampaian opini publik secara terbuka dan *real-time*. Karakteristik Twitter yang didominasi oleh teks singkat, bersifat publik, serta memiliki aliran percakapan yang cepat menghasilkan data opini dalam jumlah besar. Akan tetapi, data tersebut bersifat tidak terstruktur dan mengandung bahasa informal, sehingga menyulitkan proses analisis secara manual. Penelitian sebelumnya [2] menunjukkan bahwa model *Long Short-Term Memory* (LSTM) mampu memberikan kinerja yang baik dalam klasifikasi sentimen pada data Twitter. Meski demikian, penelitian tersebut berfokus pada isu kebijakan publik, sehingga belum membahas sentimen yang berkaitan dengan kecanduan media sosial, serta belum melakukan perbandingan menyeluruh antar arsitektur *deep learning*.

Berdasarkan kondisi tersebut, permasalahan dalam penelitian ini adalah belum adanya kajian yang secara khusus menganalisis sentimen kecanduan media sosial berbahasa Indonesia menggunakan data cuitan X (Twitter) serta membandingkan kinerja beberapa arsitektur *deep learning* secara komprehensif, terutama dengan memanfaatkan pendekatan *weak supervision* sebagai solusi atas keterbatasan data berlabel.

Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen kecanduan media sosial serta membandingkan kinerja tiga model *deep learning*, yaitu *Convolutional Neural Network* (CNN), *Bidirectional Long Short-Term Memory* (BiLSTM), dan *Gated Recurrent Unit* (GRU).

Untuk menyelesaikan permasalahan tersebut, penelitian ini dilakukan dengan memanfaatkan data

cuitan berbahasa Indonesia dari platform X (Twitter), menerapkan pendekatan *weak supervision* berbasis aturan heuristik dalam proses pelabelan data, serta mengevaluasi kinerja masing-masing model menggunakan metrik akurasi, presisi, *recall*, dan *F1-score.*

2. TINJAUAN PUSTAKA

Analisis sentimen merupakan salah satu pendekatan dalam *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi kecenderungan opini atau sikap pengguna terhadap suatu isu berdasarkan data teks, khususnya yang bersumber dari media sosial. Platform X (Twitter) banyak dimanfaatkan dalam penelitian analisis sentimen karena karakteristik datanya yang bersifat publik, *real-time*, serta mencerminkan respons masyarakat terhadap berbagai fenomena sosial. Penelitian oleh [3] menunjukkan bahwa penerapan analisis sentimen berbasis *machine learning* pada data Twitter mampu memberikan gambaran persepsi publik yang cukup akurat, termasuk pada isu kesehatan mental, meskipun masih menghadapi tantangan berupa data tidak terstruktur dan penggunaan bahasa informal.

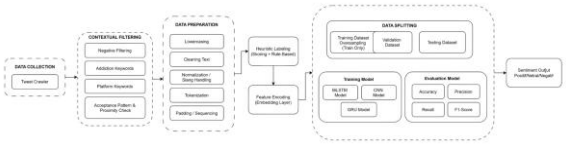
Seiring dengan meningkatnya intensitas penggunaan media sosial, fenomena kecanduan media sosial menjadi perhatian dalam berbagai kajian ilmiah karena berpotensi memengaruhi kondisi psikologis dan perilaku sosial individu. Beberapa penelitian terdahulu mengungkapkan bahwa penggunaan media sosial secara berlebihan berkorelasi dengan meningkatnya tingkat stres, kecemasan, dan gangguan emosional. Studi yang dilakukan oleh [1] memanfaatkan pendekatan machine learning berbasis data survei untuk mengidentifikasi tingkat kecanduan media sosial, dan menunjukkan bahwa metode komputasional mampu memberikan analisis yang lebih objektif dibandingkan pendekatan konvensional. Namun, penelitian tersebut belum memanfaatkan data teks yang dihasilkan langsung dari media sosial sebagai sumber utama analisis sentimen.

Dalam konteks analisis sentimen berbasis teks, penggunaan model deep learning dilaporkan mampu meningkatkan performa klasifikasi dibandingkan metode machine learning tradisional. Arsitektur *Convolutional Neural Network* (CNN) efektif dalam menangkap pola lokal pada teks, sementara *Bidirectional Long Short-Term Memory* (BiLSTM) dan *Gated Recurrent Unit* (GRU) unggul dalam memodelkan ketergantungan sekuensial dan konteks jangka panjang. Penelitian oleh [4] membandingkan beberapa arsitektur deep learning pada data media sosial dan menunjukkan bahwa setiap model memiliki karakteristik performa yang berbeda, sehingga perbandingan antar model diperlukan untuk memperoleh hasil klasifikasi sentimen yang optimal dan stabil.

Meskipun demikian, penerapan deep learning dalam analisis sentimen masih menghadapi kendala utama berupa keterbatasan data berlabel dalam jumlah besar. Untuk mengatasi permasalahan tersebut, pendekatan *weak supervision* dikembangkan dengan memanfaatkan sinyal tidak langsung, seperti emoji, kata kunci, atau pola linguistik tertentu, sebagai dasar pelabelan otomatis. Penelitian sebelumnya [5] menunjukkan bahwa pendekatan *weak supervision* mampu menghasilkan dataset berskala besar secara efisien dan meningkatkan kinerja model *deep learning* dalam klasifikasi sentimen dan emosi pada data Twitter, meskipun label yang dihasilkan bersifat lemah dan mengandung noise. Oleh karena itu, pendekatan ini dinilai relevan untuk diterapkan dalam analisis sentimen kecanduan media sosial yang memanfaatkan data teks media sosial berbahasa Indonesia.

3. METODE PENELITIAN

Metode penelitian pada studi ini disusun untuk mengkaji sentimen kecanduan media sosial dengan memanfaatkan data cuitan yang diperoleh dari platform X (Twitter) serta menerapkan pendekatan *deep* learning. Rangkaian tahapan penelitian meliputi proses pengumpulan data, pra-pemrosesan teks, dan pelabelan data secara otomatis menggunakan pendekatan *weak supervision* berbasis aturan heuristik. Data yang telah melalui tahapan tersebut selanjutnya digunakan dalam proses pemodelan dan evaluasi kinerja model.



Gambar 1. Alur Penelitian Analisis Sentimen

Gambar 1 menunjukkan alur penelitian analisis sentimen yang diawali dengan transformasi data cuitan berlabel ke dalam representasi numerik agar dapat diproses oleh model deep learning. Data tersebut selanjutnya dianalisis menggunakan tiga arsitektur, yaitu CNN, BiLSTM, dan GRU yang memiliki karakteristik berbeda dalam memodelkan pola linguistik pada teks [6]. Kinerja masing-masing model kemudian dievaluasi menggunakan metrik klasifikasi untuk menilai tingkat akurasi dan konsistensi dalam mengidentifikasi sentimen kecanduan media sosial.

3.1. Pengumpulan Data

Pada penelitian ini, pengumpulan data dilakukan dari platform X (Twitter) dengan menerapkan teknik *web crawling* melalui Twitter API menggunakan bahasa pemrograman Python. Proses akuisisi data berlangsung selama periode 1 Januari hingga 31 Desember 2024 dan difokuskan pada cuitan berbahasa Indonesia yang memuat kata kunci serta frasa yang berkaitan dengan kecanduan penggunaan media sosial. Pendekatan ini bertujuan memastikan bahwa data yang diperoleh tetap relevan dengan topik penelitian.

Hasil dari proses *crawling* menghasilkan sebanyak 12.730 cuitan yang selanjutnya

dimanfaatkan sebagai dataset penelitian. Atribut *full_text* dipilih sebagai sumber utama data karena mampu menggambarkan opini pengguna secara lebih lengkap dan kontekstual, sehingga sesuai untuk kebutuhan analisis sentimen berbasis teks[7].

3.2. Pra-Pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk membersihkan serta menyeragamkan data teks sebelum digunakan dalam proses pelatihan model *deep learning*. Proses ini mencakup pembersihan teks dari elemen yang tidak relevan, penerapan *lowercase* dengan mengonversi seluruh huruf menjadi huruf kecil, normalisasi kata tidak baku, serta tokenisasi. Tujuan dari tahapan ini adalah untuk meminimalkan noise yang umum terdapat pada data media sosial dan meningkatkan kualitas representasi teks, sebagaimana lazim diterapkan dalam penelitian analisis sentimen berbasis *Natural Language Processing* (NLP)[8].

3.3. Pelabelan Data Berbasis *Weak Supervision*

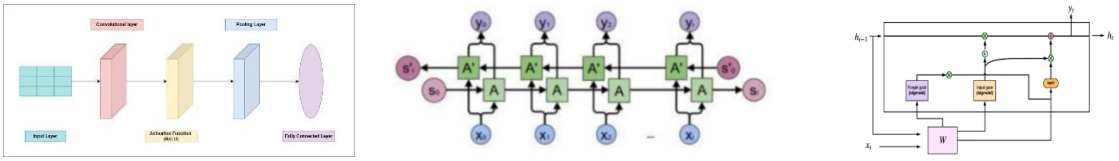
Pelabelan data dalam penelitian ini diterapkan menggunakan pendekatan *weak supervision* yang memanfaatkan aturan heuristik serta pola linguistik untuk mengatasi keterbatasan ketersediaan data berlabel manual dalam skala besar[9]. Setiap cuitan diberi label secara otomatis ke dalam kategori sentimen positif, netral, atau negatif berdasarkan keberadaan frasa tertentu yang mengindikasikan kecenderungan kecanduan media sosial. Pendekatan ini dipilih karena mampu menghasilkan data berlabel dalam jumlah besar secara efisien tanpa memerlukan proses anotasi manual yang kompleks. Selain itu, metode *weak supervision* telah banyak digunakan dan terbukti efektif dalam berbagai penelitian analisis sentimen dan klasifikasi teks [10].

3.4. Pembagian Dataset

Dataset yang telah melalui proses pelabelan selanjutnya dipisahkan menjadi dua subset, yaitu data latih dan data uji. Proporsi pembagian yang digunakan adalah 80% untuk data latih dan 20% untuk data uji, dengan tujuan memastikan model memperoleh data yang memadai selama proses pelatihan serta dapat dievaluasi secara objektif menggunakan data yang belum pernah digunakan sebelumnya.

Proses pembagian data dilakukan secara acak dengan tetap menjaga proporsi kelas sentimen pada masing-masing subset agar distribusi data tetap seimbang. Pendekatan pembagian dataset ini banyak diterapkan dalam penelitian klasifikasi teks dan analisis sentimen karena dinilai efisien serta mampu memberikan gambaran performa model yang representatif [11].

3.5. Arsitektur Model



Gambar 2. Arsitektur Model CNN, BiLSTM dan GRU

Gambar 2 menunjukkan arsitektur model CNN, BiLSTM, dan GRU yang digunakan dalam penelitian. CNN memanfaatkan lapisan konvolusi untuk mengekstraksi pola lokal pada teks, seperti kombinasi kata yang merepresentasikan sentimen pengguna secara efektif [12]. Sementara itu, BiLSTM memproses data teks secara dua arah, yaitu dari awal ke akhir urutan dan sebaliknya, sehingga memungkinkan model memahami konteks kata dengan mempertimbangkan informasi sebelum dan sesudahnya secara bersamaan [13]. GRU merupakan varian jaringan saraf rekuren dengan struktur gerbang yang lebih sederhana dibandingkan LSTM, di mana mekanisme *update gate* dan *reset gate* digunakan untuk mengatur aliran informasi yang relevan dalam urutan teks, sehingga mampu memodelkan ketergantungan sekuensial secara efisien[13].

3.6. Evaluasi Model

Tahap evaluasi dilakukan untuk menilai kinerja model CNN, BiLSTM, dan GRU dalam mengklasifikasikan sentimen kecanduan media sosial menggunakan data uji yang tidak dilibatkan pada proses pelatihan. Penilaian performa model dilakukan dengan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, *confusion matrix* digunakan sebagai pendukung untuk menggambarkan distribusi hasil prediksi pada masing-masing kelas sentimen [15].

Tabel 1. Metrik Evaluasi Model

Metrik	Rumus
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F1-score	$\frac{2\ x\ (Precision\ x\ Recall)}{Precision\ +\ Recall}$

Tabel 1 merupakan perhitungan metrik evaluasi pada penelitian ini mengacu pada rumus yang disajikan pada.

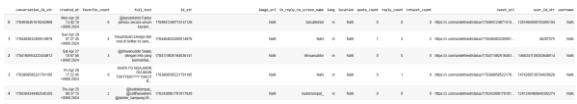
3.7. Visualisasi

Visualisasi digunakan sebagai sarana pendukung dalam proses interpretasi hasil, sehingga analisis klasifikasi tidak hanya terbatas pada nilai metrik kuantitatif, tetapi juga dapat dipahami secara kontekstual. Pada penelitian ini, hasil visualisasi disajikan dalam bentuk grafik distribusi sentimen untuk menggambarkan kecenderungan sentimen secara keseluruhan pada setiap kategori. Selain itu, *wordcloud* dimanfaatkan untuk menampilkan kata atau frasa yang paling dominan pada masing-masing kelas sentimen, sehingga membantu mengidentifikasi pola bahasa serta konteks percakapan yang muncul dari pengguna media sosial.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Tahap pengumpulan data pada penelitian ini menghasilkan sebanyak 12.730 cuitan berbahasa Indonesia yang diperoleh dari platform X (Twitter) dalam rentang waktu 1 Januari hingga 31 Desember 2024.

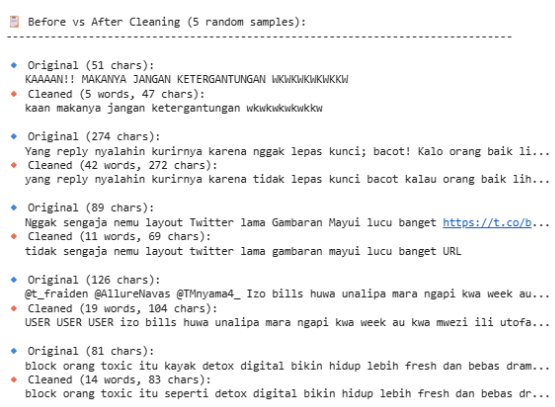


Gambar 3. Hasil Crawling Data

Gambar 3 menunjukan hasil dari *crawling data* menggunakan Twitter API, dengan pemilihan kata kunci dan frasa yang berkaitan dengan kecanduan penggunaan media sosial guna menjaga kesesuaian data dengan topik penelitian.

4.2. *Lowercase, Cleaning Text & Normalization*

Tahap ini dilakukan prepemrosesan data meliputi lowercase, cleaning dan normalizations dari hasil data set yang telah didapatkan, berikut digambarkan pada gambar 4 :

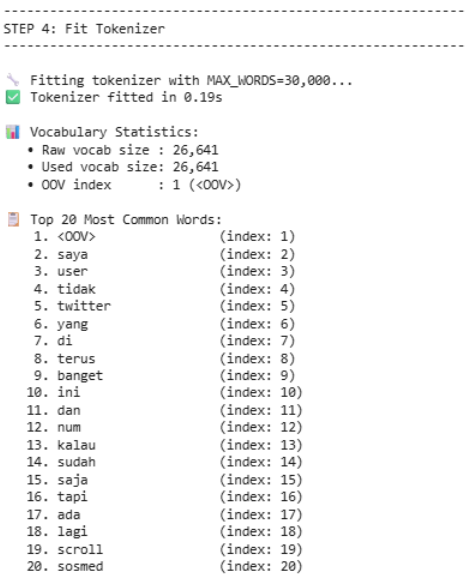


Gambar 4. Hasil *dari Lowercase, Cleaning Text & Normalization*

Pada gambar 4 menunjukan hasil dari *Lowercase, Cleaning Text dan Normalization.* Pada tahap ini, proses *lowercase* diterapkan untuk menyeragamkan seluruh karakter huruf menjadi huruf kecil guna menghindari perbedaan makna yang disebabkan oleh variasi kapitalisasi. Selanjutnya, dilakukan *cleaning text* untuk menghilangkan elemen-elemen yang tidak relevan, seperti URL, *mention*, simbol, serta karakter yang berulang dan tidak memiliki kontribusi terhadap makna sentimen. Selain itu, proses normalisasi kata dilakukan untuk mengubah kata tidak baku atau bentuk bahasa informal menjadi bentuk yang lebih standar.

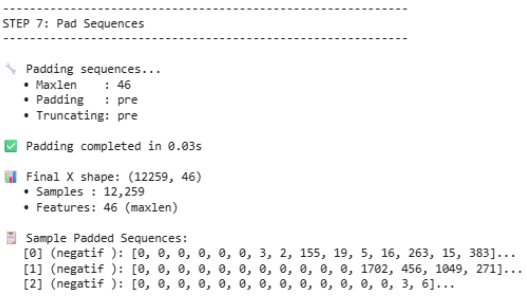
4.3. *Tokenization, Sequencing dan Padding*

Tahap tokenisasi menghasilkan proses konversi teks yang telah melalui pra-pemrosesan ke dalam bentuk representasi numerik berdasarkan indeks kata pada kosakata yang terbentuk. Melalui proses ini, setiap kata dipetakan ke dalam nilai numerik sesuai frekuensi kemunculannya, sehingga terbentuk ukuran kosakata yang merepresentasikan frasa-frasa yang sering muncul dalam cuitan.



Gambar 5. Hasil *Tokenizer*

Pada gambar 5 menunjukan hasil tokenizer yang me representasi numerik, dan selanjutnya digunakan sebagai masukan bagi model *deep learning* untuk proses pemodelan lebih lanjut.



Gambar 6. Hasil *Padding* dan *Sequence*

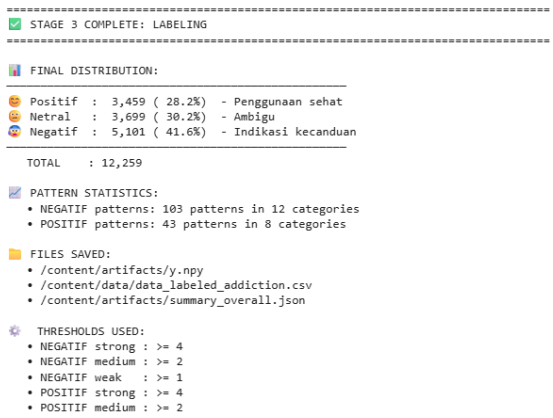
Gambar 6 menampilkan hasil proses *padding* dan *sequencing* yang dilakukan untuk menyeragamkan

panjang setiap urutan teks. Pada tahap ini, seluruh cuitan disesuaikan dengan panjang maksimum yang telah ditentukan dengan cara menambahkan nilai nol pada bagian awal urutan teks (*pre-padding*), sehingga seluruh data memiliki dimensi yang sama dan dapat diproses oleh model *deep learning.*

4.4. Hasil Labelling

Pelabelan data dalam penelitian ini diterapkan menggunakan pendekatan *weak supervision* yang memanfaatkan aturan heuristik dan mekanisme *threshold scoring.*

Pada gambar 7 merupakan hasil dari labelling, yang menunjukan pendekatan menggunakan pola linguistik yang dikelompokkan ke dalam *negative patterns*, *positive patterns*, dan *social media context patterns* sebagai dasar dalam penentuan label sentiment.



Gambar 7. Hasil Final Labelling

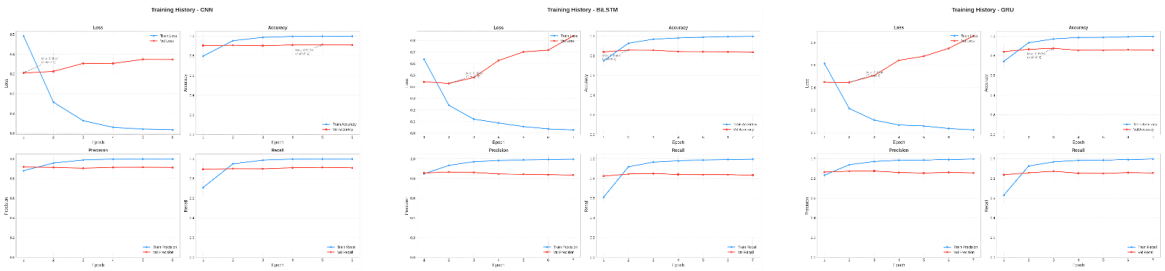
4.5. Visualisasi Wordcloud

Gambar 8. Visualisasi Wordcloud Positif, Netral dan Negatif

Gambar 8 menampilkan visualisasi *wordcloud* untuk tiga kelas sentimen, yaitu positif, netral, dan negatif. Pada kelas positif, kata-kata yang muncul didominasi oleh istilah yang menggambarkan penggunaan media sosial dalam intensitas ringan. Kelas netral ditandai oleh kemunculan kata-kata yang bersifat informatif serta tidak menunjukkan muatan emosional yang kuat. Sementara itu, *wordcloud* pada kelas negatif didominasi oleh istilah yang berkaitan dengan penggunaan media sosial secara berlebihan dan tekanan emosional, yang mencerminkan kecenderungan sentimen negatif terhadap isu kecanduan media social.

4.6. Hasil Pelatihan Model

Proses pelatihan model sebanyak 80% data yang didapat. Dengan menggunakan tiga model sebagai berikut:



Gambar 9. Grafik Histori Training Model CNN, BiLSTM dan CNN

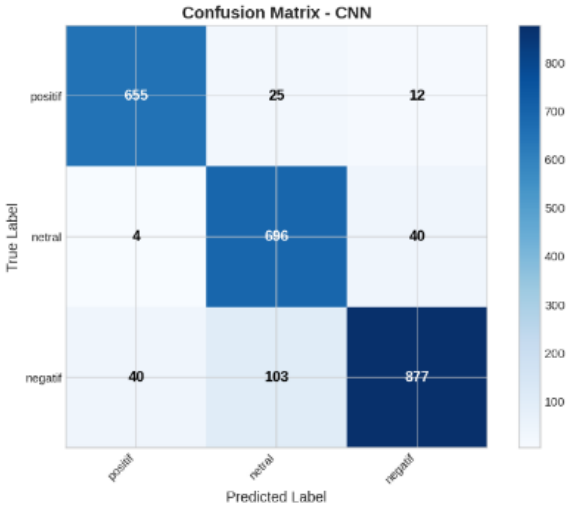
Gambar 9 menyajikan grafik riwayat pelatihan model CNN, BiLSTM, dan GRU yang menggambarkan perubahan nilai *loss*, *accuracy*, *precision*, dan *recall* selama proses pelatihan. Pada model CNN, terlihat penurunan nilai *loss* secara bertahap yang diikuti dengan peningkatan dan kestabilan metrik evaluasi, menunjukkan kemampuan model dalam mempelajari pola sentimen secara konsisten tanpa indikasi *overfitting* yang signifikan. Sementara itu, model BiLSTM juga menunjukkan penurunan *loss* dan peningkatan metrik evaluasi, meskipun terdapat perbedaan antara kurva pelatihan dan validasi yang mengindikasikan kecenderungan *overfitting* ringan seiring kompleksitas arsitekturnya. Adapun model GRU memperlihatkan pola pelatihan yang relatif stabil dengan penurunan *training loss* serta peningkatan metrik evaluasi, di mana kinerja pada data

validasi tetap menunjukkan kestabilan meskipun terdapat selisih pada nilai *loss*.

4.7. Hasil Evaluasi Model

Tahap evaluasi dilakukan menggunakan data uji sebesar 20% dari keseluruhan dataset untuk mengukur kemampuan generalisasi model terhadap data yang tidak terlibat dalam proses pelatihan. Penilaian kinerja model dilakukan dengan menerapkan metrik *accuracy, precision, recall, dan F1-score*. Selain itu, analisis confusion matrix digunakan untuk mengidentifikasi dan mengevaluasi pola kesalahan klasifikasi yang dihasilkan oleh masing-masing model.

4.8. Hasil Evaluasi CNN



Gambar 10. *Confussion Matrix* 3x3 CNN

Pada gambar 13 menunjukkan *confusion matrix* model CNN yang menggambarkan kinerja klasifikasi sentimen positif, netral, dan negatif.

Tabel 2. Metrik Evaluasi per Kelas CNN

Kelas Sentimen	Precision	Recall	F1-score	Support
Positif	0.9371	0.9465	0.9418	692
Netral	0.8447	0.9405	0.8900	740
Negatif	0.9440	0.8598	0.8999	1020
Macro Avg	0.9086	0.9156	0.9106	2452
Weighted Avg	0.9121	0.9086	0.9088	2452

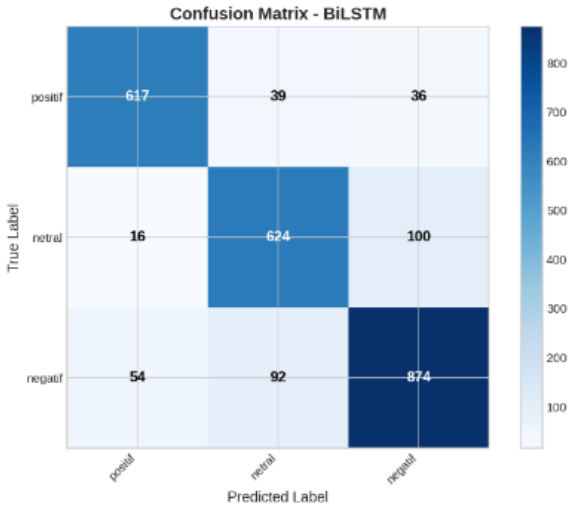
Tabel 2 menunjukkan bahwa model CNN memiliki kinerja klasifikasi yang baik pada seluruh kelas sentimen. Kelas positif dan negatif mencapai nilai *precision* dan *F1-score* yang tinggi, sedangkan kelas netral menunjukkan *recall* tinggi dengan *precision* yang lebih rendah. Nilai *macro average* dan *weighted average* yang tinggi menegaskan bahwa model CNN memiliki performa yang seimbang.

Tabel 3. Validasi *Confussion Matrix* CNN

	Negatif	Netral	Positif
Presisi	$= \frac{877}{877+12+40} = 0{,}9440$	$= \frac{696}{696+25+103} = 0{,}8447$	$= \frac{655}{655+4+40} = 0{,}9371$
Recall	$= \frac{877}{877+40+103} = 0{,}8598$	$= \frac{696}{696+40+4} = 0{,}9405$	$= \frac{655}{655+25+12} = 0{,}9465$
F1-score	$= 2 \text{ x } \frac{0{,}9440 \text{ x } 0{,}8598}{0{,}9440 + 0{,}8598} = 0{,}8999$	$= 2 \text{ x } \frac{0{,}8447 \text{ x } 0{,}9405}{0{,}8447 + 0{,}9405} = 0{,}8900$	$= 2 \text{ x } \frac{0{,}9371 \text{ x } 0{,}9465}{0{,}9371 + 0{,}9465} = 0{,}9418$
Accuracy	$= \frac{655+696+877}{2452} = \frac{2228}{2452} = 0{,}9086$		
Presisi	$= \frac{0{,}9371+0{,}8447+0{,}9440}{3} = \frac{2{,}7258}{3} = 0{,}9086$		
Recall	$= \frac{0{,}9465+0{,}9405+0{,}8598}{3} = \frac{2{,}7468}{3} = 0{,}9156$		
F1-score	$= \frac{0{,}9418+0{,}8900+0{,}8999}{3} = \frac{2{,}7468}{3} = 0{,}9106$		

Tabel 3 menunjukkan bahwa model CNN memiliki performa klasifikasi yang baik pada ketiga kelas sentimen. Kelas positif dan negatif mencapai nilai *precision* dan *F1-score* yang tinggi, sementara kelas netral memiliki *recall* lebih tinggi dengan *precision* yang relatif lebih rendah. Secara keseluruhan, nilai *accuracy* dan *macro average* yang tinggi mengindikasikan kinerja CNN yang seimbang dalam mengklasifikasikan sentimen kecanduan media sosial.

4.9. Hasil Evaluasi BiLSTM



Gambar 11. *Confussion Matrix 3x3* BiLSTM

Gambar 14 menampilkan nilai pada diagonal utama merepresentasikan jumlah prediksi yang benar untuk masing-masing kelas, sedangkan nilai di luar diagonal menunjukkan kesalahan klasifikasi antar kelas. Hasil ini menunjukkan bahwa model BiLSTM mampu melakukan klasifikasi sentimen dengan cukup baik, meskipun masih terdapat kesalahan prediksi terutama antara kelas netral dan negatif.

Tabel 4. Metrik Evaluasi per Kelas BiLSTM

Kelas Sentimen	Precision	Recall	F1-score	Support
Positif	0.8981	0.8916	0.8949	692
Netral	0.8265	0.8432	0.8348	740
Negatif	0.8653	0.8569	0.8611	1020
Macro Avg	0.8633	0.8639	0.8636	2452
Weighted Avg	0.8629	0.8626	0.8627	2452

Tabel 4 kelas positif memiliki nilai *precision*, *recall*, dan *F1-score* tertinggi, menandakan kemampuan model dalam mengenali sentimen positif dengan baik. Kelas netral menunjukkan performa yang relatif lebih rendah dibandingkan kelas lainnya, yang mengindikasikan adanya kesulitan model dalam membedakan sentimen netral dari kelas lain. Secara keseluruhan, nilai *macro average* dan *weighted average* yang mendekati 0,86 menunjukkan bahwa model BiLSTM memiliki performa yang cukup baik dan seimbang pada seluruh kelas sentimen.

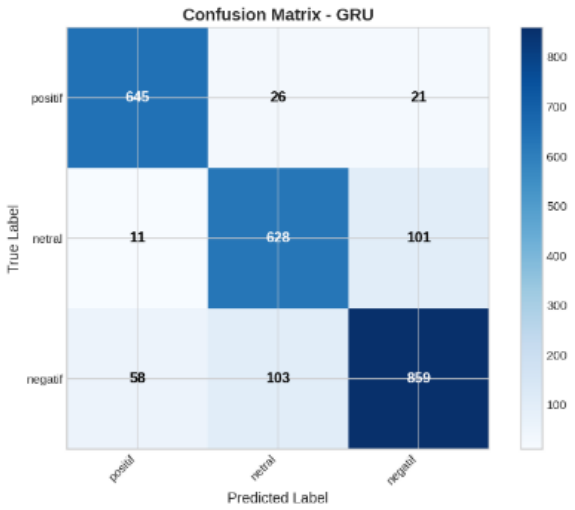
Tabel 5. Validasi *Confussion Matrix* BiLSTM

	Negatif	Netral	Positif
Presisi	$=\frac{874}{874+36+100}$ = 0,8653	$=\frac{624}{624+39+92}$ = 0,8265	$=\frac{617}{617+16+54}$ = 0,8981
Recall	$=\frac{874}{874+36+100}$ = 0,8569	$=\frac{624}{624+16+100}$ = 0,8432	$=\frac{617}{617+39+36}$ = 0,8916
F1-score	$=2 \times \frac{0,8653 \times 0,8569}{0,8653+0,8569}$ = 0,8611	$=2 \times \frac{0,8265 \times 0,8432}{0,8265+0,8432}$ = 0,8348	$=2 \times \frac{0,8981 \times 0,8916}{0,8981+0,8916}$ = 0,8949
Accuracy	$=\frac{617+624+874}{2452}=\frac{2115}{2452}$ = 0,8626		
Presisi	$=\frac{0,8981+0,8265+0,8653}{3}=\frac{2,5899}{3}$ = 0,8633		
Recall	$=\frac{0,8916+0,8432+0,8569}{3}=\frac{2,5917}{3}$ = 0,8639		
F1-score	$=\frac{0,8949+0,8348+0,8611}{3}=\frac{2,5908}{3}$ = 0,8636		

Tabel 5 menunjukkan kelas positif memiliki nilai precision dan *F1-score* tertinggi, menunjukkan bahwa model mampu mengenali sentimen positif dengan baik. Kelas netral memiliki nilai precision paling rendah, yang mengindikasikan adanya kesulitan dalam membedakan sentimen netral dari kelas lainnya. Secara keseluruhan, nilai *accuracy* dan *macro average* yang berada di kisaran 0,86 menunjukkan bahwa model BiLSTM memiliki performa yang cukup baik dan relatif seimbang dalam klasifikasi sentimen.

4.10. Hasil Evaluasi GRU

Gambar 15 menampilkan nilai pada diagonal utama menunjukkan jumlah prediksi yang benar pada masing-masing kelas, sedangkan nilai di luar diagonal menunjukkan kesalahan klasifikasi antar kelas. Secara umum, model GRU mampu mengklasifikasikan sentimen dengan cukup baik, namun masih terdapat kesalahan prediksi yang relatif lebih tinggi pada kelas netral, terutama yang diprediksi sebagai sentimen negatif.



Gambar 12. *Confussion Matrix 3x3* GRU

Tabel 6. Metrik Evaluasi per Kelas GRU

Kelas Sentimen	Precision	Recall	F1-score	Support
Positif	0.9034	0.9321	0.9175	692
Netral	0.8296	0.8486	0.8390	740
Negatif	0.8756	0.8422	0.8586	1020
Macro Avg	0.8695	0.8743	0.8717	2452
Weighted Avg	0.8696	0.8695	0.8693	2452

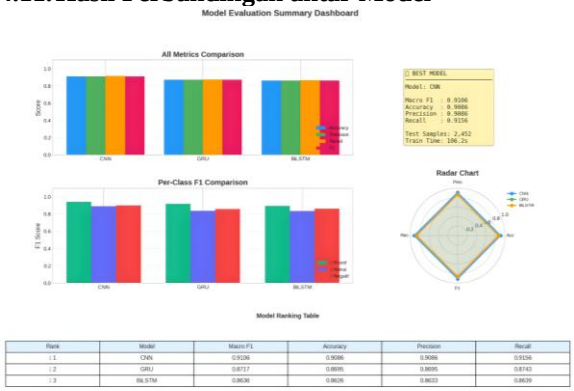
Tabel 6 menunjukkan kelas positif memiliki nilai *recall* dan *F1-score* tertinggi, menandakan kemampuan model dalam mengenali sentimen positif dengan baik. Kelas netral menunjukkan nilai *precision* dan *recall* yang lebih rendah dibandingkan kelas lainnya, yang mengindikasikan adanya kesulitan dalam membedakan sentimen netral dari kelas lain. Secara keseluruhan, nilai *macro average* dan *weighted average* yang berada di kisaran 0,87 menunjukkan bahwa model GRU memiliki performa yang cukup baik dan relatif seimbang dalam klasifikasi sentiment.

Tabel 7. Validasi *Confussion Matrix* GRU

	Negatif	Netral	Positif
Presisi	$=\frac{859}{859+21+101}$ = 0,8756	$=\frac{628}{628+26+103}$ = 0,8296	$=\frac{645}{645+11+58}$ = 0,9034
Recall	$=\frac{859}{859+58+103}$ = 0,8422	$=\frac{628}{628+11+101}$ = 0,8486	$=\frac{645}{645+26+21}$ = 0,9321
F1-score	$=2 \times \frac{0,8756 \times 0,8422}{0,8756+0,8422}$ = 0,8586	$=2 \times \frac{0,8296 \times 0,8486}{0,8296+0,8486}$ = 0,8390	$=2 \times \frac{0,9034 \times 0,9321}{0,9034+0,9321}$ = 0,9175
Accuracy	$=\frac{645+628+859}{2452}=\frac{2132}{2452}$ = 0,8695		
Presisi	$=\frac{0,9034+0,8296+0,8756}{3}=\frac{2,5752}{3}$ = 0,8695		
Recall	$=\frac{0,9321+0,8486+0,8422}{3}=\frac{2,6229}{3}$ = 0,8743		
F1-score	$=\frac{0,9175+0,8390+0,8586}{3}=\frac{2,6151}{3}$ = 0,8717		

Tabel 7 menunjukkan kelas positif memiliki nilai recall dan *F1-score* tertinggi, yang menunjukkan bahwa model GRU mampu mengenali sentimen positif dengan baik. Kelas netral memiliki nilai *precision* dan *recall* yang relatif lebih rendah, mengindikasikan adanya tumpang tindih prediksi dengan kelas lain. Secara keseluruhan, nilai *accuracy* dan macro average yang berada di kisaran 0,87 menunjukkan bahwa model GRU memiliki performa yang cukup baik dan stabil dalam mengklasifikasikan sentimen.

4.11. Hasil Perbandingan antar Model



Gambar 13. Perbandingan Hasil CNN, BiLSTM dan GRU

Gambar 16 menyajikan ringkasan perbandingan kinerja antara model CNN, GRU, dan BiLSTM berdasarkan metrik evaluasi utama. Hasil analisis menunjukkan bahwa model CNN mencapai performa paling unggul dengan nilai *macro F1-score* tertinggi, kemudian diikuti oleh model GRU dan BiLSTM. Evaluasi per kelas sentimen serta visualisasi *radar chart* memperlihatkan bahwa CNN memiliki keseimbangan performa yang lebih konsisten pada seluruh metrik yang digunakan. Oleh karena itu, model CNN ditetapkan sebagai arsitektur yang paling optimal untuk analisis sentimen kecanduan media sosial pada penelitian ini.

5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa model Convolutional Neural Network (CNN) memiliki kinerja paling optimal dalam analisis sentimen kecanduan media sosial dibandingkan BiLSTM dan GRU, dengan akurasi tertinggi sebesar 90,86%, diikuti oleh GRU (86,95%) dan BiLSTM (86,26%). Nilai *macro F1-score* dan hasil *confusion matrix* mengindikasikan bahwa CNN mampu menghasilkan klasifikasi yang lebih konsisten pada seluruh kelas sentimen. Distribusi sentimen menunjukkan dominasi sentimen negatif sebesar 38,4%, diikuti sentimen netral 33,5% dan positif 28,1%, yang mencerminkan kecenderungan respons pengguna terhadap isu kecanduan media sosial. Temuan ini menegaskan efektivitas CNN dalam menangkap pola lokal pada teks serta keberhasilan penerapan pendekatan *deep learning* berbasis *weak supervision* pada data cuitan berbahasa Indonesia.

Sebagai pengembangan selanjutnya, disarankan penerapan pendekatan *semi-supervised learning* serta eksplorasi arsitektur berbasis *attention* atau model *transformer* seperti IndoBERT untuk menangkap konteks semantik yang lebih kompleks.

DAFTAR PUSTAKA

[1] M. N. Mim, M. Firoz, M. M. Islam, M. Hasan, and M. T. Habib, “A study on social media addiction analysis on the people of Bangladesh using machine learning algorithms,” *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 5, pp. 3493–3502, Oct. 2024, doi: 10.11591/eei.v13i5.5680.

[2] F. M. Naufal and S. F. Kusuma, “JEPIN (Jurnal Edukasi dan Penelitian Informatika) Analisis Sentimen pada Media Sosial Twitter Terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat Berbasis Deep Learning,” 2022.

[3] P. A. Permatasari, Linawati, and L. Jasa, “Survei Tentang Analisis Sentimen Pada Media Sosial,” *Majalah Ilmiah Teknologi Elektro*, vol. 20, no. 2, 2021, doi: 10.24843/MITE.2021.v20i02.P01.

[4] K. L. Tan, C. P. Lee, K. M. Lim, and K. S. M. Anbananthen, “Sentiment Analysis With Ensemble Hybrid Deep Learning Model,” *IEEE Access*, vol. 10, pp. 103694–103704, 2022, doi: 10.1109/ACCESS.2022.3210182.

[5] M. Kastrati, Z. Kastrati, A. Shariq Imran, and M. Biba, “Leveraging distant supervision and deep learning for twitter sentiment and emotion classification,” *J Intell Inf Syst*, vol. 62, no. 4, pp. 1045–1070, Aug. 2024, doi: 10.1007/s10844-024-00845-0.

[6] S. Li and B. Gong, “Word embedding and text classification based on deep learning methods,” *MATEC Web of Conferences*, vol. 336, p. 06022, 2021, doi: 10.1051/matecconf/202133606022.

[7] A. Z. R. Adam and E. B. Setiawan, “Social Media Sentiment Analysis Using Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU),” *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 9, no. 1, pp. 119–131, Feb. 2023, doi: 10.26555/jiteki.v9i1.25813.

[8] Y. Kang, Z. Cai, C. W. Tan, Q. Huang, and H. Liu, “Natural language processing (NLP) in management research: A literature review,” Apr. 02, 2020, *Taylor and Francis Ltd.* doi: 10.1080/23270012.2020.1756939.

[9] H. Utama, E. Daniati, And A. Masruro, “Weak Supervision Dengan Pendekatan Labeling Function Untuk Analisis Sentimen Pada Twitter,” 2024. [Online]. Available: Https://Subset.Id/Index.Php/Ijcsr

[10] J. Zhang, L. Song, and A. Ratner, “Leveraging Instance Features for Label Aggregation in Programmatic Weak Supervision,” 2023. [Online]. Available: https://github.com/

[11] Q. H. Nguyen *et al.*, "Influence of data splitting on performance of machine learning models in prediction of shear strength of soil," *Math Probl Eng*, vol. 2021, 2021, doi: 10.1155/2021/4832864.

[12] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.

[13] M. M. Agüero-Torales, J. I. Abreu Salas, and A. G. López-Herrera, "Deep learning and multilingual sentiment analysis on social media data: An overview," *Appl Soft Comput*, vol. 107, Aug. 2021, doi: 10.1016/j.asoc.2021.107373