Perbandingan Metode Support Vector Machine dan Convolutional Neural Network untuk Analisis Sentimen pada Google Classroom di Google Play Store

Kerenhapukh Priskila Waworuntu[1] , Kevin Bryan Kalengkongan[2], Muh. Fakhril Huri[3], Sary Paturusi[4]
[1,2,3,4]Universitas Sam Ratulangi, Manado
Email : wkerenhapukh@gmail.com[1], kvnbryank@gmail.com[2], fakhrilhuri73@gmail.com[3], sarypaturusi@unsrat.ac.id[4]

Abstrak

Google Classroom merupakan salah satu platform pembelajaran daring yang banyak digunakan dan menyediakan fitur ulasan pada Google Play Store sebagai sarana bagi pengguna untuk menyampaikan pengalaman, keluhan, maupun penilaian terhadap aplikasi. Ulasan tersebut dapat dimanfaatkan untuk memperoleh informasi mengenai persepsi pengguna melalui analisis sentimen. Penelitian ini bertujuan membandingkan kinerja Support Vector Machine (SVM) dan Convolutional Neural Network (CNN) pada analisis sentimen ulasan Google Classroom. SVM dipilih karena merupakan metode klasifikasi teks yang banyak digunakan dan memiliki performa yang baik pada data berdimensi tinggi, sedangkan CNN dipilih karena mampu mempelajari representasi fitur secara otomatis dari data teks. Data penelitian diperoleh melalui proses *web scraping* dan menghasilkan 1806 ulasan, yang setelah melalui tahap *preprocessing* menjadi 1735 ulasan. Hasil pelabelan sentimen menunjukkan distribusi kelas yang tidak seimbang, yaitu 788 ulasan netral, 641 ulasan negatif, dan 306 ulasan positif. Proses klasifikasi dilakukan menggunakan tiga skenario, yaitu TF-IDF + SVM, TF-IDF + Chi-Square + SVM, dan CNN. Hasil penelitian menunjukkan bahwa penerapan Chi-Square mampu meningkatkan akurasi SVM dari 93% menjadi 94% melalui pemilihan fitur yang lebih relevan. Sementara itu, CNN menghasilkan performa terbaik dengan akurasi sebesar 96%. Temuan ini menunjukkan bahwa CNN lebih efektif dalam menangkap pola sentimen pada ulasan pengguna dibandingkan pendekatan berbasis SVM pada dataset yang digunakan.
Kata Kunci: Analisis Sentimen, Google Classroom, SVM, CNN, TF-IDF

Abstract

Google Classroom is one of the most widely used online learning platforms and provides a review feature on Google Play Store that allows users to express their experiences, complaints, and evaluations of the application. These reviews can be utilized to understand user perceptions through sentiment analysis. This study aims to compare the performance of Support Vector Machine (SVM) and Convolutional Neural Network (CNN) in sentiment analysis of Google Classroom reviews. SVM was selected because it is a widely used text classification method with strong performance on high-dimensional data, while CNN was chosen for its ability to automatically learn feature representations from textual data. The dataset was collected through a web scraping process, resulting in 1806 reviews, which were reduced to 1735 reviews after preprocessing. Sentiment labeling produced an imbalanced class distribution consisting of 788 neutral reviews, 641 negative reviews, and 306 positive reviews. Three experimental scenarios were evaluated: TF-IDF + SVM, TF-IDF + Chi-Square + SVM, and CNN. The results show that the application of Chi-Square feature selection improved SVM accuracy from 93% to 94% by selecting more relevant features. Meanwhile, CNN achieved the highest performance with an accuracy of 96%. These findings indicate that CNN is more effective in capturing sentiment patterns from user reviews than SVM on the dataset used in this study.
Key Words: CNN, Google Classroom, Sentiment Analysis, SVM, TF-IDF

PENDAHULUAN

Google Classroom merupakan salah satu platform pembelajaran daring yang umum digunakan, dengan total lebih dari 2.08 juta ulasan di Google Play Store (Google LLC, 2026). Aplikasi ini dirancang untuk memudahkan guru dan siswa dalam terhubung, mengerjakan tugas, serta mengelola aktivitas pembelajaran kapan saja dan di mana saja (Google, 2024). Google Classroom terus berkembang dan menjadi salah satu sarana untuk mendukung digitalisasi pendidikan. Namun berdasarkan data dari Google LLC (2026), menunjukkan bahwa rating Google Classroom di Google Play Store hanya berada 2.7-2.8 dari 5 bintang.

Analisis sentimen terdahulu terhadap ulasan pengguna mengindikasikan bahwa 64% sentimen bersifat negatif dan hanya 36% bersifat positif (Amany Mursianto et al., 2022). Hal ini mengindikasikan adanya kesenjangan antara harapan pengguna terhadap platform pembelajaran yang efisien dan pengalaman aktual yang didapatkan saat menggunakan aplikasi. Maka dari itu, analisis sentimen pengguna diperlukan untuk memahami berbagai permasalahan yang dialami pengguna serta menjadi dasar dalam penyusunan rekomendasi perbaikan yang tepat sasaran.

Beberapa penelitian telah dilakukan untuk menganalisis sentimen pengguna Google Classroom. Amany Mursianto et al. (2022) mengklasifikasikan sentimen ulasan Google Classroom menggunakan metode Support Vector Machine (SVM) dengan seleksi fitur Particle Swarm Optimization (PSO) pada data ulasan periode November-Desember 2021 yang menghasilkan akurasi 83%. Penelitian selanjutnya yang dilakukan oleh Nurwahidah et al. (2023) memanfaatkan algoritma Naive Bayes pada data ulasan Google Classroom mencapai akurasi, presisi, recall 82% serta F1-score 81%.

Meskipun penelitian-penelitian tersebut memberikan kontribusi penting dalam bidang analisis sentimen pada aplikasi pendidikan, masih terdapat beberapa gap yang perlu diperhatikan. Pertama, penelitian terdahulu dominan menggunakan data pada periode tahun 2021-2022, yaitu masa pandemi COVID-19 dengan pembelajaran jarak jauh penuh (Amany Mursianto et al., 2022; Nurwahidah et al., 2023). Saat ini, konteks penggunaan Google Classroom telah mengalami perubahan menuju pembelajaran hybrid yang menggabungkan antara tatap muka langsung dengan aktivitas digital, yang kemungkinan besar menghasilkan dinamika sentimen yang berbeda dibandingkan periode pandemi (Google, 2024). Kedua, penelitian terdahulu cenderung menggunakan algoritma klasifikasi seperti SVM atau Naive Bayes tanpa membandingkan dengan pendekatan deep learning seperti Convolutional Neural Npetwork (CNN) yang telah terbukti efektif untuk tugas klasifikasi teks (Amany Mursianto et al., 2022; Nurwahidah et al., 2023). Dwiyansaputra et al. (2025) menunjukkan bahwa penerapan CNN dengan TF-IDF mencapai akurasi 84% pada analisis sentimen, lebih tinggi ketika dikomparasi dengan machine learning tradisional. Ketiga, penelitian terdahulu belum secara eksplisit mengatasi masalah *class imbalance*, yang dapat menyebabkan model bias terhadap kelas mayoritas (Ginting et al., 2025). Keempat, variasi *feature selection* seperti Chi-Square belum diterapkan pada analisis sentimen Google Classroom, padahal metode ini terbukti meningkatkan akurasi SVM dengan mengurangi dimensi dan noise pada penelitian lain (Madasu & E, 2020).

Meninjau keempat hal tersebut, penelitian ini berusaha mengisi kesenjangan dengan mengadopsi pendekatan komparatif untuk membandingkan tiga model klasifikasi sentimen, yaitu TF-IDF + SVM, TF-IDF + Chi-Square + SVM, dan CNN dengan *class weighting.* Data penelitian akan menggunakan ulasan pengguna Google Classroom di Google Play Store dari

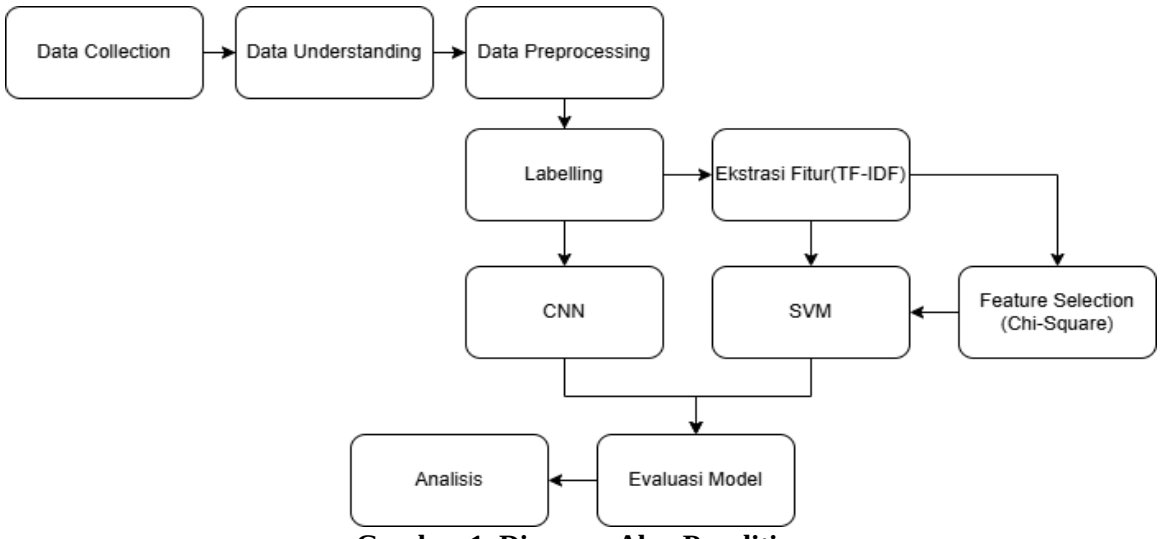
rentang waktu tahun 2024 hingga 2026 yang mencakup periode setelah pandemi di mana pola penggunaan Google Classroom telah mengalami transformasi. Penggunaan CNN untuk analisis sentimen Google Classroom merupakan kontribusi pada eksplorasi pendekatan deep learning dalam domain ini, yang hingga saat ini masih didominasi oleh metode machine learning tradisional (Amany Mursianto et al., 2022; Nurwahidah et al., 2023). Selain itu, penggunaan *class weighting* untuk menangani ketidakseimbangan kelas mengadopsi metode yang telah terbukti efektif dalam mengurangi bias terhadap kelas mayoritas pada penelitian analisis sentimen lainnya (Kowsari et al., 2019).

Metode TF-IDF dipilih sebagai teknik ekstrasi fitur karena efektif dalam menangani kompleksitas morfologi bahasa Indonesia dan memberikan representasi vektor yang informatif (Shafira Akbar Rizki et al., 2025). SVM digunakan sebagai *baseline* karena efektivitasnya pada data teks berdimensi tinggi (Amany Mursianto et al., 2022). Chi-Square digunakan untuk proses *feature selection* karena kemampuannya dalam mengidentifikasi fitur yang paling diskriminatif terhadap kelas sentimen, sehingga dapat meningkatkan performa klasifikasi (Al-Harbi1, 2019). Sementara itu, CNN dipilih karena arsitekturnya yang mampu menangkap pola (n-gram) dalam teks melalui convolutional filters, yang relevan untuk mengidentifikasi kata atau frasa yang berkontribusi terhadap pembentukan sentimen dalam teks (Dwiyansaputra et al., 2025).

Berdasarkan uraian tersebut, kebaruan dari penelitian ini terletak pada perbandingan komprehensif antara pendekatan *machine learning* tradisional dan *deep learning* pada analisis sentimen ulasan Google Classroom berbahasa Indonesia. Selain membandingkan kinerja SVM dan CNN, penelitian ini juga mengevaluasi pengaruh Chi-Square sebagai *feature selection* terhadap performa SVM. Dengan demikian, penelitian tidak hanya bertujuan untuk mengidentifikasi model dengan performa terbaik dalam mengklasifikasikan sentimen pengguna Google Classroom di Google Play Store tetapi juga bertujuan memberikan pemahaman mengenai kontribusi *feature selection* terhadap klasifikasi sentimen pada domain aplikasi pembelajaran.

METODE

Pada bagian ini akan dijelaskan alur penelitian yang dilakukan. Proses dimulai dari tahap *data collection* hingga analisis terhadap hasil perbandingan performa klasifikasi yang didapatkan. Diagram alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

Data Collection

Pada penelitian ini, data yang digunakan berupa ulasan pengguna Google Classroom pada Google Play Store dalam bahasa Indonesia. Data diperoleh melalui *web scraping* menggunakan *library* google-play-scraper pada Python dalam rentang waktu Januari 2024 hingga Mei 2026.

Setiap sampel ulasan terdiri dari sepuluh atribut yaitu, reviewId, userName, content, score, thumbsUpCount, reviewCreatedVersion, date, replyContent, repliedAt, appVersion. Content adalah atribut utama untuk analisis sentimen, sedangkan score digunakan sebagai validasi silang untuk pelabelan *lexicon-based.*

Data Understanding

Data Understanding dilakukan untuk memahami karakteristik dan kualitas data yang digunakan sebelum memasuki tahap *preprocessing.* Tahap ini mencakup analisis struktur dataset yang meliputi jumlah data dan atribut yang diperoleh dari proses *scraping*, serta pemeriksaan kualitas data melalui identifikasi *missing values* dan data duplikat. Selain itu, dilakukan analisis distribusi rating pengguna, distribusi waktu ulasan, dan distribusi panjang ulasan untuk memperoleh gambaran umum mengenai pola data yang digunakan dalam penelitian.

Data Preprocessing

Tahap ini dilakukan untuk mengubah teks mentah menjadi format yang siap diproses oleh model *machine learning.* Proses ini terdiri dari enam tahap yaitu, *data cleaning, case folding, tokenization, normalization, stopword removal,* dan *stemming.*

- *Data Cleaning*, mencakup pemilihan kolom yang relevan untuk proses analisis. Selain itu, karakter yang tidak berkontribusi pada tahap *modelling* seperti, URL, emoji, tanda baca, angka, karakter khusus, serta angka dihapus untuk mengurangi *noise* pada data teks.
- *Case Folding*, mengubah seluruh huruf menjadi huruf kecil untuk menyeragamkan representasi kata.
- *Tokenization*, yaitu proses memecah kalimat menjadi token.
- *Normalization*, yaitu proses mengubah kata tidak baku, singkatan, dan bahasa informal menjadi bentuk kata baku sesuai kamus normalisasi yang digunakan.
- *Stopword Removal*, bertujuan menghilangkan kata-kata yang memiliki kontribusi rendah terhadap proses klasifikasi sentimen. Pada penelitian ini dilakukan penyesuaian daftar *stopword* dengan mempertahankan beberapa kata yang berpotensi memengaruhi polaritas sentimen, seperti “tidak”, “belum”, dan “kurang”.
- *Stemming*, dilakukan untuk mengubah kata berimbuhan menjadi bentuk kata dasar menggunakan algoritma stemming Bahasa Indonesia, tujuannya mengurangi variasi kata yang memiliki makna serupa sehingga representasi fitur menjadi lebih konsisten.

Labelling

Pada tahap *labelling* digunakan metode *lexicon-based sentiment analysis* dengan *lexicon InSet (Indonesian Sentiment Word List).* Pendekatan ini dipilih karena efisiensi dan konsistensinya dalam memberikan label sentimen tanpa memerlukan anotasi manual yang memakan waktu. Proses *labelling* dilakukan dengan menghitung total skor sentimen dari setiap kata yang terdapat pada ulasan hasil *preprocessing.* Ulasan dengan skor positif diberi label positif, ulasan dengan skor negatif diberi label negatif, sedangkan ulasan dengan skor sama dengan nol diberi label netral. Namun, proses *labelling* pada penelitian ini tidak melibatkan proses validasi manual sehingga hasil pelabelan masih berpotensi mengandung kesalahan pada ulasan yang memiliki konteks bahasa yang kompleks.

Ekstrasi dan Pemilihan Fitur

Pada penelitian ini, ekstrasi fitur dilakukan menggunakan metode TF-IDF untuk mengubah data teks yang telah di-*preprocessing* menjadi representasi numerik yang dapat diproses oleh model

klasifikasi. TF-IDF menghitung bobot setiap kata berdasarkan frekuensi kemunculannya dalam suatu dokumen serta tingkat kemunculannya pada seluruh korpus.

Sementara itu, pemilihan fitur yang digunakan khusus untuk konfigurasi model kedua menggunakan metode Chi-Square. Metode ini mengidentifikasi fitur yang memiliki hubungan paling signifikan terhadap kelas sentimen. Tujuan dari proses pemilihan fitur untuk mengurangi jumlah fitur yang kurang relevan sehingga dapat menurunkan kompleksitas data dan mengurangi pengaruh *noise* pada proses klasifikasi.

Modelling

Pada tahap *modelling*, digunakan dua konfigurasi model berbasis SVM yaitu, TF-IDF + SVM, TF-IDF + Chi-Square + SVM, serta satu model berbasis *deep learning* berupa CNN. Sebelum proses *modelling*, dataset dibagi ke dalam *data training* dan *data testing* dengan rasio 80:20 menggunakan teknik *stratified sampling* untuk mempertahankan setiap proporsi setiap kelas sentimen. Khusus pada model CNN, sebanyak 10% dari data pelatihan digunakan sebagai *data validation* selama proses pelatihan untuk memantau performa model dan mencegah terjadinya *overfitting* melalui mekanisme *early stopping* sehingga rasio pembagian data menjadi 72:8:20. Pembagian dataset untuk konfigurasi model SVM dan model CNN dapat dilihat pada Tabel 1.

Tabel 1. Distribusi Pembagian Dataset

Konfigurasi	Kelas	Train	Validation	Test
SVM	Positif	245	-	61
	Netral	630	-	158
	Negatif	513	-	128
CNN	Positif	221	24	61
	Netral	567	63	158
	Negatif	462	51	128

Konfigurasi *hyperparameter* SVM dan CNN yang digunakan disajikan pada Tabel 2, sedangkan untuk arsitektur CNN yang digunakan ditampilkan pada Gambar 2.

Tabel 2. Hyperparameter pada SVM dan CNN

Konfigurasi	Hyperparameter	Nilai
SVM	Kernel	Linear
	C	1.0
CNN	Optimizer	Adam
	Loss	Categorical Crossentropy
	Batch Size	16
	Epoch	20
	Learning rate	0.001

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 100, 128)	1,280,000
conv1d (Conv1D)	(None, 98, 64)	24,640
global_max_pooling1d (GlobalMaxPooling1D)	(None, 64)	0
dropout (Dropout)	(None, 64)	0
dense (Dense)	(None, 64)	4,160
dropout_1 (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 3)	195

Total params: 1,308,995 (4.99 MB)
Trainable params: 1,308,995 (4.99 MB)
Non-trainable params: 0 (0.00 B)

Gambar 2. Arsitektur CNN yang digunakan

HASIL DAN PEMBAHASAN

Hasil Data Understanding

Proses *web scraping* menghasilkan 1.806 ulasan pengguna Google Classroom dengan 10 atribut. Hasil eksplorasi data menunjukkan bahwa tidak terdapat data duplikat pada dataset. Beberapa atribut memiliki nilai kosong, yaitu *reviewCreatedVersion* dan *appVersion* sebesar 33,5%, serta *replyContent* dan *repliedAt* sebesar 99,0%. Namun, atribut tersebut tidak digunakan dalam proses analisis sentimen karena penelitian ini hanya memanfaatkan atribut content sebagai sumber data utama.

Analisis distribusi rating menunjukkan bahwa rating 1 bintang memiliki jumlah ulasan terbanyak, diikuti oleh rating 5 bintang. Sementara itu, jumlah ulasan pada rating 2, 3, dan 4 bintang relatif lebih sedikit. Distribusi ulasan berdasarkan tahun menunjukkan bahwa sebagian besar ulasan berasal dari tahun 2024, diikuti tahun 2025 dan tahun 2026. Selain itu, distribusi panjang ulasan menunjukkan

bahwa sebagian besar ulasan memiliki panjang kurang dari 100 karakter. Hasil identifikasi data juga menunjukkan terdapat 19 ulasan yang hanya terdiri atas emoji tanpa informasi tekstual. Ulasan tersebut kemudian dipertimbangkan sebagai data yang tidak dapat dianalisis pada tahap pemrosesan teks selanjutnya.

Hasil Data Preprocessing

Setelah tahap *data understanding*, dilakukan proses *data preprocessing* yang meliputi *data cleaning*, *case folding*, *tokenization*, *normalization*, *stopword removal*, dan *stemming*. Tahapan ini dilakukan untuk mempersiapkan data sebelum proses pelabelan sentimen dan pemodelan.

Pada tahap *data cleaning*, elemen yang tidak digunakan dalam analisis sentimen seperti URL, *mention*, *hashtag*, emoji, angka, karakter khusus, dan spasi berlebih dihapus dari teks ulasan. Hasil pemeriksaan menunjukkan terdapat 19 ulasan yang hanya berisi emoji atau simbol tanpa informasi tekstual sehingga tidak dapat diproses lebih lanjut. Selain itu, atribut yang tidak digunakan dalam penelitian juga dihapus dari dataset.

Tahap *case folding* mengubah seluruh teks menjadi huruf kecil, kemudian dilanjutkan dengan *tokenization* untuk memecah teks menjadi token kata. Selanjutnya dilakukan *normalization* untuk menyeragamkan variasi penulisan kata tidak baku dan singkatan yang ditemukan pada ulasan pengguna. Setelah itu, dilakukan *stopword removal* menggunakan daftar stopword yang telah disesuaikan dan dilanjutkan dengan proses *stemming* menggunakan pustaka Sastrawi.

Setelah seluruh tahapan *preprocessing* selesai dilakukan, token hasil pengolahan digabung kembali menjadi dokumen teks yang digunakan sebagai masukan pada tahap pelabelan sentimen dan pemodelan. Jumlah data berkurang dari 1806 ulasan menjadi 1.735 ulasan karena sebagian data tidak lagi memiliki informasi tekstual setelah proses *preprocessing*. Perbandingan hasil sebelum dan sesudah *preprocessing* ditampilkan pada Tabel 3.

Tabel 3. Perbandingan Ulasan sebelum dan sesudah Preprocessing

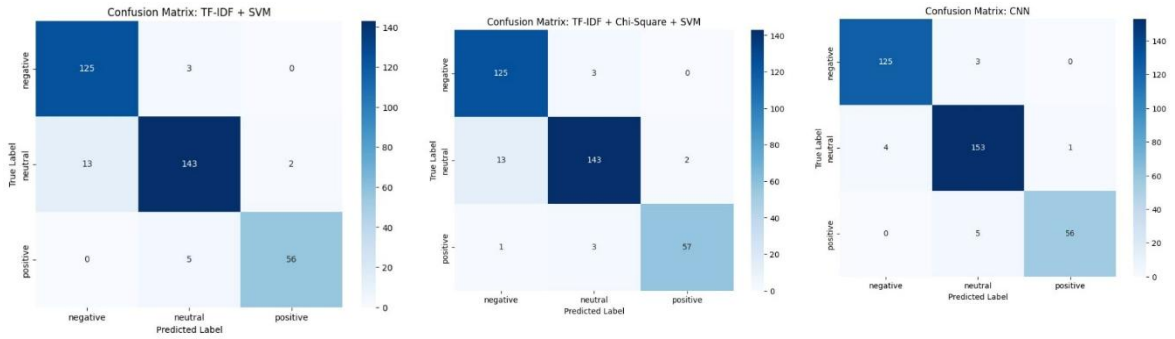
Ulasan sebelum preprocessing	Ulasan setelah preprocessing
sangat membantu dalam proses pembelajaran di era digital	bantu proses ajar era digital
kadang ga bisa masuk	kadang tidak masuk
terjadi bug ketika mau gabung kekelas malah loadingnya ga selesai selesai	ganggu gabung kelas loadingnya tidak selesai selesai

Hasil Labelling

Pada penelitian ini, pelabelan dilakukan menggunakan pendekatan *Lexicon-Based Sentiment Analysis* dengan memanfaatkan kamus sentimen InSet (*Indonesian Sentiment Word List*). Melalui pendekatan ini, setiap kata pada ulasan dicocokkan dengan kamus sentimen yang telah memiliki skor polaritas. Selanjutnya, seluruh skor kata dalam satu ulasan dijumlahkan untuk memperoleh skor sentimen akhir yang digunakan sebagai dasar penentuan label sentimen. Proses pelabelan menghasilkan tiga kategori sentimen, yaitu positif, negatif, dan netral. Melalui proses tersebut, seluruh 1735 ulasan berhasil diberi label sentimen dengan distribusi sentimen positif 306 ulasan, sentimen netral 788 ulasan, dan sentimen negatif 641 ulasan.

Hasil Evaluasi Model

Tahap evaluasi model dilakukan untuk mengukur kemampuan TF-IDF + SVM, TF-IDF + Chi-Square + SVM, dan CNN dalam mengklasifikasikan sentimen ulasan Google Classroom. Evaluasi dilakukan menggunakan *confusion matrix* yang menghasilkan metrik akurasi, presisi, *recall*, dan F1-score. *Confusion matrix* untuk ketiga model ditampilkan pada Gambar 3.



Gambar 3. Confusion Matrix setiap Model

Hasil pengujian TF-IDF + SVM menunjukkan nilai akurasi sebesar 93%. Nilai presisi, *recall*, dan F1-score untuk kelas negatif masing-masing sebesar 0.91, 0.98, dan 0.94. Pada kelas netral diperoleh nilai presisi 0.95, *recall* 0.91, dan F1-score 0.93, sedangkan pada kelas positif diperoleh nilai presisi 0.97, *recall* 0.92, dan F1-score 0.94.

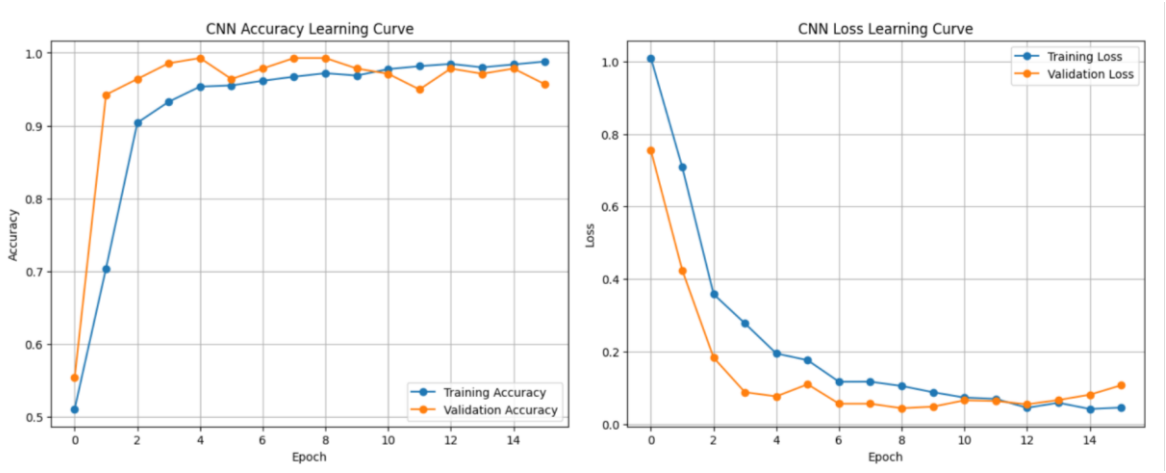
Model TF-IDF + Chi-Square + SVM memperoleh akurasi sebesar 94%. Pada kelas negatif diperoleh nilai presisi 0.90, *recall* 0.98, dan F1-score 0.94. Pada kelas netral diperoleh nilai presisi 0.96, *recall* 0.91, dan F1-score 0.93, sedangkan pada kelas positif diperoleh nilai presisi 0.97, *recall* 0.93, dan F1-score 0.95.

Model CNN menghasilkan akurasi sebesar 96%. Pada kelas negatif diperoleh nilai presisi 0.97, *recall* 0.98, dan F1-score 0.97. Pada kelas netral diperoleh nilai presisi 0.95, *recall* 0.97, dan F1-score 0.96. Sementara itu, pada kelas positif diperoleh nilai presisi 0.98, *recall* 0.92, dan F1-score 0.95. Perbandingan nilai akurasi, presisi, *recall*, dan F1-score dari ketiga model ditampilkan pada Tabel 4.

Tabel 4. Perbandingan Metrik Performa setiap Model

Model	**Kelas**	**Presisi**	**Recall**	**F1-Score**	**Akurasi**
TF-IDF + SVM	Positif	0.97	0.92	0.94	0.93
	Netral	0.95	0.91	0.93	
	Negatif	0.91	0.98	0.94	
TF-IDF + Chi-Square + SVM	Positif	0.97	0.93	0.95	0.94
	Netral	0.96	0.91	0.93	
	Negatif	0.90	0.98	0.94	
CNN	Positif	0.98	0.92	0.95	0.96
	Netral	0.95	0.97	0.96	
	Negatif	0.97	0.98	0.97	

Selain itu, *learning* pada proses *training* CNN *curve* yang ditampilkan pada Gambar 4, berlangsung cukup stabil dengan kemampuan generalisasi yang baik. Grafik akurasi menunjukkan bahwa *training accuracy* meningkat secara bertahap dengan gap yang kecil pada *validation accuracy*. Kemudian , grafik *loss* menunjukkan bahwa *training loss* mengalami penurunan secara konsisten seiring bertambahnya *epoch*, begitu juga dengan *validation loss*. Kondisi ini menunjukkan bahwa model mampu mempelajari pola dari *data training* dengan baik sekaligus mempertahankan performa pada *data validation*.



Gambar 4. Learning Curve Model CNN

Untuk menguji signifikansi perbedaan performa antar model, dilakukan pengujian menggunakan McNemar Test. Hasil pengujian pada Tabel 5 menunjukkan bahwa seluruh pasangan model menghasilkan nilai *p-value* lebih besar dari 0.05. Dengan demikian, tidak terdapat perbedaan performa yang signifikan secara statistik antara model yang dibandingkan pada tingkat signifikansi 5%.

Tabel 5. Hasil Pengujian McNemmar Test

PerbandinganModel	Statistik McNemmar	p-value
SVM vs Chi-Square + SVM	0.00	1.00
SVM vs CNN	3.68	0.05
Chi-Square + SVM vs CNN	3.04	0.08

Analisis Pembahasan

Hasil penelitian menunjukkan bahwa penerapan Chi-Square memberikan peningkatan performa pada model SVM. Peningkatan ini mengindikasikan bahwa tidak seluruh fitur yang dihasilkan oleh TF-IDF memiliki kontribusi yang sama dalam membedakan kelas sentimen. Chi-Square membantu memilih fitur yang memiliki hubungan paling kuat dengan kelas target sehingga dapat mengurangi fitur yang kurang relevan dan menekan pengaruh *noise* pada proses klasifikasi. Temuan ini sejalan dengan konsep *feature selection* yang menyatakan bahwa pemilihan fitur yang relevan dapat meningkatkan kemampuan generalisasi model sekaligus mengurangi kompleksitas data (Salsabilla et al., 2025).

Meskipun CNN menghasilkan nilai akurasi tertinggi dibandingkan kedua konfigurasi SVM, hasil tersebut perlu diinterpretasikan bersama hasil McNemar Test. Pengujian statistik menunjukkan bahwa perbedaan performa antara SVM dan CNN (p = 0,055) maupun antara Chi-Square + SVM dan CNN (p = 0,081) tidak signifikan pada tingkat signifikansi 5%. Dengan demikian, peningkatan performa yang diperoleh CNN belum cukup untuk menyimpulkan adanya keunggulan yang signifikan secara statistik. Namun demikian, hasil evaluasi tetap menunjukkan bahwa CNN secara konsisten menghasilkan nilai akurasi dan F1-score yang lebih tinggi dibandingkan model berbasis SVM

Perbedaan performa tersebut dapat dijelaskan dari cara masing-masing model merepresentasikan teks. SVM bergantung pada representasi TF-IDF yang memperlakukan kata sebagai fitur yang relatif independen sehingga informasi mengenai urutan dan konteks kata menjadi terbatas. Sebaliknya, CNN mampu mempelajari representasi fitur secara otomatis melalui *embedding layer* dan operasi konvolusi. Kemampuan tersebut memungkinkan model mengenali pola lokal antar kata dan menangkap hubungan kontekstual yang penting dalam menentukan sentimen. Kemampuan generalisasi CNN juga didukung oleh pola

learning curve yang stabil selama proses pelatihan, di mana performa data pelatihan dan data validasi menunjukkan pola yang relatif konsisten tanpa indikasi *overfitting* yang signifikan.

Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan. Pertama, dataset yang digunakan berjumlah 1735 ulasan setelah tahap *preprocessing*. Penambahan jumlah data berpotensi menghasilkan representasi sentimen yang lebih beragam dan meningkatkan kemampuan generalisasi model. Kedua, proses pelabelan sentimen dilakukan secara otomatis menggunakan pendekatan *lexicon-based sentiment analysis* tanpa validasi manual sehingga masih terdapat kemungkinan kesalahan pelabelan pada ulasan yang mengandung konteks ambigu atau makna yang bergantung pada konteks kalimat. Terlepas dari keterbatasan tersebut, hasil penelitian menunjukkan bahwa strategi representasi fitur dan metode klasifikasi memengaruhi performa analisis sentimen. Penerapan Chi-Square mampu meningkatkan performa SVM, sedangkan CNN secara konsisten menghasilkan performa tertinggi pada dataset yang digunakan. Temuan ini menunjukkan potensi pendekatan *deep learning* untuk diterapkan pada analisis sentimen ulasan aplikasi pendidikan.

SIMPULAN

Penelitian ini berhasil membandingkan kinerja Support Vector Machine (SVM) dan Convolutional Neural Network (CNN) pada analisis sentimen ulasan Google Classroom di Google Play Store. Hasil penelitian menunjukkan bahwa penerapan seleksi fitur Chi-Square mampu meningkatkan performa model SVM dari 93% menjadi 94%, yang mengindikasikan bahwa pemilihan fitur yang relevan berkontribusi terhadap peningkatan kemampuan klasifikasi. Namun demikian, CNN menghasilkan performa tertinggi dengan akurasi sebesar 96%, sehingga menunjukkan kemampuan yang lebih baik dalam mempelajari pola sentimen pada data teks dibandingkan pendekatan berbasis SVM. Hasil penelitian dapat dimanfaatkan oleh pengembang Google Classroom untuk mengidentifikasi kecenderungan sentiment pengguna secara otomatis, sehingga masukan dan keluhan pengguna dapat dianalisis secara lebih efisien sebagai bahan evaluasi dalam pengembangan fitur serta peningkatan kualitas layanan.

Penelitian selanjutnya dapat menggunakan metode pelabelan yang melibatkan validasi manual untuk meningkatkan kualitas label sentimen yang digunakan pada proses klasifikasi. Selain itu, penelitian dapat dikembangkan dengan membandingkan model berbasis Transformer.

DAFTAR PUSTAKA

Agustin Fitriana, Lady, Sugiyarto, I., & Faddillah, U. (2026). Penerapan Model LSTM dan CNN Untuk Klasifikasi Sentimen Pada Ulasan Aplikasi Roblox. *Infotek: Jurnal Informatika Dan Teknologi*, *9*(1), 71–82. https://doi.org/10.29408/jit.v9i1.32848

Al-Harbi1, O. (2019). A Comparative Study of Feature Selection Methods for Dialectal Arabic Sentiment Classification Using Support Vector Machine. In *IJCSNS International Journal of Computer Science and Network Security* (Vol. 19, Number 1). https://bit.ly/2FEGs1B

Amany Mursianto, G., Widiyanto, D., Tri Wahyono, B., Ilmu Komputer, F., Pembangunan Nasional Veteran Jakarta, U., & Fatmawati Raya, J. R. (2022). Analisis Sentimen Ulasan Pengguna Pada Aplikasi Google Classroom Menggunakan Metode SVM Dan Seleksi Fitur PSO. *JURNAL INFORMATIK Edisi Ke*, *18*, 2022.

Dwiyansaputra, R., Ika Murpratiwi, S., & Aranta, A. (2025). ANALISIS SENTIMEN PADA PENGGUNA APLIKASI X TERHADAP PEMILIHAN UMUM PRESIDEN 2024 MENGGUNAKAN METODE CONVOLUTIONAL NEURAL

NETWORK (CNN). In *Jurnal Mahasiswa Teknik Informatika)* (Vol. 9, Number 1).

Ginting, M. B., Surbakti, A. B., Ilham, S., Utomo, P., & Ginting, R. B. (2025). SMOTE and BERT Approaches for Handling Class Imbalance in Sentiment Analysis of the CoreTax Application on Big Data. *Technology and Science (BITS)*, *7*(2), 1467–1476. https://doi.org/10.47065/bits.v7i2.8310

Google. (2024). *10 Tahun Classroom: Apa yang Kami Pelajari dan Apa yang Akan Datang*.

Google LLC. (2026). *Google Classroom*.

Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. In *Information (Switzerland)* (Vol. 10, Number 4). MDPI AG. https://doi.org/10.3390/info10040150

Madasu, A., & E, S. (2020). *Efficient Feature Selection techniques for Sentiment Analysis*. http://arxiv.org/abs/1911.00288

Nurwahidah, D., Dwilestari, G., Nuris, N. D., Narasati, R., & Informatika, T. (2023). ANALISIS SENTIMEN DATA ULASAN PENGGUNA APLIKASI GOOGLE KELAS PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAÏVE BAYES. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Number 6).

Salsabilla, A. R., Sani, R. R., & Dewi, I. N. (2025). Perbandingan Metode Seleksi Fitur Chi-Square dan Information Gain untuk Peningkatan Interpretabilitas dan Optimasi Kinerja Model TabNet. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *11*(3), 253–262. https://doi.org/10.25077/teknosi.v11i3.2025.253-262

Shafira Akbar Rizki, D., Syaiful Khabib, M., Rahmayuna, N., & Gayuh Utomo, V. (2025). *Klasifikasi Sentimen Ulasan Pengguna Aplikasi Layanan Publik Google Play Store Menggunakan NLP dan ML*. *20*(1).