



Analisis Sentimen Tingkat Depresi Karyawan Generasi Z melalui Teks Media Sosial X menggunakan Metode 1d Cnn (Studi Kasus: Karyawan Gen Z)

Novi Madelena^{1*}, Donny Maulana², M. Zubair Abdurrohman³

¹⁻³Universitas Pelita Bangsa, Indonesia

*Penulis Korespondensi: novitamadelena@gmail.com

Abstract. *The mental health of Generation Z employees has become an important issue in the modern workplace due to increasing psychological pressure. Social media X (formerly Twitter) is often used as an anonymous space to express stress, complaints, and emotional conditions openly. This study aims to develop an early detection system for depression risk among Generation Z employees using Natural Language Processing (NLP) and Deep Learning approaches. The proposed method integrates a 1D-Convolutional Neural Network (1D-CNN) model with Word2Vec word representations to capture linguistic patterns that indicate emotional states. The dataset consists of text from social media X, processed through preprocessing stages such as cleaning, normalization, tokenization, and stemming to improve data quality and analysis accuracy. The model is then trained and tested to classify depression risk levels. Performance evaluation is conducted using accuracy, precision, and recall metrics to assess the model's effectiveness. The results are expected to contribute to the development of early mental health detection systems for companies and healthcare professionals, serving as a preventive tool to maintain the psychological well-being of Generation Z employees in increasingly complex digital work environments.*

Keywords: *Depression; Generation Z; NLP; Social Media X; Word2Vec.*

Abstrak. Kesehatan mental karyawan Generasi Z saat ini menjadi isu penting dalam dunia kerja modern, terutama karena tingginya tekanan psikologis yang mereka alami. Media sosial X (sebelumnya Twitter) sering digunakan sebagai ruang ekspresi anonim untuk menyampaikan keluhan, stres, dan kondisi emosional secara terbuka. Penelitian ini bertujuan untuk mengembangkan sistem deteksi dini risiko depresi pada karyawan Generasi Z dengan memanfaatkan pendekatan Natural Language Processing (NLP) dan Deep Learning. Metode yang digunakan mengintegrasikan model 1D-Convolutional Neural Network (1D-CNN) dengan representasi kata Word2Vec untuk menangkap pola bahasa yang mencerminkan kondisi emosional pengguna. Data yang digunakan berupa teks dari media sosial X yang melalui tahapan preprocessing, seperti pembersihan data, normalisasi, tokenisasi, dan stemming, guna meningkatkan kualitas data dan akurasi analisis. Model kemudian dilatih dan diuji menggunakan data tersebut untuk mengklasifikasikan tingkat risiko depresi. Evaluasi kinerja dilakukan dengan menggunakan metrik akurasi, presisi, dan recall untuk mengukur efektivitas model secara menyeluruh. Hasil penelitian diharapkan mampu memberikan kontribusi dalam pengembangan sistem deteksi dini kesehatan mental yang dapat digunakan oleh perusahaan maupun tenaga profesional kesehatan. Sistem ini juga diharapkan menjadi alat bantu preventif dalam menjaga kesejahteraan psikologis karyawan Generasi Z di lingkungan kerja digital yang semakin kompleks.

Kata kunci: Depresi; Generasi Z; Media Sosial X; NLP; Word2Vec.

1. LATAR BELAKANG

Generasi Z (Gen Z) merupakan generasi pertama yang tumbuh sebagai digital native dalam lingkungan yang dipenuhi teknologi, media sosial, dan konektivitas digital yang sangat tinggi. Paparan terhadap teknologi sejak usia dini, ditambah dengan faktor lingkungan dan pola asuh, telah membentuk karakteristik psikologis yang berbeda dibandingkan generasi sebelumnya (Chaniago, 2025; Luo et al., 2025). Berbagai penelitian menunjukkan bahwa intensitas penggunaan media sosial memiliki hubungan yang signifikan dengan kondisi kesehatan mental Gen Z, termasuk meningkatnya kecemasan dan berbagai gangguan psikologis lainnya (Novnariza et al., 2025). Di Indonesia, fenomena tersebut juga terlihat pada

kalangan mahasiswa dan generasi muda yang mengalami tantangan kesehatan mental akibat penggunaan media sosial yang tinggi (Rahmawati et al., 2024).

Seiring bertambahnya usia dan masuknya Gen Z ke dunia kerja, tantangan kesehatan mental menjadi semakin kompleks. Tuntutan pekerjaan, tekanan untuk mencapai target, beban kerja yang tinggi, serta stres kerja menjadi faktor yang dapat meningkatkan risiko gangguan psikologis pada kelompok ini (Khairil & Rinaldi, 2025). Selain itu, ketidakseimbangan antara kehidupan pribadi dan pekerjaan serta lingkungan kerja yang kurang mendukung juga berkontribusi terhadap penurunan kondisi mental karyawan Gen Z (Prahadiva et al., 2025). Dampak dari permasalahan tersebut tidak hanya dirasakan oleh individu, tetapi juga berpengaruh terhadap produktivitas, motivasi kerja, dan kinerja organisasi secara keseluruhan (Farida et al., 2026).

Di sisi lain, Gen Z dikenal sebagai generasi yang aktif dan ekspresif dalam menggunakan media sosial sebagai sarana komunikasi dan aktualisasi diri (Haniifah et al., 2025). Platform seperti Twitter (X) sering dimanfaatkan untuk menyampaikan opini, pengalaman pribadi, keluhan pekerjaan, hingga kondisi emosional yang sedang dialami (Situmorang, 2024). Aktivitas tersebut menghasilkan data tekstual dalam jumlah besar yang dapat menggambarkan kondisi psikologis pengguna secara langsung dan *real-time*, sehingga berpotensi dimanfaatkan sebagai sumber informasi untuk mendeteksi risiko kesehatan mental.

Perkembangan teknologi *Artificial Intelligence* (AI), khususnya pada bidang *Deep Learning* dan *Natural Language Processing* (NLP), membuka peluang baru dalam menganalisis data teks yang berasal dari media sosial. Berbagai penelitian telah menunjukkan bahwa model *Deep Learning* mampu mendeteksi indikasi depresi dari unggahan pengguna media sosial berbahasa Indonesia dengan tingkat performa yang baik (Holle & Maharani, 2024). Selain itu, arsitektur *Convolutional Neural Network* (CNN) terbukti efektif dalam melakukan klasifikasi teks dan mengidentifikasi pola-pola linguistik yang berkaitan dengan kondisi depresi (Tasci, 2025).

Beberapa penelitian terdahulu telah berhasil menerapkan teknik NLP dan *Deep Learning* untuk mendeteksi depresi melalui data media sosial. Namun, sebagian besar penelitian masih berfokus pada populasi pengguna media sosial secara umum dan belum secara khusus mengkaji kelompok karyawan Gen Z yang memiliki karakteristik serta tantangan psikologis yang berbeda. Padahal, kelompok ini merupakan bagian penting dari angkatan kerja masa kini yang rentan terhadap tekanan pekerjaan dan paparan media sosial secara bersamaan.

Selain itu, penelitian sebelumnya lebih banyak menitikberatkan pada penggunaan model *Deep Learning* tertentu tanpa mengoptimalkan representasi makna kata yang mampu menangkap konteks bahasa secara lebih baik. Padahal, pemahaman konteks menjadi aspek penting dalam analisis teks kesehatan mental karena ekspresi depresi sering kali disampaikan secara implisit melalui pilihan kata, kalimat, maupun gaya bahasa tertentu yang digunakan pengguna media sosial.

Berdasarkan kondisi tersebut, terdapat kesenjangan penelitian (*research gap*) berupa terbatasnya kajian yang mengintegrasikan metode Word2Vec sebagai teknik representasi kata dengan arsitektur 1D *Convolutional Neural Network* (1D-CNN) untuk mendeteksi risiko depresi pada karyawan Gen Z berdasarkan komentar berbahasa Indonesia di Twitter (X). Kesenjangan ini menunjukkan adanya kebutuhan untuk mengembangkan pendekatan yang lebih spesifik, akurat, dan sesuai dengan karakteristik data teks yang dihasilkan oleh kelompok Gen Z dalam konteks dunia kerja. Kebaruan penelitian ini terletak pada fokus objek penelitian, penggunaan kombinasi Word2Vec dan 1D-CNN, serta pemanfaatan data komentar media sosial yang berkaitan dengan pekerjaan dan kesehatan mental.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengembangkan model *Deep Learning* berbasis *Natural Language Processing* yang mampu mendeteksi risiko depresi pada karyawan Gen Z melalui analisis komentar media sosial Twitter (X). Penelitian ini juga bertujuan untuk menerapkan metode Word2Vec sebagai representasi kata dan arsitektur 1D-CNN sebagai model klasifikasi, serta mengevaluasi performa model menggunakan indikator akurasi, presisi, dan *recall*. Hasil penelitian diharapkan dapat menjadi dasar pengembangan sistem deteksi dini kesehatan mental berbasis *Artificial Intelligence* yang dapat mendukung upaya pencegahan dan penanganan masalah kesehatan mental di lingkungan kerja.

2. KAJIAN TEORITIS

Generasi Z dan Kesehatan Mental

Generasi Z merupakan kelompok yang tumbuh dan berkembang di tengah lingkungan digital yang sangat dinamis sehingga aktivitas sehari-hari mereka tidak terlepas dari penggunaan teknologi dan media sosial (Kullolli & Trebicka, 2023). Paparan informasi yang berlangsung secara terus-menerus menjadikan Generasi Z memiliki karakteristik yang berbeda dibandingkan generasi sebelumnya, terutama dalam aspek psikologis dan emosional. Beberapa penelitian menunjukkan bahwa kelompok ini memiliki tingkat sensitivitas emosional yang relatif tinggi akibat tekanan sosial, akademik, maupun profesional yang semakin kompleks (Sharma et al., 2023).

Penggunaan media sosial yang intensif juga berpotensi meningkatkan risiko gangguan kesehatan mental melalui fenomena perbandingan sosial, tekanan pencitraan diri, serta paparan konten negatif secara berulang. Dalam konteks dunia kerja, Generasi Z menghadapi berbagai tantangan seperti ketidakpastian karier, tuntutan produktivitas, dan kesulitan menjaga keseimbangan antara pekerjaan dan kehidupan pribadi. Kondisi tersebut menjadikan Generasi Z sebagai kelompok yang rentan mengalami tekanan psikologis sehingga diperlukan upaya deteksi dini terhadap risiko gangguan kesehatan mental, termasuk depresi (Jeshna, 2025).

Depresi dan Ekspresi Linguistik di Media Sosial

Depresi merupakan gangguan suasana hati yang ditandai oleh perasaan sedih berkepanjangan, kehilangan minat terhadap aktivitas sehari-hari, kelelahan emosional, serta menurunnya fungsi sosial individu (Chen et al., 2023). Berbagai penelitian menunjukkan bahwa kondisi depresi sering kali tercermin melalui pola bahasa yang digunakan seseorang dalam komunikasi sehari-hari, termasuk pada media sosial (Kelley & Gillan, 2022). Individu yang mengalami tekanan psikologis cenderung menggunakan kata-kata bernada negatif, pesimis, dan menunjukkan ekspresi keputusan maupun isolasi sosial (Cha et al., 2022).

Media sosial menjadi sarana yang memungkinkan pengguna mengekspresikan kondisi emosional secara lebih terbuka dan spontan. Pada platform seperti Twitter (X), berbagai unggahan teks dapat digunakan sebagai indikator untuk memahami kondisi psikologis pengguna. Penelitian sebelumnya menunjukkan bahwa pola linguistik tertentu dapat dimanfaatkan untuk mengidentifikasi kecenderungan depresi, sehingga analisis teks menjadi salah satu pendekatan yang efektif dalam mendukung deteksi dini kesehatan mental tanpa memerlukan intervensi langsung kepada individu (Beierle et al., 2023).

Analisis Sentimen

Analisis sentimen merupakan teknik yang digunakan untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini, emosi, maupun sikap yang terkandung dalam suatu teks (Palomino & Aider, 2022). Metode ini banyak diterapkan untuk memahami persepsi publik terhadap suatu topik, produk, layanan, maupun fenomena sosial. Dalam perkembangannya, analisis sentimen tidak hanya digunakan pada bidang pemasaran, tetapi juga pada penelitian sosial dan kesehatan mental untuk mengidentifikasi kondisi emosional pengguna berdasarkan teks yang dihasilkan (Surbakti et al., 2024).

Proses analisis sentimen umumnya diawali dengan tahap pra-pemrosesan untuk membersihkan dan menyiapkan data teks sebelum dilakukan ekstraksi fitur dan klasifikasi (Palomino & Aider, 2022). Pada penelitian ini, analisis sentimen digunakan untuk membantu mengidentifikasi kecenderungan emosi negatif yang berkaitan dengan depresi melalui

unggahan media sosial. Dengan demikian, analisis sentimen menjadi salah satu landasan penting dalam proses pengolahan dan interpretasi data teks yang digunakan untuk mendeteksi tingkat depresi pada karyawan Generasi Z.

Natural Language Processing (NLP)

Natural Language Processing (NLP) merupakan cabang kecerdasan buatan yang berfokus pada pemrosesan bahasa alami agar dapat dipahami dan dianalisis oleh komputer (Machová et al., 2023). NLP berperan penting dalam mengubah bahasa manusia menjadi representasi yang dapat diolah oleh algoritma *machine learning* maupun *deep learning* (Siino et al., 2024). Dalam penelitian berbasis teks, NLP digunakan untuk mengekstraksi informasi penting dari data yang tidak terstruktur sehingga dapat dimanfaatkan untuk proses analisis dan klasifikasi.

Pada penelitian kesehatan mental berbasis media sosial, NLP digunakan untuk melakukan berbagai tahapan seperti *text preprocessing*, tokenisasi, normalisasi, penghapusan *stopword*, serta transformasi teks ke dalam bentuk numerik menggunakan teknik representasi kata. Penelitian terdahulu menunjukkan bahwa penerapan NLP mampu meningkatkan kualitas analisis emosional dan klasifikasi pada data teks pendek seperti tweet (Machová et al., 2023). Oleh karena itu, NLP menjadi fondasi utama dalam penelitian ini karena berfungsi mempersiapkan data teks sebelum diproses lebih lanjut menggunakan metode Word2Vec dan arsitektur 1D-CNN.

3. METODE PENELITIAN

Pengumpulan dan Pelabelan Data

Penelitian ini menggunakan data berupa unggahan dan komentar publik berbahasa Indonesia pada platform X (Twitter) yang berkaitan dengan kondisi pekerjaan, tekanan kerja, stres emosional, dan indikasi depresi pada karyawan Generasi Z. Data dikumpulkan melalui proses *crawling* menggunakan kata kunci yang berkaitan dengan pekerjaan dan kondisi mental. Adapun kata kunci yang digunakan ditunjukkan pada Tabel 1.

Tabel 1. Kata Kunci *Crawling* Data

Jenis Keyword	Keyword
Keyword Utama	capek kerja, stres kerja, lembur, beban kerja, tekanan kerja, kerja terus, kerja rodi, kerja nggak berhenti, keluar kerja, berhenti kerja, pindah kerja, kerja berlebihan, kerja tiap hari
Keyword Emosi dan Kondisi Mental	lelah mental, capek mental, pikiran capek, mental capek, mental drop, mental kena, capek banget, putus asa kerja, pikiran kacau, kerjaan bikin tertekan

Data yang diperoleh selanjutnya diberi label secara manual ke dalam dua kategori, yaitu depresi dan tidak depresi. Penentuan label dilakukan berdasarkan konteks kalimat, kata kunci emosional, serta referensi literatur psikologi yang relevan. Untuk meningkatkan konsistensi hasil, proses pelabelan dapat dilakukan oleh lebih dari satu *annotator*.

Preprocessing Data

Tahap *preprocessing* dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam proses pelatihan model. Tahapan yang diterapkan meliputi *data cleansing*, *case folding*, *tokenizing*, *normalization*, *stopword removal*, dan *stemming*. Pada tahap *data cleansing*, elemen yang tidak relevan seperti URL, *mention*, *hashtag*, *emoji*, dan tanda baca dihapus. Selanjutnya seluruh teks diubah menjadi huruf kecil (*case folding*), dipecah menjadi token kata (*tokenizing*), dinormalisasi ke bentuk baku (*normalization*), dihapus kata umum yang tidak memiliki makna penting (*stopword removal*), dan dikembalikan ke bentuk kata dasar menggunakan metode *stemming*.

Representasi Teks Menggunakan Word2Vec

Representasi teks dilakukan menggunakan Word2Vec dengan pendekatan Skip-Gram. Metode ini bertujuan mempelajari hubungan semantik antar kata melalui prediksi kata konteks berdasarkan kata target. Sebagai contoh, pada kalimat “capek kerja lembur” dengan *window size* = 1, pasangan Skip-Gram yang terbentuk adalah: (capek, kerja), (kerja, capek), (kerja, lembur), (lembur, kerja)

Setiap kata direpresentasikan dalam bentuk vektor numerik dan digunakan sebagai masukan pada model klasifikasi. Proses pembelajaran dilakukan dengan memaksimalkan probabilitas kemunculan kata konteks terhadap kata target menggunakan fungsi objektif Word2Vec.

Klasifikasi Menggunakan 1D-CNN

Vektor hasil Word2Vec digunakan sebagai fitur masukan pada model *One-Dimensional Convolutional Neural Network* (1D-CNN). Model ini berfungsi untuk mengekstraksi pola linguistik yang berkaitan dengan indikasi depresi pada teks media sosial. Arsitektur model terdiri atas *embedding layer*, *convolution layer*, *pooling layer*, *fully connected layer*, dan *output layer*. Dataset dibagi menjadi data pelatihan sebesar 80% dan data pengujian sebesar 20%. Proses pelatihan dilakukan melalui tahapan *forward propagation*, perhitungan *loss*, *backpropagation*, dan pembaruan bobot hingga model mencapai performa yang optimal.

Evaluasi Model

Evaluasi dilakukan menggunakan *confusion matrix* untuk mengukur kemampuan model dalam mengklasifikasikan data ke dalam kategori depresi dan tidak depresi. Struktur *confusion matrix* ditunjukkan pada Tabel 2.

Tabel 2. *Confusion Matrix*

	Prediksi Depresi	Prediksi Tidak Depresi
Aktual Depresi	TP	FN
Aktual Tidak Depresi	FP	TN

Berdasarkan nilai pada confusion matrix, performa model diukur menggunakan *metrik Accuracy, Precision, Recall, dan F1-Score*. Rumus yang digunakan adalah sebagai berikut.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

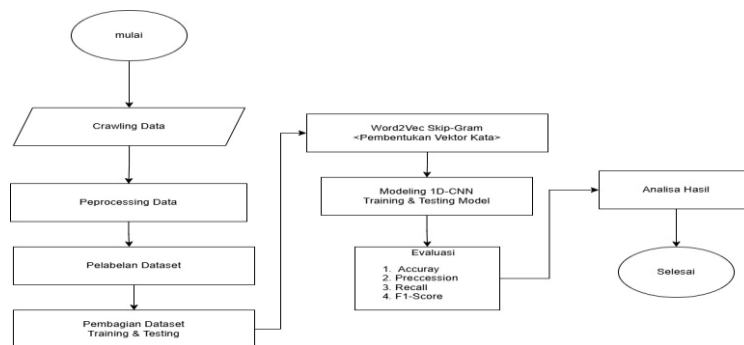
$$Recall = \frac{TP}{TP + FN}$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Gambar 1. Rumus

Nilai dari keempat metrik tersebut digunakan untuk mengevaluasi tingkat ketepatan dan kemampuan model dalam mendeteksi risiko depresi berdasarkan teks media sosial.

Model penelitian



Gambar 2. Rancangan model dalam penelitian

Gambar 1 menampilkan rancangan model yang dikembangkan dalam penelitian ini. *Output* model adalah klasifikasi risiko depresi yang dapat digunakan sebagai saran dan rekomendasi untuk mendukung pengambilan keputusan yang adaptif terkait minimalisasi risiko depresi dalam cakupan pekerjaan.

4. HASIL DAN PEMBAHASAN

Hasil Preprocessing dan Pelabelan Data

Tahap *preprocessing* dilakukan untuk meningkatkan kualitas data hasil crawling sebelum digunakan dalam proses pelatihan model. Data mentah yang diperoleh dari media sosial X masih mengandung berbagai *noise* seperti URL, *hashtag*, *mention*, simbol, angka, dan karakter khusus yang dapat mempengaruhi proses pembelajaran model. Oleh karena itu dilakukan serangkaian proses pembersihan yang meliputi *case folding*, *cleaning text*, penghapusan karakter khusus, dan penghapusan data duplikat. Tahapan ini menghasilkan *dataset* yang lebih bersih, konsisten, dan siap digunakan pada proses klasifikasi.

Hasil preprocessing menunjukkan bahwa seluruh data berhasil dikonversi ke format yang seragam sehingga dapat meningkatkan kualitas representasi teks pada tahap berikutnya. Selain mengurangi *noise*, *preprocessing* juga membantu model dalam mengenali pola linguistik yang berkaitan dengan kondisi psikologis pengguna. *Input preprocessing* berasal dari file *dataset.csv* hasil *crawling*, sedangkan *output preprocessing* berupa file *clean_dataset.csv*.

```
SISTEM DETEKSI DEPRESI
=====
1 Crawling
2 Preprocess
3 Training
4 Prediksi
0 Exit
pilih: 2

PREPROCESSING...
TOTAL: 1843

SELESAI PREPROCESS
FILE: output/clean_dataset.csv

text keyword clean
0      stres bnrn depresi dah gua kyny depresi      stres bnrn depresi dah gua kyny clean
1 gagal masuk ptn bukan akhir dari segalanya ya ... depresi gagal masuk ptn bukan akhir dari segalanya ya ...
2 depresi aku cok liat ini depresi      depresi aku cok liat ini clean
3 lagi suka bgt nonton one piece, iya iya gue te... depresi lagi suka bgt nonton one piece iya iya gue te...
4 perasaan baru kemaren deh depresi gara gara ni... depresi perasaan baru kemaren deh depresi gara gara ni...
```

Gambar 3. Tahapan preprocessing

Proses pelabelan dilakukan menggunakan pendekatan *rule-based labeling* berdasarkan kemunculan kata kunci yang berkaitan dengan indikasi depresi. Data kemudian diklasifikasikan ke dalam dua kategori, yaitu Indikasi Depresi dan Tidak Depresi. Hasil pelabelan menunjukkan bahwa dataset terdiri atas 886 data kategori indikasi depresi dan 153 data kategori tidak depresi, sehingga diperlukan proses data *balancing* sebelum pelatihan model.

Tabel 3. Distribusi Label D ataset

Kategori	Jumlah Data
Indikasi Depresi	886
Tidak Depresi	153
Total	1039

Berdasarkan hasil pelabelan, distribusi data menunjukkan ketidakseimbangan kelas (*imbalanced dataset*), di mana kelas indikasi depresi memiliki jumlah data yang jauh lebih besar dibandingkan kelas tidak depresi. Kondisi ini berpotensi menyebabkan model lebih cenderung mempelajari karakteristik kelas mayoritas. Oleh karena itu, dilakukan proses data

balancing sebelum tahap pelatihan model untuk memperoleh distribusi yang lebih seimbang dan meningkatkan kemampuan generalisasi model.

Implementasi dan Evaluasi Model CNN

Model klasifikasi dibangun menggunakan arsitektur *Convolutional Neural Network* (CNN). Sebelum data dimasukkan ke dalam model, teks terlebih dahulu dikonversi ke bentuk numerik menggunakan tokenizer dan *padding sequence*. Proses ini memungkinkan model mengolah data teks dalam bentuk representasi numerik yang sesuai untuk pembelajaran mesin.

Tabel 4. Konfigurasi Model CNN

Parameter	Nilai
<i>Embedding Size</i>	128
<i>Filter Conv1D</i>	64
<i>Kernel Size</i>	5
<i>Epoch</i>	10
<i>Batch Size</i>	32
<i>Optimizer</i>	Adam
<i>Loss Function</i>	<i>Binary Crossentropy</i>
<i>Aktivasi Output</i>	<i>Sigmoid</i>

Embedding layer digunakan untuk mengubah representasi teks menjadi vektor numerik. Selanjutnya layer *Conv1D* digunakan untuk mengekstraksi pola dari teks. *Output* hasil ekstraksi diteruskan ke *Global Max Pooling* untuk mengambil fitur dominan.

Dropout digunakan untuk mengurangi risiko *overfitting*, sedangkan *layer output* menggunakan fungsi aktivasi sigmoid karena klasifikasi terdiri dari dua kategori yaitu depresi dan tidak depresi.

```

1 # train_cnn.py
2 import os
3 import pickle
4 import pandas as pd
5
6 from sklearn.model_selection import train_test_split
7
8 from tensorflow.keras.preprocessing.text import Tokenizer
9 from tensorflow.keras.preprocessing.sequence import pad_sequences
10
11 from tensorflow.keras.models import Sequential
12 from tensorflow.keras.layers import (
13     Embedding,
14     Conv1D,
15     GlobalMaxPooling1D,
16     Dense,
17     Dropout
18 )
19
20 print("LOAD DATA...")
21
22 df = pd.read_csv(
23     "output/clean_dataset.csv"
24 )
25
26 df = df.dropna()
27
28 depress = [
29     "depress",
30     "stress",
31     "anxiety",
32     "burnout",
33     "depression",
34     "mental health",
35     "mental disorder",
36     "look happy",
37     "happy",
38     "sad",
39     "silly"
40 ]
41
42 normal = [
43     "normal",
44     "happy"
45 ]

```

Gambar 4. Implementasi model CNN

Pelatihan model dilakukan menggunakan rasio pembagian data 80:20 antara data *training* dan testing. Proses *training* berlangsung selama 10 *epoch* dengan *batch size* 32. Selama proses pelatihan, model menunjukkan peningkatan performa yang konsisten pada

setiap epoch. Hal ini mengindikasikan bahwa CNN mampu mempelajari pola-pola penting yang terdapat pada data teks terkait indikasi depresi.

Tabel 5. Hasil Evaluasi Model

Metrik	Nilai
<i>Accuracy</i>	90,38%
<i>Precision</i>	90%
<i>Recall</i>	100%
<i>F1-Score</i>	95%

Hasil evaluasi menunjukkan bahwa model memperoleh nilai *accuracy* sebesar 90,38%. Nilai ini menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar. *Precision* sebesar 90% mengindikasikan bahwa prediksi kelas depresi yang dihasilkan model memiliki tingkat ketepatan yang tinggi. *Recall* sebesar 100% menunjukkan kemampuan model dalam mendeteksi seluruh data yang termasuk kategori depresi pada data pengujian. Sementara itu, F1-Score sebesar 95% menunjukkan keseimbangan yang baik antara *precision* dan *recall*.

[Gambar 3. Grafik Akurasi Pelatihan dan Validasi – masukkan Gambar 4.9 dari skripsi]

Berdasarkan grafik pelatihan, akurasi training mengalami peningkatan secara bertahap hingga mendekati 100% pada *epoch* terakhir. Akurasi validasi juga menunjukkan tren peningkatan yang relatif stabil dan mencapai sekitar 90%. Hasil ini menunjukkan bahwa model mampu mempelajari pola data dengan baik serta memiliki kemampuan generalisasi yang cukup baik pada data yang belum pernah dilihat sebelumnya.

Pembahasan

Hasil penelitian menunjukkan bahwa pendekatan *Convolutional Neural Network* (CNN) efektif digunakan untuk mendeteksi indikasi depresi pada teks media sosial. Tingginya nilai akurasi dan F1-Score menunjukkan bahwa model mampu mengidentifikasi pola linguistik yang berkaitan dengan kondisi psikologis pengguna. Temuan ini mendukung berbagai penelitian sebelumnya yang menyatakan bahwa CNN memiliki kemampuan yang baik dalam ekstraksi fitur otomatis pada data teks.

Nilai *recall* sebesar 100% merupakan salah satu temuan penting dalam penelitian ini karena menunjukkan bahwa seluruh data yang mengandung indikasi depresi berhasil terdeteksi oleh model. Dalam konteks deteksi dini kesehatan mental, kemampuan ini sangat penting karena dapat meminimalkan risiko kasus depresi yang tidak teridentifikasi. Dengan demikian, model berpotensi digunakan sebagai alat bantu skrining awal untuk mengidentifikasi pengguna yang memerlukan perhatian lebih lanjut.

Meskipun demikian, perbedaan antara akurasi *training* dan validasi menunjukkan adanya indikasi *overfitting* ringan. Kondisi ini dapat disebabkan oleh ukuran dataset yang masih relatif terbatas, yaitu sebanyak 1.039 data. Semakin banyak data yang digunakan, semakin baik pula kemampuan model dalam mempelajari variasi bahasa yang terdapat pada media sosial.

Dalam hal ini, sistem yang dikembangkan telah berhasil melakukan proses *crawling*, *preprocessing*, pelabelan, pelatihan model, dan prediksi otomatis terhadap teks media sosial. Dengan tingkat akurasi 90,38%, model CNN menunjukkan potensi yang baik untuk diterapkan sebagai sistem pendukung deteksi dini risiko depresi pada karyawan Generasi Z berdasarkan unggahan media sosial.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, model *Convolutional Neural Network* (CNN) yang dikembangkan berhasil digunakan untuk mendeteksi indikasi depresi pada teks media sosial X (Twitter) dan mampu mengklasifikasikan data ke dalam kategori depresi maupun tidak depresi dengan performa yang baik. Sistem yang dibangun telah mengimplementasikan seluruh tahapan penelitian mulai dari pengumpulan data, *preprocessing*, pelabelan, pelatihan model, hingga prediksi otomatis terhadap teks baru. Hasil evaluasi menunjukkan bahwa model memperoleh nilai *accuracy* sebesar 90,38%, *precision* sebesar 90%, *recall* sebesar 100%, dan F1-score sebesar 95%, sehingga menunjukkan bahwa pendekatan CNN memiliki potensi yang baik untuk mendukung deteksi dini indikasi depresi berbasis analisis teks media sosial. Meskipun demikian, hasil penelitian ini perlu diinterpretasikan secara hati-hati karena diperoleh dari dataset yang terbatas pada data media sosial berbahasa Indonesia dan belum dapat digeneralisasikan secara luas sebagai alat diagnosis klinis kesehatan mental.

Penelitian ini masih memiliki beberapa keterbatasan, antara lain jumlah dataset yang relatif terbatas serta proses pelabelan yang masih menggunakan pendekatan berbasis kata kunci sehingga belum sepenuhnya merepresentasikan kondisi psikologis pengguna secara komprehensif. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan *dataset* yang lebih besar dan beragam, melibatkan ahli psikologi dalam proses anotasi data, serta melakukan perbandingan dengan metode lain seperti LSTM, BiLSTM, maupun model berbasis *Transformer*. Selain itu, pengembangan sistem ke dalam bentuk aplikasi berbasis web atau *mobile* serta penambahan fitur analisis sentimen dan tingkat depresi dapat menjadi langkah lanjutan untuk meningkatkan kegunaan dan akurasi sistem dalam mendukung pemantauan kesehatan mental secara lebih efektif.

DAFTAR PUSTAKA

- Arisanti, T., Budiyono, A. S., Syahputra, A. Z., & Surono, A. (2025). The concept of restoring the rights of victims of mass fraud with justice. *International Journal of Environmental Sciences*, 11(5). <https://theaspd.com/index.php>
- Asy'ari, K. D., & Naibaho, N. (2024). Perampasan aset dalam upaya pemulihan aset hasil tindak pidana korupsi di Indonesia. *Jurnal Ilmu Hukum, Humaniora dan Politik*, 4(5). <https://doi.org/10.38035/jihhp.v4i5.2792>
- Beierle, F., Pryss, R., & Aizawa, A. (2023). Sentiments about mental health on Twitter before and during the COVID-19 pandemic. *Healthcare*, 11(21), 2893. <https://doi.org/10.3390/healthcare11212893>
- Calabresi, G. (1970). *The costs of accidents: A legal and economic analysis*. Yale University Press.
- Cha, J., Kim, S., & Park, E. (2022). A lexicon-based approach to examine depression detection in social media: The case of Twitter and university community. *Humanities and Social Sciences Communications*, 9(1), 325. <https://doi.org/10.1057/s41599-022-01313-2>
- Chen, P., Feng, Y., Li, X.-H., Li, J.-X., Wang, Y.-Y., Zheng, W.-Y., Su, Z., Cheung, T., Ungvari, G. S., Ng, C. H., Sha, S., & Xiang, Y.-T. (2023). Systematic reviews and meta-analyses on major depressive disorder: A bibliometric perspective. *Frontiers in Psychiatry*, 14, 1136125. <https://doi.org/10.3389/fpsyt.2023.1136125>
- Chaniago, S. N. (2025). Kesehatan mental generasi Z dalam era digital: Studi psikologis tentang kecemasan sosial dan ketergantungan media sosial.
- Cory, E. T., et al. (2025). Pengalokasian aset hasil tindak pidana korupsi dalam perspektif negara kesejahteraan. *Jurnal Inovasi Hukum dan Kebijakan*, 6(2). <https://ejournals.com/ojs/index.php/jihk/article/view/2020>
- Fadhila, Y. R. (2025). Perlindungan pihak ketiga beriktikad baik dalam non-conviction based asset forfeiture. *Jurnal USM Law Review*, 8(3). <https://doi.org/10.26623/julr.v8i3.12263>
- Farida, I., Wijono, D., Kusumawati, D. A., Muslimin, M., & Nersiwad, N. (2026). The effect of work-life integration and quiet quitting on Gen Z employee productivity: SEM equation model. *Jurnal Pengabdian Masyarakat dan Riset Pendidikan*, 4(4), 24706–24711. <https://doi.org/10.31004/jerkin.v4i4.6249>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Haniifah, N. R. N., Isbandi, F. S., Rahayu, K. A., Natasha, A. D., Thoriq, A., & Fadillah, H. (2025). Pengaruh penggunaan media sosial terhadap kesehatan mental Gen Z. *Jurnal Bisnis dan Komunikasi Digital*, 2(2), 16. <https://doi.org/10.47134/jbk.v2i2.3581>
- Holle, A. P. B., & Maharani, W. (2024). Depression detection on Twitter using gated recurrent unit. *Jurnal Teknik Informatika (JUTIF)*, 5(1), 121–128. <https://doi.org/10.52436/1.jutif.2024.5.1.1187>
- Irwansyah. (2020). *Penelitian hukum: Pilihan metode & praktik penulisan artikel*. Mirra Buana Media.

- Iksan, M., Surbakti, N., Kurnianingsih, M., Budiono, A., Al-Fatih, S., & Ramon, T. M. (2023). Fulfilling the restitution rights of crime victims: The legal practice in Indonesia. *Academic Journal of Interdisciplinary Studies*, 12(4). <https://doi.org/10.36941/ajis-2023-0101>
- Iqbal, A. D., & Edrisy, I. F. (2025). Paradigma pemidanaan baru dalam KUHP 2023: Alternatif sanksi dan keadilan restoratif dalam peradilan pidana. *Journal Evidence of Law*, 4(3).
- Jeshna, M. (2025). The impact of social media on Gen Z's self-perception and mental health.
- Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed.). Pearson.
- Kaban, B., Mulyadi, M., & Mansar, A. (2021). Ganti rugi sebagai upaya perlindungan hak korban kejahatan perspektif politik hukum pidana. *Jurnal Ilmiah Penegakan Hukum*.
- Kelley, S. W., & Gillan, C. M. (2022). Using language in social media posts to study the network dynamics of depression longitudinally. *Nature Communications*, 13(1), 870. <https://doi.org/10.1038/s41467-022-28513-3>
- Khairil, M., & Rinaldi, R. (2025). Pengaruh beban kerja, stres kerja, dan motivasi kerja terhadap produktivitas karyawan Generasi Z di Yogyakarta. *OPTIMAL: Jurnal Ekonomi dan Manajemen*, 5(3), 638–650. <https://doi.org/10.55606/optimal.v5i3.7277>
- Kullolli, T., & Trebicka, B. (2023). Generation Z and the evolution of social media: A two-decade analysis of impact and usage trends. *Interdisciplinary Journal of Research and Development*, 10(3), 77.
- Luo, Y., Choi, J.-A., & Benton, B. (2025). Why Gen Z is more anxious than ever? Turning to social chats for answers.
- Machová, K., Szabóová, M., Paralič, J., & Mičko, J. (2023). Detection of emotion by text analysis using machine learning. *Frontiers in Psychology*, 14, 1190326. <https://doi.org/10.3389/fpsyg.2023.1190326>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv*.
- Novnariza, E., Pradipta, Y., Henda, S. A. D., & Shabrina, A. J. (2025). The association between social media addiction and mental health among Generation Z. *Contagion: Scientific Periodical Journal of Public Health and Coastal Health*, 7(1), 286–298.
- Palomino, M. A., & Aider, F. (2022). Evaluating the effectiveness of text pre-processing in sentiment analysis. *Applied Sciences*, 12(17). <https://doi.org/10.3390/app12178765>
- Prahadiva, L., Yunaida, L. F., Pratiwi, G. B., Aisy, R. H., & Salsabila, Z. (2025). Work life balance dan kesehatan mental: Studi komunikasi krisis pekerja Generasi Z. *Journal of Social, Culture, and Language*, 3(2), 60–71. <https://doi.org/10.21107/jscl.v3i2.29998>
- Rahmawati, R., Rahayu, P. I., & Taufiq, M. S. (2024). Analisis dampak penggunaan media sosial terhadap kesehatan mental Gen-Z. *MASSIVE: Jurnal Ilmu Komunikasi*, 4(2), 1–13. <https://doi.org/10.35842/massive.v4i2.156>
- Scikit-Learn. (2024). *Machine learning in Python*.
- Sharma, M., Kaushal, D., & Joshi, S. (2023). Adverse effect of social media on Generation Z user's behavior. *Journal of Retailing and Consumer Services*, 72, 103256. <https://doi.org/10.1016/j.jretconser.2023.103256>
- Siino, M., Tinnirello, I., & La Cascia, M. (2024). Is text preprocessing still worth the time? *Information Systems*, 121, 102342. <https://doi.org/10.1016/j.is.2023.102342>

- Situmorang, G. F. (2024). Deteksi potensi depresi dari unggahan media sosial X menggunakan IndoBERT (Skripsi, Universitas Mikroskil).
- Surbakti, N. M., Angelyca, A., Talia, A., Perangin-Angin, C. B., Nainggolan, D. O., Friskaully, N. D., & Tumorang, S. R. B. (2024). Penggunaan bahasa pemrograman Python dalam pembelajaran kalkulus fungsi dua variabel. *Algoritma: Jurnal Matematika, Ilmu Pengetahuan Alam, Kebumian dan Angkasa*, 2(3), 98–107. <https://doi.org/10.62383/algoritma.v2i3.67>
- Tasci, B. (2025). Deep learning-based detection of depression and suicidal tendencies in social media data with feature selection. *Behavioral Sciences*.
- TensorFlow. (2024). *TensorFlow documentation*.
- Thomas, N. K., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. *ICLR*.
- X Help Center. (2024). *Twitter/X platform documentation*.