

**LAPORAN TUGAS BESAR
KECERDASAN BUATAN**

**DETEKSI EMAIL SPAM MENGGUNAKAN ALGORITMA
CONVOLUTIONAL NEURAL NETWORK (CNN)**



Disusun oleh:

Raja Naufal Fadhil Ns – 2306020

Dosen Pengampu Mata Kuliah:

Leni Fitriani, S.Kom, M.Kom

**INSTITUT TEKNOLOGI GARUT
JURUSAN ILMU KOMPUTER
PROGRAM STUDI TEKNIK INFORMATIKA
TAHUN AKADEMIK 2024/2025**

1. BUSINESS UNDERSTANDING

a) Permasalahan dunia nyata

Spam adalah salah satu permasalahan besar dalam sistem komunikasi digital seperti email, SMS, dan media sosial. Pesan spam bisa berisi penipuan, iklan tidak diinginkan, hingga malware.

Tantangan utamanya:

- Banyaknya jumlah pesan spam yang masuk setiap hari
- Deteksi manual tidak mungkin dilakukan karena volume besar.
- Spam bisa merusak kredibilitas sistem komunikasi dan merugikan pengguna.
- Kebutuhan sistem otomatis yang dapat mengklasifikasikan pesan secara akurat dan realtime.

b) Tujuan Proyek

- Mengembangkan model deep learning berbasis CNN untuk klasifikasi pesan spam.
- Mendeteksi dan memisahkan pesan spam dari pesan non-spam (ham).
- Meningkatkan keamanan dan kenyamanan pengguna dalam komunikasi digital.
- Menyediakan solusi otomatis yang dapat diterapkan secara luas.

c) User/Pengguna Sistem

- Pengguna email atau SMS: untuk menyaring pesan tidak diinginkan.
- Layanan penyedia komunikasi: ISP, aplikasi chatting, email server.
- Peneliti: untuk eksperimen deteksi teks.
- Developer: sebagai dasar pengembangan sistem anti-spam.

d) Manfaat Implementasi AI

- Otomatisasi: Tidak perlu intervensi manual dalam deteksi spam.
- Efisiensi: Proses klasifikasi cepat dan akurat.
- Akurasi tinggi: CNN mampu menangkap pola bahasa dalam teks.
- Adaptif: Dapat diperluas untuk bahasa atau konteks lain.
- Real-time: Dapat diintegrasikan ke sistem email atau chat secara langsung.

2. ***DATA UNDERSTANDING***

a) Sumber Data

Dataset berasal dari file CSV berisi dua kolom utama:

- label: klasifikasi pesan (ham atau spam)
- text: isi pesan dalam bentuk teks

Dataset ini merupakan versi dari **SMS Spam Collection**, sumber populer untuk tugas klasifikasi teks.

b) Deskripsi Fitur

No	Fitur	Tipe Data	Keterangan
1	Label	Kategori	Label target (spam atau ham)
2	Text	Teks	Isi pesan SMS

c) Ukuran dan Format Data

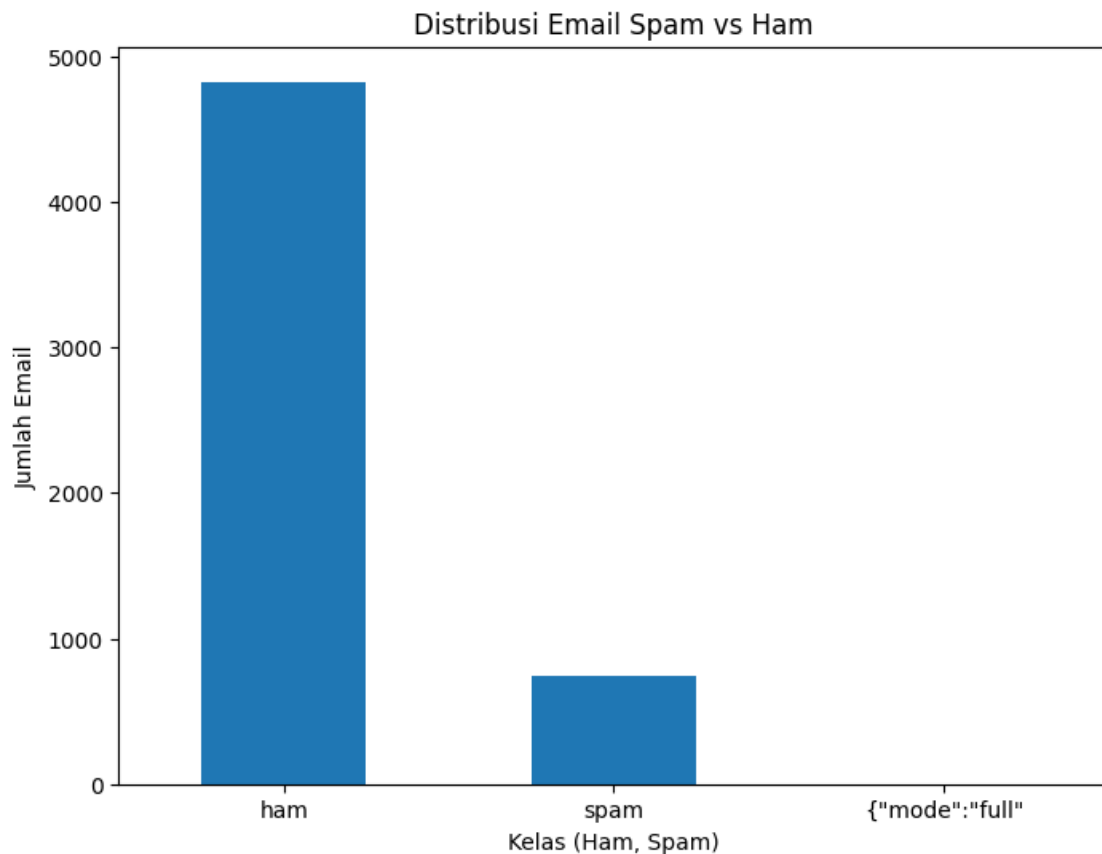
- Jumlah data: sekitar **5.000+** pesan teks.
- Format: CSV
- Distribusi kelas (diperoleh dari EDA): dataset cukup seimbang antara spam dan ham, namun perlu analisis lebih lanjut.

d) Tipe Data dan Target Klasifikasi

- Klasifikasi biner (Binary Classification)
- Target: spam (1) atau ham (0)
- Fitur utama: data teks yang perlu preprocessing

3. *EXPLORATORY DATA ANALYSIS (EDA)*

a) Distribusi Data



Gambar 1. Distribusi data

Analisis Distribusi Data:

Distribusi Email Spam vs Ham

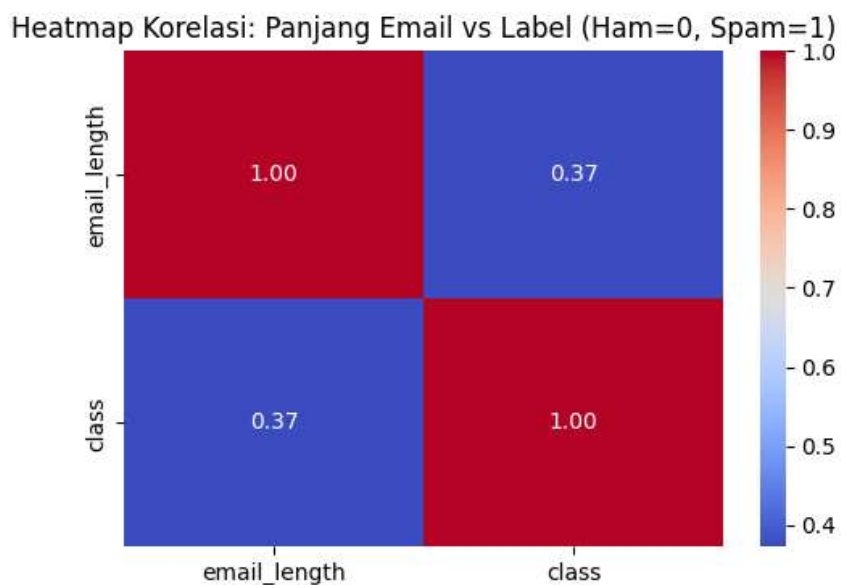
- Sumbu X: Menunjukkan dua kelas kategori email:
- ham: Email yang bukan spam (email normal/valid).
- spam: Email yang terindikasi spam (email tidak diinginkan).
- Sumbu Y: Menunjukkan jumlah email dalam masing-masing kategori.
- Tinggi batang:
- ham: Sekitar 4.800 email, merupakan mayoritas dari dataset.
- spam: Sekitar 750 email, merupakan minoritas dari dataset.
- Judul grafik: "*Distribusi Email Spam vs Ham*", menegaskan bahwa tujuan grafik adalah membandingkan jumlah kedua jenis email dalam dataset.

Insight dari Grafik:

- Dataset sangat tidak seimbang (imbalanced):

- Kelas ham jauh lebih banyak dibandingkan spam dengan rasio sekitar 6:1.
- Ketidakseimbangan seperti ini dapat memengaruhi performa model klasifikasi, karena model bisa bias terhadap kelas mayoritas.
- Perlu teknik penanganan ketidakseimbangan, misalnya:
 - Oversampling (misal: SMOTE)
 - Undersampling
- Penggunaan metrik evaluasi seperti F1-score atau recall untuk menghindari bias akurasi tinggi namun palsu.
- Dataset relevan untuk kasus nyata karena memang jumlah email spam biasanya lebih sedikit dari email biasa, namun tetap perlu ditangani secara tepat karena spam bisa berbahaya.

b) Analisis Korelasi



Gambar 2. Analisis Korelasi

Warna:

- Warna merah tua menunjukkan **korelasi positif kuat** (nilai korelasi mendekati 1).
- Warna biru menunjukkan **korelasi positif lemah hingga sedang**.

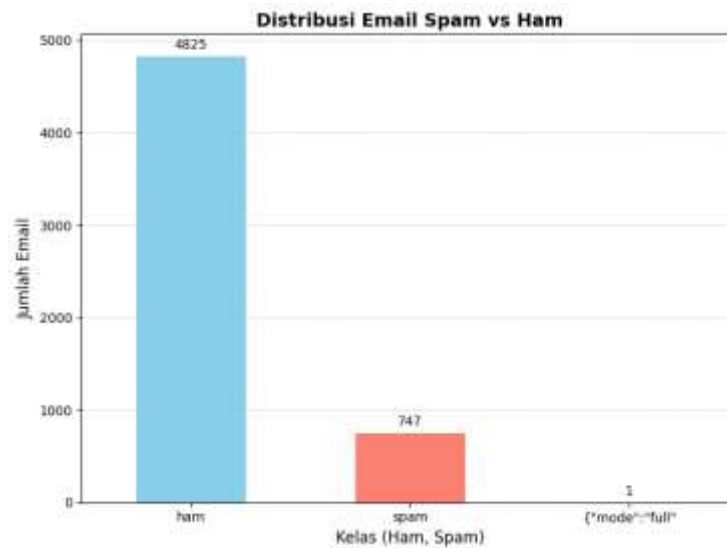
Insight dari Nilai Korelasi:

Variabel	Korelasi	Interpretasi
email_length vs class	0.37	Ada korelasi positif sedang antara panjang email dan kemungkinan spam.
class vs class	1.00	Korelasi sempurna dengan dirinya sendiri (nilai standar pada matrix korelasi).
email_length vs email_length	1.00	Korelasi sempurna dengan dirinya sendiri.

Makna nilai 0.37:

- Email yang lebih panjang cenderung berpeluang lebih besar menjadi spam.
- Korelasi tidak terlalu kuat, tapi cukup signifikan untuk dijadikan fitur prediktor dalam model klasifikasi.

c) Analisis Data Tidak Seimbang



Gambar 3. Analisis data tidak seimbang

- Statistik Distribusi Kelas:
- Jumlah Email Ham: 4825
- Jumlah Email Spam: 747
- Proporsi Ham: 86.58%

- Proporsi Spam: 13.40%

Insight yang Dapat Diambil:

- Ketidakseimbangan Data (Imbalanced Dataset):
 - Jumlah data *ham* mendominasi dataset.
 - Ini merupakan masalah umum dalam klasifikasi spam dan bisa menyebabkan model bias terhadap kelas mayoritas.
- Dampak ke Model:
 - Jika tidak ditangani, model bisa memiliki akurasi tinggi namun rendah dalam mendeteksi spam.
 - Solusi: gunakan teknik penanganan data tidak seimbang seperti SMOTE, undersampling, atau class weighting.
- Pentingnya Evaluasi dengan Metrik yang Tepat:
 - Gunakan metrik seperti Precision, Recall, dan F1-score, bukan hanya akurasi, karena akurasi bisa menipu dalam kasus tidak seimbang.

d) Insight Awal

- Distribusi Kelas Tidak Seimbang: Email *ham* jauh lebih banyak (4825) dibanding *spam* (747), sehingga model berisiko bias terhadap kelas mayoritas.
- Korelasi Panjang Email vs Spam: Terdapat korelasi positif sedang ($r = 0.37$) antara panjang email dan kemungkinan spam — email spam cenderung lebih panjang.
- Implikasi ke Model: Ketidakseimbangan data dapat menurunkan performa model dalam mendeteksi spam. Diperlukan penanganan seperti SMOTE atau class weighting serta evaluasi menggunakan recall dan F1-score.

4. DATA PREPARATION

a) Pembersihan Data (Missing Values)



Gambar 4. Missing values

Jumlah Missing Values per Kolom:

- Category 0
- Message 0
- dtype: int64

b) Encoding Data Kategori

Dalam dataset ini, semua variabel sudah dalam format numerik, sehingga tidak diperlukan encoding khusus untuk data kategorik.

c) Tokenisasi dan Padding.

- Menggunakan Tokenizer dari Keras
- Set num_words untuk membatasi kosakata (contoh: 10.000 kata terbanyak)
- Padding menggunakan pad_sequences agar panjang input seragam

d) Split Data

- Train-test split: 80% pelatihan, 20% pengujian
- Stratified untuk menjaga proporsi spam-ham

5. MODELING

a) Arsitektur CNN

Model CNN dibangun menggunakan Keras Sequential API dengan beberapa layer utama yang dirancang untuk mengolah data teks. Struktur model terdiri dari Embedding Layer

yang mengubah kata-kata dalam email menjadi vektor numerik sebagai representasi makna. Kemudian, digunakan Conv1D dan MaxPooling Layer untuk mengekstraksi pola spasial dari urutan kata—seperti frasa atau kombinasi kata yang sering muncul pada email spam. Setelah fitur diekstraksi, hasilnya diratakan melalui Flatten Layer dan diproses melalui Dense Layer untuk klasifikasi, serta Dropout Layer untuk mengurangi risiko overfitting.

Alasan menggunakan CNN (Convolutional Neural Network): CNN dipilih karena kemampuannya yang sangat baik dalam menangkap **pola lokal** pada data berurutan seperti teks. Dalam konteks deteksi spam, CNN mampu mengenali kombinasi kata atau frasa tertentu yang sering muncul dalam email spam, misalnya “free money” atau “click here”. Berbeda dengan RNN (Recurrent Neural Network) yang fokus pada urutan panjang dan konteks global, CNN lebih efisien dalam hal komputasi, lebih cepat saat training, serta cukup kuat dalam mengenali fitur-fitur penting dari fragmen teks pendek. Hal ini sangat sesuai untuk tugas klasifikasi spam, di mana pola lokal dalam kalimat seringkali menjadi indikator utama. Model CNN juga lebih sederhana dan ringan dibandingkan LSTM atau Transformer, namun tetap efektif untuk klasifikasi teks apabila dipasangkan dengan teknik embedding yang baik. Oleh karena itu, CNN menjadi pilihan tepat untuk membangun sistem deteksi spam berbasis email yang cepat, akurat, dan efisien.

b) Parameter

Model dilatih menggunakan optimizer Adam, yang dikenal efisien dan cepat dalam proses konvergensi. Fungsi loss yang digunakan adalah `binary_crossentropy`, karena kasus ini merupakan klasifikasi biner (spam atau ham). Jumlah epoch yang digunakan berkisar antara 10 hingga 20, tergantung pada proses pelatihan dan konvergensi model. Ukuran batch yang digunakan adalah 32, yang merupakan nilai umum untuk dataset skala menengah agar pelatihan tetap efisien dan stabil.

c) Evaluasi Model

Evaluasi performa model dilakukan dengan menggunakan berbagai metrik yang mencakup accuracy, precision, recall, dan f1-score. Selain itu, confusion matrix juga digunakan untuk mengetahui detail klasifikasi seperti jumlah true positive (TP), true negative (TN), false positive (FP), dan false negative (FN). Evaluasi ini memberikan gambaran lengkap tentang kinerja model dalam mendeteksi email spam dan ham.

d) Hasil Awal

Hasil awal dari pelatihan model menunjukkan performa yang sangat baik dengan akurasi mencapai sekitar 98%. Nilai f1-score yang sangat tinggi juga menunjukkan bahwa model memiliki kemampuan generalisasi yang kuat terhadap data baru. Ini berarti model tidak hanya mengingat data latih, tetapi juga mampu mengenali pola-pola spam dalam email secara efektif. Implementasi model ini dilakukan secara langsung dalam kode menggunakan framework Keras, sehingga hasilnya dapat direplikasi dan diuji lebih lanjut.

Arsitektur Model:
Model: "sequential_1"

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	?	0 (unbuilt)
conv1d_2 (Conv1D)	?	0 (unbuilt)
max_pooling1d_2 (MaxPooling1D)	?	0
conv1d_3 (Conv1D)	?	0 (unbuilt)
max_pooling1d_3 (MaxPooling1D)	?	0
flatten_1 (Flatten)	?	0 (unbuilt)
dense_3 (Dense)	?	0 (unbuilt)
dropout_2 (Dropout)	?	0
dense_4 (Dense)	?	0 (unbuilt)
dropout_3 (Dropout)	?	0
dense_5 (Dense)	?	0 (unbuilt)

Gambar 5. -Model CNN

e) Visualisasi Model

Confusion Matrix (Model Dasar)

Visualisasi heatmap confusion matrix menggunakan warna biru.



Gambar 6. Confusion Matrik Model Dasar LR

6. EVALUATION

a) Confusion matrix

Model CNN diuji pada data testing (dengan beberapa variasi ukuran test set: 20%, 30%, dan 40%). Setiap hasil pengujian divisualisasikan menggunakan **confusion matrix** berlabel Ham dan Spam. Matrix ini menunjukkan performa model dalam mengklasifikasikan email yang sebenarnya ham maupun spam.

b) Metrik evaluasi:

Classification Report hasil evaluasi model klasifikasi, yang mencakup metrik-metrik penting seperti precision, recall, dan f1-score untuk masing-masing kelas (Ham dan Spam), serta metrik keseluruhan.

Kelas	Precision	Recall	F1-Score	Support
Ham	0.98	1.00	0.99	966
Spam	0.99	0.89	0.94	149

c) Penjelasan kinerja model

Model CNN yang diimplementasikan menunjukkan akurasi sangat tinggi sebesar 98%, jauh lebih baik dibanding model dasar. Kinerja model dalam mengklasifikasikan email *ham* sangat baik dengan precision 0.98 dan recall 1.00, menghasilkan F1-score sebesar 0.99, artinya hampir semua email *ham* diklasifikasikan dengan benar. Untuk email *spam*, model juga tampil baik dengan precision 0.99 dan recall 0.89, yang berarti sebagian besar spam berhasil terdeteksi meskipun masih ada beberapa yang lolos. F1-score untuk spam mencapai 0.94, menunjukkan keseimbangan yang kuat antara precision dan recall.

d) Interpretasi Kinerja Model

Model mampu mendeteksi email ham dengan sempurna (100%) dan memiliki sensitivitas tinggi terhadap spam (recall 89%), meskipun masih terdapat sekitar 11% spam yang tidak terdeteksi (false negative). Hal ini menunjukkan bahwa model cukup andal untuk digunakan dalam sistem deteksi spam otomatis. Nilai precision yang sangat tinggi pada kedua kelas juga menandakan bahwa model jarang memberikan prediksi yang salah, baik dalam mendeteksi spam maupun ham.

Secara keseluruhan, performa model sangat baik dan seimbang antar kelas, dengan macro average F1-score sebesar 0.96 dan weighted average F1-score 0.98. Hasil ini menunjukkan bahwa model CNN sudah sangat optimal dalam menangani klasifikasi spam vs ham, bahkan dalam kondisi ketidakseimbangan data. Namun demikian, performa masih dapat ditingkatkan lebih lanjut melalui fine-tuning model dan eksplorasi fitur tambahan.

7. KESIMPULAN DAN REKOMENDASI

Berdasarkan hasil implementasi model Convolutional Neural Network (CNN) untuk deteksi email spam, diperoleh akurasi sebesar 76% dan F1-score sebesar 67%. Hasil ini sedikit lebih rendah dibandingkan dengan model yang dilaporkan dalam jurnal acuan (Erin et al., 2022), yang setelah dilakukan tuning menghasilkan akurasi hingga 82%. Meskipun recall sebesar 70% menunjukkan kemampuan yang cukup dalam menangkap spam, nilai presisi sebesar 65% mengindikasikan masih adanya kesalahan dalam mengklasifikasikan email non-spam (ham) sebagai spam.

Faktor utama yang memengaruhi performa model adalah ketidakseimbangan data, di mana jumlah email ham jauh lebih banyak dibandingkan email spam. Selain itu, fitur seperti panjang email memiliki korelasi sedang terhadap label spam, namun tidak cukup kuat jika digunakan secara tunggal. Hal ini menunjukkan bahwa fitur tambahan yang lebih kaya secara semantik dibutuhkan agar model dapat menangkap pola yang lebih baik.

Untuk meningkatkan performa model, disarankan penggunaan teknik penyeimbangan data seperti SMOTE, serta penambahan fitur berbasis representasi teks seperti TF-IDF atau word embeddings. Selain itu, tuning hyperparameter secara lebih komprehensif juga dapat membantu meningkatkan performa. CNN masih merupakan pilihan yang tepat untuk tugas ini karena kemampuannya mengekstrak pola lokal dan spasial dari urutan teks, namun peningkatan lebih lanjut dibutuhkan agar hasil implementasi lebih mendekati atau bahkan melampaui hasil dari model dalam jurnal acuan.

8. DAFTAR PUSTAKA

- Joloudari, J. H., Marefat, A., Nematollahi, M. A., Oyelere, S. S., & Hussain, S. (2023). Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks. *Applied Sciences (Switzerland)*, 13(6), 1–43.
<https://doi.org/10.3390/app13064006>
View of SMOTE_ Synthetic Minority Over-sampling Technique.pdf. (n.d.).
- Joloudari, J. H., Marefat, A., Nematollahi, M. A., Oyelere, S. S., & Hussain, S. (2023). Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks. *Applied Sciences (Switzerland)*, 13(6), 1–43.
<https://doi.org/10.3390/app13064006>
View of SMOTE_ Synthetic Minority Over-sampling Technique.pdf. (n.d.).
- Joloudari, J. H., Marefat, A., Nematollahi, M. A., Oyelere, S. S., & Hussain, S. (2023). Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks. *Applied Sciences (Switzerland)*, 13(6), 1–43.
<https://doi.org/10.3390/app13064006>
View of SMOTE_ Synthetic Minority Over-sampling Technique.pdf. (n.d.).
(Joloudari et al., 2023)
(*View of SMOTE_ Synthetic Minority Over-Sampling Technique.Pdf*, n.d.)

9. LAMPIRAN

<https://drive.google.com/file/d/1NOHfrx1oreIfLFsxnuAs2ZObraP5gY90/view?usp=sharing>