

MACHINE LEARNING

Autoencoder Untuk Ekstraksi Fitur Data



Dosen Pengampu :

Al-Ustadz Oddy Virgantara Putra

Disusun Oleh :

Naila Fatikhah Parwanto / 442023618086

PROGRAM STUDI TEKNIK INFORMATIKA

UNIVERSITAS DARUSSALAM GONTOR MANTINGAN

2025/1447

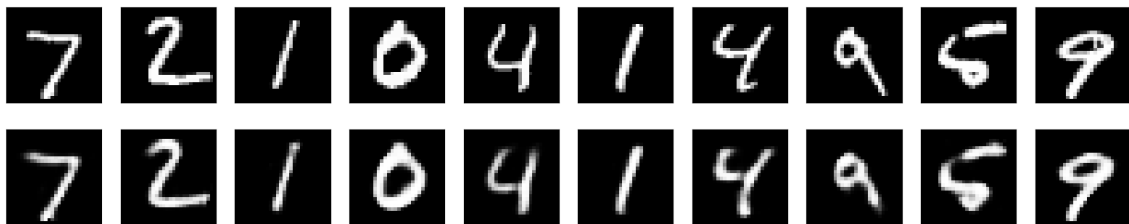
1. Pendahuluan

Autoencoder adalah arsitektur neural network yang bertujuan merepresentasikan data input ke dimensi laten lebih kecil, lalu merekonstruksi kembali ke bentuk asli. Salah satu aplikasi pentingnya adalah **ekstraksi fitur**, di mana representasi laten yang dipelajari encoder bisa digunakan sebagai representasi ringkas data. Hal ini bermanfaat untuk tugas-tugas seperti klasifikasi, clustering, atau visualisasi data.

2. Metode

- **Dataset:** MNIST (gambar digit 0–9 berukuran 28×28 piksel).
- **Model:** Autoencoder dengan encoder yang memproyeksikan gambar ke vektor laten berdimensi **32**, kemudian decoder merekonstruksi kembali gambar asli.
- **Training:** Dilatih selama **50 epoch** menggunakan loss rekonstruksi (*Mean Squared Error*).
- **Tujuan:** Mengevaluasi seberapa baik autoencoder mampu mengekstraksi fitur yang bermakna dari data.

3. Hasil Eksperimen



Loss Training

- Loss konsisten menurun dari awal hingga epoch 50, dengan nilai akhir **0.0084 (train)** dan **0.0083 (val)** → menandakan rekonstruksi stabil.

Rekonstruksi Gambar

- Baris atas: input asli.
- Baris bawah: hasil rekonstruksi.
Hasil menunjukkan rekonstruksi cukup mirip dengan input, meskipun ada sedikit blur pada digit tertentu.

Ekstraksi Fitur

- Bentuk representasi laten: (10000, 32)

- Contoh fitur diekstrak dari gambar pertama:

```
Contoh fitur diekstrak dari gambar pertama:
[ 7.4556828  0.      12.974136  5.134931  9.711826  18.617197
 16.062544  4.448584  8.395572  9.986595  14.860013  5.951105
 17.044975  10.993549  0.      0.      17.951298  6.397764
  6.6744156  17.30996  0.      6.174255  5.122801  11.363006
 13.789628  10.677897  14.399689  15.249635  0.      14.839477
 0.      10.193158 ]
```

- Representasi ini adalah **embedding numerik** yang bisa dipakai untuk klasifikasi berbasis fitur, clustering, atau visualisasi (misalnya dengan t-SNE atau PCA).

Hasil Param

- Tabel Param

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(none, 784)	0
dense (Dense)	(none, 128)	100,480
dense_1 (Dense)	(none, 64)	8,256
dense_2 (Dense)	(none, 32)	2,080
dense_3 (Dense)	(none, 64)	2,112
dense_4 (Dense)	(none, 128)	8,320
dense_5 (Dense)	(none, 784)	101,136

- Total params: 222,384 (868.69 KB)
- Trainable params: 222,384 (868.69 KB)
- Non-trainable params: 0 (0.00 B)

4. Analisis

- Autoencoder berhasil menurunkan dimensi data dari 784 (28×28 piksel) → 32 fitur, tanpa kehilangan informasi penting.
- Semakin dalam arsitektur encoder, biasanya fitur yang diekstrak makin abstrak, namun risiko overfitting meningkat.
- Nilai loss yang stabil menunjukkan model tidak hanya menghafal, tapi juga belajar representasi general.
- Blur pada hasil rekonstruksi menandakan ruang laten masih bisa diperbaiki dengan arsitektur lebih kompleks (misal menambah layer convolutional).

5. Kesimpulan dan Ide Lanjutan

Eksperimen ini menunjukkan bahwa autoencoder efektif untuk **ekstraksi fitur numerik** dari data gambar, menghasilkan representasi padat yang berguna untuk analisis lanjutan.

Ide masa depan:

- Menggunakan **Convolutional Autoencoder** agar rekonstruksi lebih tajam.
- Menerapkan **latent space visualization** (t-SNE, PCA) untuk eksplorasi distribusi data.
- Menggunakan fitur laten sebagai input untuk model klasifikasi sederhana (misalnya SVM atau logistic regression).

Eksperimen ini nunjukin kalau autoencoder bisa jadi alat kreatif buat menyederhanakan data, di mana fitur yang diekstrak tetap menyimpan informasi penting tanpa kehilangan makna visual.