

Nama: Rizki Aprilia Rahman (1103213007)

Kelas: TK-45-GAB

LAPORAN TUGAS 14 PEMBELAJARAN MESIN

- Bidirectional RNN

Dataset yang digunakan dalam eksperimen ini adalah data perbankan yang terdiri dari 45,211 entri dengan 17 kolom, mencakup informasi tentang karakteristik nasabah, riwayat interaksi dengan bank, dan hasil (kolom y) yang menunjukkan apakah nasabah tertarik pada produk tertentu (yes/no).

Eksperimen ini bertujuan untuk:

1. Melatih model **Bidirectional RNN** pada dataset tersebut.
2. Mengidentifikasi kombinasi parameter terbaik (ukuran hidden layer, pooling, optimizer, dan jumlah epoch) untuk menghasilkan akurasi prediksi tertinggi.

Preprocessing Data

- **Transformasi:**
 - Kolom kategorikal dikonversi menjadi one-hot encoding.
 - Kolom target (y) diubah menjadi format biner (1 untuk yes, 0 untuk no).
- **Normalisasi:**
 - Fitur numerik dinormalisasi menggunakan StandardScaler.
- **Pembagian Data:**
 - 80% data untuk pelatihan.
 - 20% data untuk pengujian.

Arsitektur Model

Model yang digunakan adalah **Bidirectional RNN** dengan konfigurasi sebagai berikut:

Fungsi Aktivasi

- **Output layer menggunakan fungsi aktivasi sigmoid untuk menghasilkan nilai probabilitas.**

Fungsi Loss

- **Binary Cross Entropy Loss digunakan untuk menghitung selisih antara prediksi dan label sebenarnya.**

Optimizer

- **Tiga optimizer diuji: SGD, RMSprop, dan Adam.**

Arsitektur

- **Layer RNN:**
 - Hidden layer bidirectional dengan dimensi variabel (16, 32, 64).
- **Pooling:**
 - Max pooling atau average pooling untuk mereduksi dimensi keluaran RNN.
- **Fully Connected Layer:**
 - Satu layer untuk menghasilkan output akhir.

Prosedur Pelatihan

Konfigurasi Eksperimen

- **Hidden Size:** 16, 32, 64
- **Pooling:** Max, Average
- **Optimizer:** SGD, RMSprop, Adam
- **Jumlah Epoch:** 5, 50, 100, 250, 350
- **Early Stopping:** Jika validasi loss tidak membaik selama 10 epoch berturut-turut.

Langkah Pelatihan

1. Model dilatih menggunakan data pelatihan dengan batch size 64.
2. Evaluasi dilakukan pada data pengujian di setiap epoch untuk menghitung loss dan akurasi.
3. Scheduler digunakan untuk menyesuaikan learning rate berdasarkan validasi loss.

Hasil Eksperimen

Hasil Akurasi Terbaik

Model Type	Hidden Size	Pooling	Optimizer	Epochs	Accuracy
Bidirectional RNN 64	64	Avg	SGD	100	90.88%
Bidirectional RNN 16	16	Max	SGD	100	90.81%
Bidirectional RNN 32	32	Max	RMSprop	50	90.77%
Bidirectional RNN 64	64	Max	SGD	350	90.73%
Bidirectional RNN 64	64	Avg	SGD	250	90.67%

Analisis Hasil

- Kombinasi **hidden size 64**, **average pooling**, dan **SGD optimizer** dengan **100 epoch** memberikan akurasi terbaik.
- Optimizer **SGD** menunjukkan kinerja yang stabil di berbagai konfigurasi.
- **Max pooling** memberikan hasil yang sedikit lebih baik dibandingkan **average pooling** pada hidden size kecil.
- Penggunaan epoch terlalu banyak (contoh: 350) cenderung meningkatkan risiko overfitting.

Kesimpulan

1. **Model Terbaik:**
 - **Bidirectional RNN** dengan konfigurasi:
 - Hidden Size: 64
 - Pooling: Average

- Optimizer: SGD
- Epochs: 100

2. **Akurasi Tertinggi:**

- 90.88% pada data pengujian.

3. **Rekomendasi:**

- Optimasi lebih lanjut dapat dilakukan dengan mencoba kombinasi learning rate dan jumlah hidden layers tambahan.
- Pengujian pada dataset lain untuk validasi generalisasi model.

Lampiran

Top 10 Hasil Eksperimen:

Model Type	Hidden Size	Pooling	Optimizer	Epochs	Accuracy
Bidirectional RNN 64	Avg		SGD	100	90.88%
Bidirectional RNN 16	Max		SGD	100	90.81%
Bidirectional RNN 32	Max		RMSprop	50	90.77%
Bidirectional RNN 64	Max		SGD	350	90.73%
Bidirectional RNN 64	Avg		SGD	250	90.67%