

Nama: M.Rizky Fadillah (1103210259)

Kelas: TK-45-GAB

LAPORAN TUGAS 14 PEMBELAJARAN MESIN

- RNN & deep RNN

Tujuan eksperimen ini adalah untuk mengevaluasi kinerja model **Recurrent Neural Network (RNN)** dalam klasifikasi multi-kelas pada dataset **Spambase**, yang berfokus pada pengklasifikasian data sebagai spam atau bukan spam. Model ini diuji dengan berbagai konfigurasi parameter, seperti ukuran hidden layer, jenis pooling, algoritma optimizer, dan jumlah epoch untuk pelatihan. Analisis dilakukan untuk mengidentifikasi pengaruh masing-masing faktor terhadap akurasi model.

2. Persiapan Data

- **Dataset:** Dataset yang digunakan dalam eksperimen ini adalah **Spambase**, yang diunggah melalui Google Colab. Dataset ini memiliki 57 fitur dan 1 target yang mengklasifikasikan apakah data tersebut spam atau bukan spam.
- **Preprocessing:**
 - **Pemisahan Fitur dan Target:** Fitur dan target dipisahkan menggunakan indeks kolom, dengan X sebagai fitur dan y sebagai target.
 - **Standardisasi:** Fitur distandarisasi menggunakan StandardScaler dari scikit-learn untuk memastikan bahwa data berada dalam rentang yang seragam, yang dapat mempercepat proses pelatihan model.
 - **Pembagian Data:** Data dibagi menjadi data pelatihan (80%) dan data pengujian (20%) menggunakan fungsi `train_test_split` dengan stratifikasi, memastikan distribusi label pada kedua set tersebut tetap seimbang.

3. Desain Model

Model yang digunakan adalah **RNNModel** yang mengimplementasikan RNN dengan beberapa komponen utama:

- **RNN Layer:** Menggunakan layer RNN untuk memproses data sekuensial dan menangkap pola temporal.
- **Pooling:** Dua jenis pooling diuji, yaitu:
 - **Max Pooling:** Mengambil nilai maksimum dari output RNN sepanjang dimensi waktu.
 - **Average Pooling:** Mengambil rata-rata dari output RNN sepanjang dimensi waktu.
- **Fully Connected Layer:** Lapisan fully connected digunakan untuk menghasilkan output akhir dari model.

4. Fungsi Pelatihan dan Evaluasi

- **Fungsi Pelatihan:** Model dilatih menggunakan fungsi `train_model`, yang melakukan proses pelatihan dalam beberapa epoch. Dalam setiap epoch, dilakukan perhitungan loss menggunakan `CrossEntropyLoss`, diikuti dengan pembaruan bobot menggunakan optimizers seperti Adam. Scheduler digunakan untuk mengurangi learning rate setiap beberapa epoch.
- **Fungsi Evaluasi:** Fungsi `evaluate_model` digunakan untuk mengukur akurasi model pada data pengujian. Akurasi dihitung dengan membandingkan hasil prediksi dengan label sebenarnya pada data uji.

5. Eksperimen dan Hasil

Eksperimen dilakukan dengan mengubah berbagai parameter untuk menguji pengaruhnya terhadap kinerja model. Parameter yang diuji meliputi ukuran hidden layer, jenis pooling, algoritma optimizer, dan jumlah epoch.

5.1. Pengaruh Ukuran Hidden Layer

Dua ukuran hidden layer diuji, yaitu 32 dan 64. Hasil eksperimen menunjukkan bahwa ukuran hidden layer 32 memberikan akurasi yang sedikit lebih baik (93.81%) dibandingkan dengan hidden layer 64 (92.62%).

Ukuran Hidden Layer	Akurasi
32	93.81%
64	92.62%

5.2. Pengaruh Jenis Pooling

Jenis pooling yang diuji adalah max pooling dan average pooling. Hasilnya menunjukkan bahwa max pooling menghasilkan akurasi sedikit lebih tinggi (93.92%) dibandingkan dengan average pooling (93.49%).

Jenis Pooling	Akurasi
Max Pooling	93.92%
Average Pooling	93.49%

5.3. Pengaruh Optimizer

Tiga optimizer diuji: SGD, RMSProp, dan Adam. Hasil eksperimen menunjukkan bahwa optimizer SGD menghasilkan akurasi yang lebih rendah (91.10%) dibandingkan dengan RMSProp dan Adam, yang keduanya menghasilkan akurasi yang sama (92.83%).

Optimizer	Akurasi
SGD	91.10%
RMSProp	92.83%
Adam	92.83%

5.4. Pengaruh Jumlah Epoch

Jumlah epoch diuji pada berbagai nilai: 5, 50, 100, 250, dan 350. Hasil eksperimen menunjukkan bahwa 250 epoch menghasilkan akurasi terbaik (94.14%), sementara jumlah epoch lainnya memberikan hasil yang sedikit lebih rendah.

Jumlah Epoch	Akurasi
5	93.49%
50	93.27%
100	93.27%

Jumlah Epoch	Akurasi
250	94.14%
350	93.27%

6. Kesimpulan

Dari eksperimen yang dilakukan, dapat disimpulkan bahwa:

- **Ukuran Hidden Layer:** Ukuran hidden layer 32 memberikan hasil sedikit lebih baik daripada 64.
- **Jenis Pooling:** Max pooling sedikit lebih baik daripada average pooling.
- **Optimizer:** Optimizer Adam dan RMSProp menghasilkan hasil yang lebih baik daripada SGD.
- **Jumlah Epoch:** 250 epoch memberikan akurasi terbaik.

7. Saran

Untuk eksperimen lebih lanjut, disarankan untuk menguji model dengan jumlah layer yang lebih banyak dan menggunakan teknik regularisasi seperti dropout untuk menghindari overfitting. Selain itu, penyesuaian lebih lanjut terhadap parameter optimizer, seperti learning rate, dapat meningkatkan performa model lebih jauh.