


Tugas Week 14 Machine Learning

Nama : Rizki Ramadhan

NIM : 1103213091

Dataset “**Diabetes 130-US Hospitals for Years 1999-2008**”

Bidirectional RNN

 Diabetes 130-US Hospitals for Years 1999-2008 Donated on 5/2/2014		
The dataset represents ten years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. Each row concerns hospital records of patients diagnosed with diabetes, who underwent laboratory, medications, and stayed up to 14...		
Dataset Characteristics	Subject Area	Associated Tasks
Multivariate	Health and Medicine	Classification, Clustering
Feature Type	# Instances	# Features
Categorical, Integer	101766	47

Link : [Diabetes 130-US Hospitals for Years 1999-2008 - UCI Machine Learning Repository](#)

1. Hidden size 32

A. Jumlah Layer

2 Layer:

Memberikan hasil paling stabil dengan akurasi terbaik mencapai 59.12% (Adam, pooling Avg, 350 epoch).

Adam lebih unggul dibandingkan optimizer lain pada semua kombinasi.

3 Layer:

Performa sedikit lebih rendah dibandingkan 2 layer pada pooling Max dan Avg, dengan akurasi terbaik 58.95% (Adam, pooling Max, 250 epoch).

RMSprop menunjukkan performa yang tidak stabil pada jumlah layer ini.

4 Layer:

Menunjukkan akurasi tertinggi sebesar 59.13% (Adam, pooling Avg, 350 epoch), tetapi performa optimizer lain cenderung lebih rendah dan sering terjadi overfitting.

Kesimpulan: 4 layer memberikan akurasi tertinggi, tetapi 2 layer adalah pilihan yang lebih stabil secara keseluruhan.

B. Pooling Method (Max vs Avg)

Pooling Max:

Memberikan akurasi terbaik pada epoch menengah hingga tinggi, dengan akurasi maksimum 58.95% (Adam, 3 layer, 250 epoch).

Pooling Max cocok untuk kombinasi Adam dengan jumlah layer yang lebih sedikit (2 atau 3 layer).

Pooling Avg:

Pooling Avg menunjukkan performa konsisten yang lebih tinggi dibandingkan Max, terutama pada epoch tinggi. Akurasi terbaik adalah 59.13% (Adam, 4 layer, 350 epoch).

Cocok untuk menangani data yang lebih kompleks dengan jumlah layer yang lebih besar.

Kesimpulan: Pooling Avg sedikit lebih unggul secara keseluruhan dibandingkan Max.

C. Optimizer (SGD, RMSprop, Adam)

SGD:

Peningkatan lambat dengan akurasi maksimum 56.27% (Pooling Avg, 350 epoch). Tidak direkomendasikan untuk model kompleks.

RMSprop:

Akurasi lebih tinggi dibandingkan SGD, tetapi tidak sebaik Adam. Akurasi tertinggi adalah 58.42% (Pooling Max, 2 layer, 350 epoch).

Rentan terhadap overfitting pada layer lebih tinggi.

Adam:

Konsisten memberikan akurasi terbaik di semua kombinasi. Akurasi tertinggi adalah 59.13% (Pooling Avg, 4 layer, 350 epoch).

Kesimpulan: Adam adalah optimizer terbaik untuk Bidirectional RNN dengan hidden size 32.

D. Jumlah Epoch

Performa meningkat signifikan hingga 250 epoch, dengan beberapa kombinasi mencapai puncak pada 350 epoch tanpa tanda overfitting, terutama dengan Adam.

Epoch optimal adalah 250-350 tergantung pada jumlah layer.

E. Kesimpulan hidden size

Hyperparameter Optimal untuk Bidirectional RNN dengan Hidden Size 32:

Layers: 4

Pooling: Avg

Optimizer: Adam

Epochs: 350

Akurasi Terbaik: 59.13%

2. Hidden size 64

A. Jumlah Layer

2 Layer:

Memberikan hasil paling stabil dengan akurasi terbaik mencapai 59.21% (Adam, pooling Max, 100 epoch).

Adam consistently unggul dibandingkan optimizer lain.

3 Layer:

Sedikit lebih rendah dibandingkan 2 layer, dengan akurasi tertinggi mencapai 58.89% (Adam, pooling Avg, 350 epoch).

Penambahan layer memberikan sedikit peningkatan fleksibilitas, tetapi risiko overfitting meningkat pada kombinasi tertentu.

4 Layer:

Tidak memberikan peningkatan signifikan dibandingkan 2 layer. Akurasi terbaik mencapai 58.87% (Adam, pooling Avg, 350 epoch).

Overfitting lebih sering terjadi, terutama dengan optimizer RMSprop.

Kesimpulan: 2 layer memberikan akurasi terbaik secara konsisten, dengan stabilitas lebih baik dibandingkan jumlah layer yang lebih tinggi

B. Pooling Method (Max vs Avg)

Pooling Max:

Memberikan hasil yang kompetitif pada semua jumlah layer. Akurasi terbaik adalah 59.21% (Adam, 2 layer, 100 epoch).

Konsisten unggul pada optimizer Adam dibandingkan pooling Avg.

Pooling Avg:

Lebih stabil dibandingkan Max pada jumlah layer lebih tinggi (3 atau 4 layer). Akurasi terbaik adalah 58.89% (Adam, 3 layer, 350 epoch).

Tidak cocok untuk optimizer RMSprop karena cenderung memberikan hasil yang lebih rendah.

Kesimpulan: Pooling Max adalah pilihan optimal untuk hidden size 64, terutama dengan Adam.

C. Optimizer (SGD, RMSprop, Adam)

SGD:

Memiliki peningkatan lambat dengan akurasi maksimum 56.41% (Pooling Avg, 350 epoch). Tidak direkomendasikan untuk model dengan hidden size besar.

RMSprop:

Akurasi lebih tinggi dibandingkan SGD, tetapi tidak sebaik Adam. Akurasi terbaik adalah 56.99% (Pooling Max, 2 layer, 250 epoch).

Rentan terhadap overfitting pada jumlah layer yang lebih tinggi.

Adam:

Konsisten memberikan akurasi terbaik di semua kombinasi. Akurasi tertinggi adalah 59.21% (Pooling Max, 2 layer, 100 epoch).

Kesimpulan: Adam adalah optimizer terbaik untuk Bidirectional RNN dengan hidden size 64.

D. Jumlah Epoch

Performa meningkat signifikan hingga 100 epoch, dengan beberapa kombinasi mencapai puncak pada 250 epoch, tetapi risiko overfitting meningkat setelah itu.

Epoch optimal untuk hidden size 64 adalah 100-250 epoch.

E. Kesimpulan hidden size

Hyperparameter Optimal untuk Bidirectional RNN dengan Hidden Size 64:

Layers: 2

Pooling: Max

Optimizer: Adam

Epochs: 100

Akurasi Terbaik: 59.21%

3. Hidden size 128

A. Jumlah Layer

2 Layer:

Menunjukkan performa paling stabil dengan akurasi terbaik mencapai 58.99% (Adam, pooling Max, 250 epoch).

SGD cukup kompetitif tetapi tidak melampaui performa Adam.

3 Layer:

Akurasi terbaik mencapai 59.09% (Adam, pooling Avg, 350 epoch).

Penambahan layer meningkatkan fleksibilitas tetapi tidak signifikan dalam performa keseluruhan.

4 Layer:

Hasil yang lebih rendah dibandingkan 2 atau 3 layer, dengan akurasi terbaik mencapai 58.87% (Adam, pooling Avg, 100 epoch).

RMSprop cenderung menunjukkan overfitting pada jumlah layer lebih tinggi.

Kesimpulan: 3 layer adalah pilihan terbaik untuk akurasi maksimum dengan stabilitas yang baik.

B. Pooling Method (Max vs Avg)

Pooling Max:

Memberikan hasil terbaik pada kombinasi layer dan epoch yang lebih rendah hingga menengah, dengan akurasi maksimum 58.99% (Adam, 2 layer, 250 epoch).

Cocok untuk kombinasi sederhana.

Pooling Avg:

Memberikan akurasi terbaik pada jumlah layer lebih tinggi, dengan hasil tertinggi mencapai 59.09% (Adam, 3 layer, 350 epoch).

Lebih fleksibel tetapi cenderung memerlukan epoch lebih banyak.

Kesimpulan: Pooling Avg sedikit lebih unggul, terutama pada jumlah layer yang lebih banyak.

C. Optimizer (SGD, RMSprop, Adam)

SGD:

Performa konsisten tetapi memiliki batas maksimum pada 56.59% (Pooling Avg, 350 epoch). Tidak ideal untuk hidden size besar.

RMSprop:

Performa tidak stabil dengan akurasi maksimum hanya 54.47% (Pooling Avg, 250 epoch). Cenderung overfitting pada jumlah layer tinggi.

Adam:

Memberikan performa terbaik di semua kombinasi, dengan akurasi tertinggi mencapai 59.09% (Pooling Avg, 3 layer, 350 epoch).

Kesimpulan: Adam adalah optimizer terbaik untuk Bidirectional RNN dengan hidden size 128.

D. Jumlah Epoch

Performa meningkat signifikan hingga 250-350 epoch, dengan risiko overfitting pada beberapa kombinasi.

Epoch optimal adalah 250-350 epoch, tergantung jumlah layer dan pooling type.

E. Kesimpulan hidden size

Hyperparameter Optimal untuk Bidirectional RNN dengan Hidden Size 128:

Layers: 3

Pooling: Avg

Optimizer: Adam

Epochs: 350

Akurasi Terbaik: 59.09%

KESIMPULAN MODEL BIDIRECTIONAL RNN

Hyperparameter Optimal

Hidden Size: 64

Number of Layers: 2

Pooling Type: Max

Optimizer: Adam

Epochs: 100

Best Accuracy: 59.21%

Kombinasi ini menghasilkan performa terbaik dengan keseimbangan antara kapasitas model, stabilitas, dan akurasi prediksi

- **Hidden Size:**

- Hidden size **64** menunjukkan performa paling optimal, mengungguli hidden size **32** dan **128**. Hidden size **128** cenderung lebih rentan terhadap overfitting.

- **Jumlah Layers:**

- Model dengan **2 layer** adalah yang paling stabil dan memberikan akurasi terbaik. Penambahan layer cenderung meningkatkan kompleksitas tanpa peningkatan akurasi yang signifikan.

- **Pooling :**

- **Pooling Max** unggul untuk kombinasi optimal, memberikan hasil konsisten dibandingkan pooling Avg.

- **Optimizer:**

- **Adam** secara konsisten menjadi pilihan terbaik, memberikan konvergensi lebih cepat dan akurasi lebih tinggi dibandingkan SGD dan RMSprop.
- **Epochs:**
 - **100 epoch** adalah jumlah optimal untuk model ini. Penambahan epoch memberikan peningkatan kecil tetapi berisiko overfitting, terutama pada hidden size lebih besar.