

Ujian Akhir Artificial Intelligence

Pertama-tama dataset dipanggil menggunakan pandas dataframe setelah itu diolah agar mendapatkan dataframe antara nama file dengan setiap penyakitnya. Setelah didapatkan dataframe tersebut, data gambar di *copy* menggunakan *for* loop beserta dengan *command line* : *copy* untuk membuat dataset.

Setelah itu, data image yang sudah diproses diambil menggunakan *imutils* untuk mengambil path dan dibaca menggunakan *Pillow Image*. Data image di-*resize* menjadi 128x128. Pembagian dataset menjadi Train, Test, Val menggunakan library dari *sklearn.model_selection.train_test_split*

Model pertama yang dipakai adalah Convolution Neural Network yang diambil dari referensi 4. Training dilakukan dengan dataset original pertama-tama. Batch_size sebesar 32 dan Epoch sebesar 30. Hasilnya dapat dilihat pada **Table 1**.

Table 1. Hasil dari training dengan model CNN

Datasets	Parameter	CNN
Original Dataset	Akurasi Train	0.775
	Akurasi Evaluate	0.45
	Akurasi Validasi	0.35
Inpainted Dataset	Akurasi Train	0.88
	Akurasi Evaluate	0.4
	Akurasi Validasi	0.55
Original + Inpainted Dataset	Akurasi Train	0.9
	Akurasi Evaluate	0.5
	Akurasi Validasi	0.545
Original Dataset 2 Kelas	Akurasi Train	0.9625
	Akurasi Evaluate	0.814
	Akurasi Validasi	0.85
Inpainted Dataset 2 Kelas	Akurasi Train	0.8875
	Akurasi Evaluate	0.899
	Akurasi Validasi	0.85
Original + Inpainted Dataset 2 Kelas	Akurasi Train	0.557
	Akurasi Evaluate	0.444
	Akurasi Validasi	0.444

Hasil yang didapatkan menggunakan CNN kurang memuaskan, selanjutnya saya coba untuk menggunakan file yang pernah dipakai dalam latihan klasifikasi 3Scences untuk mencoba mengklasifikasi dataset tersebut. Modelnya antara lain Random Forest, Support Vector Machine (SVM), dan Multi Layer Perceptron(MLP). Hasilnya dapat dilihat pada **Table 2**.

Table 2. Hasil dari training dengan beberapa model Classifier

Datasets	Parameter	Random Forest	SVM	MLP
Original Dataset	Precision	0.56	0.69	0.57
	Recall	0.48	0.62	0.5
	F1-Score	0.5	0.6	0.45
Inpainted Dataset	Precision	0.54	0.63	0.36
	Recall	0.47	0.52	0.42
	F1-Score	0.48	0.53	0.38
Original + Inpainted Dataset	Precision	0.59	0.51	0.52
	Recall	0.57	0.50	0.48
	F1-Score	0.66	0.44	0.45
Original Dataset 2 Kelas	Precision	0.81	0.92	0.87
	Recall	0.78	0.92	0.80
	F1-Score	0.88	0.95	0.9
Inpainted Dataset 2 Kelas	Precision	0.84	0.92	0.81
	Recall	0.73	0.92	0.78
	F1-Score	0.88	0.95	0.88
Original + Inpainted Dataset 2 Kelas	Precision	0.20	0.53	0.6
	Recall	0.17	0.53	0.6
	F1-Score	0.19	0.56	0.54

Kesimpulan : Model CNN mungkin tidak terlalu bagus untuk jumlah data yang sedikit sehingga kurang lebih menghasilkan performa yang sama dengan model Classifier lainnya

Referensi :

1. <https://github.com/jrosebr1/imutils/blob/master/imutils/paths.py>
2. <https://keras.io/models/sequential/>
3. <https://github.com/deep-learning-indaba/indaba-2018>

Special Thanks to collaborator:

Riskyana Dewi