

**TUGAS 1**  
**REGRESI BOSTON HOUSING PRICE DENGAN ANN**

**SEBAGAI TUGAS EKSPLORASI PADA**  
**MATA KULIAH PEMBELARAN MESIN LANJUT**

Oleh  
**RAHMAN INDRA KESUMA**  
**NIM: 33221026**  
**(Program Studi Doktor Teknik Elektro dan Informatika)**



**INSTITUT TEKNOLOGI BANDUNG**  
**Maret 2022**

## Beberapa Tahapan Utama pada Ekplorasi dan Percobaan

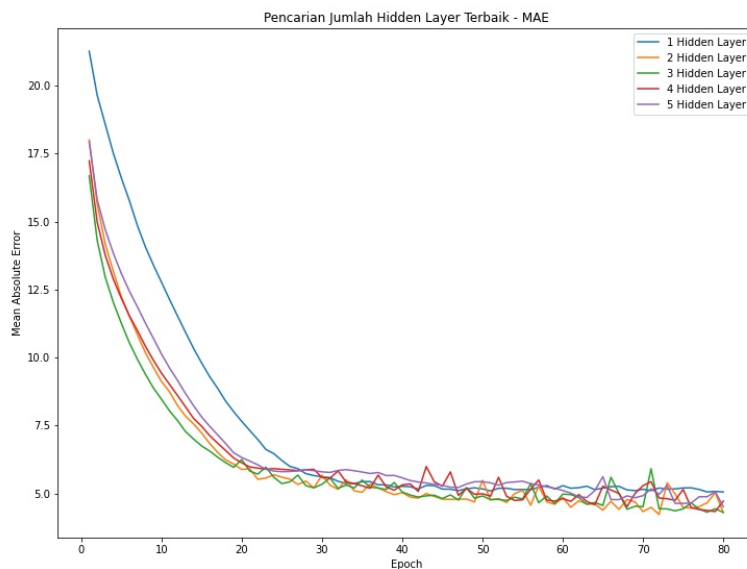
1. Melakukan pembagian dataset dalam tiga kelompok data, (i) data training untuk menghasilkan model pembelajaran; (ii) data validation digunakan untuk memiliki model dengan setting hyperparameter terbaik; serta (iii) data testing untuk menguji kualitas kinerja dari model dengan setting hyperparameter terbaik. Pembagian dataset tersebut seperti terlihat pada gambar berikut.

```
(x_train, y_train), (x_remaining, y_remaining) = boston_housing.load_data(path="boston_housing.npz", test_split=0.3, seed=113)
x_valid, x_test, y_valid, y_test = train_test_split(x_remaining, y_remaining, test_size=0.5)

print("Ukuran Training Data -> ", len(x_train), "( Max_y : ",max(y_train)," & Min_y : ",min(y_train),")")
print("Ukuran Validation Data -> ", len(x_valid), "( Max_y : ",max(y_valid)," & Min_y : ",min(y_valid),")")
print("Ukuran Testing Data -> ", len(x_test), "( Max_y : ",max(y_test)," & Min_y : ",min(y_test),")")
print("TOTAL DATA = ", len(x_train)+len(x_valid)+len(x_test))

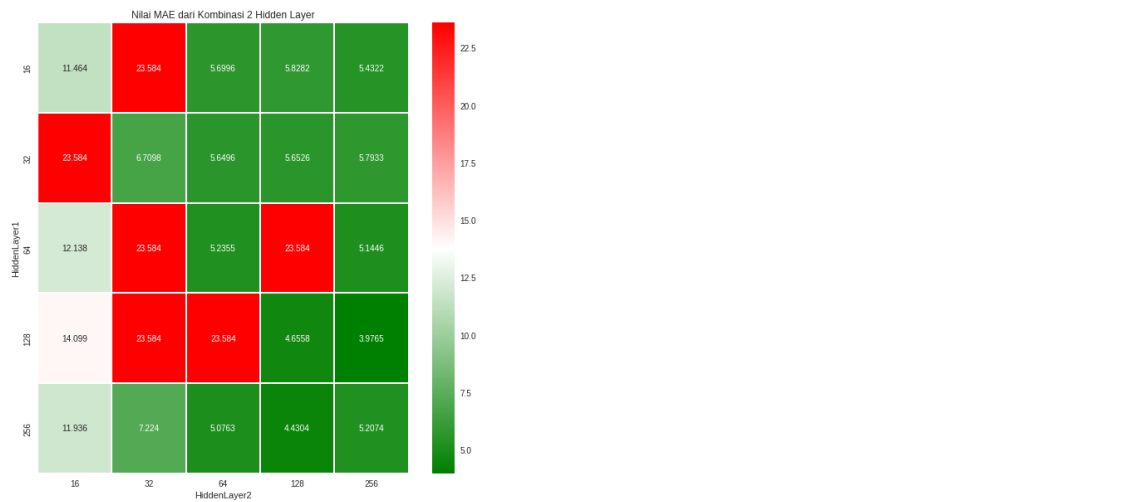
Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-datasets/boston\_housing.npz
57344/57026 [=====] - 0s 0us/step
65536/57026 [=====] - 0s 0us/step
Ukuran Training Data -> 354 ( Max_y : 50.0 & Min_y : 5.0 )
Ukuran Validation Data -> 76 ( Max_y : 50.0 & Min_y : 5.6 )
Ukuran Testing Data -> 76 ( Max_y : 50.0 & Min_y : 7.0 )
TOTAL DATA = 506
```

2. Proses eksplorasi dengan percobaan untuk mencari jumlah hidden layer terbaik (1-5 hidden layer) dalam regression model dengan tahapan dimana setiap hidden layer memiliki 128 neuron dan menggunakan fungsi aktivasi Sigmoid. Berikut hasil yang diperoleh dari hasil eksplorasi, yang mana penggunaan 80 epoch, 50 batch size, 2 hidden layer (128 neuron di setiap hidden layer), sigmoid activation function pada hidden layer, relu activation function pada output layer, RMSProp optimizer dan mean absolute error loss function, memperoleh kinerja yang baik dengan nilai loss pada validation data sebesar 4,3074.



3. Walaupun 2 hidden layer memiliki nilai stabilitas MAE yang baik, namun 3 hidden layer memiliki kualitas kinerja yang mendekati 2 hidden layer. Maka itu, pada percobaan berikut dilakukan pencarian jumlah hidden neuron dengan kombinasi nilai 16, 32, 64, 128, dan 256 dari 2 dan 3 hidden layer yang didefinisikan. Berdasarkan hasil yang diperoleh seperti yang terlihat pada heatmap berikut, diperoleh hasil bahwa arsitektur dengan 2 hidden layer

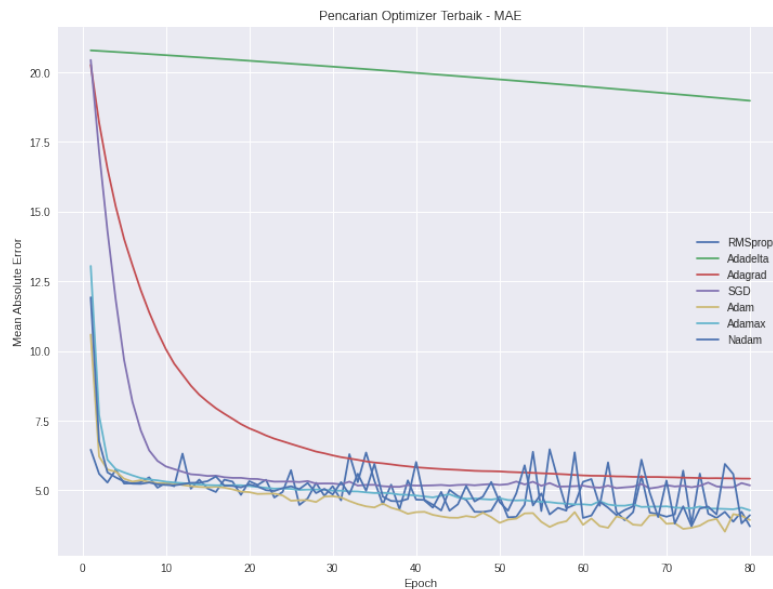
dengan komposisi 128 neuron pada hidden layer 1 dan 256 neuron pada hidden layer 2, memperoleh nilai MAE yang baik itu sebesar 3,9763.



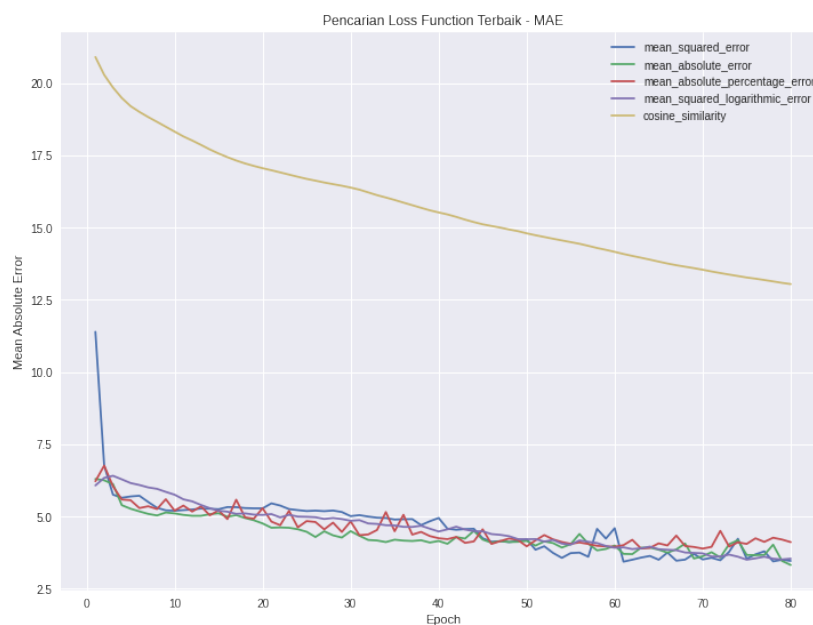
- Berikutnya eksplorasi dan percobaan dilanjutkan untuk mencari kombinasi activation function pada 2 hidden layer yang terbaik dari percobaan sebelumnya. Digunakan beberapa potensi activation function yang bisa dikombinasikan, meliputi linier, relu, sigmoid, tanh dan softsign. Diperoleh hasil dari percobaan tersebut pada kombinasi penggunaan softsign pada hidden layer ke-1 dan tanh pada hidden layer ke-2 memperoleh hasil MAE yang baik yaitu sebesar 3,7867.



5. Selanjutnya, proses eksplorasi berfokus untuk menemukan optimizer terbaik pada proses training dengan tujuan meningkatkan hasil model terhadap validation data yang telah disiapkan. Beberapa optimizer yang dilakukan percobaan meliputi RMSProp, Adadelta, Adagrad, SGD with Momentum, Adam, Adamax, dan Nadam. Berdasarkan hasil yang diperoleh dan terlihat pada gambar berikut, diperoleh bahwa model dengan Adam optimizer pada proses traininnya memiliki kinerja yang stabil dengan nilai MAE pada validation data sebesar 3,9381.



6. Pada akhirnya, proses percobaan diakhiri dengan mencari loss function yang mengarahkan pada hasil model terbaik. Dilakukan percobaan dengan beberapa loss function meliputi mean squared error, mean absolute error, mean absolute percentage error, mean squared logarithmic error, dan cosine similarity. Diperoleh hasil dari percobaan bahwa dengan menerapkan mean squared error loss function pada proses training, memperoleh nilai MAE sebesar 3,4621 terhadap validation data.



7. Dari model terbaik yang telah disimpan dengan arsitektur dan pengaturan pada proses pelatihan sebagai berikut:
  - a. Jumlah Epoch adalah 80
  - b. Ukuran Batch adalah 50
  - c. Jumlah Hidden Layer adalah 2
  - d. Kombinasi Jumlah Hidden Neuron adalah 128 di Layer 1 dan 256 di Layer 2
  - e. Kombinasi Fungsi Aktifitasi adalah Softsign di Hidden Layer 1, Tanh di Hidden Layer 2, dan Relu di Output Layer
  - f. Fungsi Optimizer adalah Adam
  - g. Fungsi Loss adalah Mean Squared Error
  - h. Nilai Learning Rate adalah 0,001

Diperoleh hasil sebagai berikut, yang mana nilai MAE sebesar 3,3800 dengan testing data.

```
# Memuat model yang tersimpan pada google drive dengan cara mengakses google drive akun pribadi
# Selanjutnya mengambil file h5 dari penyimpanan model terbaik
drive.mount('/content/drive')
best_model = models.load_model('/content/drive/MyDrive/Eksplorasi Machine Learning/Tugas Deep Learn
best_model.summary()
```

---

```
Mounted at /content/drive
Model: "sequential_114"
```

Layer (type)	Output Shape	Param #
flatten_114 (Flatten)	(None, 13)	0
dense_341 (Dense)	(None, 128)	1792
dense_342 (Dense)	(None, 256)	33024
dense_343 (Dense)	(None, 1)	257

---

```
Total params: 35,073
Trainable params: 35,073
Non-trainable params: 0
```

---

```
# Menguji model terbaik ke testing data yang telah disediakan
hasil_predict_test = best_model.predict(x_test)
Loss, MAE = best_model.evaluate(x_test, y_test, verbose=1)
print('Restored model, MAE: {:.4f}'.format(MAE))
```

```
3/3 [=====] - 0s 9ms/step - loss: 22.9975 - mean_absolute_error: 3.3800
Restored model, MAE: 3.3800
```