

PEMBELAJARAN MESIN
***K-FOLD CROSS-VALIDATION* UNTUK KLASIFIKASI PADA 2**
KELAS DATA IRIS



Vincent Michael Sutanto

16/398531/PA/17492

PROGRAM STUDI ILMU KOMPUTER
DEPARTEMEN ILMU KOMPUTER DAN ELEKTRONIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS GADJAH MADA
YOGYAKARTA

2018

K-FOLD CROSS-VALIDATION UNTUK KLASIFIKASI KELAS IRIS DENGAN DATA IRIS

Link Github : <https://github.com/vincentmichael089/ML-SLP-K-Fold>

Bahasa Pemrograman : Python (*Jupyter Notebook*)

Library : Pandas (*Data-frame*), Matplotlib (*Plotting*)

PENJELASAN

K-Fold Cross-Validation merupakan suatu prosedur dalam Pembelajaran Mesin yang digunakan untuk mengevaluasi Model Pembelajaran Mesin pada jumlah data yang terbatas. Prosedur ini menggunakan parameter K yang merepresentasikan jumlah grup yang dibentuk dari suatu data. Langkah-langkah dalam prosedur ini sebagai berikut:

1. Membagi dataset kedalam K jumlah grup
2. Pada setiap grup, tentukan mayoritas data sebagai data *training* dan sebagian lainnya sebagai data *validation*
3. Evaluasi model data training dan validasi pada setiap grup

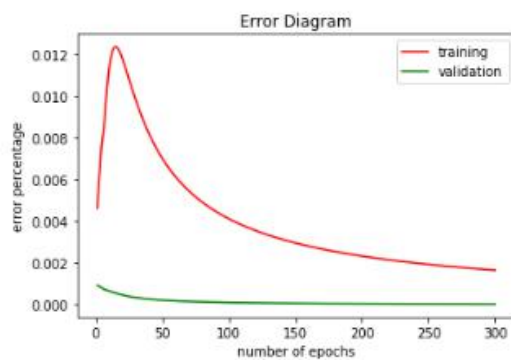
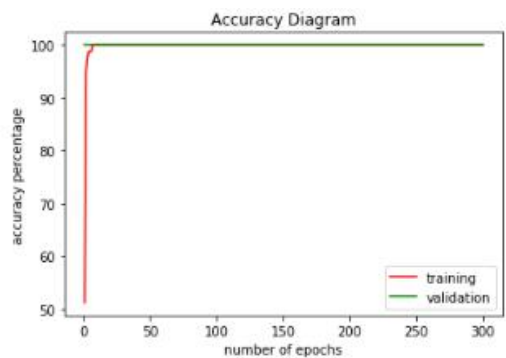
Pada tugas kali ini *K-Fold Cross-Validation* digunakan untuk mengklasifikasi data iris ke dalam 2 kelas yang telah ditentukan (*Iris Setosa* dan *Iris Versicolor*), dan menentukan perubahan prosentase dari akurasi dan error terhadap *epoch*. Nilai K yang digunakan adalah 5, dengan *Learning Rate* 0.1 dan 0.8. *Epochs* data sebanyak 300 per *K-Fold*. Perhitungan Akurasi per-*epoch* ditentukan dengan menjumlahkan *True Positive* dan *True Negative* kemudian dibagi dengan jumlah keseluruhan data. Perhitungan Error dengan kuadrat selisih *binary-type* dengan hasil fungsi aktivasi (*sigmoid*).

Langkah-langkah yang dilakukan dalam pengerjaan tugas adalah sebagai berikut:

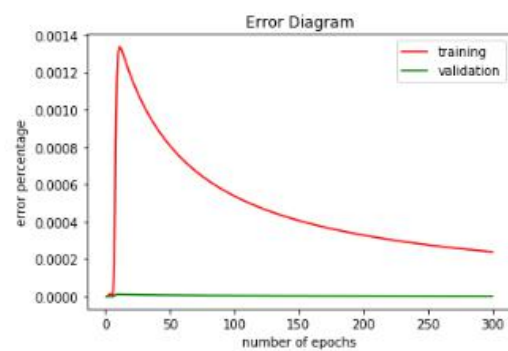
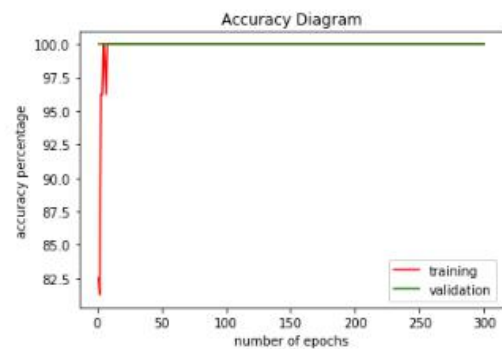
1. Inisialisasi data, mengambil 100 data pertama dari data iris, menambahkan kolom-kolom yang dibutuhkan dalam pengolahan data
2. Membagi data kedalam K=5 grup, dan dalam setiap grup dataset kemudian dipecah menjadi 2 bagian yaitu data *training* dan data *validation* dengan perbandingan 1:4
3. Dalam perputaran 300 *Epoch*, setiap iterasi dilakukan validasi model. Hasil akurasi dan error dari validasi maupun training kemudian disimpan untuk disajikan kedalam diagram plot

DIAGRAM AKURASI DAN ERROR

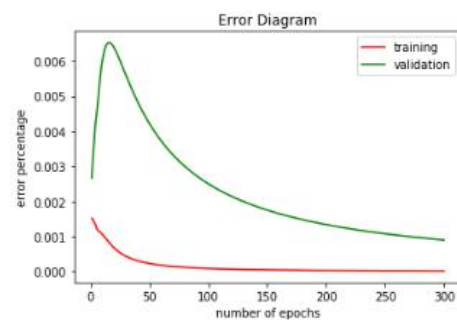
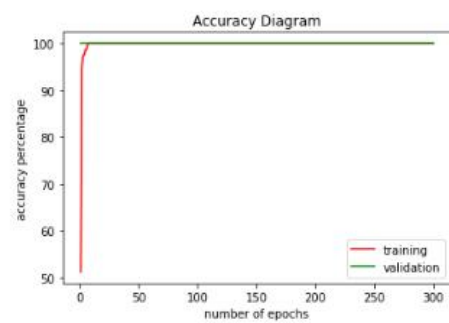
K-1 Learning Rate : 0.1 Epochs : 300



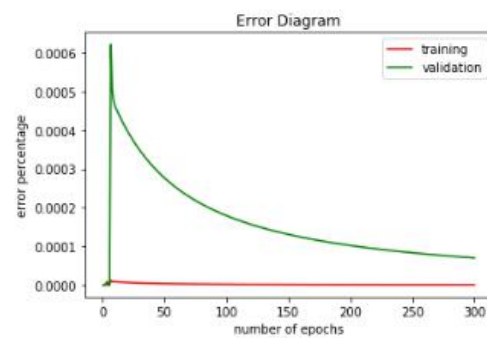
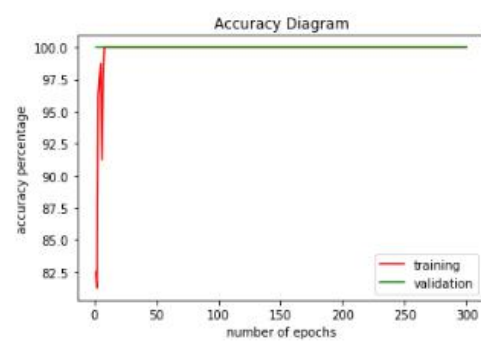
K-1 Learning Rate : 0.8 Epochs : 300



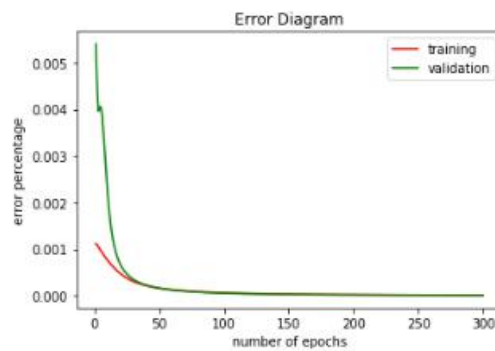
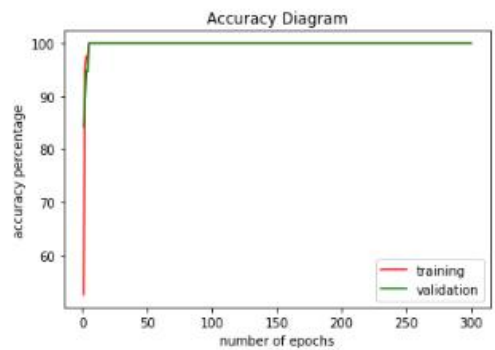
K-2 Learning Rate : 0.1 Epochs : 300



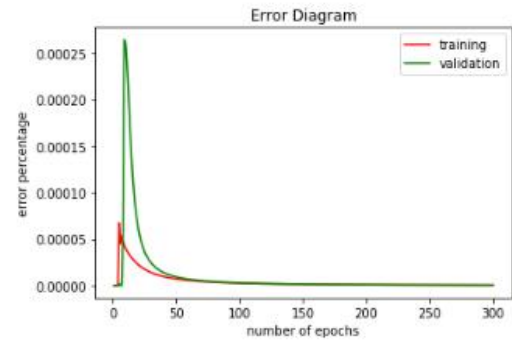
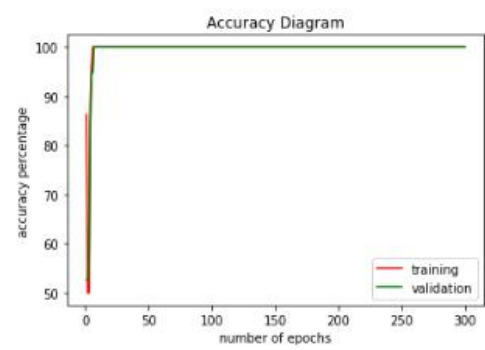
K-2 Learning Rate : 0.8 Epochs : 300



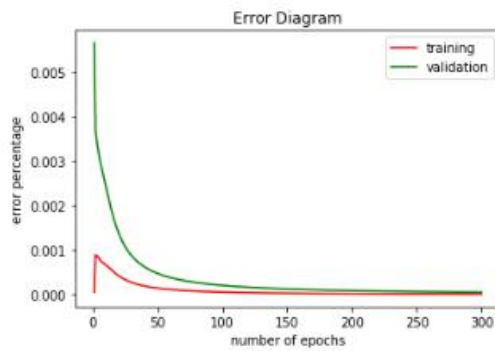
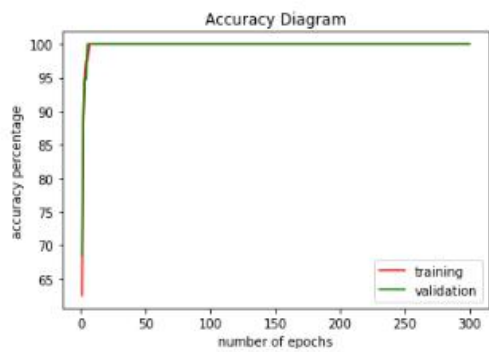
K-3 Learning Rate : 0.1 Epochs : 300



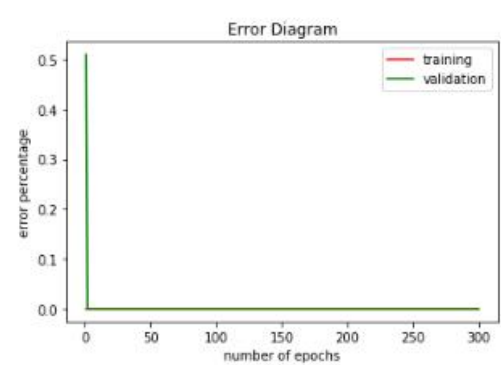
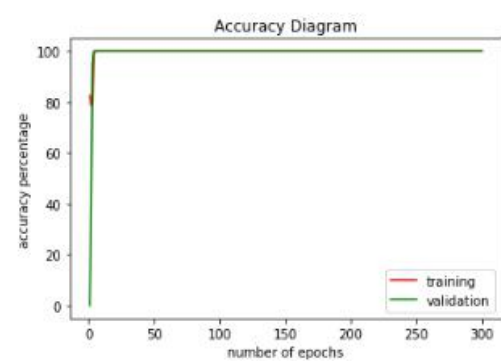
K-3 Learning Rate : 0.8 Epochs : 300



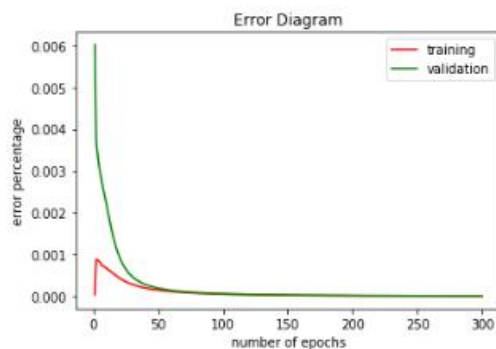
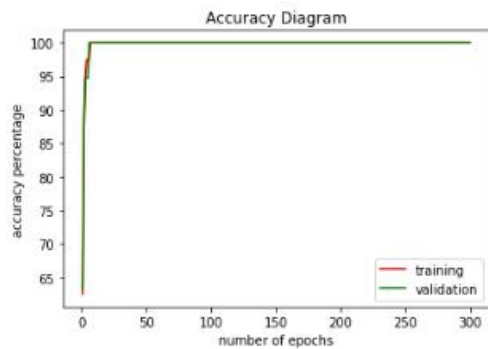
K-4 Learning Rate : 0.1 Epochs : 300



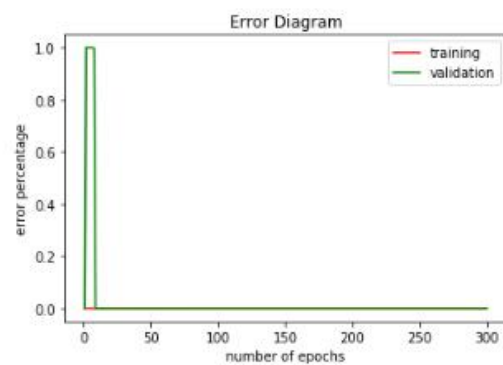
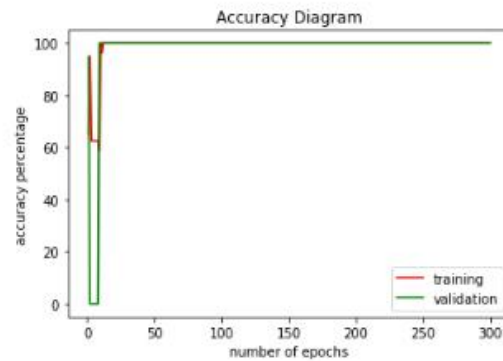
K-4 Learning Rate : 0.8 Epochs : 300



K-5 Learning Rate : 0.1 Epochs : 300



K-5 Learning Rate : 0.8 Epochs : 300



KESIMPULAN

Dari kelima diagram dapat disimpulkan suatu kesamaan yaitu:

1. Tingkat akurasi cenderung mencapai 100% dalam rentang iterasi kurang dari 25 *Epochs*
2. Prosentase Error bervariasi dengan error dibawah 1% pada seluruh K:
 - K-1 : data *training* mengalami kenaikan error yang dilanjutkan dengan penurunan pada rentang 0-50 *Epoch*, sementara data *validation* cenderung stabil dibawah 0.02 %
 - K-2 : Berkebalikan dengan K-1, data *validation* mengalami kenaikan error yang dilanjutkan dengan penurunan pada rentang 0-50 *Epoch*, sementara data *training* cenderung stabil dibawah 0.02 %
 - K-3 : data *training* tertinggi di angka 0.001% untuk setiap *Learning Rate*, dan data *validation* mengalami perbaikan error yang signifikan dalam rentang 50 *Epochs* pertama

K-4 : data *training* tertinggi di angka 0.001% untuk setiap *Learning Rate*,
dan data *validasi* mengalami perbaikan error yang signifikan dalam rentang
50 *Epochs* pertama

K-5 : error data *training* tertinggi di angka 0.001% untuk setiap *Learning Rate*,
dan data *validasi* mengalami perbaikan error yang signifikan dalam rentang
50 *Epochs* pertama