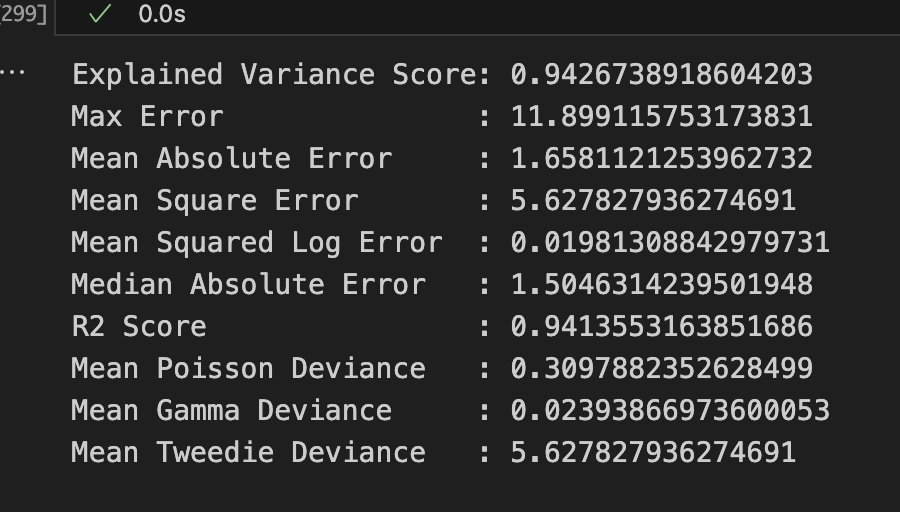
|  |  |
| --- | --- |
| **QUIZ 1 COMP8043041-MACHINE LEARNING 2023-2024** | |
| 1.A. | Sdr diminta membantu rekan Sdr untuk mengestimasi harga sebuah rumah di suatu kota tertentu di US. Kepada Sdr diberikan dataset perumahan di kota tersebut yaitu HOUSING.CSV yang memuat informasi hubungan antara 13 variabel bebas (CRIM, ZN, dstnya) dengan variabel output MEDV (PRICE) sebagai berikut :   * CRIM - per capita crime rate by town * ZN - proportion of residential land zoned for lots over 25,000 sq.ft. * INDUS - proportion of non-retail business acres per town. * CHAS - Charles River dummy variable (1 if tract bounds river; 0 otherwise) * NOX - nitric oxides concentration (parts per 10 million) * RM - average number of rooms per dwelling * AGE - proportion of owner-occupied units built prior to 1940 * DIS - weighted distances to five Boston employment centres * RAD - index of accessibility to radial highways * TAX - full-value property-tax rate per $10,000 * PTRATIO - pupil-teacher ratio by town * B - 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town * LSTAT - % lower status of the population * MEDV - Median value of owner-occupied homes in $1000's   Untuk mendapatkan hasil prediksi yang akurat, pastikan data HOUSING.CSV yang akan Sdr gunakan adalah data yang sudah “cleaned” dari *missing values* dan *outlier*. Lebih lanjut, pastikan juga Sdr hanya memilih variabel bebas yang memiliki “strong correlation” terhadap variabel independent (variabel MEDV). **Implementasikan model regresi Sdr dengan DL** dalam python notebook untuk proses-proses antara lain feature engineering, model training dan testing, serta model evaluation. Bandingkan hasil model berbasis DL Sdr dengan model MLR (40%). |
| 2. | Sebagai seorang data scientist pada sebuah perguruan tinggi Sdr diminta membuat sebuah model classifier berbasis SVM yang dapat membantu memutuskan apakah calon mahasiswa yang mendaftar dapat diterima atau ditolak berdasarkan nilai 3 atribut yang dimilikinya yaitu “gmat”, “gpa” dan “work\_experience” sebagaimana diperlihatkan oleh tabel di bawah ini (dataset untuk testing)   |  |  |  |  | | --- | --- | --- | --- | | gmat | gpa | work\_experience | admitted | | 590 | 2 | 3 | …. | | 740 | 3.7 | 4 | .... | | 680 | 3.3 | 6 | …. | | 610 | 2.3 | 1 | …. | | 710 | 3 | 5 | …. |     Sementara itu untuk training dan validasi model tsb, Sdr dapat gunakan dataset berikut ini.  Dataset in Python  Pastikan Sdr telah melakukan standarisasi data sebelum melakukan proses pelatihan, validasi dan testing. Evaluate your model (30%). |
| 3. | Bandingkan hasil Sdr pada problem no. 2 di atas bila Sdr menggunakan model Logistic Regression (15%). |
| 4. | Bandingkan hasil Sdr pada problem no. 2 di atas bila Sdr menggunakan model Multinomial Naïve Bayes (15%). |

================================Jawaban================================

1. Hasil menggunakan DL dengan ANN:



Hasil menggunakan MLR:

A screen shot of a computer program

Description automatically generated

NICOLAUS AVENITO - 2502400834