

PROJECT2 . COLLECTING DATA

Data 1

Diberikan data sebagai berikut.

“penjualan.csv.” , diakses di <https://project2koleksidata.s3.amazonaws.com/penjualan.csv>

Data tersebut menggambarkan Biaya Iklan dan Pendapatan hasil penjualan Kotor (dalam juta)

- Dari data Biaya Iklan dan penjualan tersebut deskripsikan dalam bentuk visual yg cocok (sebutkan) dan boxplot dan berikan penjelasan informasi apa yang bisa diberikan sehingga bisa membantu pengambilan keputusan
- Modelkan antara biaya iklan dan penjualan, beri penjelasan mana variable independen dan mana variable dependen , Menurut Anda apakah data di atas dapat didekati dengan model regresi linear? Jelaskan pendapat Anda,
- Terapkan model regresi linear dan beberapa model regresi non linear, manakah yang terbaik berdasarkan MAE, MSE, dan R2
- Terapkan K-Means Klustering , pilihlah klustering terbaik (banyak kluster optimal) untuk data di atas

Data 2

Diberikan data mengenai survey kebahagiaan warga yang terdiri dari 7 atribut sebagai berikut

D = decision attribute (D) with values 0 (unhappy) and 1 (happy)

X1 = the availability of information about the city services

X2 = the cost of housing

X3 = the overall quality of public schools

X4 = your trust in the local police

X5 = the maintenance of streets and sidewalks

X6 = the availability of social community events

“proyekklasifikasi.csv”, diakses di :

<https://project2koleksidata.s3.amazonaws.com/proyekklasifikasi.csv>

Sumber: <https://archive.ics.uci.edu/ml/datasets/Somerville+Happiness+Survey>

Terapkan beberapa model klasifikasi dan pilihlah manakah model yang terbaik di antara model yang Anda terapkan beserta dengan penjelasannya ambil X1-X5.