

Tugas Analisis Visualisasi Data

Metode Klasifikasi dan Clustering



Oleh :

Ferry Triwantono – 082111633094

FAKULTAS SAINS DAN TEKNOLOGI
PROGRAM STUDI SISTEM INFORMASI
UNIVERSITAS AIRLANGGA

2022

Metode Klasifikasi dan Clustering

1. Carilah suatu contoh kasus terkait trends, forecasts, atau clusters dengan Tableau dan integrasi pemanfaatan Python/Matlab/R untuk analisa lebih lanjut.

Data

Dataset yang digunakan untuk studi kasus ini adalah data terkait performa siswa dilihat berdasarkan skor yang dicapainya. Variabel yang dilihat untuk menentukan performa siswa adalah ID, gender, etnisitas, tingkat pendidikan orang tua, makan siang, persiapan test, skormatematika, skor reading, skor writing, rata-rata, total skor.

| ID | gender | ethnicity | parental level of education | lunch | test preparation course | math score | reading score | writing score | average score | total score |
|------|--------|-----------|-----------------------------|--------------|-------------------------|------------|---------------|---------------|---------------|-------------|
| 4697 | female | group B | bachelor's degree | standard | none | 72 | 72 | 74 | 72,67 | 218 |
| 4036 | female | group C | certificate's degree | standard | completed | 69 | 90 | 88 | 82,33 | 247 |
| 7055 | female | group B | master's degree | standard | none | 90 | 95 | 93 | 92,67 | 278 |
| 9991 | male | group A | associate's degree | free/reduced | none | 47 | 57 | 44 | 49,33 | 148 |
| 8104 | male | group C | certificate's degree | standard | none | 76 | 78 | 75 | 76,33 | 229 |
| 4364 | female | group B | associate's degree | standard | none | 71 | 83 | 78 | 77,33 | 232 |
| 5879 | female | group B | certificate's degree | standard | completed | 88 | 95 | 92 | 91,67 | 275 |
| 4072 | male | group B | certificate's degree | free/reduced | none | 40 | 43 | 39 | 40,67 | 122 |
| 4949 | male | group D | high school | free/reduced | completed | 64 | 64 | 67 | 65,00 | 195 |
| 1798 | female | group B | high school | free/reduced | none | 38 | 60 | 50 | 49,33 | 148 |
| 6871 | male | group C | associate's degree | standard | none | 58 | 54 | 52 | 54,67 | 164 |
| 6985 | male | group D | associate's degree | standard | none | 40 | 52 | 43 | 45,00 | 135 |
| 8509 | female | group B | high school | standard | none | 65 | 81 | 73 | 73,00 | 219 |
| 3773 | male | group A | certificate's degree | standard | completed | 78 | 72 | 70 | 73,33 | 220 |
| 9644 | female | group A | master's degree | standard | none | 50 | 53 | 58 | 53,67 | 161 |
| 2650 | female | group C | master's degree | standard | none | 69 | 75 | 78 | 74,00 | 222 |
| 8631 | male | group C | high school | standard | none | 88 | 89 | 86 | 87,67 | 263 |
| 3891 | female | group B | master's degree | free/reduced | none | 18 | 32 | 28 | 26,00 | 78 |
| 7128 | male | group C | master's degree | free/reduced | completed | 46 | 42 | 46 | 44,67 | 134 |
| 6336 | female | group C | associate's degree | free/reduced | none | 54 | 58 | 61 | 57,67 | 173 |
| 6855 | male | group D | high school | standard | none | 66 | 69 | 63 | 66,00 | 198 |

Menambah variabel korelasi

Analisis dimulai dengan membuat script py pada menu 'Analysis – Create CalculatedField'. Script ini akan menghasilkan sebuah variabel baru yang bernama 'Read_WriteCorrelation' untuk membantu mencari korelasi dari nilai reading dan writing pada dataset. Hal ini berarti mencari korelasi antara nilai reading dengan nilai writing dari siswa.

Read_Write Correlation

×

Results are computed along Table (across).

```
SCRIPT_REAL('
import numpy as np
return np.corrcoef(_arg1,_arg2)[0,1]', SUM([Reading Score]), SUM([Writing Score
])
```

Default Table Calculation

The calculation is valid.

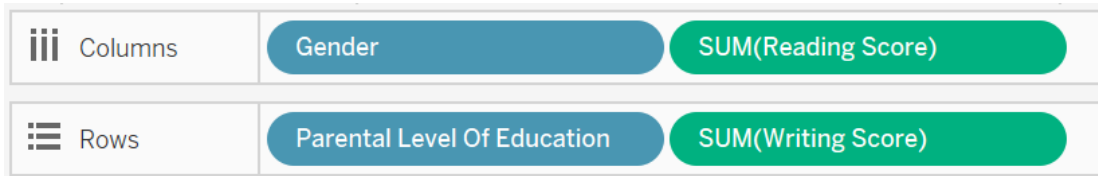
1 Dependency ▾

Apply

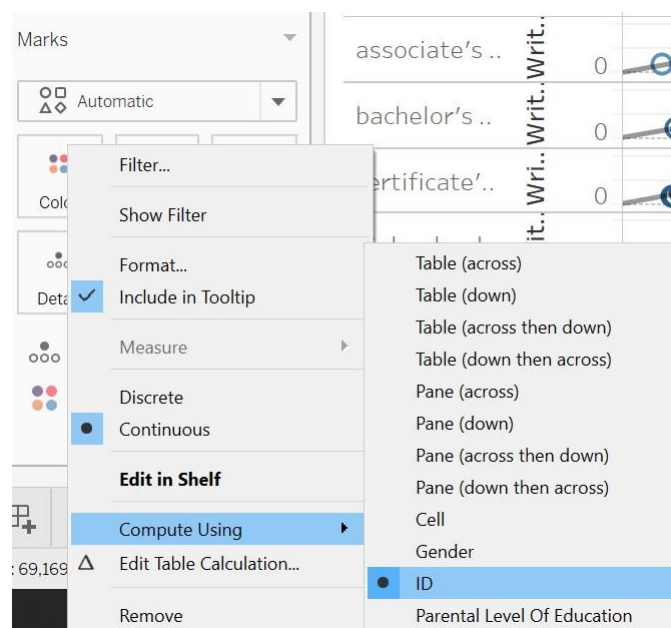
OK

Menambahkan variabel pada Tableau

Setelah membuat variabel baru dengan menu analysis tersebut, kita lanjutkan kepada visualisasi data dengan menginputkan variabel-variabel yang ada ke dalam columns dan juga rows dari sheets Tableau.

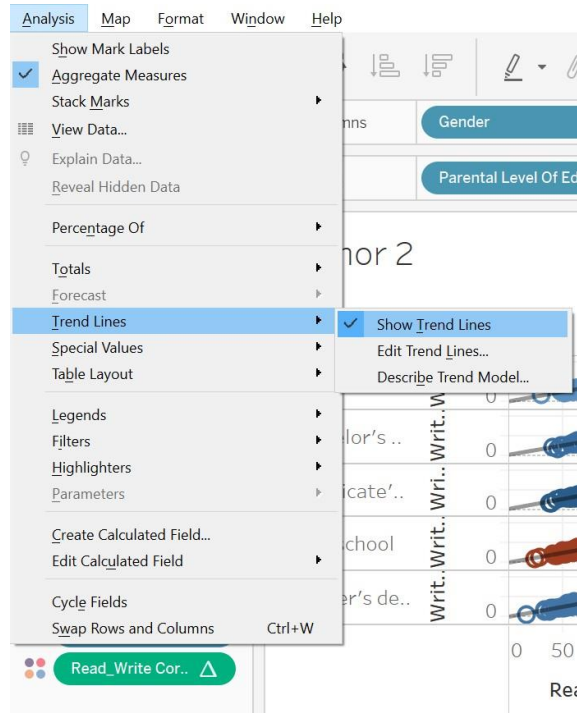


Jika ingin menambahkan filter, maka kita bisa menambahkan filter pada variabel yang digunakan (Dalam hal ini tidak menambahkan filter). Setelah itu, kita juga menginputkan id dan juga variabel perhitungan (Read_Write Correlation) ke dalam bagian marks dari Tableau. Saat melakukan hal ini, kita juga perlu memilih opsi compute using Id pada saat menginputkan Read_Write Correlation.

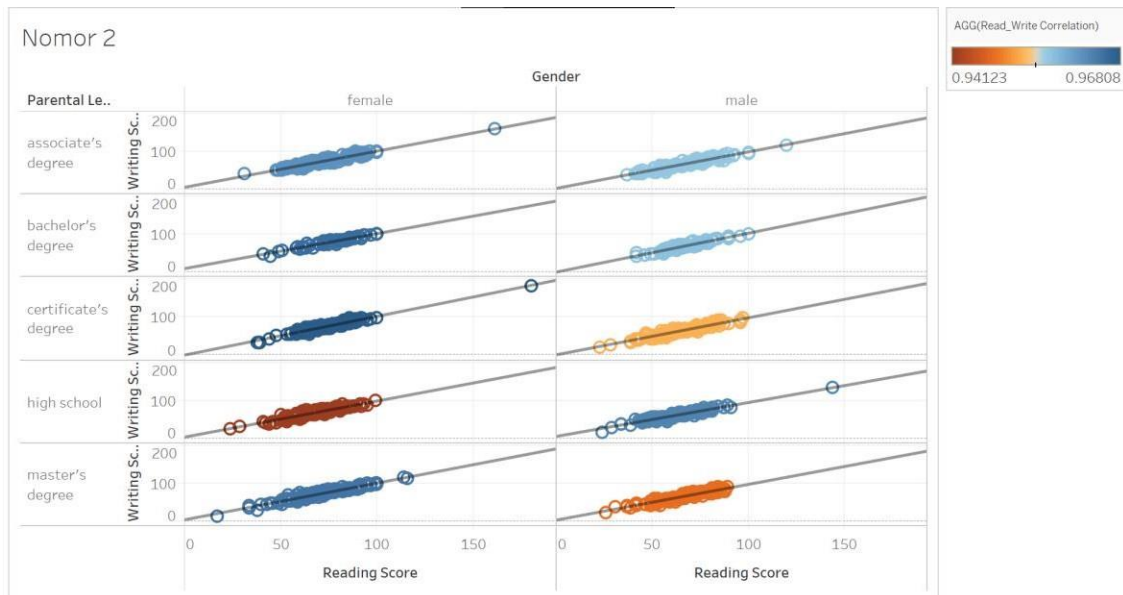


Menambahkan trendlines pada grafik

Setelah itu, untuk melihat trends kita mengaktifkan trendlines pada grafik dengan mengakses menu Analysis – Trendlines – Show Trendlines



Setelah itu, kita akan menghasilkan grafik/output akhir seperti di bawah ini :



Melalui grafik diatas, kita bisa melihat bahwa semua trend line pada grafik menunjukkan bahwa writing dan reading score memiliki korelasi yang positif, dalam artian semakin tinggi nilai reading siswa, maka semakin tinggi juga nilai writing siswa tersebut.