**Mudah Belajar Otodidak Data Science**

**(Praktek dalam bahasa python3)**

**Disusun oleh**

**Aryajaya Alasmsyah, S.Kom., M.Kom., MTA.**

**Pertemuan 2 – Visualisasi Seaborn dan Matplotlib**

**Tujuan pembelajaran**

* Mahasiswa mampu memahami konsep visualisasi data seperti *histogram*, *lineplot*, *pieplot*, *barplot*, *scatterplot*, *boxplot*, *heatmap*, dan lain-lain.
* Mahasiswa dapat menggunakan beberapa library visualisasi data seperti *matplotlib*, *seaborn*, dan *plotly-express*.

**Visualisasi data menggunakan library seaborn**

**Studi kasus: Analisis data penjualan video games (vgsales.csv)**

|  |
| --- |
| # Lib untuk manipulasi data  import numpy as np  import pandas as pd    # lib untuk analisa statistik  import scipy.stats as sc  import statsmodels.api as sm    # lib untuk visualisasi data  import seaborn as sns  import matplotlib.pyplot as plt |

1. **Akuisisi Data**

|  |
| --- |
| # load dataset vgsales  dataset = pd.read\_csv("../dataset/vgsales.csv").drop(columns=["Rank"], axis=1)  # cek metadata dataset  print(dataset.info()) |
| <class 'pandas.core.frame.DataFrame'>  RangeIndex: 16598 entries, 0 to 16597  Data columns (total 10 columns):  # Column Non-Null Count Dtype  --- ------ -------------- -----  0 Name 16598 non-null object  1 Platform 16598 non-null object  2 Year 16327 non-null float64  3 Genre 16598 non-null object  4 Publisher 16540 non-null object  5 NA\_Sales 16598 non-null float64  6 EU\_Sales 16598 non-null float64  7 JP\_Sales 16598 non-null float64  8 Other\_Sales 16598 non-null float64  9 Global\_Sales 16598 non-null float64  dtypes: float64(6), object(4)  memory usage: 1.3+ MB |

|  |
| --- |
| # menampilkan dataset  dataset.head(3) |
| Name Platform Year Genre Publisher NA\_Sales EU\_Sales .. Global Sales  0 Wii Sports Wii 2006 Sports Nintendo 41.49 29.02 .. 82.74  1 Super Mario Bros NES 1985 Platform Nintendo 29.08 3.58 .. 42.24  2 Mario Kart Wii Wii 2008 Racing Nintendo 15.85 12.88 .. 35.84 |

1. **Penjualan video games untuk semua wilayah**

|  |
| --- |
| # proses grouping berdasarkan tahun dan global sales  df = dataset.groupby(by="Year", as\_index=False)["Global\_Sales"].sum()    # proses menghapus penjualan tahun 2017, 2020  df = df.drop(labels=[37,38], axis=0)  # membuat frame  fig, ax = plt.subplots(figsize = (15,5))    # membuat time series plot  ax.plot(df["Year"], df["Global\_Sales"], color="tab:blue", label="Global Sales", linewidth=2)    # membuat label-label  ax.set\_title("History of video games sales", fontsize=12)  ax.set\_xlabel("", fontsize=10)  ax.set\_ylabel("", fontsize=10)  ax.legend(loc='upper left')  ax.grid(True)    # menampilkan plot  plt.show() |
|  |
| Gambar x. Output program |

|  |
| --- |
|  |