
Analisis Pengaruh Pandemi COVID-19 pada Pelajar

Presentasi Akhir

Kelompok Bintang Jatuh

1. Aulia Radityatama Suhendra - 1906399865
2. Fathan Muhammad - 2006486090
3. Andira Anggun Maharani - 1906298765
4. Fariz Wahyuzan Dwitilas - 1906399511

Latar Belakang

Pandemi COVID-19 yang terjadi pada 2 tahun belakangan ini mengubah rutinitas dan kebiasaan masyarakat di seluruh dunia ini. Salah satu perubahan yang paling besar dapat dirasakan oleh pelajar, di mana metode pembelajaran yang semula dilakukan dengan interaksi melalui tatap muka, berubah menjadi interaksi dua arah melalui layar secara daring. Dengan serangkaian perubahan yang terjadi, tentunya akan ada dampak-dampak lain yang timbul juga.

Tujuan

Melakukan serangkaian analisis pada data dampak COVID-19 bagi pelajar, serta keterkaitannya pada demografi, rutinitas pembelajaran saat online, kesehatan, serta pengalaman pembelajaran online.

Analisis ini akan bermanfaat bagi setiap individu, khususnya pemerintah maupun yang terkait agar dapat mengetahui seberapa besar pengaruh yang dirasakan para pelajar akibat perubahan metode belajar yang disebabkan oleh COVID-19.

Dataset

Sumber:

<https://www.kaggle.com/datasets/kunal28chaturvedi/covid19-and-its-impact-on-students>

Data Preprocessing

Unused Columns

```
[ ] # Read dataset
df = pd.read_csv("COVID-19 Survey Student Responses.csv")
```

```
[ ] df = df.drop(["ID", "Stress busters", "What you miss the most", "Region of residence", "Time spent on TV"], axis = 1)
```

[] df

[illegible]

Null Values

```
[ ] # Untuk atribut "Rating of Online Class experience", null value akan diisi dengan nilai modus
mode_rating_study = df["Rating of Online Class experience"].mode()[0]
df["Rating of Online Class experience"].fillna(mode_rating_study,inplace=True)

# Untuk atribut "Medium for online class", null value akan diisi dengan nilai modus
mode_medium = df["Medium for online class"].mode()[0]
df["Medium for online class"].fillna(mode_medium,inplace=True) #mode
```

```
[ ] # Cek nilai null
print("Count missing values untuk df\n", df.isnull().sum())
```

```
Count missing values untuk df
Age of Subject                                0
Time spent on Online Class                    0
Rating of Online Class experience             0
Medium for online class                       0
Time spent on self study                      0
Time spent on fitness                         0
Time spent on sleep                           0
Time spent on social media                    0
Prefered social media platform                0
Number of meals per day                      0
Change in your weight                        0
Health issue during lockdown                  0
Time utilized                                0
Do you find yourself more connected with your family, close friends , relatives ? 0
dtype: int64
```


Data Duplication

```
[ ] # Cek duplikasi pada data  
    print("jumlah data yang redundan adalah ", df.duplicated().sum())
```

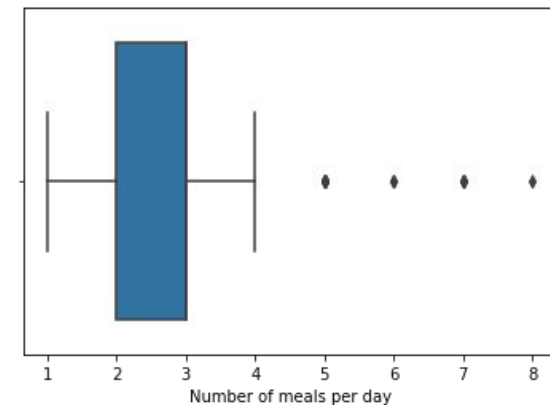
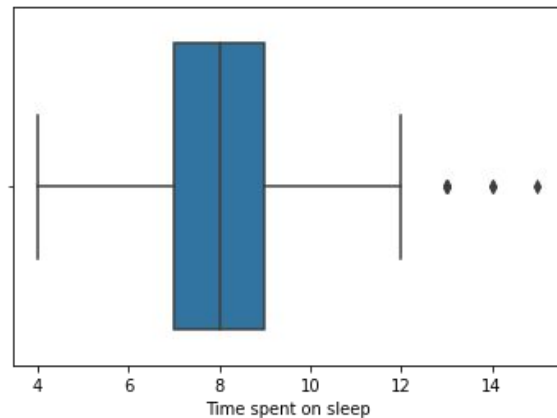
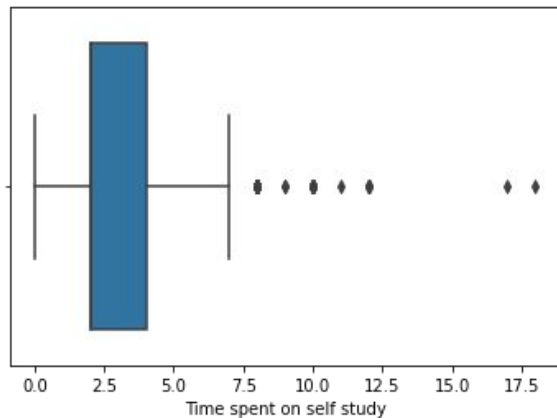
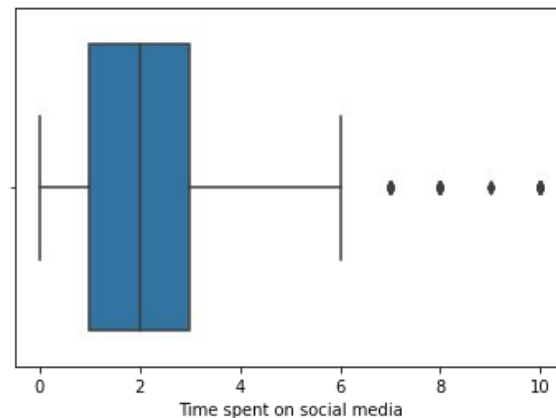
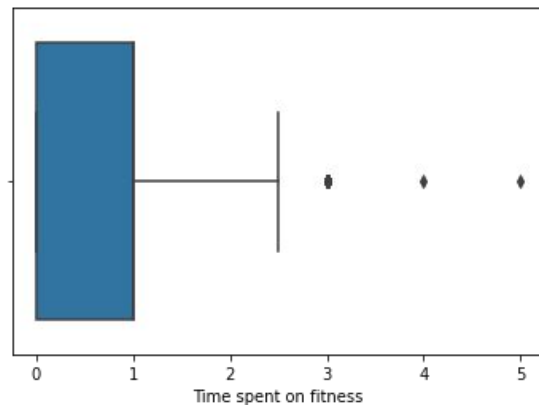
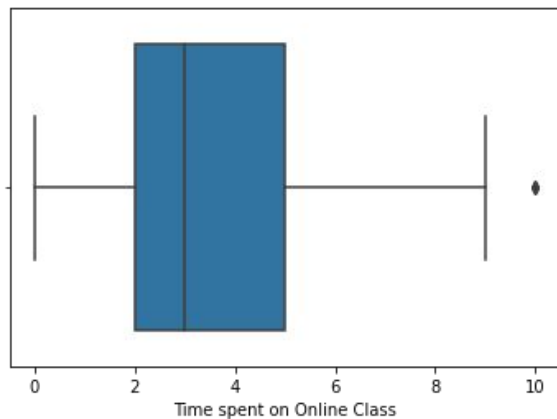
```
jumlah data yang redundan adalah 4
```

```
[ ] #handle duplicate  
    df = df.drop_duplicates()
```

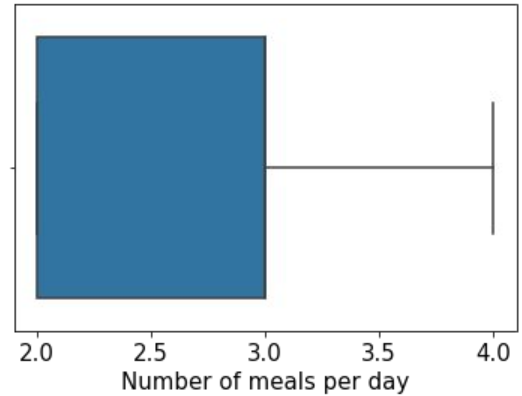
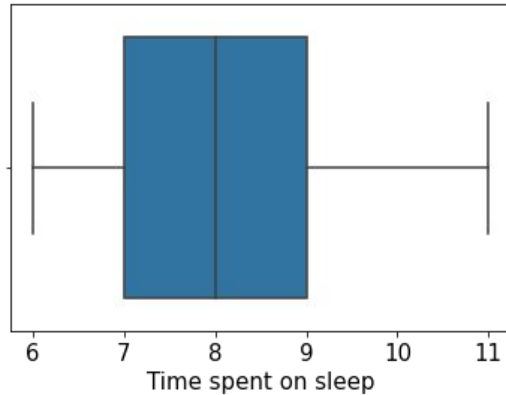
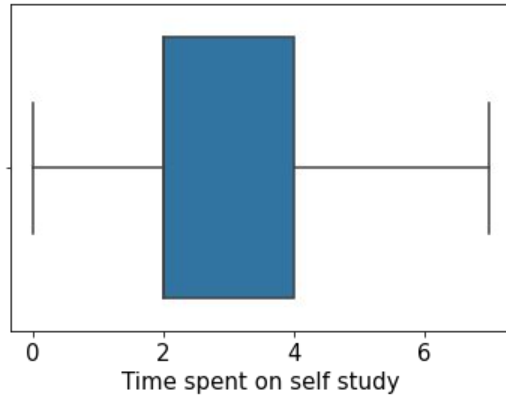
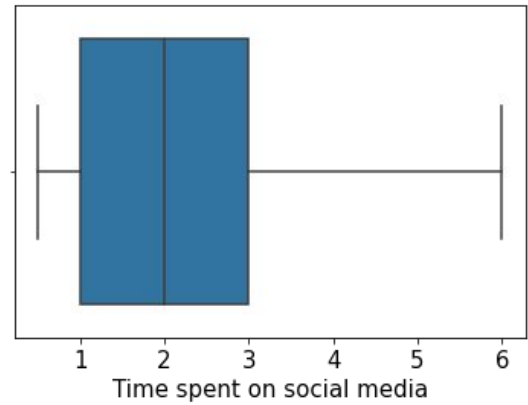
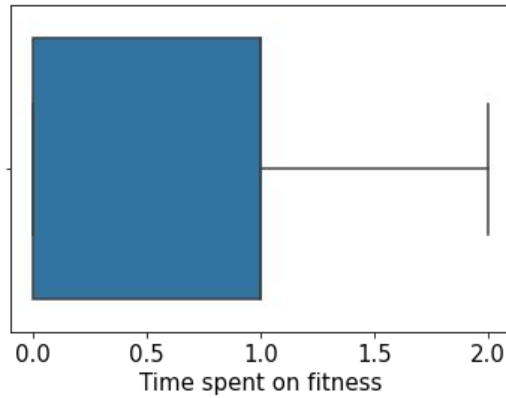
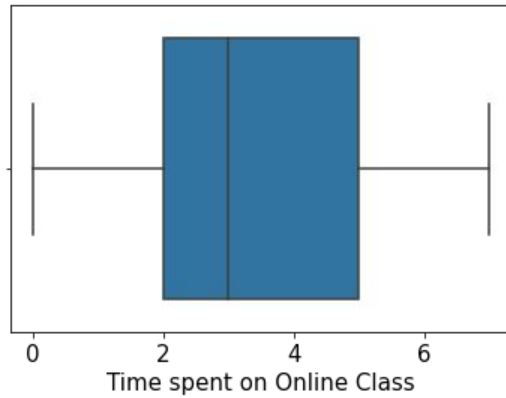
```
[ ] # Cek duplikasi pada data  
    print("jumlah data yang redundan adalah ", df.duplicated().sum())
```

```
jumlah data yang redundan adalah 0
```

Outliers (Before Winsorized)



Outliers (After Winsorized)



Data Formatting

```
[ ] # Handle preferred medium 'None' dan 'None '  
    # Handle preferred social media platform 'Whatsapp' dan 'WhatsApp'  
  
[ ] mapping = {'None ': 'None', 'Whatsapp': 'WhatsApp'}  
    df = df.replace({'Prefered social media platform': mapping})  
  
    print(df['Prefered social media platform'].unique())  
    print(df['Medium for online class'].unique())  
  
['Linkedin' 'Youtube' 'Instagram' 'WhatsApp' 'None' 'Reddit' 'Snapchat'  
 'Omegle' 'Twitter' 'Telegram' 'Facebook' 'Elyment' 'Quora' 'Talklife']  
['Laptop/Desktop' 'Smartphone' 'Tablet' 'Any Gadget'  
 'Smartphone or Laptop/Desktop']
```

Encoding Nominal Data

```
mapping = {'Decreased': 0, 'Remain Constant': 1, 'Increased': 2}
df = df.replace({'Change in your weight': mapping})

mapping = {'Very poor': 0, 'Poor': 1, 'Average': 2, 'Good': 3, 'Excellent': 4}
df = df.replace({'Rating of Online Class experience': mapping})

mapping = {'NO': 0, 'YES': 1}
df = df.replace({'Health issue during lockdown': mapping, 'Do you find yourself more connected with your family, close friends , relatives ?': mapping, 'Time utilized': mapping})

df.head()
```

	Age of Subject	Time spent on Online Class	Rating of Online Class experience	Medium for online class	Time spent on self study	Time spent on fitness	Time spent on sleep	Time spent on social media	Preferred social media platform	Number of meals per day	Change in your weight	Health issue during lockdown	Time utilized	Do you find yourself more connected with your family, close friends , relatives ?
0	21	2.0	3	Laptop/Desktop	4.0	0.0	7.0	3.0	Linkedin	4	2	0	1	1
1	21	0.0	4	Smartphone	0.0	2.0	10.0	3.0	Youtube	3	0	0	1	0
2	20	7.0	0	Laptop/Desktop	3.0	0.0	6.0	2.0	Linkedin	3	1	0	0	1
3	20	3.0	0	Smartphone	2.0	1.0	6.0	5.0	Instagram	3	0	0	0	0
4	21	3.0	3	Laptop/Desktop	3.0	1.0	8.0	3.0	Instagram	4	1	0	0	0

Encoding Categorical Data

```
df_dummy = pd.get_dummies(df[['Medium for online class', 'Prefered social media platform']])  
df = df.drop(['Medium for online class', 'Prefered social media platform'], axis=1)  
df = pd.concat([df, df_dummy], axis=1)  
df.head()  
df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1178 entries, 0 to 1181
Data columns (total 31 columns):
#   Column                                                                 Non-Null Count  Dtype
---  -
0   Age of Subject                                                         1178 non-null   int64
1   Time spent on Online Class                                             1178 non-null   float64
2   Rating of Online Class experience                                     1178 non-null   int64
3   Time spent on self study                                              1178 non-null   float64
4   Time spent on fitness                                                 1178 non-null   float64
5   Time spent on sleep                                                   1178 non-null   float64
6   Time spent on social media                                            1178 non-null   float64
7   Number of meals per day                                               1178 non-null   int64
8   Change in your weight                                                 1178 non-null   int64
9   Health issue during lockdown                                          1178 non-null   int64
10  Time utilized                                                         1178 non-null   int64
11  Do you find yourself more connected with your family, close friends , relatives ? 1178 non-null   int64
12  Medium for online class_Any Gadget                                    1178 non-null   uint8
13  Medium for online class_Laptop/Desktop                                1178 non-null   uint8
14  Medium for online class_Smartphone                                    1178 non-null   uint8
15  Medium for online class_Smartphone or Laptop/Desktop                 1178 non-null   uint8
16  Medium for online class_Tablet                                         1178 non-null   uint8
17  Preferred social media platform_Flyment                              1178 non-null   uint8
18  Preferred social media platform_Facebook                             1178 non-null   uint8
19  Preferred social media platform_Instagram                            1178 non-null   uint8
20  Preferred social media platform_Linkedin                              1178 non-null   uint8
21  Preferred social media platform_None                                  1178 non-null   uint8
22  Preferred social media platform_Omegle                               1178 non-null   uint8
23  Preferred social media platform_Quora                                 1178 non-null   uint8
24  Preferred social media platform_Reddit                                1178 non-null   uint8
25  Preferred social media platform_Snapchat                              1178 non-null   uint8
26  Preferred social media platform_Talklife                              1178 non-null   uint8
27  Preferred social media platform_Telegram                              1178 non-null   uint8
28  Preferred social media platform_Twitter                              1178 non-null   uint8
29  Preferred social media platform_WhatsApp                             1178 non-null   uint8
30  Preferred social media platform_Youtube                              1178 non-null   uint8
dtypes: float64(5), int64(7), uint8(19)
memory usage: 141.5 KB

```

Medium for online class_Any Gadget	Medium for online class_Laptop/Desktop	Medium for online class_Smartphone	...	Prefered social media platform_Talklife	Prefered social media platform_Telegram	Prefered social media platform_Twitter	Prefered social media platform_WhatsApp	Prefered social media platform_Youtube
0	1	0	...	0	0	0	0	0
0	0	1	...	0	0	0	0	1
0	1	0	...	0	0	0	0	0
0	0	1	...	0	0	0	0	0
0	1	0	...	0	0	0	0	0
...
0	0	1	...	0	0	0	0	0
0	0	1	...	0	0	0	1	0
0	0	1	...	0	0	0	0	1
0	1	0	...	0	0	0	0	1
0	0	0	...	0	0	0	1	0

Normalization

```
[ ] df = df.reset_index(drop=True)

[ ] from sklearn.preprocessing import Normalizer

df_norm = df.copy()

norm = Normalizer()
normed = norm.fit_transform(df_norm[['Time spent on Online Class', 'Time spent on self study', 'Time spent on fitness', 'Time spent on sleep', 'Time spent on social media']])
newdf = pd.DataFrame(normed)
newdf.columns = df_norm[['Time spent on Online Class', 'Time spent on self study', 'Time spent on fitness', 'Time spent on sleep', 'Time spent on social media']].columns

df_norm = df_norm.drop(['Time spent on Online Class', 'Time spent on self study', 'Time spent on fitness', 'Time spent on sleep', 'Time spent on social media'], axis=1)
df_norm = pd.concat([df_norm, newdf], axis=1)
df_norm
```

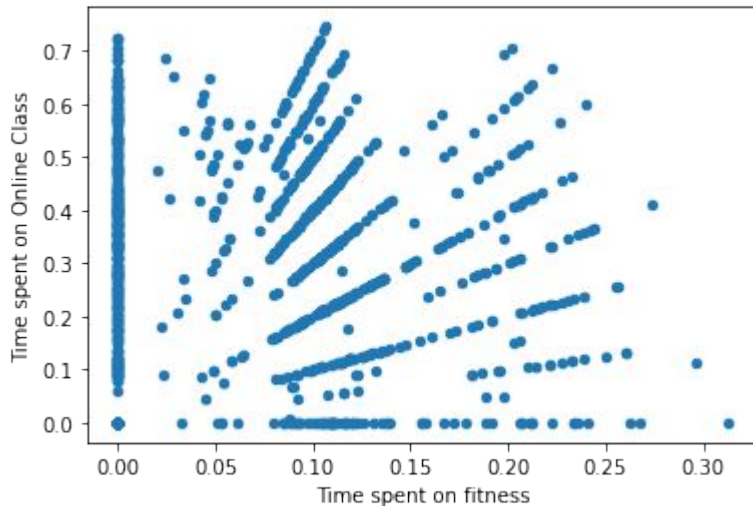
Data numerik yang ternormalisasi

Time spent on Online Class	Time spent on self study	Time spent on fitness	Time spent on sleep	Time spent on social media
0.226455	0.452911	0.000000	0.792594	0.339683
0.000000	0.000000	0.188144	0.940721	0.282216
0.707107	0.303046	0.000000	0.606092	0.202031
0.346410	0.230940	0.115470	0.692820	0.577350
0.312772	0.312772	0.104257	0.834058	0.312772
...
0.314485	0.419314	0.104828	0.838628	0.104828
0.516398	0.344265	0.086066	0.774597	0.086066
0.423405	0.000000	0.052926	0.846810	0.317554
0.493865	0.345705	0.098773	0.790184	0.049386
0.561656	0.224662	0.056166	0.786318	0.112331

Task

EDA

Hubungan waktu yang digunakan untuk mengikuti kelas daring dengan waktu untuk berolahraga



Hasil Observasi :

Ada kecenderungan bahwa semakin lama online class akan semakin sedikit waktu yang dihabiskan untuk berolahraga.

Kesimpulan:

Tidak ada korelasi yang kuat antara waktu yang digunakan untuk mengikuti kelas daring dengan waktu yang digunakan untuk berolahraga.

EDA

Berapa rata-rata waktu belajar mandiri jika pelajar bisa menggunakan waktunya dengan efektif dan tidak efektif

```
df_utilized = df.loc[df['Time utilized'] == 1]
print("Rerata waktu belajar mandiri, jika siswa mampu memanfaatkan waktu belajar dengan efektif : {:.2f} Jam".format(df_utilized["Time spent on self study"].mean()))

df_utilized = df.loc[df['Time utilized'] == 0]
print("Rerata waktu belajar mandiri, jika siswa tidak memanfaatkan waktu belajar dengan efektif : {:.2f} Jam".format(df_utilized["Time spent on self study"].mean()))
```

Rerata waktu belajar mandiri, jika siswa mampu memanfaatkan waktu belajar dengan efektif : 3.16 Jam
Rerata waktu belajar mandiri, jika siswa tidak memanfaatkan waktu belajar dengan efektif : 2.47 Jam

Membandingkan rata-rata waktu belajar mandiri jika pelajar bisa menggunakan waktunya dengan efektif dan tidak efektif

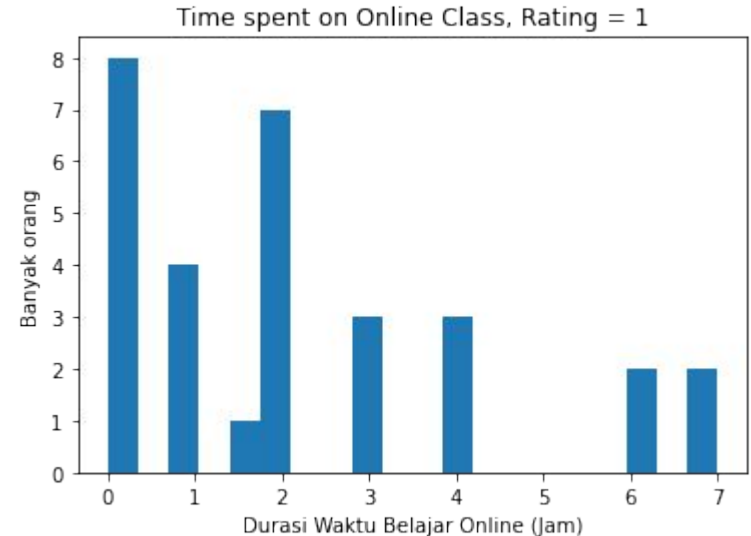
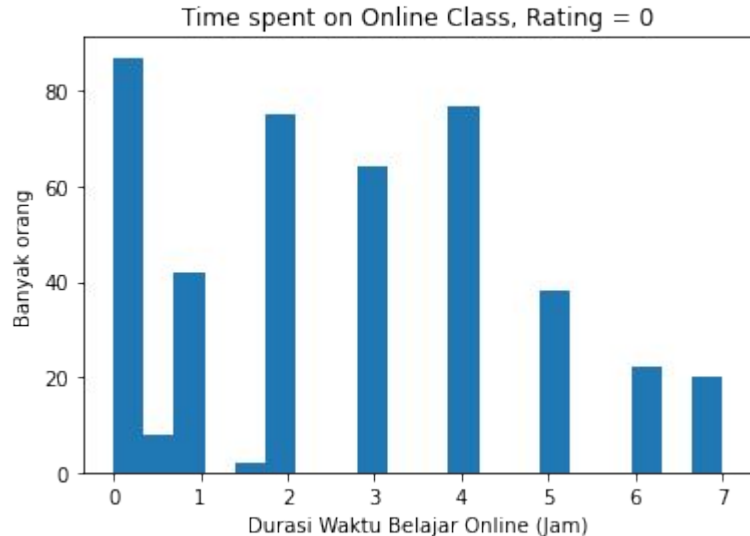
Rerata waktu belajar mandiri yang dihabiskan oleh seorang siswa ketika ia mampu memanfaatkan waktu belajarnya dengan efektif adalah **3.16 jam**, sedangkan jika ia tidak mampu memanfaatkan waktu belajarnya dengan efektif adalah **2.47 jam**.

Hal ini menunjukkan bahwa :

- Siswa akan cenderung menggunakan waktu belajar yang lebih banyak, ketika ia bisa memanfaatkan waktu yang dimilikinya selama masa pandemi dengan baik dan efektif
- Keefektifan penggunaan waktu di masa pandemi, memiliki pengaruh pada rerata waktu belajar siswa

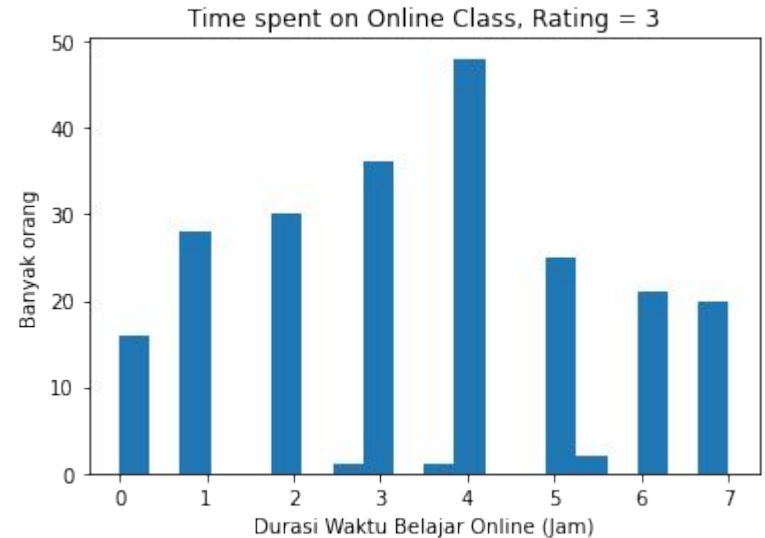
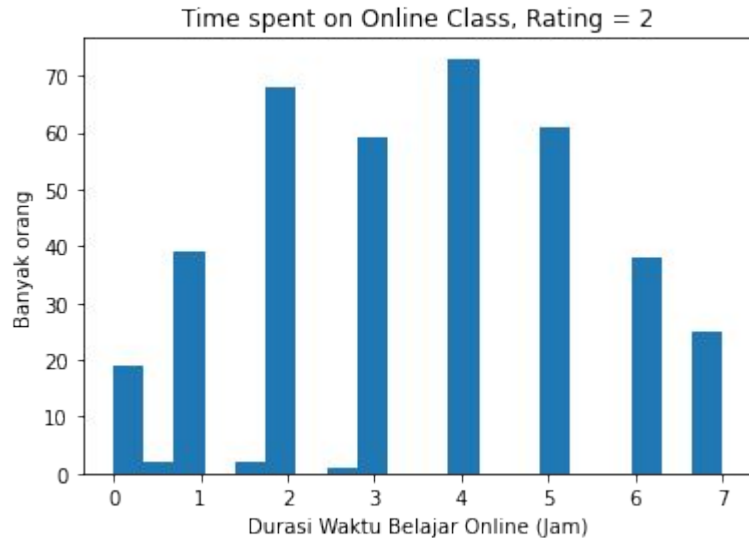
EDA

Membandingkan waktu mengikuti kelas online berdasarkan rating yang berbeda (1)

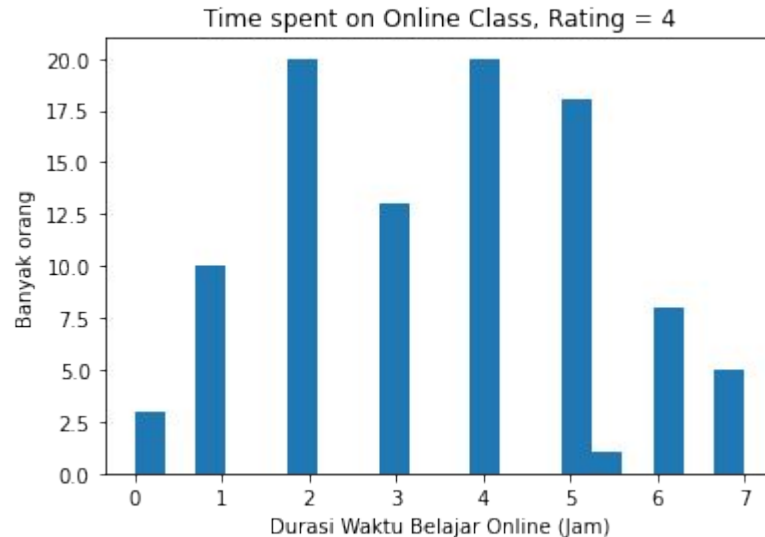


EDA

Membandingkan waktu mengikuti kelas online berdasarkan rating yang berbeda (2)



Membandingkan waktu mengikuti kelas online berdasarkan rating yang berbeda (3)

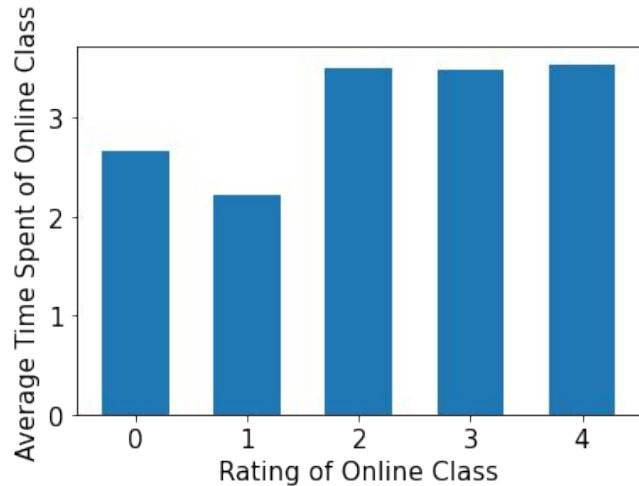


Hasil Observasi:

Siswa dengan waktu belajar online di antara 4-7 jam cenderung memberikan rating 2. Sedangkan siswa dengan waktu belajar online < 3 jam cenderung memberikan rating 0.

EDA

Membandingkan rata-rata waktu mengikuti kelas online berdasarkan rating yang berbeda

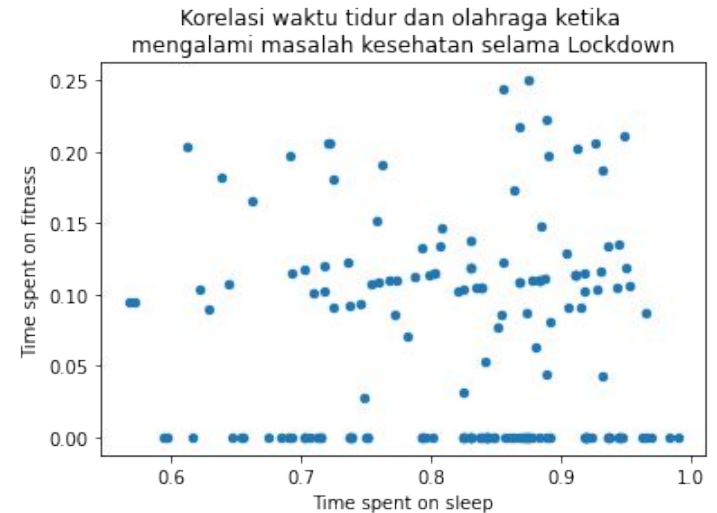
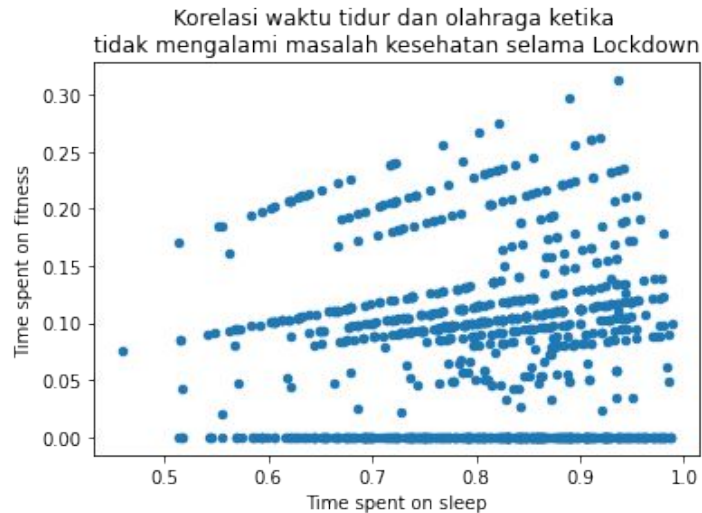


Kesimpulan:

Semakin sedikit rata-rata waktu mengikuti kelas daring, semakin kecil rating yang diberikan, begitupun juga sebaliknya

EDA

Korelasi waktu tidur, waktu berolahraga dan kondisi kesehatan selama *lockdown*



Korelasi waktu tidur, waktu berolahraga dan kondisi kesehatan selama *lockdown*

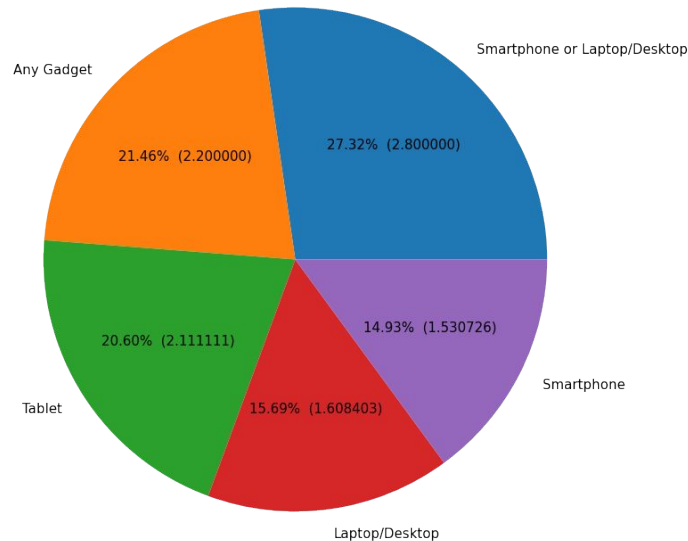
Hasil Observasi :

Pada siswa yang tidak mengalami masalah kesehatan, terdapat kecenderungan bahwa, semakin sedikit waktu tidur yang dimiliki, maka waktu untuk melakukan fitness juga akan berkurang.

Kesimpulan:

Tidak ada korelasi yang kuat antara waktu yang digunakan untuk tidur dengan kondisi kesehatan siswa selama masa *lockdown*.

Visualisasi *rating of online class* berdasarkan medium yang digunakan



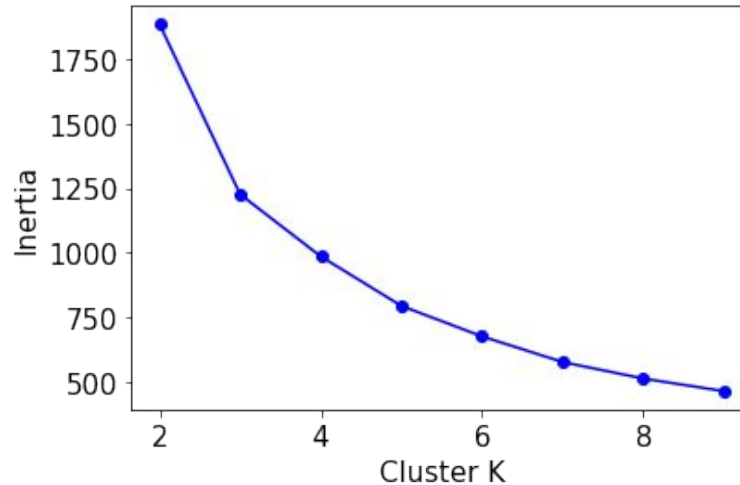
Kesimpulan:

Media terbanyak yang digunakan adalah Smartphone atau laptop/desktop, kemudian disusul oleh gadget lainnya, dan Tablet.

Clustering

Melakukan clustering pelajar berdasarkan umur, waktu yang digunakan untuk kelas daring, waktu untuk tidur, jumlah porsi makan, dan perubahan berat badan

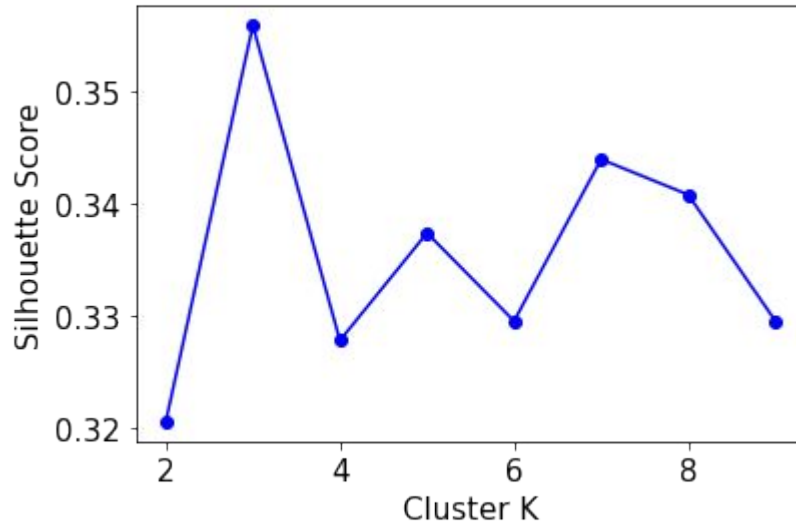
- Inertia:



Clustering

Melakukan clustering pelajar berdasarkan umur, waktu yang digunakan untuk kelas daring, waktu untuk tidur, jumlah porsi makan, dan perubahan berat badan

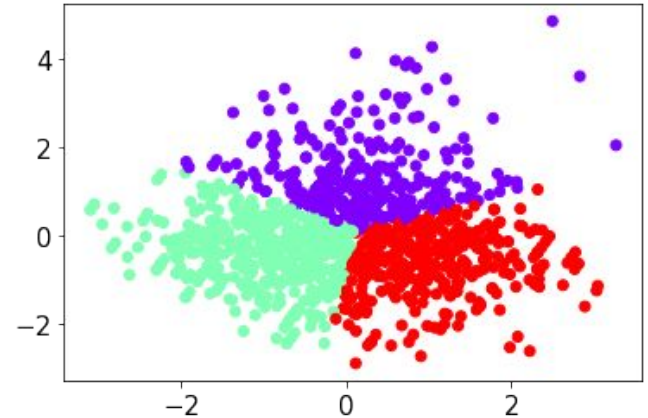
- Silhouette score:



Clustering

Melakukan clustering pelajar berdasarkan umur, waktu yang digunakan untuk kelas daring, waktu untuk tidur, jumlah porsi makan, dan perubahan berat badan

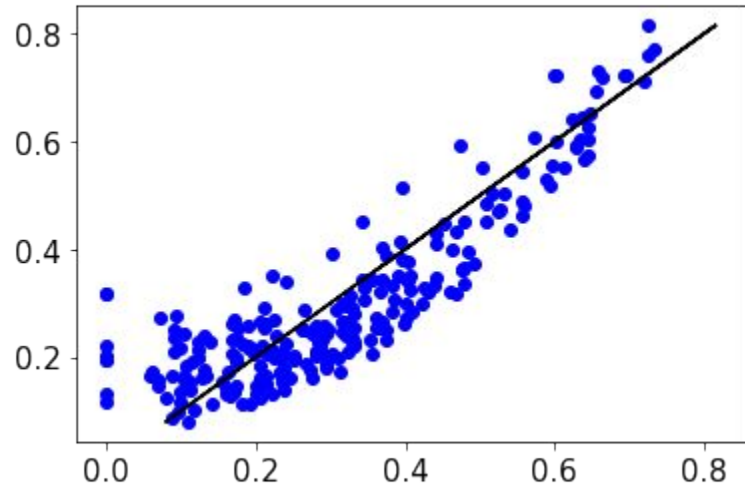
- Hasil clustering dengan jumlah cluster = 3:



Regresi

Memprediksikan rata-rata waktu yang dihabiskan oleh pelajar untuk belajar mandiri per-harinya

- Hasil linear regression:



Regresi

Memprediksikan rata-rata waktu yang dihabiskan oleh pelajar untuk belajar mandiri per-harinya

- Hasil metrik evaluasi:

```
[194] from sklearn.metrics import mean_squared_error
      from sklearn.metrics import mean_absolute_error
      import math
      from sklearn.metrics import r2_score

      mse_reg = mean_squared_error(yy_test, reg.predict(xx_test))
      mae_reg = mean_absolute_error(yy_test, reg.predict(xx_test))
      rmse_reg = math.sqrt(mean_squared_error(yy_test, reg.predict(xx_test)))
      r_squared_reg = r2_score(yy_test, reg.predict(xx_test))

      print("MSE = ", mse_reg)
      print("MAE = ", mae_reg)
      print("RMSE = ", rmse_reg)
      print("r_squared = ", r_squared_reg)

      MSE = 0.007269999875517462
      MAE = 0.06910594839001892
      RMSE = 0.08526429425918837
      r_squared = 0.766115041677758
```

Klasifikasi

Klasifikasi perasaan keterhubungan pelajar berdasarkan fitur-fitur lainnya

- Kondisi sebelum oversampling:
 - class pada kolom “perasaan keterhubungan pelajar”: 1= 828 data, 0= 350 data
 - Mean ROC AUC klasifikasi: 0.545
- Kondisi setelah oversampling menggunakan SMOTE:
 - class pada kolom “perasaan keterhubungan pelajar”: 1= 828 data, 0= 828 data
 - Mean ROC AUC klasifikasi: 0.777

Classification Report

	precision	recall	f1-score	support
0	0.71	0.67	0.69	178
1	0.64	0.69	0.66	154
accuracy			0.68	332
macro avg	0.68	0.68	0.68	332
weighted avg	0.68	0.68	0.68	332

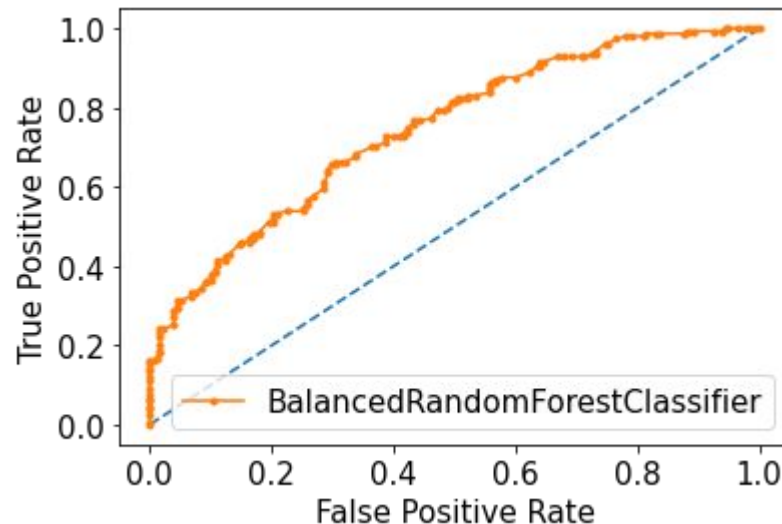
Model Evaluation

- ROC-AUC Curve

- ROC AUC of
BalancedRandomForestClassifier =
0.7457865168539326

- Error Metrics

- MSE = 0.31626506024096385
- MAE = 0.31626506024096385
- RMSE = 0.562374483988173
- r_squared = -0.2717058222676201



Terima Kasih