

**MODEL PEMBELAJARAN DAN LAPORAN AKHIR
PROJECT-BASED LEARNING
MATA KULIAH ANALISIS DATA EKSPLORATIF
KELAS A**



"ANALISIS HUBUNGAN ANTARA PANDEMI COVID-19 DAN KESEHATAN MENTAL PELAJAR DENGAN MEMPERTIMBANGKAN FAKTOR-FAKTOR PENYEBABNYA SERTA STRATEGI UNTUK MENGATASINYA"

DISUSUN OLEH KELOMPOK 5 :

- | | |
|------------------------------------|-------------------------|
| 1. REZA PUTRI ANGGA | (22083010006) - KETUA |
| 2. KANESSA JASMINE PRISHEILA A.Z.S | (22083010016) - ANGGOTA |
| 3. SHARLEEN AGUSTINE | (22083010030) - ANGGOTA |
| 4. ANNABEL GRACIA PURYANI | (22083010048) - ANGGOTA |
| 5. HAUZAN HANIFAH ZAHRA | (22083010075) - ANGGOTA |

DOSEN PENGAMPU:

TRESNA MAULANA FAHRUDIN, S.S.T., MT (199305012022031007)

PROGRAM STUDI SAINS DATA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2023

DAFTAR ISI

DAFTAR ISI.....	2
DAFTAR TABEL.....	3
DAFTAR GAMBAR	4
BAB I PENDAHULUAN.....	6
1.1. Latar Belakang	6
1.2. Rumusan Masalah	6
BAB II TINJAUAN PUSTAKA	8
2.1. Teori Penunjang	8
2.1.1. COVID-19	8
2.1.2. Lock-Down COVID-19.....	8
2.1.3. Pelajar	8
2.1.4. Kesehatan Mental	8
2.1.5. K-Means	9
2.1.6. Analisis Data Eksploratif.....	9
2.2. Penelitian Terkait	10
BAB III METODOLOGI.....	12
BAB IV HASIL DAN PEMBAHASAN	13
4.1. Deskripsi Dataset.....	13
4.2. Persiapan / <i>Preparing</i> Data	15
4.3. Transformasi Data	19
4.4. Statistika Deskriptif	26
4.5. Normalisasi Data	38
4.6. Pengujian Hipotesis Analisis Korelasi Bivariate Dan Multivariate	41
4.7. Pengembangan Model Dan Evaluasi.....	55
4.8. Visualisasi Data Interaktif	56
4.9. Hasil Analisis Kasus.....	58
4.9.1. Analisis Statistika Deskriptif	58
4.9.2. Analisis Bivariate	59
4.9.3. Analisis Multivariate	60
4.9.4. Strategi dan Cara Untuk Mengatasi Dampak Pandemi COVID-19	60
BAB V KESIMPULAN.....	61
DAFTAR PUSTAKA	62
LAMPIRAN.....	63

DAFTAR TABEL

Tabel 2.1. Penelitian Terkait.....	11
Tabel 4 1. Tabel Time Spent on Online Class	32
Tabel 4 2. Tabel Time Spent on Self-Study.....	33
Tabel 4.3. Time Spent on Fitness	33
Tabel 4.4. Tabel Spent on Sleep	34
Tabel 4.5. Tabel Spent on Social Medua.....	34
Tabel 4 6. Tabel Time Spent on TV	34

DAFTAR GAMBAR

Gambar 4.1. Library yang Digunakan	16
Gambar 4.2. Load Dataset	17
Gambar 4.3. Pemilihan Dataset Usia 5-20 Tahun	17
Gambar 4.4. Output Pemilihan Dataset Usia 5-20 Tahun.....	17
Gambar 4.5. Membuat Salinan DataFrame.....	18
Gambar 4.6. Output Salinan DataFrame.....	18
Gambar 4.7. Menampilkan Jumlah Baris dan Kolom.....	18
Gambar 4.8. Output Jumlah Baris dan Kolom.....	18
Gambar 4.9. Pengecekan Nama Kolom dan Tipe Data	18
Gambar 4.10. Output Nama Kolom dan Tipe Data	19
Gambar 4.11. Menghapus Kolom	19
Gambar 4.12. Output Penghapusan Kolom	19
Gambar 4.13. Proses Encoding.....	20
Gambar 4.14. Output Proses Encoding.....	20
Gambar 4.15. Pengecekan Nama Kolom dan Tipe Data	21
Gambar 4.16. Output Pengecekan Nama Kolom dan Tipe Data	21
Gambar 4.17. Pengecekan Nilai Kolom ‘Time spent on TV’	21
Gambar 4.18. Nilai Unik dan Konversi Data Kolom ‘Time spent on TV’	22
Gambar 4.19. Pengecekan Missing Value.....	22
Gambar 4.20. Output Pengecekan Missing Value.....	22
Gambar 4.21. Imputasi Missing Value	22
Gambar 4.22. Output Imputasi Missing Value.....	23
Gambar 4.23. Pengecekan Missing Value.....	23
Gambar 4.24. Output Pengecekan Missing Value.....	23
Gambar 4.25. Pengecekan Data Duplikat	24
Gambar 4.26. Pengecekan Kolom dan Tipe Data	24
Gambar 4.27. Output Pengecekan Kolom dan Tipe Data.....	24
Gambar 4.28. Deteksi Outlier	25
Gambar 4.29. Output Deteksi Outlier	25
Gambar 4.30. Visualisasi Outlier	25
Gambar 4.31. Boxplot Visualisasi Outlier	26
Gambar 4.32. Statistika Deskriptif Dataset.....	26
Gambar 4.33. Pengecekan Distribusi Dataset.....	27
Gambar 4.34. Persentase Usia Responden dalam Dataset	29
Gambar 4.35. Jumlah Pelajar SD, SMP, dan SMA	30
Gambar 4.36. Persentase Asal Daerah Responden	30
Gambar 4.37. Persentase Device yang Digunakan Responden	31
Gambar 4.38. Visualisasi Lama Waktu yang Dihabiskan dalam Beraktivitas	32
Gambar 4.39. Numbers of Meals Per-Day.....	35
Gambar 4.40. Visualisasi Stress Busters	36
Gambar 4.41. Visualisasi Change in Your Weight	37
Gambar 4.42. Preferred Social Media Platform	38
Gambar 4.43. Normalisasi Data.....	39
Gambar 4.44. Visualisasi Z-Score.....	40

Gambar 4.45. Visualisasi Normalisasi	40
Gambar 4.46. Korelasi Antara Kolom Age of Subject dan Time spent on Online Class.....	41
Gambar 4.47. Korelasi antara kolom ‘Age of Subject’ dan ‘Time spent on self study’	42
Gambar 4.48. Korelasi antara kolom ‘Age of Subject’ dan ‘Time spent on fitness’	43
Gambar 4.49. Korelasi antara kolom ‘Age of Subject’ dan ‘Time spent on sleep’	44
Gambar 4.50. Korelasi antara kolom ‘Age of Subject’ dan ‘Time spent on social media’	45
Gambar 4.51. Korelasi antara kolom ‘Age of Subject’ dan ‘Time spent on TV’	46
Gambar 4.52. Korelasi antara kolom ‘Age of Subject’ dan ‘Number of meals per day’	47
Gambar 4.53. Korelasi antara kolom ‘Age of Subject’ dan ‘Stress Busters’	48
Gambar 4.54. Korelasi Antara Semua Kolom dengan ‘Health Issue During Lockdown’	49
Gambar 4.55. Visualisasi Heatmap Hubungan Linier Antar Kedua Kolom	50
Gambar 4.56. OLS Pemgaruh ‘Age Of Subject’ dan ‘Time Spent On Online Class’ terhadap	51
Gambar 4.57. OLS Pengaruh ‘Age Of Subject’ dan ‘Number Of Meals Per-Day’	52
Gambar 4.58. OLS Pengaruh Semua Kolom terhadap ‘Health Issue During Lockdown’	53
Gambar 4.59. Visualisasi Heatmap Untuk Menentukan Hubungan Linier Lebih Dari Dua Kolom	54
Gambar 4.60. Penentuan Jumlah Cluster dengan Metode Elbow.....	55
Gambar 4.61. Jumlah Masing-Masing Cluster	56
Gambar 4.62. Visualisasi Data Interaktif	57

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pandemi COVID-19 telah memberikan dampak yang signifikan terhadap berbagai aspek kehidupan. Berbagai kebijakan telah dibuat sebagai upaya untuk memutus penyebaran COVID-19, akibatnya hal ini menyebabkan terjadinya perubahan – perubahan yang signifikan dalam kehidupan sehari-hari, seperti pembatasan sosial, pembelajaran jarak jauh, larangan untuk berkumpul, dan penutupan tempat umum seperti sekolah, kantor, supermarket, dan tempat bermain. Perubahan – perubahan ini menimbulkan berbagai macam stres dan kecemasan yang dapat berdampak negatif pada kesehatan mental. Pelajar merupakan salah satu kelompok yang rentan mengalami perubahan emosi dan perilaku yang dapat mengganggu kesehatan mental. Perubahan ini merupakan reaksi terhadap rasa takut, khawatir, marah, dan sedih akibat situasi pandemi COVID-19¹.

Selain itu, selama pandemi, pelajar menghadapi berbagai tantangan tambahan seperti kesulitan belajar secara daring, kesepian, dan adanya pembatasan sosial. Pandemi COVID-19 berdampak terhadap kesehatan dan kesejahteraan mental remaja memburuk. UNICEF menunjukkan bahwa, secara global 1 dari 7 anak mengalami dampak pembatasan secara langsung². Berdasarkan penelitian I-NAMHS, selama pandemi COVID-19, 4.6% remaja melaporkan bahwa mereka sering merasa lebih cemas, lebih depresi, lebih kesepian, atau lebih sulit berkonsentrasi dari biasanya³. Oleh karena itu, perlu dilakukan analisis mendalam untuk memahami hubungan antara pandemi COVID-19 dan kesehatan mental pelajar, serta mengidentifikasi faktor-faktor penyebabnya dan merancang strategi yang efektif untuk mengatasi dampak negatifnya.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang di atas, dapat dirumuskan beberapa permasalahan, diantaranya yakni :

1. Apakah terjadinya pandemi COVID-19 dapat memengaruhi kondisi kesehatan mental pelajar?
2. Apa saja faktor yang dapat menyebabkan terjadinya risiko peningkatan gangguan kesehatan mental pada pelajar di masa pandemi COVID-19?
3. Bagaimana strategi yang dapat dilakukan untuk mengatasi dampak pandemi COVID-19 terhadap gangguan kesehatan mental pelajar?

¹ UNICEF. Resources for supporting children's emotional well-being during the COVID-19 pandemic [Internet]. 2020. [cited 2020 August 17]. Available from: 17 Agustus 2020. Didapat dari:
<https://www.unicef.org/ukraine/en/stories/resources-supporting-childrens-emotional-well-being-during-covid-19-pandemic>.

²

³ Gloria. 2022. *Hasil Survei I-NAMHS: Satu dari Tiga Remaja Indonesia Memiliki Masalah Kesehatan Mental*. Universitas Gadjah Mada. <https://ugm.ac.id/id/berita/23086-hasil-survei-i-namhs-satu-dari-tiga-remaja-indonesia-memiliki-masalah-kesehatan-mental/>

1.1. Tujuan

Adapun tujuan dari pembahasan *project* ini, yakni :

1. Untuk mengetahui hubungan antara terjadinya pandemi COVID-19 dengan kondisi kesehatan mental pelajar.
2. Untuk mengetahui dan menganalisis faktor-faktor yang berkontribusi pada risiko terjadinya gangguan kesehatan mental pada pelajar.
3. Untuk mengetahui dan merumuskan strategi yang dapat dilakukan untuk mengatasi dampak pandemi COVID-19 terhadap gangguan kesehatan mental pelajar.

1.2. Manfaat

Adapun manfaat dari pembahasan *project* ini, yakni :

1. Untuk penulis, pembahasan *project* ini bertujuan untuk menambah wawasan dan pengetahuan mengenai dampak pandemi COVID-19 terhadap kesehatan mental pelajar dan faktor-faktor penyebabnya.
2. Untuk pelajar, pembahasan *project* ini bertujuan untuk mengetahui hubungan antara pandemi COVID-19 dengan kesehatan mental mereka dan bagaimana strategi atau cara yang dapat dilakukan untuk mengatasi risiko terjadinya gangguan kesehatan mental.
3. Untuk pembaca, pembahasan *project* ini bertujuan untuk memberikan pemahaman, pengetahuan, dan strategi mengenai bagaimana pandemi COVID-19 dapat memengaruhi kesehatan mental yang dapat diterapkan di berbagai bidang lingkup kehidupan.

BAB II

TINJAUAN PUSTAKA

Pembahasan mengenai teori penunjang dan penelitian terkait analisis hubungan antara pandemi COVID-19 dan kesehatan mental pelajar dengan mempertimbangkan faktor-faktor penyebabnya serta strategi untuk mengatasinya.

2.1. Teori Penunjang

2.1.1. COVID-19

COVID-19, atau "*Coronavirus Disease 2019*" adalah penyakit yang disebabkan oleh jenis coronavirus yang dikenal sebagai SARS-CoV-2 (*Severe Acute Respiratory Syndrome Coronavirus 2*) yang dapat menyebabkan gangguan pernapasan dan radang paru.⁴ Virus ini pertama kali diidentifikasi di kota Wuhan, Provinsi Hubei, Tiongkok, pada akhir tahun 2019 dan menyebar secara global, menyebabkan pandemi. Gejala klinis yang didapatkan dari virus sangat beragam, mulai dari yang ringan seperti batuk, pilek, nyeri tenggorokan, nyeri otot, nyeri kepala, hingga berat seperti pneumonia. Virus ini dapat dengan cepat menular dan menyebar antar manusia baik secara langsung maupun tidak langsung (melalui benda atau permukaan yang terkontaminasi). Orang-orang yang berada dalam jarak dekat dengan orang yang terinfeksi dapat terpapar COVID-19 ketika percikan infeksi masuk ke mulut, hidung, atau mata mereka.

2.1.2. Lock-Down COVID-19

Lock-Down COVID-19 merupakan salah satu upaya yang dilakukan pemerintah untuk menghentikan rantai penyebaran COVID-19 dengan membatasi gerak dan interaksi sosial masyarakat secara menyeluruh. Penerapan *lock-down* ini berdampak pada semua tempat umum, seperti sekolah, kantor, dan tempat perbelanjaan serta berdampak pada semua lapisan masyarakat baik pelajar maupun pekerja. Hal ini menyebabkan adanya dampak positif dan negatif akibat penerapan *lock-down*. Dampak positif yang terjadi, seperti pengendalian penyebaran virus COVID-19 yang bisa menurunkan adanya tingkat bertambahnya kasus dan kematian COVID-19. Sementara itu, dampak negatif yang terjadi yakni dalam bidang ekonomi, sosial, dan khususnya psikologis.

2.1.3. Pelajar

Pelajar merupakan individu atau seseorang yang sedang mengikuti proses pembelajaran di suatu institusi pendidikan, baik sekolah maupun perguruan tinggi meliputi tingkat dasar, menengah, atas, dan tinggi, termasuk pendidikan formal dan non formal. Pelajar terlibat dalam kegiatan pembelajaran yang berguna untuk meningkatkan pengetahuan, keterampilan, dan pemahaman tentang berbagai bidang studi yang melibatkan partisipasi dalam baik dalam kelas, penelitian, ujian, dan kegiatan lainnya yang mendukung pengembangan intelektual dan sosial.

2.1.4. Kesehatan Mental

Kesehatan mental merupakan kesehatan yang berkaitan dengan kondisi emosi, kejiwaan, dan psikis seseorang. Seseorang dapat dinyatakan memiliki kesehatan mental

⁴ Infeksi Semarang. 2020. *Apakah Coronavirus dan COVID-19 itu?*.

<https://infeksiemerging.kemkes.go.id/uncategorized/apakah-coronavirus-dan-covid-19-itu>

yang baik jika orang tersebut memiliki kondisi batin yang tenteram dan tenang sehingga dapat menikmati kehidupan sehari-hari dan menghargai dirinya serta orang sekitar. Seseorang dengan kesehatan mental yang baik mampu mengoptimalkan kemampuan dan potensinya dalam menghadapi tantangan yang ada dalam hidup, serta menjalin hubungan positif dengan orang lain.⁵ Sebaliknya, individu yang mengalami gangguan kesehatan mental akan menghadapi perubahan suasana hati, terganggunya kemampuan berpikir, kehilangan pengendalian emosi, yang akhirnya dapat menyebabkan perilaku yang tidak diinginkan. Beberapa gangguan mental yang secara umum sering terjadi di sekitar kita adalah stres, gangguan kecemasan, dan depresi.

2.1.5. K-Means

K-Means adalah algoritma *clustering* yang digunakan untuk mengelompokkan data yang tidak berlabel ke dalam k kelompok yang berbeda. Algoritma ini bekerja dengan cara memilih secara acak k titik data awal sebagai pusat *cluster* dan mengelompokkan seluruh data ke dalam *cluster* dengan pusat terdekat. Pusat *cluster* kemudian diperbarui menjadi rata-rata dari seluruh data dalam cluster tersebut. Proses ini diulangi hingga pusat *cluster* tidak lagi berubah. K-means adalah algoritma *clustering* yang sederhana dan efisien yang dapat memberikan hasil yang baik untuk berbagai tipe data.

2.1.6. Analisis Data Eksploratif

Analisis Data Eksploratif adalah suatu pendekatan statistik yang memanfaatkan berbagai teknik statistik dan visualisasi untuk merinci dan merangkum sifat-sifat kunci dalam suatu dataset. Proses ini bertujuan mengungkap pola, hubungan, dan struktur yang terdapat dalam data. Dalam EDA, perhitungan statistika digunakan untuk memberikan gambaran tentang pemusatan dan sebaran data, sementara grafik eksploratif seperti *histogram*, *scatter plot*, *box plot*, dan *line plot* menjadi alat penting untuk menggambarkan distribusi, hubungan, dan tren data. Transformasi data, seperti transformasi logaritmik atau skala kuartil, digunakan untuk menangani asimetri atau outlier. Uji hipotesis awal, seperti uji normalitas dan uji korelasi, turut mendukung pemahaman mendalam tentang struktur dan karakteristik dataset. Secara keseluruhan, EDA tidak hanya memberikan insight mendalam, tetapi juga membantu merencanakan langkah-langkah analisis selanjutnya dan menjawab pertanyaan penelitian yang lebih mendalam.

⁵ Kementerian Kesehatan Republik Indonesia. 2018. Pengertian Kesehatan Mental. <https://ayosehat.kemkes.go.id/pengertian-kesehatan-mental>

2.2. Penelitian Terkait

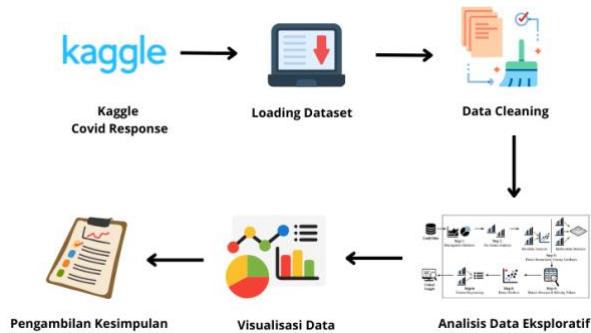
No	Nama Peneliti	Judul	Metode	Dataset	Hasil
1.	Natasha Y. Sheikhan, MPH, Lisa D. Hawke, PhD, Clement Ma, PhD, Darren Courtney, MD, Peter Szatmari, MD, Kristin Cleverley, RN, PhD , Aristotle Voineskos, MD, PhD, Amy Cheung, MD , and Joanna Henderson, PhD	<i>A Longitudinal Cohort Study of Youth Mental Health and Substance use Before and During the COVID-19 Pandemic in Ontario, Canada: An Exploratory Analysis</i>	Metode yang digunakan adalah analisis statistik dengan menggunakan perhitungan melalui SPSS 25	Dataset yang dipakai adalah dataset yang dikumpulkan dalam rentang waktu April 2018 - April 2021 dari 168 partisipan yang ada di Ontario, Canada.	Berdasarkan hasil analisis statistika, ditemukan bahwa terdapat beberapa faktor yang dapat menjadi penyebab gangguan kesehatan mental pada mahasiswa, baik dari aspek internal maupun eksternal. Faktor-faktor tersebut melibatkan perbedaan dalam waktu, tempat lahir, tingkat pendidikan, dan status pekerjaan. Selain itu, hasil analisis juga menunjukkan adanya penurunan penggunaan narkoba pada mahasiswa selama rentang waktu yang diamati. Temuan ini memberikan wawasan penting tentang dinamika kesehatan mental dan perilaku penggunaan zat di kalangan mahasiswa, serta potensi pengaruh beberapa faktor tertentu terhadap kondisi mereka.
2.	Rinda Indriyani Shindi, Ayu Qomaria Putri, Zafira Dea Natasari	Analisis Dampak COVID-19 Terhadap Kehidupan Sosial dan Kesehatan Mental Siswa	Metode yang digunakan menggunakan teknik Analisis Data Eksplorasi (EDA)	<i>COVID-19 and its Impact on Education, Social Life and Mental Health of Students</i> Link : https://www.kaggle.com/datasets/kunal28chaturvedi/covid19-and-its-impact-on-students	Berdasarkan hasil pengolahan data dengan menggunakan metode Analisis Data Eksplorasi (EDA), diketahui bahwa persentase hasil hubungan antara umur dan waktu tidur dapat memengaruhi interaksi sosial dan kesehatan mental individu selama masa pandemi COVID-19. Semakin tua dari semakin banyak waktu yang dihabiskan untuk tidur selama pandemi dapat mengurangi interaksi sosial. Sebaliknya, meningkatnya interaksi sosial dapat terjadi ketika waktu dihabiskan untuk bekerja selama masa pandemi.

3.	Annisa Rahmadani, Casi Setianingsih, Fussy Mentari Tahun 2022	Tes Gangguan Depresi, Kecemasan, dan Stres pada Mahasiswa Selama Masa COVID-19 Menggunakan Algoritma Naive Bayes	<p>Dataset ini digunakan dalam proses membangun model Naïve Bayes untuk memetakan tingkat keparahan depresi, kecemasan dan stres dengan rekomendasi pengobatan. Dataset skala depresi kecemasan stres berupa jawaban atas 42 item pertanyaan. Dataset pengujian DASS telah diperbarui pada tahun 2018 dan terdiri dari 39.775 baris.</p> <p>Link : https://openpsychometrics.org/</p>	Akurasi sistem klasifikasi yang diperoleh sebesar 86,44% dengan 90% data latih dan 10% data uji partisi. Output berupa <i>website</i> yaitu halaman tes DASS-42 yang berisi 5 items pertanyaan dalam satu halaman.

Tabel 2.1. Penelitian Terkait

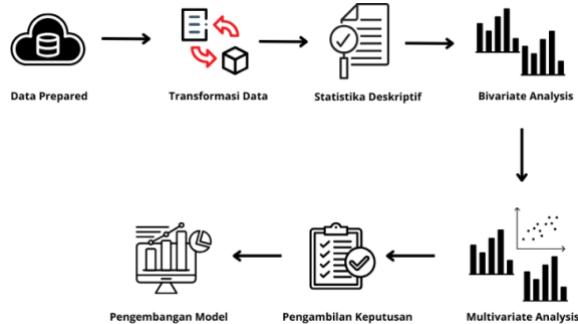
BAB III

METODOLOGI PENELITIAN



Gambar 3.1. Diagram Alir Metodologi Penelitian

Metode Penelitian ini dimulai dengan pengambilan *dataset* yang berasal dari *Kaggle* dengan judul "*COVID-19 Survey Student Responses*". Proses selanjutnya yaitu dilakukan pemuatan dan pemrosesan data dengan menggunakan *Python*. Proses pembersihan data dilakukan untuk memastikan kebersihan dan kevalidan data, termasuk penghapusan kolom yang tidak diperlukan, pengisian nilai yang hilang dengan median, dan pemilihan subset data yang relevan. Setelah tahap pembersihan data, dilanjutkan dengan analisis data eksploratif untuk memahami karakteristik *dataset*. Statistik deskriptif seperti *mean* dan *median* digunakan untuk memberikan gambaran awal tentang distribusi variabel-variabel kunci. Visualisasi data menggunakan *boxplot* membantu mengidentifikasi *outliers* pada kolom numerik tertentu. Berikut merupakan proses Analisis Data Eksploratif yang dilakukan



Gambar 3.2. Diagram Alir Proses Analisis Data Eksploratif

Proses Analisis Data Eksploratif melibatkan beberapa tahapan, dimulai dari persiapan data hingga transformasi, analisis statistik deskriptif, normalisasi, pengujian hipotesis *bivariate* dan *multivariate* menggunakan korelasi Pearson serta analisis korelasi regresi. Pada tahap akhir, jika relevan dengan tujuan penelitian, dilakukan pengembangan model prediktif. Kesimpulan dari analisis data disajikan, dengan membahas tentang faktor-faktor, dampak, dan memberikan opsi solusi pada pelajar terkait kesehatan mental selama pandemi COVID-19. Metodologi penelitian ini dirancang secara sistematis, mencakup pengambilan data, analisis eksploratif, dan visualisasi, untuk mendapatkan pemahaman menyeluruh terkait respons mahasiswa terhadap situasi pandemi.

BAB IV

HASIL DAN PEMBAHASAN

Untuk mengimplementasikan metode K-Means yang dipergunakan dengan tujuan untuk melakukan analisis hubungan antara pandemi COVID-19 dan kesehatan mental pelajar ditingkat SD, SMP, dan SMA dengan mempertimbangkan faktor-faktor penyebab atau pendukung serta strategi untuk mengatasinya diperlukan beberapa tahapan untuk melakukan pemahaman data dan melakukan analisis dari *dataset* yang tersedia.

Dimana proses analisis data ini, yakni dimulai dari data *prepared* meliputi melakukan proses *load library* yang diperlukan, *load dataset* yang akan diproses, pemilihan data dalam *dataset* dengan rentang usia 5-20 tahun, melakukan salinan *dataset* untuk keperluan pemberian informasi statistika deskriptif tanpa menggunakan data yang telah *diencoding*, dan informasi awal mengenai *dataset*. Proses transformasi data meliputi penghapusan kolom, *encoding*, pengecekan *missing values* dan imputasi, pengecekan data duplikat, dan deteksi *outliers* menggunakan IQR. Proses analisis statistika deskriptif meliputi pengecekan apakah data pada masing-masing kolom berdistribusi normal atau *uniform* atau *skewness* atau kurtosis dan memberikan informasi mengenai data.

Proses pengujian hipotesis dengan menggunakan analisis korelasi *bivariate* dan *multivariate* yang dipergunakan untuk mengetahui apakah suatu kolom memengaruhi kolom yang lain. Proses pengembangan model dan evaluasi menggunakan K-Means. Dan yang terakhir adalah proses visualisasi interaktif dari model K-Means yang telah dibuat untuk mendapatkan *output* apakah antar satu kolom dengan kolom yang lainnya saling mempengaruhi.

4.1. Deskripsi Dataset

Pada *project* ini, dilakukan serangkaian proses penerapan analisis data eksploratif dengan menggunakan dataset *COVID-19 and Its Impact on Education, Social Life, and Mental Health of Student*. Sebelum melakukan proses analisis lebih lanjut, tentunya diperlukan pemahaman informasi mengenai data dan kolom yang terdapat dalam *dataset*. Berikut informasi mengenai *dataset* lebih lanjut,

- a. Nama *Dataset* : *COVID-19 Survey Student Responses*
- b. Sumber *Dataset* : *Kaggle*
<https://www.kaggle.com/datasets/kunal28chaturvedi/covid19-and-its-impact-on-students>
- c. Deskripsi *Dataset* : Kumpulan data yang berisi informasi mengenai dampak kesehatan mental selama pandemi *COVID-19* yang dialami oleh pelajar. Dalam *dataset* ini mencakup faktor-faktor akibat dan perubahan perilaku yang berdampak terhadap kesehatan mental.
- d. Ukuran *Dataset* : 1182 baris 19 kolom
- e. Format *Dataset* : *Comma-Separated Values (CSV)*

Disajikan kolom *dataset* sebagaimana berikut,

- a. ID : Berisi ID responden dalam pengisian dataset, dengan tipe data *string/object*
- b. *Region of residence* : Berisi informasi mengenai wilayah negara dari responden, dengan tipe data *string/object*
- c. *Age of Subject* : Berisi informasi usia responden, dengan tipe data *integer*
- d. *Time spent on Online Class* : Berisikan informasi mengenai jumlah waktu yang dihabiskan oleh responden dalam mengikuti kelas *online*
- e. *Rating of Online Class* : Berisikan informasi mengenai penilaian yang diberikan responden terhadap kelas *online*, dengan tipe data *string* berupa teks
- f. *Medium for Online Class* : Berisikan informasi mengenai media atau *device* yang digunakan oleh responden dalam pelaksanaan kelas *online*, dengan tipe data *string* berupa teks
- g. *Time spent on self study* : Berisikan informasi mengenai jumlah waktu yang dihabiskan oleh responden untuk belajar mandiri di luar kelas, dengan tipe data *float*
- h. *Time spent of fitness* : Berisikan informasi mengenai jumlah waktu yang dihabiskan oleh responden untuk aktivitas kebugaran fisik, dengan tipe data *float*
- i. *Time spent on sleep* : Berisikan informasi mengenai jumlah waktu tidur yang dihabiskan oleh responden, dengan tipe data *float*
- j. *Time spent on social media* : Berisikan informasi mengenai jumlah waktu yang dihabiskan oleh responden dalam penggunaan media sosial, dengan tipe data *float*
- k. *Preferred social media platform* : Berisikan informasi mengenai platform media sosial yang sering digunakan oleh responden, dengan tipe *string* berupa teks
- l. *Time spent on TV* : Berisikan informasi mengenai jumlah waktu yang dihabiskan oleh responden untuk menonton televisi, dengan tipe data *string*

- m. *Number of meals per day* : Berupa informasi mengenai jumlah makanan yang dikonsumsi oleh responden setiap harinya, dengan tipe data *integer*
- n. *Change in your weight* : Berisikan informasi mengenai perubahan berat badan responden selama pandemi, dengan tipe data *string*
- o. *Health issue during lockdown* : Berisikan informasi mengenai masalah kesehatan yang dialami oleh responden selama pandemi, dengan tipe data *string*
- p. *Stress busters* : Berisikan informasi mengenai cara atau strategi yang dilakukan responden dalam menghadapi stress, dengan tipe data *string*
- q. *Time utilized* : Berisikan informasi mengenai jumlah waktu yang dihabiskan oleh responden untuk aktivitas atau kegiatan tertentu selama pandemi, dengan tipe data *string*
- r. *Do you find yourself more connected with your family, close friends, relatives?* : Berisikan informasi mengenai apakah responden apakah terhubung dengan keluarga, teman dekat, atau kerabat selama pandemi, dengan tipe data *string*
- s. What you miss the most : Berisikan informasi tentang hal yang paling dirindukan oleh responden selama pandemi, dengan tipe data *string*

4.2. Persiapan / Preparing Data

Pada tahapan ini dilakukan proses *load*/penemuan/*preparing* dari *library* dan *dataset* yang akan diproses analisis lebih lanjut. Kemudian, dilakukan proses pemilihan data dengan usia 5 tahun hingga 20 tahun berdasarkan usia pelajar Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP), dan Sekolah Menengah Atas (SMA). Serta, dilakukan pengecekan jumlah baris dan kolom dari *dataset* awal sebelum dilakukan proses analisis

4.2.1. Library yang Digunakan

```
#dipergunakan untuk mengabaikan peringatan
import warnings

#dipergunakan untuk membaca dan manipulasi data
import pandas as pd

#dipergunakan untuk visualisasi data
import matplotlib.pyplot as plt

#dipergunakan untuk perhitungan
import numpy as np

#dipergunakan untuk visualisasi data
import seaborn as sns

#dipergunakan untuk encode
from sklearn.preprocessing import LabelEncoder

#dipergunakan untuk mengecek distribusi normal, uniform, skewness, kurtosis
from scipy.stats import norm, uniform, skew, kurtosis

#dipergunakan untuk analisis korelasi
from scipy.stats import pearsonr

#dipergunakan untuk analisis korelasi regresi
import statsmodels.api as sm

#dipergunakan untuk normalisasi
from sklearn.preprocessing import StandardScaler

#dipergunakan untuk clustering
from sklearn.cluster import KMeans

#dipergunakan untuk visualisasi interaktif
import dash
```

Gambar 4.1. Library yang Digunakan

Pada tahap ini dilakukan *import library* yang akan dipergunakan untuk proses analisis. *Library warnings* untuk mengabaikan peringatan yang mungkin muncul saat menjalankan program. *Library pandas* untuk membaca dan manipulasi data dalam bentuk *DataFrame*. *Library matplotlib.pyplot* untuk membuat visualisasi data dalam bentuk grafik atau plot. *Library seaborn* untuk membuat visualisasi data statistik yang indah dan informatif.

Library LabelEncoder dari *sklearn.preprocessing* untuk mengubah label kategori menjadi numerik dalam proses *encoding*. *Library scipy.stats* untuk melakukan pengecekan distribusi normal, *uniform*, *skewness*, dan *kurtosis*. *Library pearsonr* dari *scipy.stats* untuk menghitung nilai koefisien korelasi Pearson. *Library StandarScaler* dari *sklearn.preprocessing* untuk melakukan normalisasi data yaitu mengubah distribusi data sehingga memiliki rata-rata nol dan varians satu. *Library Kmeans* dari *sklearn.cluster* untuk melakukan *clustering* data. *Library dash* untuk membuat aplikasi web interaktif dan membuat *dashboard* interaktif untuk visualisasi data

4.2.2. Load Dataset

```
import pandas as pd

print('Dataset COVID-19 And Impact On Students : ')

data_covid_1 = pd.read_csv('COVID-19 Survey Student Response.csv')
data_covid_1
```

Dataset COVID-19 And Impact On Students :																			
ID	Region of residence	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	Time spent on TV	Number of meals per day	Change in your weight	Health issue during lockdown	Stress busters	Time utilized	Do you find yourself more connected with your family, close friends , relatives ?	what you miss the most	
0	R1	Delhi-NCR	21	2.0	Good	Laptop/Desktop	4.0	0.0	7.0	3.0	LinkedIn	1	4	Increased	NO	Cooking	YES	YES	School,college
1	R2	Delhi-NCR	21	0.0	Excellent	Smartphone	0.0	2.0	10.0	3.0	Youtube	0	3	Decreased	NO	Scrolling through social media	YES	NO	Roaming around freely
2	R3	Delhi-NCR	20	7.0	Very poor	Laptop/Desktop	3.0	0.0	6.0	2.0	LinkedIn	0	3	Remain Constant	NO	Listening to music	NO	YES	Traveling
3	R4	Delhi-NCR	20	3.0	Very poor	Smartphone	2.0	1.0	6.0	5.0	Instagram	0	3	Decreased	NO	Watching web series	NO	NO	Friends, relatives
4	R5	Delhi-NCR	21	3.0	Good	Laptop/Desktop	3.0	1.0	8.0	3.0	Instagram	1	4	Remain Constant	NO	Social Media	NO	NO	Traveling
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
1177	R1191	Delhi-NCR	12	3.0	Good	Smartphone	4.0	1.0	8.0	1.0	Instagram	2	3	Decreased	NO	Dancing	YES	YES	Traveling
1178	R1192	Delhi-NCR	14	6.0	Average	Smartphone	4.0	1.0	9.0	1.0	Whatsapp	1	4	Remain Constant	NO	Listening to music	YES	YES	Friends, relatives
1179	R1193	Delhi-NCR	13	4.0	Average	Smartphone	0.0	0.5	8.0	3.0	Youtube	2	4	Decreased	NO	Online gaming	NO	YES	School,college
1180	R1194	Delhi-NCR	14	5.0	Excellent	Laptop/Desktop	3.5	1.0	8.0	0.5	Youtube	1	4	Remain Constant	NO	Reading books	YES	YES	School,college
1181	R1195	Delhi-NCR	13	5.0	Good	Tablet	2.0	0.5	7.0	1.0	Whatsapp	1	3	Remain Constant	NO	Talking	YES	YES	School,college

Gambar 4.2. Load Dataset

Penjelasan :

Kode ini bertujuan untuk membaca dan menampilkan *dataset COVID-19 Survey Student Responde* yang disimpan dalam format csv, dengan menggunakan library pandas sebagai pd.

Interpretasi Hasil :

Output dari kode ini adalah informasi data pada semua kolom. Didapatkan record data sejumlah 1182 baris dengan 19 kolom.

4.2.3. Pemilihan Data dengan Kriteria Usia 5-20 Tahun

```
#menampilkan dataset dengan usia 5-20 tahun
data_covid_a = data_covid_1[(data_covid_1['Age of Subject'] >= 5) & (data_covid_1['Age of Subject'] <= 20)]

print('Dataset Setelah Pemilihan : ')
data_covid_a
```

Gambar 4.3. Pemilihan Dataset Usia 5-20 Tahun

Dataset Setelah Pemilihan :																			
ID	Region of Subject	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	Time spent on TV	Number of meals per day	Change in your weight	Health issue during lockdown	Stress busters	Time utilized	Do you find yourself more connected with your family, close friends , relatives ?	what you miss the most	
2	R3	Delhi-NCR	20	7.0	Very poor	Laptop/Desktop	3.0	0.0	6.0	2.0	LinkedIn	0	3	Remain Constant	NO	Listening to music	NO	YES	Traveling
3	R4	Delhi-NCR	20	3.0	Very poor	Smartphone	2.0	1.0	6.0	5.0	Instagram	0	3	Decreased	NO	Watching web series	NO	NO	Friends, relatives
6	R7	Delhi-NCR	19	2.0	Very poor	Smartphone	2.0	1.0	5.0	4.0	Instagram	0	3	Increased	NO	Watching web series	NO	YES	Friends, relatives
7	R8	Outside Delhi-NCR	19	2.0	Very poor	Tablet	1.0	1.0	10.0	5.0	Instagram	0	3	Increased	YES	Scrolling through social media	NO	YES	Eating outside
9	R10	Outside Delhi-NCR	20	0.0	Very poor	Laptop/Desktop	1.0	0.5	8.0	5.0	Instagram	3	3	Decreased	YES	live stream watching	NO	NO	School,college
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
1177	R1191	Delhi-NCR	12	3.0	Good	Smartphone	4.0	1.0	8.0	1.0	Instagram	2	3	Decreased	NO	Dancing	YES	YES	Traveling
1178	R1192	Delhi-NCR	14	6.0	Average	Smartphone	4.0	1.0	9.0	1.0	Whatsapp	1	4	Remain Constant	NO	Listening to music	YES	YES	Friends, relatives
1179	R1193	Delhi-NCR	13	4.0	Average	Smartphone	0.0	0.5	8.0	3.0	Youtube	2	4	Decreased	NO	Online gaming	NO	YES	School,college
1180	R1194	Delhi-NCR	14	5.0	Excellent	Laptop/Desktop	3.5	1.0	8.0	0.5	Youtube	1	4	Remain Constant	NO	Reading books	YES	YES	School,college
1181	R1195	Delhi-NCR	13	5.0	Good	Tablet	2.0	0.5	7.0	1.0	Whatsapp	1	3	Remain Constant	NO	Talking	YES	YES	School,college

Gambar 4.4. Output Pemilihan Dataset Usia 5-20 Tahun

Penjelasan :

Kode ini bertujuan untuk pemilihan *dataset* dengan usia 5 – 20 tahun. Pemilihan umur ini didasarkan pada usia pelajar jenjang pendidikan Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP), dan Sekolah Menengah Atas (SMA). Dengan melakukan pemilihan data pada kolom *Age of Subject*, dan menggunakan kondisi lebih besar sama dengan (≥ 5) dan kurang dari sama dengan 20 (≤ 20).

Interpretasi Hasil :

Output dari kode ini adalah pemilihan *dataset* tersebut disimpan dalam variabel *data_covid_a*. Diperoleh *record* data sebesar 717 baris dengan 19 kolom.

4.2.4. Membuat Salinan DataFrame dari Dataset

```
#melakukan pemutuan salinan dari dataframe pada dataset, yang akan digunakan untuk proses encode
data_covid = data_covid_a.copy()
data_covid
```

Gambar 4.5. Membuat Salinan DataFrame

Dataset Setelah Pemilihan :																			
ID	Region of residence	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	Time spent on TV	Number of meals per day	Change in your weight	Health issue during lockdown	Stress busters	Time utilized	Do you find yourself more connected with your family , close friends , relatives ?	What you miss the most	
2	R3	Delhi-NCR	20	7.0	Very poor	Laptop/Desktop	3.0	0.0	6.0	2.0	LinkedIn	0	3	Remain Constant	NO	Listening to music	NO	YES	Travelling
3	R4	Delhi-NCR	20	3.0	Very poor	Smartphone	2.0	1.0	6.0	5.0	Instagram	0	3	Decreased	NO	Watching web series	NO	NO	Friends , relatives
6	R7	Delhi-NCR	19	2.0	Very poor	Smartphone	2.0	1.0	5.0	4.0	Instagram	0	3	Increased	NO	Watching web series	NO	YES	Friends , relatives
7	R8	Outside Delhi-NCR	19	2.0	Very poor	Tablet	1.0	1.0	10.0	5.0	Instagram	0	3	Increased	YES	Reading through social media	NO	YES	Eating outside
9	R10	Outside Delhi-NCR	20	0.0	Very poor	Laptop/Desktop	1.0	0.5	8.0	5.0	Instagram	3	3	Decreased	YES	live stream watching	NO	NO	School/college
...	
1177	R1191	Delhi-NCR	12	3.0	Good	Smartphone	4.0	1.0	8.0	1.0	Instagram	2	3	Decreased	NO	Dancing	YES	YES	Travelling
1178	R1192	Delhi-NCR	14	6.0	Average	Smartphone	4.0	1.0	9.0	1.0	Whatsapp	1	4	Remain Constant	NO	Listening to music	YES	YES	Friends , relatives
1179	R1193	Delhi-NCR	13	4.0	Average	Smartphone	0.0	0.5	8.0	3.0	Youtube	2	4	Decreased	NO	Online gaming	NO	YES	School/college
1180	R1194	Delhi-NCR	14	5.0	Excellent	Laptop/Desktop	3.5	1.0	8.0	0.5	Youtube	1	4	Remain Constant	NO	Reading books	YES	YES	School/college
1181	R1195	Delhi-NCR	13	5.0	Good	Tablet	2.0	0.5	7.0	1.0	Whatsapp	1	3	Remain Constant	NO	Talking	YES	YES	School/college

Gambar 4.6. Output Salinan DataFrame

Penjelasan:

Kode ini bertujuan untuk penyalinan *DataFrame* dari dataset *data_covid_a* yang disimpan dalam variabel *data_covid*.

Interpretasi Hasil :

Output dari kode ini adalah salinan *DataFrame* yang akan dipergunakan untuk proses *encoding* tanpa memengaruhi *DataFrame* asli.

4.2.5. Menampilkan Jumlah Baris dan Kolom

```
print("Jumlah Baris Dan Kolom Dari Dataset COVID-19 And Impact On Students")
print('Jumlah Baris : ', data_covid.shape[0])
print('Jumlah Kolom : ', data_covid.shape[1])
```

Gambar 4.7. Menampilkan Jumlah Baris dan Kolom

```
'Jumlah Baris Dan Kolom Dari Dataset COVID-19 And Impact On Students'
Jumlah Baris :  717
Jumlah Kolom :  19
```

Gambar 4.8. Output Jumlah Baris dan Kolom

Penjelasan :

Kode ini bertujuan untuk melakukan perhitungan dan mencetak jumlah baris dan kolom dari dataset *data_covid*. Dengan menggunakan *.shape[indeks]* untuk menghitung jumlah kolom.

Interpretasi Hasil :

Output dari kode ini adalah *record* data sebesar 717 baris dan sebanyak 19 kolom.

4.2.6. Pengecekan Nama Kolom dan Tipe Data

```
print('Di Tampilkan Kolom Dan Tipe Data Pada Setiap Kolom Data')
data_covid.info()
```

Gambar 4.9. Pengecekan Nama Kolom dan Tipe Data

```

Di Tampilkan Kolom Dan Tipe Data Pada Setiap Kolom Data
<class 'pandas.core.frame.DataFrame'>
Index: 717 entries, 2 to 1181
Data columns (total 19 columns):
 #   Column          Non-Null Count Dtype
 --- 
 0   ID              717 non-null    object
 1   Region of residence 717 non-null    object
 2   Age of Subject    717 non-null    int64
 3   Time spent on Online Class 717 non-null    float64
 4   Rating of Online Class experience 706 non-null    object
 5   Medium for online class 694 non-null    object
 6   Time spent on self study 717 non-null    float64
 7   Time spent on fitness 717 non-null    float64
 8   Time spent on sleep 717 non-null    float64
 9   Time spent on social media 717 non-null    float64
 10  Preferred social media platform 705 non-null    object
 11  Time spent on TV 717 non-null    object
 12  Number of meals per day 717 non-null    int64
 13  Change in your weight 717 non-null    object
 14  Health issue during lockdown 717 non-null    object
 15  Stress busters 717 non-null    object
 16  Time utilized 717 non-null    object
 17  Do you find yourself more connected with your family, close friends , relatives ? 717 non-null    object
 18  What you miss the most 717 non-null    object
dtypes: float64(5), int64(2), object(12)
memory usage: 112.0+ KB

```

Gambar 4.10. Output Nama Kolom dan Tipe Data

Penjelasan :

Kode ini bertujuan untuk pengecekan nama kolom dan tipe data dari dengan menggunakan .info().

Interpretasi Hasil :

Output dari kode ini adalah mengetahui nama-nama kolom untuk melakukan proses *encoding* pada kolom-kolom yang memiliki tipe data selain *integer* dan *float* yang dilakukan pada proses transformasi data.

4.3. Transformasi Data

Pada tahapan ini dilakukan proses untuk mengubah dan memanipulasi data menjadi format yang lebih sesuai dan relevan sebelum dilakukan analisis lebih lanjut. Beberapa proses yang dilakukan di tahap ini, yakni melakukan penghapusan kolom ‘*Do you find yourself more connected with your family,*’ proses *encode*, pengecekan tipe data pada masing-masing kolom, pengecekan duplikasi data, pengecekan *missing values*, dan deteksi *outlier* menggunakan IQR.

4.3.1. Penghapusan Kolom ‘*Do You Find Yourself More Connected With Your Family, Close Friends, Relatives?*’

```

print('Di lakukan Penghapusan Kolom : ')
data_covid.drop(data_covid.columns[17], axis=1, inplace=True)
data_covid

```

Gambar 4.11. Menghapus Kolom

Di lakukan Penghapusan Kolom :															
ID	Region of residence	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	Time spent on TV	Number of meals per day	Change in your weight	Health issue during lockdown	Stress busters
2	R3	Delhi-NCR	20	7.0	Very poor	Laptop/Desktop	3.0	0.0	6.0	2.0	LinkedIn	0	3	Remain Constant	NO Listening to music
3	R4	Delhi-NCR	20	3.0	Very poor	Smartphone	2.0	1.0	6.0	5.0	Instagram	0	3	Decreased	NO Watching web series
6	R7	Delhi-NCR	19	2.0	Very poor	Smartphone	2.0	1.0	5.0	4.0	Instagram	0	3	Increased	NO Watching web series
7	R8	Outside Delhi-NCR	19	2.0	Very poor	Tablet	1.0	1.0	10.0	5.0	Instagram	0	3	Increased	YES Scrolling through social media
9	R10	Outside Delhi-NCR	20	0.0	Very poor	Laptop/Desktop	1.0	0.5	8.0	5.0	Instagram	3	3	Decreased	YES live stream watching
...	
1177	R1191	Delhi-NCR	12	3.0	Good	Smartphone	4.0	1.0	8.0	1.0	Instagram	2	3	Decreased	NO Dancing
1178	R1192	Delhi-NCR	14	6.0	Average	Smartphone	4.0	1.0	9.0	1.0	Whatsapp	1	4	Remain Constant	NO Listening to music
1179	R1193	Delhi-NCR	13	4.0	Average	Smartphone	0.0	0.5	8.0	3.0	Youtube	2	4	Decreased	NO Online gaming
1180	R1194	Delhi-NCR	14	5.0	Excellent	Laptop/Desktop	3.5	1.0	8.0	0.5	Youtube	1	4	Remain Constant	NO Reading books
1181	R1195	Delhi-NCR	13	5.0	Good	Tablet	2.0	0.5	7.0	1.0	Whatsapp	1	3	Remain Constant	NO Talking

717 rows x 10 columns

Gambar 4.12. Output Penghapusan Kolom

Penjelasan :

Kode ini bertujuan untuk melakukan penghapusan kolom pada indeks ke 17 atau kolom ‘*Do You Find Yourself More Connected With Your Family, Close Friends, Relatives?*’ karena tidak dapat dibaca dan tidak dapat digunakan dalam proses analisis.

Interpretasi Hasil :

Output dari kode ini adalah *DataFrame* berjumlah 717 baris dan 18 kolom.

4.3.2. Proses *Encoding* pada Kolom dengan Tipe Data *Object* atau *String*.

```
#library sklearn untuk encoding
from sklearn.preprocessing import LabelEncoder

#pemilihan kolom yang ingin di-encode dengan label encoding
columns_to_label_encode = ["Region of residence", "Rating of Online Class experience",
                            "Medium for online class", "Prefered social media platform",
                            "Change in your weight", "Stress busters", "What you miss the most"]

#menggunakan label encoding
label_encoder = LabelEncoder()
data_covid[columns_to_label_encode] = data_covid[columns_to_label_encode].apply(lambda col:
                                         label_encoder.fit_transform(col))

#melakukan salin dataframe ke dataframe baru untuk encoding nilai tertentu
data_covid_encoded = data_covid.copy()

#ganti yes/no dengan nilai tertentu dengan angka biner (0 atau 1)
mapping_binary = {"Health issue during lockdown": {'YES': 1, 'NO': 0}, "Time utilized": {'YES': 1, 'NO': 0}}

# Gunakan replace untuk mengganti nilai sesuai dengan mapping_binary
data_covid_encoded.replace(mapping_binary, inplace = True)

#menghapus kolom asli yang sudah diencode dari data_covid_encoded
data_covid_encoded = data_covid_encoded.drop(columns = columns_to_label_encode, axis = 1)

#melakukan penggabungan datafarme dengan kolom yang diencode dan kolom yang belum diencode
data_covid_fitted = pd.concat([data_covid_encoded, data_covid[columns_to_label_encode]], axis=1)

#hapus kolom yang berisi data yang sama
data_covid_fitted = data_covid_fitted.loc[:,~data_covid_fitted.columns.duplicated()]

#penampilan hasil
print('Di Tampilkan DataFrame Dari Hasil Proses Encoding : ')
data_covid_fitted
```

Gambar 4.13. Proses *Encoding*

Di Tampilkan DataFrame Dari Hasil Proses Encoding :																	
ID	Age of Subject	Time spent on Online Class	Time spent on self study	Time spent on fitness	Time spent on sleep	Time spent on social media	Time spent on TV	Number of meals per day	Health issue during lockdown	Time utilized	Region of residence	Rating of Online Class experience	Medium for online class	Prefered social media platform	Change in your weight	Stress busters	What you miss the most
2	R3	20	7.0	3.0	0.0	6.0	2.0	0	3	0	0	0	0	0	0	0	
3	R4	20	3.0	2.0	1.0	6.0	5.0	0	3	0	0	0	0	0	0	0	
6	R7	19	2.0	2.0	1.0	5.0	4.0	0	3	0	0	0	0	0	0	0	
7	R8	19	2.0	1.0	1.0	10.0	5.0	0	3	1	0	0	1	0	0	1	
9	R10	20	0.0	1.0	0.5	8.0	5.0	3	3	1	0	0	1	0	0	1	
...	
1177	R1191	12	3.0	4.0	1.0	8.0	1.0	2	3	0	1	0	0	0	0	0	
1178	R1192	14	6.0	4.0	1.0	9.0	1.0	1	4	0	1	0	0	0	0	0	
1179	R1193	13	4.0	0.0	0.5	8.0	3.0	2	4	0	0	0	0	0	0	0	
1180	R1194	14	5.0	3.5	1.0	8.0	0.5	1	4	0	1	0	0	0	0	0	
1181	R1195	13	5.0	2.0	0.5	7.0	1.0	1	3	0	1	0	0	0	0	0	

Gambar 4.14. Output Proses *Encoding*

Penjelasan :

Kode ini bertujuan untuk proses *encoding* pada semua kolom kecuali kolom yang telah memiliki tipe data numerik dan kolom ‘ID’ dalam *DataFrame* *data_covid*. Dengan modul *LabelEncoder* dari *scikit-learn* untuk proses label *encoding*. Kolom-kolom yang akan *diencode* diinisiasi dalam variabel *columns_to_label_encode*. Kemudian dilakukan pada kolom-kolom tersebut menggunakan metode *apply* dan fungsi *lambda*. Setiap kolom diproses secara terpisah menggunakan objek *LabelEncoder*, yang mengonversi nilai kategorikal menjadi nilai numerik.

Setelah *encoding* label, *DataFrame* asli data_covid disalin ke dalam *DataFrame* baru yang disebut data_covid_encoded. Selanjutnya, nilai 'YES' dan 'NO' pada beberapa kolom diubah menjadi angka biner (1 dan 0) sesuai dengan *mapping* yang telah ditentukan dalam variabel mapping_binary.

Interpretasi Hasil :

Output dari kode ini adalah pembentukan kolom asli yang sudah *diencode* kemudian dihapus dari *DataFrame* data_covid_encoded, dan *DataFrame* ini digabungkan kembali dengan kolom-kolom yang belum *di encode* dari *DataFrame* asli. Kolom-kolom yang berisi data yang sama dihapus untuk memastikan tidak ada kolom yang berganda setelah penggabungan. Sehingga akan ditampilkan *record* data sebesar 717 baris dan 18 kolom.

4.3.3. Pengecekan Nama Kolom dan Tipe Data

```
print('Di Tampilkan Kolom Dan Tipe Data Pada Setiap Kolom Data : ')
data_covid_fitted.info()
```

Gambar 4.15. Pengecekan Nama Kolom dan Tipe Data

```
Di Tampilkan Kolom Dan Tipe Data Pada Setiap Kolom Data :
<class 'pandas.core.frame.DataFrame'>
Index: 717 entries, 2 to 1181
Data columns (total 18 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   ID               717 non-null    object 
 1   Age of Subject   717 non-null    int64  
 2   Time spent on Online Class 717 non-null    float64 
 3   Time spent on self study 717 non-null    float64 
 4   Time spent on fitness 717 non-null    float64 
 5   Time spent on sleep   717 non-null    float64 
 6   Time spent on social media 717 non-null    float64 
 7   Time spent on TV     717 non-null    object 
 8   Number of meals per day 717 non-null    int64  
 9   Health issue during lockdown 717 non-null    int64  
 10  Time utilized      717 non-null    int64  
 11  Region of residence 717 non-null    int64  
 12  Rating of Online Class experience 717 non-null    int64  
 13  Medium for online class 717 non-null    int64  
 14  Preferred social media platform 717 non-null    int64  
 15  Change in your weight 717 non-null    int64  
 16  Stress busters     717 non-null    int64  
 17  What you miss the most 717 non-null    int64  
dtypes: float64(5), int64(11), object(2)
memory usage: 106.4+ KB
```

Gambar 4.16. Output Pengecekan Nama Kolom dan Tipe Data

Penjelasan :

Kode di atas bertujuan untuk pengecekan nama kolom dan tipe data dari *dataset* data_covid_fitted dengan menggunakan .info().

Interpretasi Hasil :

Output dari kode ini adalah untuk melakukan pengecekan ulang dan menentukan kolom yang akan diproses lebih lanjut. Didapatkan pada kolom 'Time spent on TV' masih memiliki tipe data *object*. Oleh karena itu diperlukan pemrosesan data lebih lanjut.

4.3.4. Pengecekan Nilai pada Kolom 'Time spent on TV'

```
#Karena pada time on spent on sleep terdapat tipe data object
#maka, dilakukan pengecekan nilai unik pada kolom tersebut
print('Nilai Unik Pada Kolom Time Spent On Sleep : \n')

for unique_value in data_covid_fitted['Time spent on TV'].unique():
    print(unique_value)
    print('\n')

#dan ditemukan "No Tv" lalu dilakukan konversi kolom tersebut ke tipe data numerik, dengan mengabaikan nilai
#engkonversi kolom "Time spent on TV" ke tipe data float dengan mengabaikan nilai yang tidak dapat diubah
print('Di tampilkan Hasil Konversi Tipe Data : ')

data_covid_fitted['Time spent on TV'] = pd.to_numeric(data_covid_fitted['Time spent on TV'], errors='coerce').astype(float)
data_covid_fitted['Time spent on TV']
```

Gambar 4.17. Pengecekan Nilai Kolom 'Time spent on TV'

```

Di tampilkan Hasil Konversi Tipe Data :
: 2    0.0
3    0.0
6    0.0
7    0.0
9    3.0
      ...
1177  2.0
1178  1.0
1179  2.0
1180  1.0
1181  1.0
Name: Time spent on TV, Length: 717, dtype: float64

Nilai Unik Pada Kolom Time Spent On TV :
8
3
1
2
0.5
4.5
1.5
N
4
0.3
5
0.1
0.6
8
6
15
No tv
2.5
3.5
0.4

```

Gambar 4.18. Nilai Unik dan Konversi Data Kolom ‘Time spent on TV’

Penjelasan :

Kode di atas bertujuan untuk mengecek nilai unik dalam kolom ‘Time spent on TV’ pada DataFrame data_covid_fitted. Dengan iterasi pencarian nilai unik dalam kolom tersebut menggunakan *loop for*, dan mencetak setiap nilai unik. Hasil pengecekan menunjukkan bahwa terdapat nilai ‘N’ dan ‘No TV’ dalam kolom tersebut. Selanjutnya, karena kolom ‘Time spent on TV’ memiliki tipe data objek kemudian di ubah menjadi tipe data numerik dengan pd.to_numeric.

Interpretasi Hasil :

Output dari kode ini adalah nilai seperti ‘N’ dan ‘No TV’ akan diubah menjadi nilai ‘nan’ dan dilakukan proses lebih lanjut.

4.3.5. Pengecekan Missing Value

```

#pengecekan missing values
print('Hasil Cek Missing Value : ')
null_value = data_covid_fitted.isnull().sum()
null_value

```

Gambar 4.19. Pengecekan Missing Value

Hasil Cek Missing Value :	
ID	0
Age of Subject	0
Time spent on 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
Time spent on TV	5
Number of meals per day	0
Health issue during lockdown	0
Time utilized	0
Region of residence	0
Rating of Online Class experience	0
Medium for online class	0
Preferred social media platform	0
Change in your weight	0
Stress busters	0
What you miss the most	0

Gambar 4.20. Output Pengecekan Missing Value

Penjelasan :

Kode di atas bertujuan untuk melakukan pengecekan *missing values* atau nilai yang hilang dalam DataFrame data_covid_fitted dengan menggunakan metode .isnull dan metode .sum() untuk menghitung jumlah *missing value* pada setiap kolom.

Interpretasi Hasil :

Output dari kode ini adalah diketahui adanya *missing value* pada kolom ‘Time spent on TV’ sebesar 5.

4.3.6. Imputasi Missing Value pada Kolom ‘Time spent on TV’

```

#mengisi nilai nan pada kolom 'time spent on tv' dengan nilai median
median_value = data_covid_fitted["Time spent on TV"].median()
data_covid_fitted[["Time spent on TV"]].fillna(median_value, inplace=True)

print('Di Lakukan Imputasi Missing Value : ')
data_covid_fitted

```

Gambar 4.21. Imputasi Missing Value

Di Lakukan Imputasi Missing Value :												
ID	Age of Subject	Time spent on Online Class	Time spent on self study	Time spent on fitness	Time spent on sleep	Time spent on social media	Time spent on TV	Number of meals per day	Health issue during lockdown	Time utilized		
2	R3	20	7.0	3.0	0.0	6.0	2.0	0.0	3	0	0	
3	R4	20	3.0	2.0	1.0	6.0	5.0	0.0	3	0	0	
6	R7	19	2.0	2.0	1.0	5.0	4.0	0.0	3	0	0	
7	R8	19	2.0	1.0	1.0	10.0	5.0	0.0	3	1	0	
9	R10	20	0.0	1.0	0.5	8.0	5.0	3.0	3	1	0	
...	
1177	R1191	12	3.0	4.0	1.0	8.0	1.0	2.0	3	0	1	
1178	R1192	14	6.0	4.0	1.0	9.0	1.0	1.0	4	0	1	
1179	R1193	13	4.0	0.0	0.5	8.0	3.0	2.0	4	0	0	
1180	R1194	14	5.0	3.5	1.0	8.0	0.5	1.0	4	0	1	
1181	R1195	13	5.0	2.0	0.5	7.0	1.0	1.0	3	0	1	

717 rows × 18 columns

Gambar 4.22. Output Imputasi Missing Value

Penjelasan :

Kode di atas bertujuan untuk mengisi *missing value* pada kolom ‘*Time spent on TV*’ dengan menghitung nilai median dari kolom tersebut menggunakan metode *.median()* dan disimpan dalam variabel *median_value*.

Interpretasi Hasil :

Output dari kode ini adalah pengisian *missing value* dalam kolom ‘*Time spent on TV*’ dengan nilai median yang telah dihitung sebelumnya menggunakan metode *fillna()*.

4.3.7. Pengecekan Missing Value Setelah Imputasi

```
#pengecekan missing values
print('Hasil Cek Missing Value Setelah Imputasi : ')

null_value = data_covid_fitted.isnull().sum()
null_value
```

Gambar 4.23. Pengecekan Missing Value

```
Hasil Cek Missing Value Setelah Imputasi :
ID 0
Age of Subject 0
Time spent on 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
Time spent on TV 0
Number of meals per day 0
Health issue during lockdown 0
Time utilized 0
Region of residence 0
Rating of Online Class experience 0
Medium for online class 0
Preferred social media platform 0
Change in your weight 0
Stress busters 0
What you miss the most 0
dtype: int64
```

Gambar 4.24. Output Pengecekan Missing Value

Penjelasan :

Kode di atas bertujuan untuk melakukan pengecekan kembali terhadap *missing value* setelah proses imputasi dilakukan pada kolom ‘*Time spent on TV*’ dalam *DataFrame* *data_covid_fitted*.

Interpretasi Hasil :

Output dari kode ini disimpan dalam variabel *null_value*, yang berisi jumlah *missing value* untuk setiap kolom setelah dilakukan imputasi. Didapatkan bahwa sudah tidak terdapat *missing value* pada *DataFrame*.

4.3.8. Pengecekan Data Duplikat

```
#memeriksa duplikasi data berdasarkan kolom 'ID'  
duplicated_rows = data_covid_fitted[data_covid_fitted.duplicated(subset = 'ID')]  
  
#pembuatan kondisi untuk pengecekan duplikasi data  
if duplicated_rows.empty:  
    print('Tidak Terdapat Duplikasi Data Dalam Dataset.')  
else:  
    print('Duplikasi Data DiTemukan. Baris Duplikat : ')  
    print(duplicated_rows)  
  
Tidak Terdapat Duplikasi Data Dalam Dataset.
```

Gambar 4.25. Pengecekan Data Duplikat

Penjelasan :

Kode di atas bertujuan untuk memeriksa keberadaan duplikasi data berdasarkan kolom 'ID' dalam *DataFrame* *data_covid*. Dengan menggunakan metode *.duplicated()* dengan argumen *subset='ID'*. Hasil dari operasi ini disimpan dalam variabel *duplicated_rows*, yang berisi baris-baris yang memiliki duplikat pada kolom 'ID'.

Interpretasi Hasil :

Output dari kode ini adalah tidak terdapat duplikasi data dalam setiap kolom yang terdapat dalam *dataset*.

4.3.9. Pengecekan Kolom dan Tipe Data

```
print('Di Tampilkan Kolom Dan Tipe Data Pada Setiap Kolom Data : ')  
  
data_covid_fitted.info()
```

Gambar 4.26. Pengecekan Kolom dan Tipe Data

```
Di Tampilkan Kolom Dan Tipe Data Pada Setiap Kolom Data :  
<class 'pandas.core.frame.DataFrame'>  
Index: 717 entries, 2 to 1181  
Data columns (total 18 columns):  
 #   Column           Non-Null Count  Dtype     
 ---  --     
 0   ID               717 non-null    object    
 1   Age of Subject  717 non-null    int64     
 2   Time spent on Online Class 717 non-null    float64   
 3   Time spent on self study 717 non-null    float64   
 4   Time spent on fitness 717 non-null    float64   
 5   Time spent on sleep   717 non-null    float64   
 6   Time spent on social media 717 non-null    float64   
 7   Time spent on TV     717 non-null    float64   
 8   Number of meals per day 717 non-null    int64     
 9   Health issue during lockdown 717 non-null    int64     
 10  Time utilized      717 non-null    int64     
 11  Region of residence 717 non-null    int64     
 12  Rating of Online Class experience 717 non-null    int64     
 13  Medium for online class 717 non-null    int64     
 14  Preferred social media platform 717 non-null    int64     
 15  Change in your weight   717 non-null    int64     
 16  Stress busters       717 non-null    int64     
 17  What you miss the most 717 non-null    int64     
dtypes: float64(6), int64(11), object(1)  
memory usage: 108.4+ KB
```

Gambar 4.27. Output Pengecekan Kolom dan Tipe Data

Penjelasan :

Kode diatas bertujuan untuk pengecekan nama kolom dan tipe data dari *dataset* *data_covid_fitted* dengan menggunakan *.info()* setelah proses *encoding*, pengecekan *missing value*, dan data duplikat.

Interpretasi Hasil :

Output dari kode di atas diperoleh bahwa semua kolom dan tipe data sudah sesuai dan dapat dilakukan untuk proses berikutnya.

4.3.10. Integrasi Data

a) Deteksi Outlier

```
#melakukan detecting outliers pada kolom dengan tipe data numerik
numerical_columns = data_covid_fitted.select_dtypes(include=['float64', 'int64']).columns

#dipergunakan 1000 baris sampel dari kolom yang dipilih
#karna data kurang dari 100, dipergunakan semua sampel
sample_size = min(1000, len(data_covid_fitted[numerical_columns]))

data_covid_outliers = data_covid_fitted[numerical_columns].sample(n=sample_size, random_state=42, replace=True).copy()

#perhitungan IQR untuk setiap kolom
Q1 = data_covid_outliers.quantile(0.25)
Q3 = data_covid_outliers.quantile(0.75)
IQR = Q3 - Q1

#batas atas dan batas bawah untuk deteksi outliers
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

#baris yang diluar batas atas atau batas bawah
outlier_indexes = ((data_covid_outliers < lower_bound) | (data_covid_outliers > upper_bound)).any(axis=1)

#penampilan DataFrame yang berisi outliers
outliers = data_covid_outliers[outlier_indexes]

print('Di Tampilkan DataFrame Yang Mengandung Outliers : ')
outliers
```

Gambar 4.28. Deteksi Outlier

Age of Subject	Time spent on online class	Time spent on self study	Time spent on fitness	Time spent on sleep	Time spent on social media	Time spent on TV	Number of meals per day	Health issue during lockdown	Time utilized	Region of residence	Rating of	Medium	Preferred	Change in your weight	
											Online Class experience	for online class	social media platform		
712	20	1.5	2.5	2.0	9.0	1.0	1.5	2	1	0	0	4	2	12	0
170	20	0.0	3.0	2.0	8.0	2.0	1.0	3	1	1	1	4	2	11	2
47	20	2.0	3.0	0.0	6.0	2.0	0.0	3	1	0	0	4	2	12	0
1076	16	8.0	4.0	0.0	5.0	8.0	3.0	3	0	1	0	0	2	11	2
206	16	7.0	4.0	1.0	7.0	1.0	0.0	3	0	1	0	0	4	11	2
...
634	16	2.0	5.0	1.0	9.0	1.0	1.0	1	0	1	1	2	5	1	0
9	20	0.0	1.0	0.5	8.0	5.0	3.0	3	1	0	1	4	1	1	0
241	16	7.0	3.0	2.0	11.0	1.5	1.0	3	1	0	1	0	2	1	0
356	13	2.0	1.0	3.0	11.0	3.0	2.0	4	0	0	0	1	2	12	1
441	20	0.0	2.0	0.0	7.0	5.0	0.0	2	1	0	1	4	1	1	0

243 rows × 17 columns

Gambar 4.29. Output Deteksi Outlier

Penjelasan :

Kode di atas bertujuan untuk melakukan deteksi *outlier* pada kolom-kolom dengan tipe data numerik dalam. Tahap pertama yaitu mengidentifikasi kolom-kolom numerik dan menyimpan nama-nama kolom tersebut dalam variabel numerical_columns. Selanjutnya, dibuat *DataFrame* baru yang berisi sampel acak dari data numerik, dengan ukuran sampel tergantung pada apakah jumlah data kurang dari 1000 atau tidak.

Proses deteksi *outliers* dilakukan dengan menghitung *Interquartile Range* (IQR) untuk setiap kolom. IQR dihitung sebagai selisih antara kuartil ketiga (Q3) dan kuartil pertama (Q1). Selanjutnya, batas atas dan batas bawah ditentukan untuk masing-masing kolom, di mana nilai di luar batas ini dianggap sebagai *outliers*.

b) Visualisasi Outlier

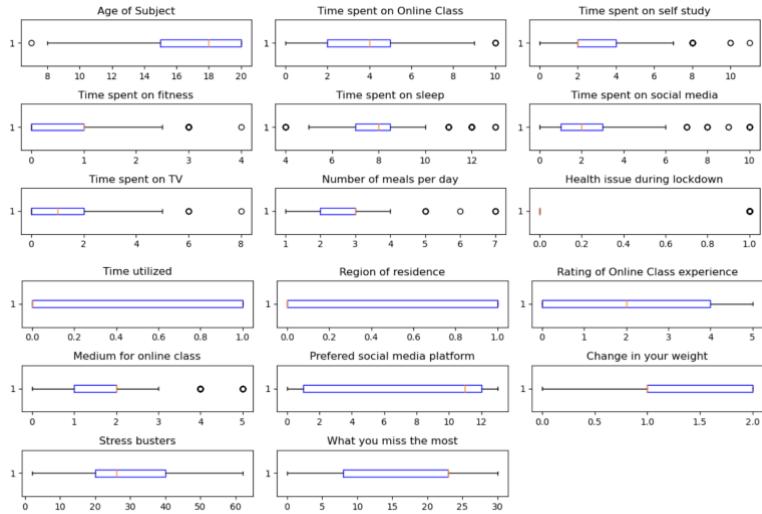
```
import matplotlib.pyplot as plt

#melakukan visualisasi IQR
#melakukan perhitungan jumlah kolom numerik
num_plots = len(numerical_columns)
num_rows = (num_plots - 1) // 3 + 1
num_cols = min(num_plots, 3)

#visualisasi boxplot untuk semua kolom numerik
plt.figure(figsize=(12, 8))
for i, column in enumerate(numerical_columns, 1):
    plt.subplot(num_rows, num_cols, i)
    plt.boxplot(data_covid_outliers[column], vert=False, boxprops=dict(color='blue'))
    plt.title(column)

#ditampilkan hasil visualisasi
plt.tight_layout()
plt.show()
```

Gambar 4.30. Visualisasi Outlier



Gambar 4.31. Boxplot Visualisasi Outlier

Penjelasan :

Kode di atas bertujuan untuk membuat deteksi *outlier* menggunakan IQR dan memvisualisasikannya pada kolom-kolom dengan tipe data numerik. Kode menghitung jumlah kolom numerik yang akan divisualisasikan, kemudian menentukan jumlah baris dan kolom pada diagram *boxplot* berdasarkan jumlah kolom numerik. Selanjutnya, gambar *boxplot* dibuat untuk setiap kolom numerik menggunakan iterasi. Setiap *subplot* dalam gambar menampilkan *boxplot* untuk satu kolom numerik, dengan judul *subplot* sesuai dengan nama kolom dalam *dataset*.

Interpretasi Hasil :

Output dari kode diatas adalah diperoleh bahwa semua data disetiap kolom tidak memiliki *outlier* walaupun beberapa visualisasi terlihat memiliki *outlier*, namun itu merupakan bagian dari data yang akan diproses dan tidak termasuk *outlier*.

4.4. Statistika Deskriptif

Pada tahapan ini dilakukan proses analisis statistika deskriptif meliputi pengecekan apakah data pada masing-masing kolom berdistribusi normal atau *uniform* atau *skewness* atau *kurtosis* dan memberikan informasi yang *informatif* mengenai data-data di semua kolom.

4.4.1. Informasi Statistika Deskriptif Pada Setiap Kolom Dalam *Dataset*

```
print('Di Tampilkan Hasil Statistika Deskriptif : ')
#statistika deskriptif pada tiap kolom yang ada dalam dataset
descriptive_stat = data_covid_fitted.describe()
descriptive_stat
```

	Age of Subject	Time spent on Online Class	Time spent on self study	Time spent on fitness	Time spent on sleep	Time spent on social media	Time spent on TV	Number of meals per day	Health issue during lockdown	Time utilized	Region of residence	Rating of Online Class experience
count	717.000000	717.000000	717.000000	717.000000	717.000000	717.000000	717.000000	717.000000	717.000000	717.000000	717.000000	717.000000
mean	17.341702	3.534519	2.816597	0.794421	7.931381	2.269317	1.067922	2.912134	0.114365	0.486750	0.355649	2.086471
std	2.739960	2.145693	1.911671	0.731368	1.621677	1.771677	1.263388	0.850127	0.318477	0.500173	0.479044	1.704676
min	7.000000	0.000000	0.000000	0.000000	4.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000
25%	16.000000	2.000000	2.000000	0.000000	7.000000	1.000000	0.000000	2.000000	0.000000	0.000000	0.000000	0.000000
50%	18.000000	4.000000	2.000000	1.000000	8.000000	2.000000	1.000000	3.000000	0.000000	0.000000	0.000000	2.000000
75%	20.000000	5.000000	4.000000	1.000000	9.000000	3.000000	2.000000	3.000000	0.000000	1.000000	1.000000	4.000000
max	20.000000	10.000000	12.000000	4.000000	15.000000	10.000000	15.000000	7.000000	1.000000	1.000000	1.000000	5.000000

Gambar 4.32. Statistika Deskriptif Dataset

Penjelasan :

Kode di atas bertujuan untuk menampilkan hasil statistika deskriptif dari setiap kolom *DataFrame* yang telah dilakukan *encoding* dan hasilnya akan disimpan dalam *descriptive_stat* yang di dalamnya terdapat fungsi *data_covid.describe()* untuk menghasilkan ringkasan statistika deskriptif.

Interpretasi Hasil :

Output dari kode di atas meliputi jumlah data (*count*), rata-rata (*mean*), standar deviasi (*std*), nilai minimum (*min*), kuartil bawah (25%), median (50%), kuartil atas (75%), dan nilai maksimum (*max*) dari 17 kolom yang dipergunakan untuk informasi tambahan.

4.4.2. Pengecekan Distribusi Normal, Distribusi Uniform, Skewness, Dan Kurtosis

Pada Masing-Masing Kolom

```
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import norm, uniform, skew, kurtosis

#distribusi untuk semua kolom
plt.figure(figsize=(30,22))

#Kolom yang akan di visualisasikan
variables_to_plot = ['Age of Subject', 'Time spent on Online Class', 'Time spent on self study', 'Time spent on sleep',
                     'Time spent on social media', 'Time spent on TV', 'Number of meal',
                     'Health issue during lockdown', 'Time utilized', 'Region of residence',
                     'Rating of Online Class experience', 'Medium for online class', 'Preferred social media p',
                     'Change in your weight', 'Stress busters', 'What you miss the most']

#tata letak plot distribusi
rows = 4
cols = 5

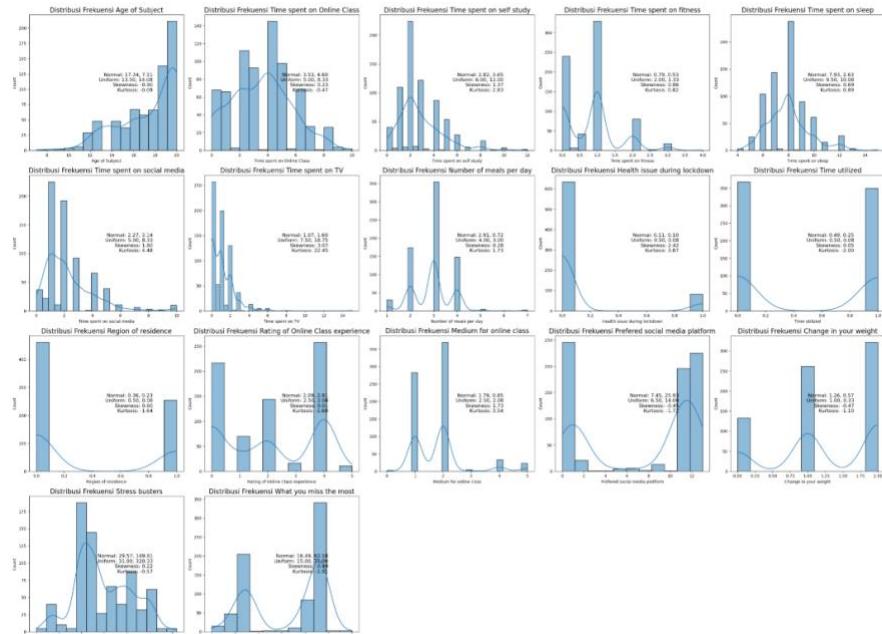
#looping for untuk membuat plot
for i, variable in enumerate(variables_to_plot, 1):
    plt.subplot(rows, cols, i)

    #plot distribusi
    sns.histplot(data_covid_fitted[variable], kde=True, edgecolor='black')
    plt.title(f'Distribusi Frekuensi {variable}', fontsize=16)

    #informasi distribusi
    normal_dist = norm(data_covid_fitted[variable].mean(), data_covid_fitted[variable].std())
    uniform_dist = uniform(data_covid_fitted[variable].min(), data_covid_fitted[variable].max() - data_covid_fitted[variable].min())
    skewness = skew(data_covid_fitted[variable])
    kurt = kurtosis(data_covid_fitted[variable])

    #penampilan informasi distribusi
    plt.text(0.80, 0.50, f"Normal: {normal_dist.mean():.2f}, {normal_dist.var():.2f}\n"
                         f"Uniform: {uniform_dist.mean():.2f}, {uniform_dist.var():.2f}\n"
                         f"Skewness: {skewness:.2f}\n"
                         f"Kurtosis: {kurt:.2f}", fontsize=12,
            horizontalalignment='right', verticalalignment='center', transform=plt.gca().transAxes)

plt.tight_layout()
plt.show()
```



Gambar 4.33. Pengecekan Distribusi Dataset

Penjelasan :

Kode ini bertujuan untuk membuat plot distribusi baik distribusi normal, *uniform*, *skewness*, maupun kurtosis untuk semua kolom yang ada dalam DataFrame kecuali kolom “ID”. Dengan menggunakan *library matplotlib* dan *seaborn* untuk membuat dan menyempurnakan visualisasi dengan mengambil daftar kolom yang akan divisualisasikan dalam variabel *variables_to_plot* kemudian melooping setiap variabel yang akan divisualisasikan menggunakan fungsi *enumerate* dan memvisualisasikan plot dari masing-masing kolom yang telah diambil dari data tersebut.

Interpretasi Hasil :

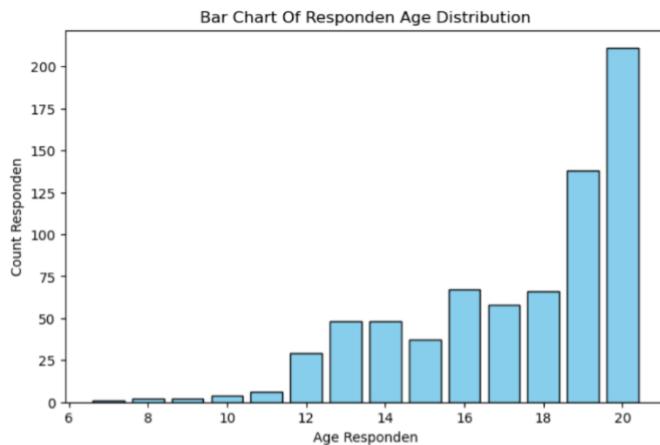
Waktu belajar *online* cenderung mendekati distribusi normal, menunjukkan pola waktu belajar yang sama. Distribusi untuk aktivitas mandiri, berolahraga, menonton TV, dan media sosial memiliki *skewness* positif, menandakan konsentrasi nilai di sebelah kanan rata-rata. Jumlah makanan per hari mendekati distribusi uniform, mencerminkan sebaran merata.

Analisis kurtosis menunjukkan distribusi masalah kesehatan, waktu yang dimanfaatkan, dan perubahan berat badan dengan kurtosis positif, menunjukkan konsentrasi nilai di sekitar rata-rata.

Kesimpulannya bahwa waktu belajar *online* yang merata mungkin dipengaruhi kebijakan pemerintah dan teknologi. Aktivitas mandiri cenderung terkonsentrasi pada nilai-nilai tinggi. Jumlah makanan per hari tersebar merata, mencerminkan perhatian terhadap kesehatan. Masalah kesehatan dan perubahan berat badan mungkin terkait dengan stres dan adaptasi terhadap situasi.

4.4.3. Distribusi Usia Responden Dalam Dataset

```
import matplotlib.pyplot as plt
# Menghitung jumlah responden pada usia tertentu
age_counts = data_covid_a['Age of Subject'].value_counts()
# Menghitung persentase
age_percentages = (age_counts / age_counts.sum()) * 100
# Membuat diagram batang distribusi usia responden
plt.figure(figsize=(8, 5))
plt.bar(age_counts.index, age_counts.values, color='skyblue', edgecolor='black')
plt.xlabel('Age Responden')
plt.ylabel('Count Responden')
plt.title('Bar Chart Of Responden Age Distribution')
# Menampilkan plot
plt.show()
# Menampilkan persentase dan jumlah responden dengan kata-kata sesuai umur
print("Dari Visualisasi Bar Chart DiAtas Didapatkan Kesimpulan Bahwa Distribusi Usia Responden, yakni:")
for age, percentage in age_percentages.items():
    count = age_counts.loc[age]
    print(f"Umur {age} tahun memiliki jumlah responden sebanyak {count} ({percentage:.2f}%)")
```



Dari Visualisasi Bar Chart DiAtas Didapatkan Kesimpulan Bahwa Distribusi Usia Responden, yakni:
 Umur 20 tahun memiliki jumlah responden sebanyak 211 (29.43%)
 Umur 19 tahun memiliki jumlah responden sebanyak 138 (19.25%)
 Umur 16 tahun memiliki jumlah responden sebanyak 67 (9.34%)
 Umur 18 tahun memiliki jumlah responden sebanyak 66 (9.21%)
 Umur 17 tahun memiliki jumlah responden sebanyak 58 (8.09%)
 Umur 14 tahun memiliki jumlah responden sebanyak 48 (6.69%)
 Umur 13 tahun memiliki jumlah responden sebanyak 48 (6.69%)
 Umur 15 tahun memiliki jumlah responden sebanyak 37 (5.16%)
 Umur 12 tahun memiliki jumlah responden sebanyak 29 (4.04%)
 Umur 11 tahun memiliki jumlah responden sebanyak 6 (0.84%)
 Umur 10 tahun memiliki jumlah responden sebanyak 4 (0.56%)
 Umur 9 tahun memiliki jumlah responden sebanyak 2 (0.28%)
 Umur 8 tahun memiliki jumlah responden sebanyak 2 (0.28%)
 Umur 7 tahun memiliki jumlah responden sebanyak 1 (0.14%)

Gambar 4.34. Persentase Usia Responden dalam Dataset

Penjelasan :

Kode ini bertujuan untuk mengetahui distribusi usia dari tiap responden dengan *bar chart* untuk visualisasi usia dari responden menggunakan *library matplotlib*. Dilakukan proses untuk mengetahui persentase dari tiap usia dengan perhitungan membagi jumlah responden pada setiap usia dengan jumlah total responden dan mengalikannya dengan 100. Dan ditampilkan visualisasi bar *chart*, persentase , dan jumlah responden pada setiap usia.

Interpretasi Hasil :

Output dari kode di atas adalah usia responden 20 tahun dengan jumlah responden 211 (29.43%), usia 19 tahun dengan jumlah responden 138 (19.25%), usia 16 tahun dengan jumlah responden 67 (9.34%), usia 18 tahun dengan jumlah responden 66 (9.21%), usia 17 tahun dengan jumlah responden 58 (8.09%), usia 14 tahun dengan jumlah responden 48 (6.69%), usia 13 tahun dengan jumlah responden 48 (6.69%), usia 15 tahun dengan jumlah responden 37 (5.16%), usia 12 tahun dengan jumlah responden 29 (4.04%), usia 11 tahun dengan jumlah responden 6 (0.84%), usia 10 tahun dengan jumlah responden 4 (0.56%), usia 9 tahun dengan jumlah responden 2 (0.28%), usia 8 tahun dengan jumlah responden 2 (0.28%), dan usia 7 tahun dengan jumlah responden 1 (0.14%)

4.4.4. Jumlah Usia Pelajar Di Tingkat SD, SMP, SMA

```
#batasan usia untuk setiap kelompok pendidikan
bins = [5, 12, 15, 20]
labels = ['SD', 'SMP', 'SMA']

data_covid['Education Level'] = pd.cut(data_covid['Age of Subject'], bins=bins, labels=labels, right=False)

#groupby untuk mengelompokkan berdasarkan 'Education Level' dan perhitungan masing-masing jumlah
education_level_counts = data_covid.groupby('Education Level').size().reset_index(name='Total')
education_level_counts
```

	Education Level	Total
0	SD	15
1	SMP	125
2	SMA	366

Gambar 4.35. Jumlah Pelajar SD, SMP, dan SMA

Penjelasan :

Kode ini bertujuan untuk menghitung jumlah usia pelajar di tingkat SD, SMP, dan SMA dengan cara menentukan label dan batas-batas usia pada tiap kelompok pendidikan jika berusia 5-12 tahun akan masuk dalam kelompok SD, 13-15 tahun akan masuk dalam kelompok SMP, dan 16-20 akan masuk dalam kelompok SMA.

Interpretasi Hasil :

Output dari kode di atas adalah diperoleh hasil sebesar 15 orang responden di tingkat SD, 125 responden di tingkat SMP, dan 356 responden di tingkat SMA.

4.4.5. Persentase Asal Daerah Dari Masing-Masing Responden

```
import matplotlib.pyplot as plt

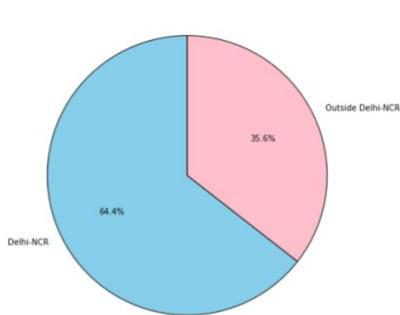
#menghitung jumlah responden per asal daerah
region_counts = data_covid_a['Region of residence'].value_counts()

#menghitung persentase
region_percentages = (region_counts / region_counts.sum()) * 100

#membuat diagram pie
plt.figure(figsize(8, 8))
colors = ['skyblue', 'pink']
wedges, texts, autotexts = plt.pie(region_percentages, labels=region_percentages.index, autopct='%1.1f%%', startangle=90, colors=colors)
for wedge in wedges:
    wedge.set_edgecolor('black')

plt.title('Percentage Of Respondents Places Of Region')
plt.show()

#menampilkan persentase dengan kata-kata sesuai asal daerah
print("Dari Visualisasi Bar Chart DiAtas Didapatkan Presentase Asal Daerah Responden, yakni : ")
for region, percentage in region_percentages.items():
    print(f"({region}) Sebanyak : {percentage:.2f}%")
```



Dari Visualisasi Bar Chart DiAtas Didapatkan Presentase Asal Daerah Responden, yakni :
Delhi-NCR Sebanyak : 64.44%
Outside Delhi-NCR Sebanyak : 35.56%

Gambar 4.36. Persentase Asal Daerah Responden

Penjelasan :

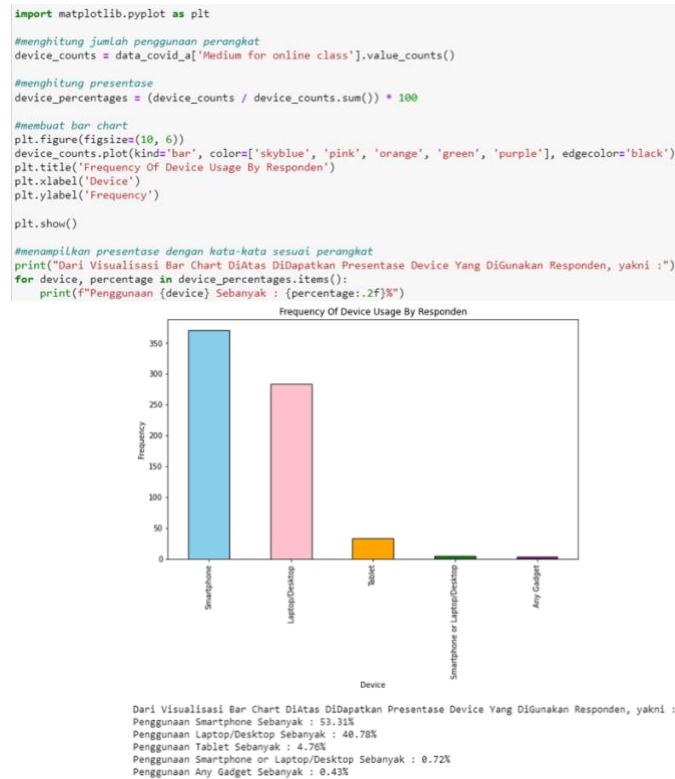
Kode di atas digunakan untuk menghitung dan memvisualisasikan persentase responden berdasarkan asal daerah ‘Region of Residence’ dengan menghitung jumlah responden per-asal negara dengan output dalam bentuk persentase dan

memvisualisasikannya dalam bentuk *diagram pie* dengan warna yang berbeda untuk menunjukkan asal negara yang berbeda.

Interpretasi Hasil :

Output dari kode di atas adalah diperoleh sebesar 64,44% responden berasal dari Dehli-NCR dan sisanya sebanyak 35,56% berada di luar Dehli-NCR. Sehingga dapat disimpulkan bahwa responden dengan asal dari Delhi-NCR lebih mendominasi.

4.4.6. Persentase Device Yang Di Pergunakan Oleh Responden



Gambar 4.37. Persentase Device yang Digunakan Responden

Penjelasan :

Kode di atas digunakan untuk mengetahui persentase perangkat (*device*) yang digunakan oleh responden. Visualisasi dilakukan menggunakan bar *chart* dengan menggunakan data dari kolom '*Medium For Online Class*'. Dengan *output* yang ditampilkan adalah *bar chart* dan perhitungan persentase penggunaan *device* oleh responden.

Interpretasi Hasil :

Output dari kode di atas adalah visualisasi *bar chart* dengan penggunaan *device smartphone* sebesar 53.31%, *device laptop/desktop* sebesar 40.78%, *device tablet* sebesar 4.76%, *device smartphone/laptop/desktop* sebesar 0.72%, dan penggunaan *device any gadget* (selain *smartphone/laptop/device/tablet*) sebesar 0.43%

4.4.7. Lama Waktu Menghabiskan Waktu Untuk Masing-Masing Aktivitas (*Time Spent On Online Class, Time Spent On Self-Study, Time Spent On Fitness, Time Spent On Sleep, Time Spent On Social Media, Dan Time Spent On TV*)



Gambar 4.38. Visualisasi Lama Waktu yang Dihabiskan dalam Beraktivitas

Penjelasan :

Kode di atas digunakan untuk memvisualisasikan tiap-tiap waktu yang dihabiskan oleh responden dari berbagai aktivitas dengan nama kolom yang telah diinisiasi dalam variabel `colom_to_visualize` dan divisualisasikan dalam satu *output* dengan menggunakan *loop* untuk setiap aktivitas dengan menghitung jumlah waktu yang dipergunakan oleh responden dalam kategori tertentu.

Untuk pembahasan lebih lanjut akan dibahas lebih lanjut, seperti berikut.

a) Time Spent on Online Class

Interpretasi Hasil :

Total Waktu (Jam)	Frekuensi (Orang)	Total Waktu (Jam)	Frekuensi (Orang)
0.0	62	3.5	1
0.5	6	4.0	145
0.7	1	5.0	98
0.75	7	5.5	2
0.8	1	6.0	69
1.0	57	7.0	27
1.5	3	7.5	1
2.0	110	8.0	26
2.5	2	9.0	4
3.0	93	10.0	2

Tabel 4.1. Tabel Time Spent on Online Class

Berdasarkan analisis yang dilakukan, sebagian besar responden menghabiskan waktu belajar *online* selama 4 jam dengan total responden 145 orang, 4 jam dengan total responden 110 orang, dan 5 jam dengan total responden sebanyak 98 orang. Dapat

disimpulkan bahwa sebagian besar menghabiskan waktu untuk belajar *online* selama 2 – 5 jam, dengan total responden 449 responden atau 62,62% responden.

b) *Time Spent On Self-Study*

Interpretasi Hasil :

Total Waktu (Jam)	Frekuensi (Orang)	Total Waktu (Jam)	Frekuensi (Orang)
0.0	40	4.0	87
0.5	4	4.5	1
1.0	109	5.0	54
1.2	1	6.0	27
1.5	6	7.0	6
2.0	224	8.0	17
2.3	1	9.0	1
2.5	6	10.0	5
3.0	122	11.0	1
3.5	3	12.0	2

Tabel 4.2. Tabel Time Spent on Self-Study

Berdasarkan analisis yang dilakukan, waktu yang dihabiskan responden untuk belajar mandiri, sebagian besar dilakukan selama 1 jam dengan total responden sebanyak 109 orang, 2 jam sebanyak 224 orang, dan 3 jam sebanyak 122 orang. Dapat disimpulkan bahwa sebagian besar responden menghabiskan waktu untuk belajar mandiri selama 1 – 3 jam dengan total responden 469 responden atau sebanyak 65,41% responden.

c) *Time Spent On Fitness*

Interpretasi Hasil :

Total Waktu (Jam)	Frekuensi (Orang)	Total Waktu (Jam)	Frekuensi (Orang)
0.0	240	1.0	329
0.25	4	2.0	80
0.3	3	2.5	1
0.5	40	3.0	17
0.6	2	4.0	1

Tabel 4.3. Time Spent on Fitness

Berdasarkan analisis yang dilakukan pada kolom ‘*Time Spent On Fitness*’ didapatkan bahwa sebagian besar responden menghabiskan waktu untuk aktivitas fitness. Didapat bahwa jumlah responden yang tidak menghabiskan waktu untuk aktivitas fitness sebanyak 240 orang , sedangkan yang melakukan aktivitas fitness sebanyak 477 orang atau 66,53% responden.

d) *Time Spent On Sleep*

Interpretasi Hasil :

Total Waktu (Jam)	Frekuensi (Orang)	Total Waktu (Jam)	Frekuensi (Orang)
4.0	3	8.5	1
5.0	25	9.0	90
6.0	104	10.0	61
6.5	1	11.0	15
7.0	144	12.0	27

7.5	3	13.0	4
8.0	237	15.0	1
8.3	1	-	-

Tabel 4.4. Tabel Spent on Sleep

Berdasarkan analisis yang dilakukan pada kolom ‘Time Spent On Sleep’, didapatkan bahwa sebagian besar responden menghabiskan waktu untuk tidur selama 8 jam sebanyak 237 orang, 7 jam sebanyak 144 orang, dan 6 jam sebanyak 104 orang. Sedangkan waktu ideal yang diperlukan pelajar adalah 7 – 10 jam sehari, maka jumlah responden yang telah memenuhi waktu ideal untuk tidur sebanyak 537 responden atau 74,90% responden.

e) *Time Spent On Social Media*

Interpretasi Hasil :

Total Waktu (Jam)	Frekuensi (Orang)	Total Waktu (Jam)	Frekuensi (Orang)
0.0	30	2.0	192
0.1	2	3.0	92
0.15	1	3.5	1
0.25	1	4.0	66
0.3	1	5.0	39
0.4	2	6.0	11
0.5	21	7.0	6
0.6	1	8.0	3
1,0	224	9.0	2
1.3	1	10.0	10
1.5	11	-	-

Tabel 4.5. Tabel Spent on Social Medua

Berdasarkan analisis yang dilakukan pada kolom ‘Time Spent On Social Media’, didapatkan bahwa waktu yang dihabiskan oleh responden untuk melakukan aktivitas sosial media yaitu selama 1 jam sebanyak 224 responden dan 2 jam sebanyak 192 responden. Waktu yang ideal dalam melakukan aktivitas sosial media yaitu selama 2 jam. Maka dapat disimpulkan bahwa total responden yang memenuhi waktu ideal sebanyak 487 responden atau 67,92 % dan sebanyak 230 responden atau 32,08% melebihi waktu ideal.

f) *Time Spent On TV*

Interpretasi Hasil :

Total Waktu (Jam)	Frekuensi (Orang)	Total Waktu (Jam)	Frekuensi (Orang)
0.0	241	2.5	2
0.1	4	3.0	37
0.3	11	3.5	1
0.4	1	4.0	13
0.5	51	4.5	2
0.6	1	5.0	5
1.0	199	6.0	5
1.5	12	8.0	1
2	130	15.0	1

Tabel 4 6. Time Spent on TV

Berdasarkan hasil analisis kolom ‘*Time Spent On TV*’ didapatkan bahwa sebagian besar responden menghabiskan waktu untuk menonton televisi selama 1 jam sebanyak 199 responden dan selama 2 jam sebanyak 130 responden, adapun sebanyak 241 responden tidak menghabiskan waktunya untuk menonton televisi.

4.4.8. Frekuensi Responden Makan Dalam Sehari



Gambar 4.39. *Numbers Of Meals Per-Day*

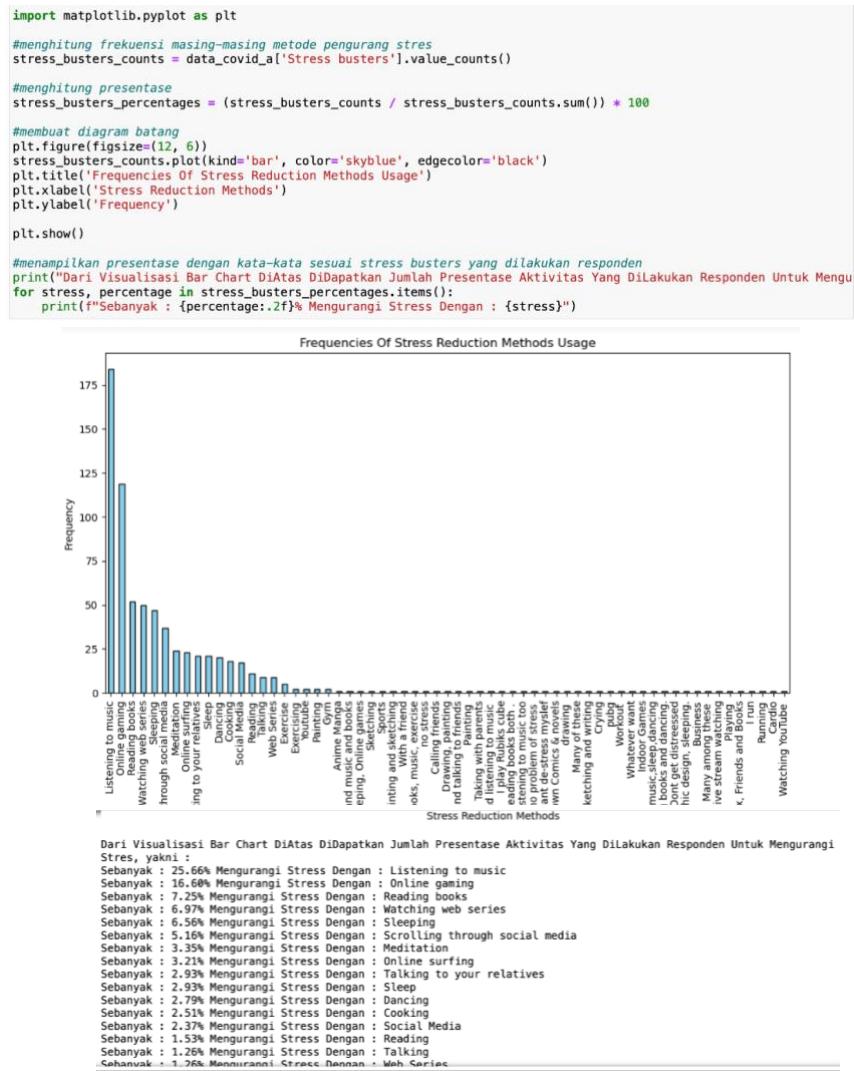
Penjelasan :

Kode di atas digunakan untuk memvisualisasikan dan mengetahui jumlah makan responden di setiap harinya. Dengan menggunakan kolom ‘*Numbers Of Meals Per-Day*’ dapat diketahui dan diperhitungkan masing-masing jumlah makanan responden selama satu hari dan dilakukan perhitungan persentase jumlah makan responden dalam satu hari. Visualisasi *bar chart* dipergunakan untuk memperlihatkan distribusi frekuensi jumlah makanan per-hari, dengan warna yang membedakan jumlah makan setiap responden per hari.

Interpretasi Hasil :

Output dari kode di atas adalah visualisasi *bar chart* dengan sebesar kurang dari 50 responden (4,32%) makan sebanyak 1 kali per-hari, lebih dari 150 dan kurang dari 200 responden (24,27%) makan sebanyak 2 kali per-hari, 350 responden (49,51%) makan sebanyak 3 kali per-hari, 150 responden (20,64%) makan sebanyak 4 kali per-hari, kurang dari 50 responden (0,70%) makan 5 kali per-hari, kurang dari 50 responden (0,14%) makan sebanyak 6 kali perhari, dan kurang dari 50 responden (0,42%) responden makan sebanyak 7 kali-perhari.

4.4.9. Cara Mereka Mengurangi Tingkat Stress



Gambar 4.40. Visualisasi Stress Busters

Penjelasan :

Kode di atas bertujuan untuk memvisualisasikan dan mengetahui persentase dan cara untuk mengurangi rasa stres selama pandemi COVID-19. Dengan menggunakan kolom ‘*Stress Busters*’ yang belum di encoding hal ini dipergunakan agar hasil yang dijabarkan lebih informatif. Selanjutnya, dilakukan perhitungan persentase, jumlah dan penghilang stres.

Interpretasi Hasil :

Output hasil visualisasi *bar chart*, dapat diinterpretasikan bahwa aktivitas yang paling umum adalah mendengarkan musik dengan persentase sebanyak 25.66%, diikuti oleh bermain *game online* dengan persentase 16.60%. Terdapat pula beragam metode lainnya seperti membaca buku, menonton *web series*, tidur, menggulir media sosial, meditasi, berselancar *online*, berbicara dengan keluarga, menari, memasak, dan banyak lagi, masing-masing dengan persentase yang berbeda.

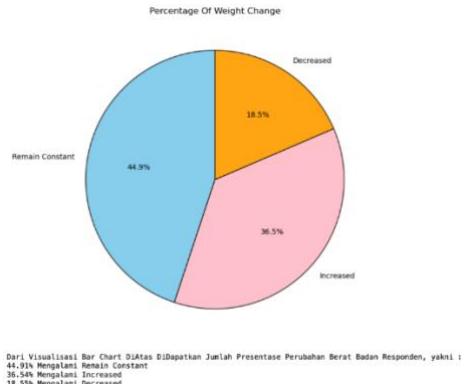
4.4.10. Dampak Covid-19 Dalam Mempengaruhi Berat Badan

```
: import matplotlib.pyplot as plt
#menghitung frekuensi perubahan berat badan
weight_counts = data_covid_a['Change in your weight'].value_counts()

#menghitung persentase
weight_percentages = (weight_counts / weight_counts.sum()) * 100

#membuat diagram pie
fig, ax = plt.subplots(figsize=(8, 8))
wedges, texts, autotexts = ax.pie(weight_percentages, labels=weight_percentages.index, autopct='%.1f%%', startangle=90)
# Menambahkan garis tepi pada setiap "wedge"
for wedge in wedges:
    wedge.set_edgecolor('black')
ax.set_title('Percentage Of Weight Change')
plt.show()

# Menampilkan persentase dengan kata-kata sesuai perubahan berat badan
print("Dari Visualisasi Bar Chart Diatas Didapatkan Jumlah Presentase Perubahan Berat Badan Responen, yakni : ")
for weight, percentage in weight_percentages.items():
    print(f"{percentage:.2f}% Mengalami {weight}")
```



Gambar 4.41. Visualisasi *Change in Your Weight*

Penjelasan :

Kode di atas bertujuan untuk mengetahui persentase dan memvisualisasikan data perubahan berat badan responden selama pandemi COVID-19. Dengan menggunakan kolom '*Change In Your Weight*' melakukan perhitungan frekuensi kemunculan di setiap nilai perubahan berat badan, kemudian diubah menjadi persentase dan dilanjutkan dengan visualisasi *pie chart* untuk mengetahui proporsi perubahan berat badan.

Interpretasi Hasil :

Output dari kode di atas adalah sebesar 44,91% responden tidak mengalami perubahan berat badan, sebesar 36,54% responden mengalami kenaikan berat badan, dan 18,55% responden mengalami penurunan berat badan.

4.4.11. Sosial Media yang Sering Digunakan

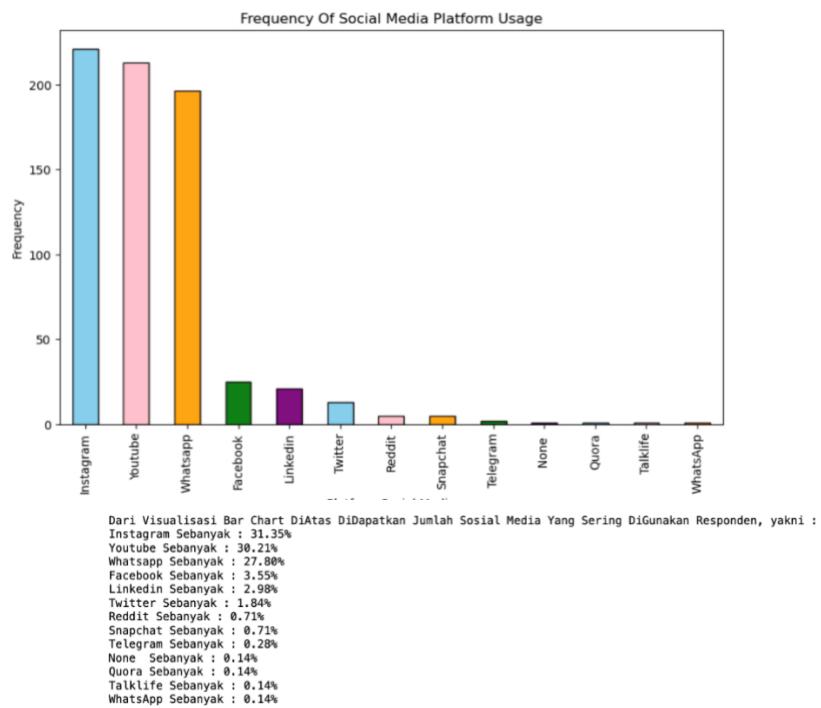
```
: import matplotlib.pyplot as plt
#menghitung frekuensi penggunaan platform sosial media
sosmed_counts = data_covid_a['Preferred social media platform'].value_counts()

#menghitung persentase
sosmed_percentages = (sosmed_counts / sosmed_counts.sum()) * 100

#membuat bar chart
plt.figure(figsize=(10, 6))
sosmed_counts.plot(kind='bar', color=['skyblue', 'pink', 'orange', 'green', 'purple'], edgecolor='black')
plt.title('Frequency Of Social Media Platform Usage')
plt.xlabel('Platform Sosial Media')
plt.ylabel('Frequency')

plt.show()

#menampilkan persentase dengan kata-kata sesuai penggunaan sosial media yang sering digunakan responden
print("Dari Visualisasi Bar Chart Diatas Didapatkan Jumlah Sosial Media Yang Sering Digunakan Responen, yakni : ")
for sosmed, percentage in sosmed_percentages.items():
    print(f"{percentage:.2f} Sebanyak : {percentage:.2f}%")
```



Gambar 4.42. Preferred Social Media Platform

Penjelasan :

Kode di atas bertujuan untuk mengetahui dan menganalisis data penggunaan sosial media oleh responden dengan visualisasi *bar chart* yang dapat mengetahui jumlah responden yang menggunakan sosial media tersebut beserta persentasenya. Dengan menggunakan kolom ‘*Preferred Social Media Platform*’. Dilakukan perhitungan frekuensi kemunculan setiap *platform* media sosial dihitung dan hasilnya diubah menjadi persentase.

Interpretasi Hasil :

Output dari kode di atas adalah visualisasi *bar chart* yang menunjukkan bahwa *Instagram* menjadi *platform* media sosial yang paling sering digunakan oleh responden dengan persentase sebesar lebih dari 200 (31,35%), 200 responden (30,21%) menggunakan *YouTube*, 180 responden (27,80%) menggunakan *WhatsApp*, kemudian dilanjutkan dengan persentase penggunaan *Facebook*, *LinkedIn*, *Twitter*, *Reddit*, *Snapchat*, *Telegram*, *None*, *Quora*, dan *Talklife* juga tercatat dengan persentase yang lebih rendah, menunjukkan variasi penggunaan sosial media di kalangan responden.

4.5. Normalisasi Data

Pada tahapan ini dilakukan normalisasi data, pada semua kolom kecuali ‘ID’ dengan menggunakan nilai *z-score*. Hal ini bertujuan untuk memastikan konsistensi skala antar variabel numerik dalam *dataset* dan menghilangkan dampak *outlier* pada analisis statistik, dan meningkatkan performa model, untuk penggunaan metode yang *sensitive* terhadap skala seperti K-Means.

4.5.1. DataFrame Setelah Normalisasi

```

from sklearn.preprocessing import StandardScaler

#kolom yang akan dinormalisasi (misalnya, kolom numerik)
numeric_columns = data_covid_fitted.select_dtypes(include=['float64', 'int64']).columns

#objek StandardScaler
scaler = StandardScaler()

#normalisasi Z-score pada data numerik
data_covid_fitted[numeric_columns] = scaler.fit_transform(data_covid_fitted[numeric_columns])

#DataFrame setelah normalisasi
print("DataFrame Setelah Normalisasi Z-score : ")
data_covid_fitted

```

ID	Age of Subject	Time spent on Online Class	Time spent on self study	Time spent on fitness	Time spent on sleep	Time spent on social media	Time spent on TV	Number of meals per day	Health issue during lockdown	Time utilized	Region of residence	Rating of Online Class experience	Medium for online class	P	
2	R3	0.970873	1.616215	0.096006	-1.086971	-1.191808	-0.152118	-0.845875	0.103429	-0.359352	-0.973843	-0.742932	1.123301	-0.860249	-1.
3	R4	0.970873	-0.249286	-0.427462	0.281284	-1.191808	1.542374	-0.845875	0.103429	-0.359352	-0.973843	-0.742932	1.123301	0.225664	-1.
6	R7	0.605650	-0.715662	-0.427462	0.281284	-1.008884	0.977543	-0.845875	0.103429	-0.359352	-0.973843	-0.742932	1.123301	0.225664	-1.
7	R8	0.605650	-0.715662	-0.950930	0.281284	1.276495	1.542374	-0.845875	0.103429	2.782787	-0.973843	1.346018	1.123301	2.397489	-1.
9	R10	0.970873	-1.648412	-0.950930	-0.402843	0.042343	1.542374	1.530351	0.103429	2.782787	-0.973843	1.346018	1.123301	-0.860249	-1.
...
1177	R1191	-1.950916	-0.249286	0.619473	0.281284	0.042343	-0.716949	0.738276	0.103429	-0.359352	1.026860	-0.742932	-0.050761	0.225664	-1.
1178	R1192	-1.220468	1.149839	0.619473	0.281284	0.659419	-0.716949	-0.053799	1.280545	-0.359352	1.026860	-0.742932	-1.224824	0.225664	0.
1179	R1193	-1.585692	0.217089	-1.474397	-0.402843	0.042343	0.412713	0.738276	1.280545	-0.359352	-0.973843	-0.742932	-1.224824	0.225664	0.
1180	R1194	-1.220468	0.683464	0.357739	0.281284	0.042343	-0.999365	-0.053799	1.280545	-0.359352	1.026860	-0.742932	-0.637793	-0.860249	0.
1181	R1195	-1.585692	0.683464	-0.427462	-0.402843	-0.574733	-0.716949	-0.053799	0.103429	-0.359352	1.026860	-0.742932	-0.050761	2.397489	0.

717 rows × 18 columns

Gambar 4.43. Normalisasi Data

Penjelasan :

Kode di atas digunakan untuk mempersiapkan data numerik dalam sebuah *DataFrame*. Dengan menggunakan objek *StandardScaler* dari *scikit-learn* dan menghitung kolom numerik untuk diubah ke skala yang berpusat di sekitar rata-rata dengan deviasi standar yang diperhitungkan dan melakukan standarisasi data numerik sehingga memiliki rata-rata 0 dan standar deviasi 1, menghilangkan bias dan memudahkan interpretasi serta meningkatkan akurasi analisis statistik.

Interpretasi Hasil :

Output dari kode di atas adalah nilai-nilai dari setiap kolom yang telah di normalisasi dan nantinya akan dipergunakan untuk melakukan pengujian hipotesis, pengembangan model dan evaluasi, serta visualisasi interaktif.

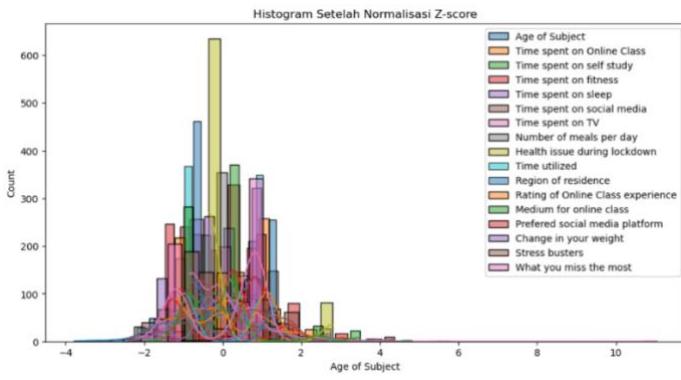
4.5.2. Visualisasi Data Normalisasi Visualisasi Histogram

```

import seaborn as sns

#visualisasi Histogram
plt.figure(figsize=(12, 6))
for column in numeric_columns:
    sns.histplot(data_covid_fitted[column], kde=True, label=column)
plt.title('Histogram Setelah Normalisasi Z-score')
plt.legend()
plt.show()

```



Gambar 4.44. Visualisasi Z-Score

Penjelasan :

Kode di atas digunakan untuk membuat visualisasi histogram dari kolom-kolom numerik dalam *DataFrame*. Dengan menggunakan *looping for* digunakan untuk membuat histogram dari nilai di setiap kolom yang telah di normalisasi. Dengan tujuan untuk memberikan gambaran distribusi data setelah normalisasi *z-score*.

Interpretasi Hasil:

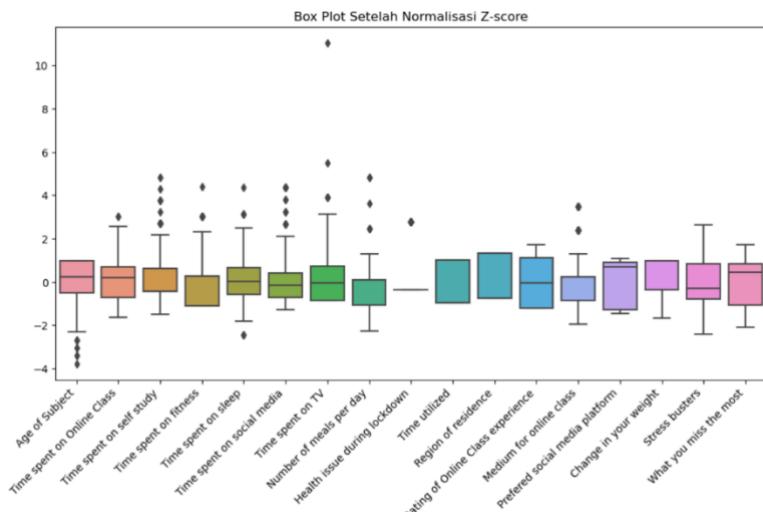
Output dari histogram tersebut memberikan hasil tentang semua kolom yang telah di normalisasi, di mana dapat merepresentasikan hubungan antara satu kolom dengan yang lainnya dan memberikan pola distribusi dari setiap kolom.

4.5.3. Visualisasi Data Normalisasi Dengan Box Plot

```
#visualisasi box plot
plt.figure(figsize=(12, 6))
sns.boxplot(data=data_covid_fitted[numeric_columns])

# Rotate x-axis tick labels
plt.xticks(rotation=45, ha='right') # You can adjust the rotation angle and alignment as needed

plt.title('Box Plot Setelah Normalisasi Z-score')
plt.show()
```



Gambar 4.45. Visualisasi Normalisasi

Penjelasan :

Kode tersebut digunakan untuk membuat *box plot* dari data yang telah dinormalisasi menggunakan *z-score*. *Box plot* tersebut menunjukkan distribusi statistik deskriptif seperti kuartil, median, dan pencilan dalam satu tampilan grafis.

Interpretasi Hasil:

Output dari kode dia tas adalah *box plot* menunjukkan bahwa setelah normalisasi *z-score*, distribusi data menjadi lebih seragam dan merepresentasikan pola distribusi masing-masing kolom.

4.6. Pengujian Hipotesis Analisis Korelasi Bivariate Dan Multivariate

Pada tahapan ini dilakukan proses pengujian hipotesis dengan menggunakan analisis korelasi *bivariate* dan *multivariate* pada beberapa kolom. Proses pengujian hipotesis ini bertujuan untuk memahami pola atau hubungan antara data di suatu kolom dengan kolom yang lainnya. Selain itu, dalam proses analisis ini pengujian hipotesis bertujuan untuk menentukan faktor-faktor yang mungkin mempengaruhi dan mengevaluasi hubungan antara variabel-variabel tertentu.

Dilakukan proses analisis korelasi *bivariate* yang dipergunakan untuk mengetahui hubungan antar dua variabel. Dan dilakukan proses analisis korelasi *multivariate* yang dipergunakan untuk mengetahui hubungan lebih dari dua variabel yang akan dijabarkan lebih lanjut, seperti berikut.

4.6.1. Korelasi antara kolom ‘Age of Subject’ dan ‘Time spent on Online Class’

```
from scipy.stats import pearsonr
import seaborn as sns
import matplotlib.pyplot as plt

#menginisiasi column 1 dan column 2
column1, column2 = 'Age of Subject', 'Time spent on Online Class'

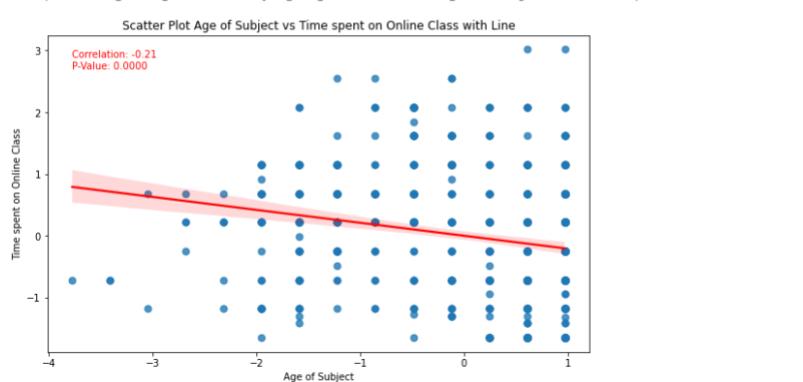
#menghitung Korelasi Pearson dan P-Value
correlation, p_value = pearsonr(data_covid_fitted[column1], data_covid_fitted[column2])

#menampilkan nilai korelasi dan nilai p-value
print(f"Korelasi Pearson: {correlation:.2f}, p-value: {p_value:.4f}")

#menentukan hipotesis
print("Kesimpulan berdasarkan P-Value :")
print("Hipotesis nol (H0) dapat ditolak." if p_value < 0.05 else "Tidak cukup bukti untuk menolak hipotesis nol (H0).")

#kesimpulan berdasarkan korelasi
print("Kesimpulan berdasarkan Nilai Korelasi :")
print(f"Terdapat hubungan {'positif' if correlation > 0 else 'negatif' if correlation < 0 else 'tidak terdapat'} linear")

#membuat scatter plot
plt.figure(figsize=(10, 6))
sns.regplot(x=column1, y=column2, data=data_covid_fitted, scatter_kws={'s': 50}, line_kws={'color': 'red'})
#menampilkan nilai korelasi dan p-value sebagai teks pada plot
plt.text(min(data_covid_fitted[column1]), max(data_covid_fitted[column2]), f'Correlation: {correlation:.2f}\nP-Value: {p_value:.4f}', color='red')
plt.title(f'Scatter Plot {column1} vs {column2} with Line')
plt.xlabel(column1)
plt.ylabel(column2)
plt.show()
```



Gambar 4.46. Korelasi Antara Kolom Age of Subject dan Time spent on Online Class
Interpretasi Hasil :

Hasilnya menunjukkan hubungan linear negatif yang signifikan antara kedua variabel dengan korelasi *Pearson* sebesar -0.21 dan *P-Value* sebesar 0.000 . *Scatter plot* dengan garis regresi memvisualisasikan hubungan negatif tersebut, di mana titik-titik

data tersebar di sekitar garis regresi yang menurun. Berdasarkan P-Value, terdapat hubungan yang signifikan. Kesimpulan dari analisis ini adalah adanya hubungan negatif yang signifikan antara 'Age of Subject' dan 'Time Spent on Online Class'.

4.6.2. Korelasi antara kolom 'Age of Subject' dan 'Time spent on self study'

```

import pandas as pd
from scipy.stats import pearsonr
import seaborn as sns
import matplotlib.pyplot as plt

#menginisiasi column 1 dan column 2
column1, column2 = 'Age of Subject', 'Time spent on self study'

#menghitung Korelasi Pearson dan P-Value
correlation, p_value = pearsonr(data_covid_fitted[column1], data_covid_fitted[column2])

#menampilkan nilai korelasi dan nilai p-value
print(f"Korelasi Pearson: {correlation:.2f}, p-value: {p_value:.4f}")

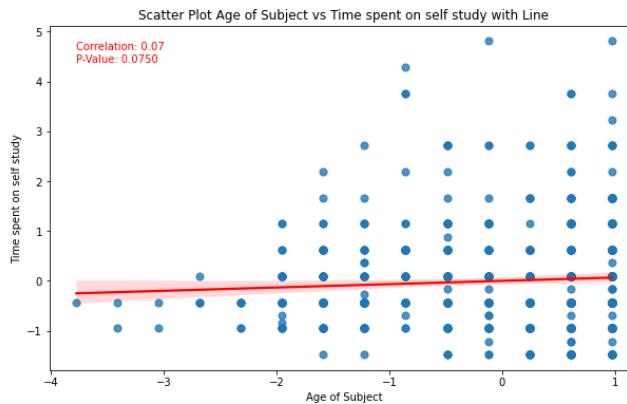
#menentukan hipotesis
print("Kesimpulan berdasarkan P-Value :")
print("Hipotesis nol ( $H_0$ ) dapat ditolak." if p_value < 0.05 else "Tidak cukup bukti untuk menolak hipotesis nol ( $H_0$ ).")

#kesimpulan berdasarkan korelasi
print("Kesimpulan berdasarkan Nilai Korelasi :")
print("Terdapat hubungan {'positif' if correlation > 0 else 'negatif' if correlation < 0 else 'tidak terdapat'} linear")

#membuat scatter plot
plt.figure(figsize=(10, 6))
sns.replot(x=column1, y=column2, data=data_covid_fitted, scatter_kws={'s': 50}, line_kws={'color': 'red'})
#menampilkan nilai korelasi dan p-value sebagai teks pada plot
plt.text(min(data_covid_fitted[column1]), max(data_covid_fitted[column2]), f'Correlation: {correlation:.2f}\n\nP-Value: {p_value:.4f}', color='red')
plt.title(f'Scatter Plot {column1} vs {column2} with Line')
plt.xlabel(column1)
plt.ylabel(column2)
plt.show()

```

Korelasi Pearson: 0.07, p-value: 0.0750
 Kesimpulan berdasarkan P-Value :
 Tidak cukup bukti untuk menolak hipotesis nol (H_0).
 Kesimpulan berdasarkan Nilai Korelasi :
 Terdapat hubungan positif linear yang signifikan antara Age of Subject dan Time spent on self study.



Gambar 4.47. Korelasi antara kolom 'Age of Subject' dan 'Time spent on self study'

Interpretasi Hasil :

Hasilnya menunjukkan hubungan linear positif yang signifikan antara kedua variabel dengan korelasi Pearson sebesar 0.07 dan P-Value sebesar 0.0750. berdasarkan P-Value, tidak terdapat hubungan signifikan. Scatter plot dengan garis regresi memvisualisasikan hubungan positif tersebut, di mana titik-titik data tersebar di sekitar garis regresi yang menurun. Kesimpulan dari analisis ini adalah adanya hubungan positif yang signifikan antara 'Age of Subject' dan 'Time Spent on Online Class'.

4.6.3. Korelasi antara kolom ‘Age of Subject’ dan ‘Time spent on fitness’

```

import pandas as pd
from scipy.stats import pearsonr
import seaborn as sns
import matplotlib.pyplot as plt

#menginisiasi column 1 dan column 2
column1, column2 = 'Age of Subject', 'Time spent on fitness'

#menghitung Korelasi Pearson dan P-Value
correlation, p_value = pearsonr(data_covid_fitted[column1], data_covid_fitted[column2])

#menampilkan nilai korelasi dan nilai p-value
print(f"Korelasi Pearson: {correlation:.2f}, p-value: {p_value:.4f}")

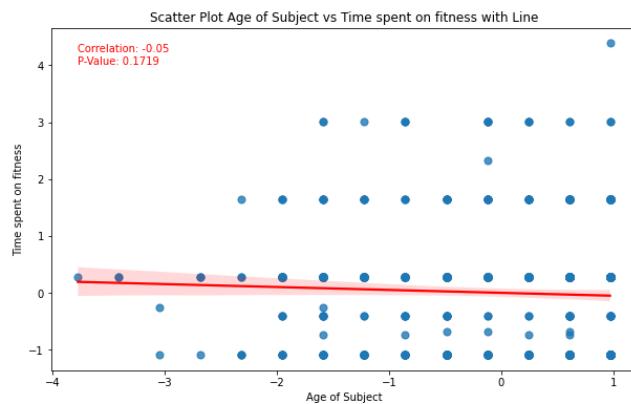
#menentukan hipotesis
print("Kesimpulan berdasarkan P-Value : ")
print("Hipotesis nol (H0) dapat ditolak." if p_value < 0.05 else "Tidak cukup bukti untuk menolak hipotesis nol (H0).")

#kesimpulan berdasarkan korelasi
print("Kesimpulan berdasarkan Nilai Korelasi :")
print(f"Terdapat hubungan {'positif' if correlation > 0 else 'negatif' if correlation < 0 else 'tidak terdapat'} linear")

#membuat scatter plot
plt.figure(figsize=(10, 6))
sns.regplot(x=column1, y=column2, data=data_covid_fitted, scatter_kws={'s': 50}, line_kws={'color': 'red'})
#menampilkan nilai korelasi dan p-value sebagai teks pada plot
plt.text(min(data_covid_fitted[column1]), max(data_covid_fitted[column2]), f'Correlation: {correlation:.2f}\nP-Value: {p_value:.4f}', color='red')
plt.title('Scatter Plot {column1} vs {column2} with Line')
plt.xlabel(column1)
plt.ylabel(column2)
plt.show()

```

Korelasi Pearson: -0.05, p-value: 0.1719
 Kesimpulan berdasarkan P-Value :
 Tidak cukup bukti untuk menolak hipotesis nol (H0).
 Kesimpulan berdasarkan Nilai Korelasi :
 Terdapat hubungan negatif linear yang signifikan antara Age of Subject dan Time spent on fitness.



Gambar 4.48. Korelasi antara kolom ‘Age of Subject’ dan ‘Time spent on fitness’

Interpretasi Hasil :

Hasilnya menunjukkan hubungan linear negatif yang signifikan antara kedua variabel dengan korelasi Pearson sebesar -0.05 dan P-Value sebesar 0.1719. Scatter plot dengan garis regresi memvisualisasikan hubungan negatif tersebut, di mana titik-titik data tersebar di sekitar garis regresi yang menurun. Kesimpulan dari analisis ini adalah adanya hubungan negatif yang signifikan antara ‘Age of Subject’ dan ‘Time Spent on Fitness’.

4.6.4. Korelasi antara kolom ‘Age of Subject’ dan ‘Time spent on sleep’

```

import pandas as pd
from scipy.stats import pearsonr
import seaborn as sns
import matplotlib.pyplot as plt

#menginisiasi column 1 dan column 2
column1, column2 = 'Age of Subject', 'Time spent on sleep'

#menghitung Korelasi Pearson dan P-Value
correlation, p_value = pearsonr(data_covid_fitted[column1], data_covid_fitted[column2])

#menampilkan nilai korelasi dan nilai p-value
print(f'Korelasi Pearson: {correlation:.2f}, p-value: {p_value:.4f}')

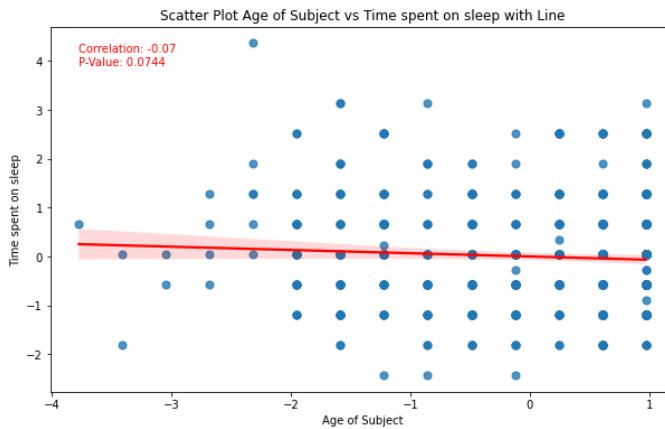
#menentukan hipotesis
print("Kesimpulan berdasarkan P-Value : ")
print("Hipotesis nol (H0) dapat ditolak." if p_value < 0.05 else "Tidak cukup bukti untuk menolak hipotesis nol (H0).")

#kesimpulan berdasarkan korelasi
print("Kesimpulan berdasarkan Nilai Korelasi :")
printf("Terdapat hubungan {'positif' if correlation > 0 else 'negatif' if correlation < 0 else 'tidak terdapat'} linear")

#membuat scatter plot
plt.figure(figsize=(10, 6))
sns.regressionplot(x=column1, y=column2, data=data_covid_fitted, scatter_kws={'s': 50}, line_kws={'color': 'red'})
#memplikon nilai korelasi dan p-value sebagai teks pada plot
plt.text(min(data_covid_fitted[column1]), max(data_covid_fitted[column2]), f'Correlation: {correlation:.2f}\nP-Value: {p_value:.4f}', color='red')
plt.title(f'Scatter Plot {column1} vs {column2} with Line')
plt.xlabel(column1)
plt.ylabel(column2)

#tampilkan
Korelasi Pearson: -0.07, p-value: 0.0744
Kesimpulan berdasarkan P-Value :
Tidak cukup bukti untuk menolak hipotesis nol (H0).
Kesimpulan berdasarkan Nilai Korelasi :
Terdapat hubungan negatif linear yang signifikan antara Age of Subject dan Time spent on sleep.

```



Gambar 4.49. Korelasi antara kolom ‘Age of Subject’ dan ‘Time spent on sleep’

Interpretasi Hasil :

Hasilnya menunjukkan hubungan linear negatif yang signifikan antara kedua variabel dengan korelasi Pearson sebesar -0.07 dan P-Value sebesar 0.0744. Berdasarkan P-Value, terdapat hubungan yang signifikan. Scatter plot dengan garis regresi memvisualisasikan hubungan negatif tersebut, di mana titik-titik data tersebar di sekitar garis regresi yang menurun. Kesimpulan dari analisis ini adalah adanya hubungan negatif yang signifikan antara ‘Age of Subject’ dan ‘Time Spent on sleep’.

4.6.5. Korelasi antara kolom ‘Age of Subject’ dan ‘Time spent on social media’

```

import pandas as pd
from scipy.stats import pearsonr
import seaborn as sns
import matplotlib.pyplot as plt

#menginisiasi column 1 dan column 2
column1, column2 = 'Age of Subject', 'Time spent on social media'

#menghitung Korelasi Pearson dan P-Value
correlation, p_value = pearsonr(data_covid_fitted[column1], data_covid_fitted[column2])

#menampilkan nilai korelasi dan nilai p-value
print(f'Korelasi Pearson: {correlation:.2f}, p-value: {p_value:.4f}')

#menentukan hipotesis
print("Kesimpulan berdasarkan P-Value : ")
print("Hipotesis nol ( $H_0$ ) dapat ditolak." if p_value < 0.05 else "Tidak cukup bukti untuk menolak hipotesis nol ( $H_0$ ).")

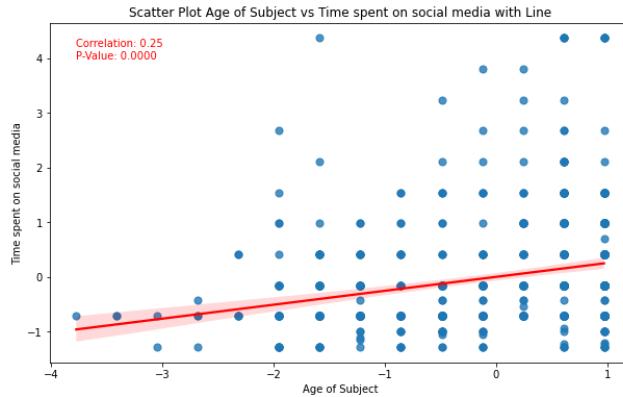
#kesimpulan berdasarkan korelasi
print("Kesimpulan berdasarkan Nilai Korelasi :")
print(f'Terdapat hubungan {{"positif" if correlation > 0 else "negatif" if correlation < 0 else "tidak terdapat"} linear')

#membuat scatter plot
plt.figure(figsize=(10, 6))
sns.regplot(x=column1, y=column2, data=data_covid_fitted, scatter_kws={'s': 50}, line_kws={'color': 'red'})
#memperlihatkan nilai korelasi dan p-value sebagai teks pada plot
plt.text(min(data_covid_fitted[column1]), max(data_covid_fitted[column2]), f'Correlation: {correlation:.2f}\nP-Value: {p_value:.4f}', color='red')
plt.title(f'Scatter Plot {column1} vs {column2} with Line')
plt.xlabel(column1)
plt.ylabel(column2)

plt.show()

Korelasi Pearson: 0.25, p-value: 0.0000
Kesimpulan berdasarkan P-Value :
Hipotesis nol ( $H_0$ ) dapat ditolak.
Kesimpulan berdasarkan Nilai Korelasi :
Terdapat hubungan positif linear yang signifikan antara Age of Subject dan Time spent on social media.

```



Gambar 4.50. Korelasi antara kolom ‘Age of Subject’ dan ‘Time spent on social media’

Interpretasi Hasil :

Hasilnya menunjukkan hubungan linear positif yang signifikan antara kedua variabel dengan korelasi *Pearson* sebesar 0.25 dan *P-Value* sebesar 0.000. Berdasarkan *P-Value*, tidak terdapat hubungan yang signifikan. Scatter plot dengan garis regresi memvisualisasikan hubungan positif tersebut, di mana titik-titik data tersebar di sekitar garis regresi yang menaik. Kesimpulan dari analisis ini adalah adanya hubungan positif yang signifikan antara ‘Age of Subject’ dan ‘Time Spent on social media’.

4.6.6. Korelasi antara kolom ‘Age of Subject’ dan ‘Time spent on TV’

```

import pandas as pd
from scipy.stats import pearsonr
import seaborn as sns
import matplotlib.pyplot as plt

#menginisiasi column 1 dan column 2
column1, column2 = 'Age of Subject', 'Time spent on TV'

#menghitung Korelasi Pearson dan P-Value
correlation, p_value = pearsonr(data_covid_fitted[column1], data_covid_fitted[column2])

#menampilkan nilai korelasi dan nilai p-value
print(f"Korelasi Pearson: {correlation:.2f}, p-value: {p_value:.4f}")

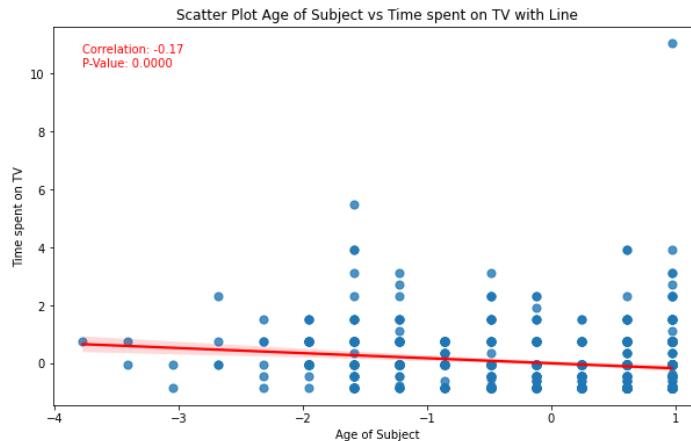
#menentukan hipotesis
print("Kesimpulan berdasarkan P-Value : ")
print("Hipotesis nol (H0) dapat ditolak." if p_value < 0.05 else "Tidak cukup bukti untuk menolak hipotesis nol (H0).")

#kesimpulan berdasarkan korelasi
print("Kesimpulan berdasarkan Nilai Korelasi :")
print(f"Terdapat hubungan {'positif' if correlation > 0 else 'negatif' if correlation < 0 else 'tidak terdapat'} linear")

#membuat scatter plot
plt.figure(figsize=(10, 6))
sns.regplot(x=column1, y=column2, data=data_covid_fitted, scatter_kws={'s': 50}, line_kws={'color': 'red'})
# Menampilkan nilai korelasi dan p-value sebagai teks pada plot
plt.text(min(data_covid_fitted[column1]), max(data_covid_fitted[column2]), f'Correlation: {correlation:.2f}\nP-Value: {p_value:.4f}', color='red')
plt.xlabel(column1)
plt.ylabel(column2)
plt.show()

Korelasi Pearson: -0.17, p-value: 0.0000
Kesimpulan berdasarkan P-Value :
Hipotesis nol (H0) dapat ditolak.
Kesimpulan berdasarkan Nilai Korelasi :
Terdapat hubungan negatif linear yang signifikan antara Age of Subject dan Time spent on TV.

```



Gambar 4.51. Korelasi antara kolom ‘Age of Subject’ dan ‘Time spent on TV’

Interpretasi Hasil :

Hasilnya menunjukkan hubungan linear negatif yang signifikan antara kedua variabel dengan korelasi *Pearson* sebesar -0.17 dan *P-Value* sebesar 0.000 . Berdasarkan *P-Value*, terdapat hubungan yang signifikan. Scatter plot dengan garis regresi memvisualisasikan hubungan negatif tersebut, di mana titik-titik data tersebar di sekitar garis regresi yang menurun. Kesimpulan dari analisis ini adalah adanya hubungan negatif yang signifikan antara ‘Age of Subject’ dan ‘Time Spent on TV’.

4.6.7. Korelasi antara kolom ‘Age of Subject’ dan ‘Number of meals per day’

```

import pandas as pd
from scipy.stats import pearsonr
import seaborn as sns
import matplotlib.pyplot as plt

#menginisiasi column 1 dan column 2
column1, column2 = 'Age of Subject', 'Number of meals per day'

#menghitung Korelasi Pearson dan P-Value
correlation, p_value = pearsonr(data_covid_fitted[column1], data_covid_fitted[column2])

#menampilkan nilai korelasi dan nilai p-value
print(f'Korelasi Pearson: {correlation:.2f}, p-value: {p_value:.4f}')

#menentukan hipotesis
print("Kesimpulan berdasarkan P-Value : ")
print("Hipotesis nol (H0) dapat ditolak." if p_value < 0.05 else "Tidak cukup bukti untuk menolak hipotesis nol (H0).")

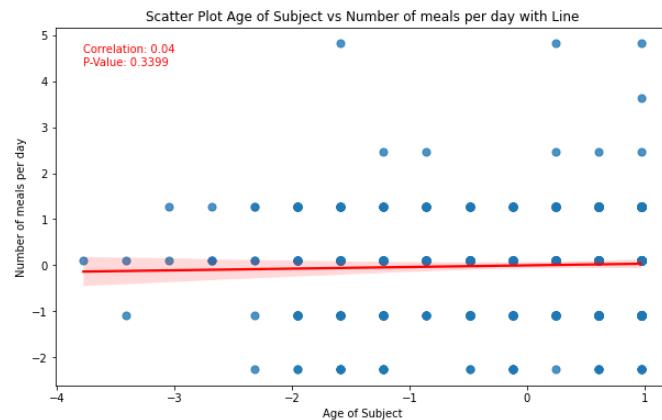
#kesimpulan berdasarkan korelasi
print("Kesimpulan berdasarkan Nilai Korelasi :")
print(f'Terdapat hubungan {"positif" if correlation > 0 else "negatif" if correlation < 0 else "tidak terdapat"} linear')

#membuat scatter plot
plt.figure(figsize=(10, 6))
sns.regplot(x=column1, y=column2, data=data_covid_fitted, scatter_kws={'s': 50}, line_kws={'color': 'red'})
# Menampilkan nilai korelasi dan p-value sebagai teks pada plot
plt.text(min(data_covid_fitted[column1]), max(data_covid_fitted[column2]), f'Correlation: {correlation:.2f}\nP-Value: {p_value:.4f}', color='red')
plt.title(f'Scatter Plot {column1} vs {column2} with Line')
plt.xlabel(column1)
plt.ylabel(column2)

plt.show()

```

Korelasi Pearson: 0.04, p-value: 0.3399
 Kesimpulan berdasarkan P-Value :
 Tidak cukup bukti untuk menolak hipotesis nol (H0).
 Kesimpulan berdasarkan Nilai Korelasi :
 Terdapat hubungan positif linear yang signifikan antara Age of Subject dan Number of meals per day.



Gambar 4.52. Korelasi antara kolom ‘Age of Subject’ dan ‘Number of meals per day’

Interpretasi Hasil :

Hasilnya menunjukkan hubungan linear positif yang signifikan antara kedua variabel dengan korelasi Pearson sebesar 0.04 dan P-Value sebesar 0.3399. Scatter plot dengan garis regresi memvisualisasikan hubungan positif tersebut, di mana titik-titik data tersebar di sekitar garis regresi yang menaik. Kesimpulan dari analisis ini adalah adanya hubungan positif yang signifikan antara ‘Age of Subject’ dan ‘Time Spent on Number of meals per day’.

4.6.8. Korelasi antara kolom ‘Age of Subject’ dan ‘Stress Busters’

```

import pandas as pd
from scipy.stats import pearsonr
import seaborn as sns
import matplotlib.pyplot as plt

#menginisiasi column 1 dan column 2
column1, column2 = 'Age of Subject', 'Stress busters'

#menghitung Korelasi Pearson dan P-Value
correlation, p_value = pearsonr(data_covid_fitted[column1], data_covid_fitted[column2])

#menampilkan nilai korelasi dan nilai p-value
print(f'Korelasi Pearson: {correlation:.2f}, p-value: {p_value:.4f}')

#menentukan hipotesis
print("Kesimpulan berdasarkan P-Value : ")
print("Hipotesis nol (H0) dapat ditolak." if p_value < 0.05 else "Tidak cukup bukti untuk menolak hipotesis nol (H0).")

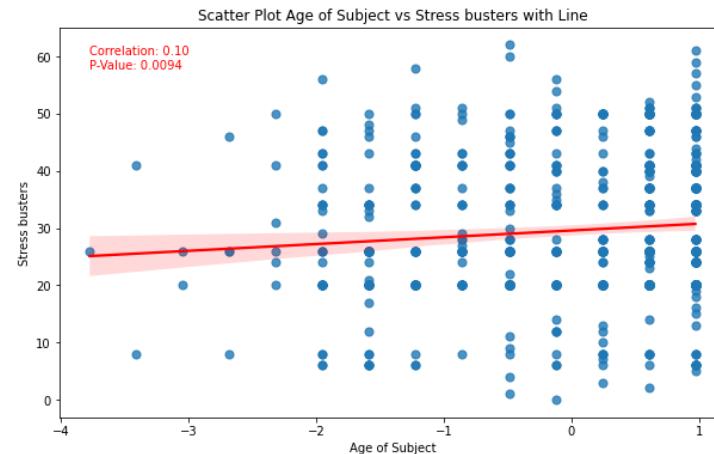
#kesimpulan berdasarkan korelasi
print("Kesimpulan berdasarkan Nilai Korelasi :")
print(f'Terdapat hubungan {'positif' if correlation > 0 else 'negatif' if correlation < 0 else 'tidak terdapat'} linear')

#membuat scatter plot
plt.figure(figsize=(10, 6))
sns.regplot(x=column1, y=column2, data=data_covid_fitted, scatter_kws={'s': 50}, line_kws={'color': 'red'})
# Menampilkan nilai korelasi dan p-value sebagai teks pada plot
plt.text(min(data_covid_fitted[column1]), max(data_covid_fitted[column2]), f'Correlation: {correlation:.2f}\n\nP-Value: {p_value:.4f}', color='red')
plt.title(f'Scatter Plot {column1} vs {column2} with Line')
plt.xlabel(column1)
plt.ylabel(column2)

plt.show()

```

Korelasi Pearson: 0.10, p-value: 0.0094
 Kesimpulan berdasarkan P-Value :
 Hipotesis nol (H0) dapat ditolak.
 Kesimpulan berdasarkan Nilai Korelasi :
 Terdapat hubungan positif linear yang signifikan antara Age of Subject dan Stress busters.



Gambar 4.53. Korelasi antara kolom ‘Age of Subject’ dan ‘Stress Busters’

Interpretasi Hasil :

Hasilnya menunjukkan hubungan linear positif yang signifikan antara kedua variabel dengan korelasi Pearson sebesar 0.10 dan P-Value sebesar 0.0094. Scatter plot dengan garis regresi memvisualisasikan hubungan positif tersebut, di mana titik-titik data tersebar di sekitar garis regresi yang menaik. Kesimpulan dari analisis ini adalah adanya hubungan positif yang signifikan antara ‘Age of Subject’ dan ‘Stress Busters’.

4.6.9. Korelasi Antar Semua Kolom dengan ‘Health Issue During Lockdown’

```

import pandas as pd
from scipy.stats import pearsonr

#variabel dependen (y)
y_column = 'Health issue during lockdown'
y = data_covid_fitted[y_column]

#variabel independen (x) selain variabel y dan kolom "ID"
x = data_covid_fitted.drop(columns=[y_column, 'ID'])

#tingkat signifikansi (alpha)
alpha = 0.05

#list untuk menyimpan hasil
results = []

#perhitungan korelasi dan p-value untuk setiap variabel
for column in x.columns:
    correlation_coefficient, p_value = pearsonr(x[column], y)

    #penentuan berdasarkan hasil uji hipotesis
    hypothesis_result = "Signifikan" if p_value < alpha else "Tidak Signifikan"

    # Menambahkan hasil ke dalam list
    result = {
        'Kolom': column,
        'Nilai Korelasi': correlation_coefficient,
        'P-Value': p_value,
        'Hasil Hipotesis': hypothesis_result,
        'Keterangan Hasil': f'{'' if p_value < alpha else 'Tidak'} Mempengaruhi Variabel Y"
    }
    results.append(result)

#DataFrame dari list hasil
result_df = pd.DataFrame(results)
result_df

```

	Kolom	Nilai Korelasi	P-Value	Hasil Hipotesis	Keterangan Hasil
0	Age of Subject	0.164824	0.000009	Signifikan	Mempengaruhi Variabel Y
1	Time spent on Online Class	-0.143232	0.000119	Signifikan	Mempengaruhi Variabel Y
2	Time spent on self study	-0.013674	0.714710	Tidak Signifikan	Tidak Mempengaruhi Variabel Y
3	Time spent on fitness	-0.029037	0.437562	Tidak Signifikan	Tidak Mempengaruhi Variabel Y
4	Time spent on sleep	0.009808	0.793193	Tidak Signifikan	Tidak Mempengaruhi Variabel Y
5	Time spent on social media	0.055734	0.135981	Tidak Signifikan	Tidak Mempengaruhi Variabel Y
6	Time spent on TV	-0.033217	0.374459	Tidak Signifikan	Tidak Mempengaruhi Variabel Y
7	Number of meals per day	-0.091796	0.013936	Signifikan	Mempengaruhi Variabel Y
8	Time utilized	-0.069384	0.063328	Tidak Signifikan	Tidak Mempengaruhi Variabel Y
9	Region of residence	0.090051	0.015866	Signifikan	Mempengaruhi Variabel Y
10	Rating of Online Class experience	0.130968	0.000438	Signifikan	Mempengaruhi Variabel Y
11	Medium for online class	-0.028361	0.448311	Tidak Signifikan	Tidak Mempengaruhi Variabel Y
12	Preferred social media platform	-0.000047	0.999001	Tidak Signifikan	Tidak Mempengaruhi Variabel Y
13	Change in your weight	-0.149327	0.000060	Signifikan	Mempengaruhi Variabel Y
14	Stress busters	0.004145	0.911787	Tidak Signifikan	Tidak Mempengaruhi Variabel Y
15	What you miss the most	0.070360	0.059693	Tidak Signifikan	Tidak Mempengaruhi Variabel Y

Gambar 4.54. Korelasi Antara Semua Kolom dengan ‘Health Issue During Lockdown’

Penjelasan

Kode di atas bertujuan untuk melakukan analisis *bivariate* dengan variabel y adalah ‘Health Issue During Lockdown’ dan variabel x adalah semua kolom selain variabel y. Melakukan inisiasi tingkat signifikansi sebesar 0,05 kemudian dilakukan *looping for* disetiap kolom. Pada proses analisis ini dipergunakan *library* pearson dan scipy.stats. Akan ditampilkan beberapa output, yakni nilai korelasi, p-value, hasil hipotesis, dan keterangan hasil. Dimana untuk menghasilkan hasil hipotesis dan keterangan hasil diperoleh dari kondisi jika $p\text{-value} < \alpha$, maka H_0 ditolak variabel x mempengaruhi variabel y.

Interpretasi Hasil :

Output dari kode ini menunjukkan bahwa kolom ‘Age Of Subject’, ‘Time Spent On Online Class’, ‘Numbers Of Meals Per-Day’, ‘Region Of Residence’, ‘Rating Of Online Class Experience’, dan ‘Change In Your Weight’ mempengaruhi variabel y. Selain kolom tersebut, tidak mempengaruhi variabel y.

4.6.10. Visualisasi *Heatmap* untuk Menentukan Hubungan Linier Antar Kedua Kolom

```

import seaborn as sns
import matplotlib.pyplot as plt

#pemilihan kolom
selected_columns = [
    'Age of Subject', 'Time spent on Online Class', 'Time spent on self study',
    'Time spent on fitness', 'Time spent on sleep', 'Time spent on social media',
    'Time spent on TV', 'Number of meals per day', 'Health issue during lockdown',
    'Time utilized', 'Rating of Online Class experience', 'Change in your weight'
]

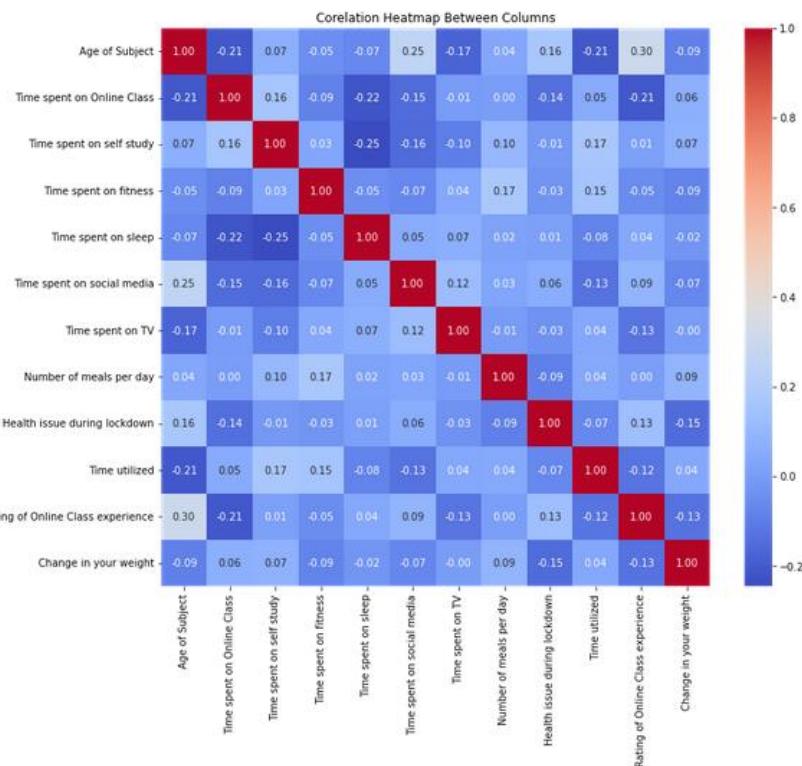
#pembuatan dataframe dari kolom yang dipilih
selected_data = data_covid_fitted[selected_columns]

#perhitungan korelasi antar kolom
correlation_matrix = selected_data.corr()

#visualisasi heatmap
plt.figure(figsize=(12, 10))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Corelation Heatmap Between Columns')

plt.show()

```



Gambar 4.55. Visualisasi *Heatmap* Hubungan Linier Antar Kedua Kolom

Penjelasan :

Kode ini menggunakan pustaka *seaborn* dan *matplotlib* untuk membuat *heatmap* korelasi antar kolom. Kemudian, memilih kolom-kolom tertentu untuk pembuatan *DataFrame* dan disimpan pada variabel ‘*selected_data*’. Perhitungan korelasi antar kolom menggunakan fungsi bawaan yaitu ‘*.corr()*’. Lalu, *heatmap* divisualisasikan.

Interpretasi Hasil :

Heatmap ini memberikan gambaran korelasi antar variabel. Korelasi positif yang signifikan terlihat antara usia dan waktu belajar *online* (koefisien korelasi 0,8), sedangkan korelasi negatif kuat terlihat antara waktu belajar *online* dan waktu belajar mandiri (koefisien -0,8), waktu berolahraga (koefisien -0,6), serta korelasi positif dengan waktu menggunakan media sosial (koefisien 0,4). Kesimpulannya, faktor-faktor seperti usia, waktu belajar mandiri, waktu berolahraga, dan waktu menggunakan media sosial mempengaruhi waktu yang dihabiskan untuk belajar *online*.

4.6.11. Korelasi Pengaruh ‘Age Of Subject’ dan ‘Time Spent On Online Class’ terhadap ‘Health Issue During Lockdown’

```

: import statsmodels.api as sm
#inisiasi y
y = data_covid_fitted['Health issue during lockdown']
#inisiasi x
x = data_covid_fitted[['Age of Subject', 'Time spent on Online Class']]
#konstanta untuk model regresi
x = sm.add_constant(x)
#pembuatan model regresi
model = sm.OLS(y, x).fit()
print(model.summary())
#perhitungan p-value
p_values = model.pvalues
print("Hasil p-value:")
print(p_values)
#pembuatan kondisi apakah H0 diterima/ditolak
alpha = 0.05
tolak_H0 = any(p_values < alpha)
if tolak_H0:
    print("\nHipotesis Nol (H0) ditolak. Setidaknya satu variabel independen memiliki pengaruh yang signifikan")
else:
    print("\nHipotesis Nol (H0) diterima. Tidak ada bukti yang cukup untuk menyatakan adanya pengaruh yang sig

```

OLS Regression Results

Dep. Variable:	Health issue during lockdown	R-squared:	0.840			
Model:	OLS	Adj. R-squared:	0.837			
Method:	Least Squares	F-statistic:	14.69			
Date:	Mon, 25 Dec 2023	Prob (F-statistic):	5.58e-07			
Time:	11:17:55	Log-Likelihood:	-1882.9			
No. Observations:	717	AIC:	2012.			
Df Model:	2	BIC:	2026.			
Df Residuals:	714					
Covariance Type:	nonrobust					
coef	std err	t	P> t	[0.025	0.975]	
const	7.116e-17	0.037	1.94e-15	1.000	-0.072	0.072
Age of Subject	0.1410	0.038	3.760	0.000	0.067	0.215
Time spent on Online Class	-0.1137	0.038	-3.032	0.003	-0.187	-0.040
====						
Omnibus:	316.108	Durbin-Watson:	2.020			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	985.183			
Skew:	2.271	Prob(JB):	1.18e-214			
Kurtosis:	6.514	Cond. No.	1.24			
====						

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Hasil p-value:

const	1.000000
Age of Subject	0.000184
Time spent on Online Class	0.002518

dtype: float64

Hipotesis Nol (H0) ditolak. Setidaknya satu variabel independen memiliki pengaruh yang signifikan.

Gambar 4.56. OLS Pemgaruh ‘Age Of Subject’ dan ‘Time Spent On Online Class’ terhadap ‘Health Issue During Lockdown’

Penjelasan :

Kode di atas bertujuan untuk mengetahui korelasi pengaruh antara kolom ‘Age Of Subject’ dan ‘Time Spent On Online Class’ terhadap ‘Health Issue During Lockdown’. Untuk melakukan analisis korelasi *multivariate* dipergunakan analisis regresi linier sederhana dengan menggunakan *library* statsmodel. Dengan inisiasi variabel x atau variabel independen adalah ‘Age Of Subject’ dan ‘Time Spent On Online Class’ dan variabel y atau variabel dependen adalah ‘Health Issue During Lockdown’. Akan ditampilkan hasil *ordinary least squares* (OLS) dengan nilai p-value masing-masing kolom. Dilakukan penetapan nilai alpha sebesar 0,05 dan dibuat kondisi jika nilai p-value < alpha maka pernyataan H0 ditolak, salah satu variabel independen memiliki pengaruh yang signifikan terhadap variabel dependen demikian sebaliknya.

Interpretasi Hasil :

Output dari kode ini menunjukkan bahwa diperoleh nilai p-value sebesar 0,001 dan 0,025 dimana nilai ini < dari nilai alpha. Maka, dapat ditarik hasil dan kesimpulan bahwa pernyataan H0 ditolak, setidaknya terdapat satu variabel antara ‘Age Of Subject’ atau ‘Time Spent On Online Class’ yang mempengaruhi ‘Health Issue During Lockdown’.

4.6.12. Korelasi Pengaruh ‘Age Of Subject’ dan ‘Number Of Meals Per-Day’ terhadap ‘Change In Your Weight’

```
In [65]: import statsmodels.api as sm
#inisiasi y
y = data_covid_fitted['Change in your weight']
#inisiasi x
x = data_covid_fitted[['Age of Subject', 'Number of meals per day']]
#konstanta untuk model regresi
x = sm.add_constant(x)
#pendekatan model regresi
model = sm.OLS(y,x).fit()
print(model.summary())
#perhitungan p-value
p_values = model.pvalues
print("\nHasil p-value:")
print(p_values)
#menentukan kondisi apakah H0 diterima/ditolak
alpha = 0.05
tolak_h0 = any(p_values < alpha)
if tolak_h0:
    print("\nHipotesis Nol (H0) ditolak. Setidaknya satu variabel independen memiliki pengaruh yang signifikan")
else:
    print("\nHipotesis Nol (H0) diterima. Tidak ada bukti yang cukup untuk menyatakan adanya pengaruh yang sig")

OLS Regression Results
=====
Dep. Variable: Change in your weight R-squared: 0.017
Model: OLS Adj. R-squared: 0.014
Method: Least Squares F-statistic: 6.188
Date: Mon, 25 Dec 2023 Prob (F-statistic): 0.00217
Time: 11:18:08 Log-Likelihood: -1011.2
No. Observations: 714 AIC: 2028.
Df Residuals: 713 BIC: 2042.
Df Model: 2 Cond. No.: 1.04
Covariance Type: nonrobust
=====
            coef  std err      t  P>|t|   [0.025   0.975]
=====
const     -9.222e-17  0.037  -2.49e-15  1.000  -0.073   0.073
Age of Subject  -8.8958  0.037  -2.581  0.018  -0.169   -0.023
Number of meals per day  0.0921  0.037  2.481  0.013  0.019   0.165
=====
Omnibus: 253.693 Durbin-Watson: 2.024
Prob(Omnibus): 0.000 Jarque-Bera (JB): 59.526
Skew: -0.453 Prob(JB): 1.19e-13
Kurtosis: 1.918 Cond. No.: 1.04
=====
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Hasil p-value:
Const          1.000000
Age of Subject 0.010043
Number of meals per day 0.013342
dtype: float64
```

Hipotesis Nol (H0) ditolak. Setidaknya satu variabel independen memiliki pengaruh yang signifikan.

Gambar 4.57. OLS Pengaruh ‘Age Of Subject’ dan ‘Number Of Meals Per-Day’ Terhadap ‘Change In Your Weight’

Penjelasan :

Kode di atas bertujuan untuk mengetahui korelasi pengaruh antara kolom ‘Age Of Subject’ dan ‘Number Of Meals Perday’ terhadap ‘Change In Your Weight’. Untuk melakukan analisis korelasi *multivariate* dipergunakan analisis regresi linier sederhana dengan menggunakan *library* statsmodel. Dengan inisiasi variabel x atau variabel independen adalah ‘Age Of Subject’ dan ‘Number Of Meals Perday’ dan variabel y atau variabel dependen adalah ‘Change In Your Weight’. Akan ditampilkan hasil *ordinary least squares* (OLS) dengan nilai p-value masing-masing kolom. Dilakukan penetapan nilai alpha sebesar 0,05 dan dibuat kondisi jika nilai p-value < alpha maka pernyataan H0 ditolak, salah satu variabel independen memiliki pengaruh yang signifikan terhadap variabel dependen demikian sebaliknya.

Interpretasi Hasil :

Output dari kode ini menunjukkan bahwa diperoleh nilai p-value sebesar 0,001 dan 0,013 dimana nilai ini < dari nilai alpha. Maka, dapat ditarik hasil dan kesimpulan bahwa pernyataan H0 ditolak, setidaknya terdapat satu variabel antara ‘Age Of Subject’ atau ‘Number Of Meals Perday’ yang mempengaruhi ‘Change In Your Weight’.

4.6.13. Pengaruh Semua Kolom terhadap ‘Health Issue During Lockdown’

```

import statsmodels.api as sm
#inisiasi y
y = data_covid_fitted['Health issue during lockdown']

#inisiasi x
x = data_covid_fitted[['Age of Subject', 'Time spent on Online Class', 'Time spent on self study', 'Time spent on sleep', 'Time utilized', 'Region of residence', 'Rating of Online Class experience', 'Medium for Preferred social media platform', 'Change in your weight', 'Stress busters', 'What you miss the most']]

#konstanta untuk model regresi
x = sm.add_constant(x)

#pembuatan model regresi
model = sm.OLS(y, x).fit()
print(model.summary())

#perhitungan p-value
p_values = model.pvalues
print("\nHasil p-value:")
print(p_values)

#pembuatan kondisi apakah H0 diterima/ditolak
alpha = 0.05
tolak_h0 = any(p_values < alpha)

if tolak_h0:
    print("\nHipotesis Nol (H0) ditolak. Setidaknya satu variabel independen memiliki pengaruh yang signifikan")
else:
    print("\nHipotesis Nol (H0) diterima. Tidak ada bukti yang cukup untuk menyatakan adanya pengaruh yang signifikan")

OLS Regression Results
=====
Dep. Variable: Health issue during lockdown R-squared: 0.878
Model: OLS Adj. R-squared: 0.877
Method: Least Squares F-statistic: 3.710
Date: Tue, 26 Dec 2023 Prob (F-statistic): 1.35e-06
Time: 13:22:57 Log-Likelihood: -988.20
No. Observations: 700 AIC: 2016.
Df Residuals: 700 BIC: 2088.
Df Model: 16 Covariance Type: nonrobust
=====
coef std err t P>|t| [0.025 0.975]
const 8.51e-17 0.036 2.35e-15 1.000 -0.071 0.071
Age of Subject 0.0112 0.045 2.479 0.813 0.135 0.199
Time spent on Online Class -0.1849 0.049 -2.11 0.009 -0.184 0.177
Time spent on self study 0.0994 0.039 0.238 0.812 0.068 0.087
Time spent on fitness -0.0173 0.038 -0.447 0.655 -0.092 0.058
Time spent on sleep -0.0076 0.039 -0.198 0.843 -0.084 0.068
Time spent on social media 0.0001 0.039 0.017 0.977 -0.121 0.133
Time spent on TV -0.0067 0.038 -0.176 0.860 -0.081 0.068
Number of meals per day -0.0797 0.038 -2.120 0.034 -0.153 -0.006
Time utilized -0.0242 0.039 -0.628 0.531 -0.100 0.052
Region of residence 0.0107 0.036 1.921 0.355 -0.020 0.142
Rating of Online Class experience 0.0577 0.049 1.268 0.145 -0.020 0.125
Medium for online class -0.0272 0.038 -0.707 0.480 -0.103 0.048
Preferred social media platform 0.0571 0.039 1.453 0.147 -0.020 0.134
Change in your weight -0.1194 0.037 -3.106 0.000 -0.193 -0.046
Stress busters -0.0008 0.037 -0.262 0.794 -0.082 0.063
What you miss the most 0.0355 0.037 0.956 0.339 -0.037 0.188
=====
Omnibus: 296.607 Durbin-Watson: 2.013
Prob(Omnibus): 0.000 Jarque-Bera (JB): 868.92
Skew: 2.144 Prob(JB): 2.41e-189
Kurtosis: 6.270 Cond. No. 2.11
=====

Hasil p-value:
const 1.000000
Age of Subject 0.013411
Time spent on Online Class 0.008462
Time spent on self study 0.811968
Time spent on fitness 0.000605
Time spent on sleep 0.843421
Time spent on social media 0.875567
Time spent on TV 0.866198
Number of meals per day 0.034368
Time utilized 0.530526
Region of residence 0.055135
Rating of Online Class experience 0.145225
Medium for online class 0.010113
Preferred social media platform 0.146781
Change in your weight 0.001484
Stress busters 0.793519
What you miss the most 0.339415
dtype: float64

Hipotesis Nol (H0) ditolak. Setidaknya satu variabel independen memiliki pengaruh yang signifikan.

```

Gambar 4.58. OLS Pengaruh Semua Kolom terhadap ‘Health Issue During Lockdown’

Penjelasan :

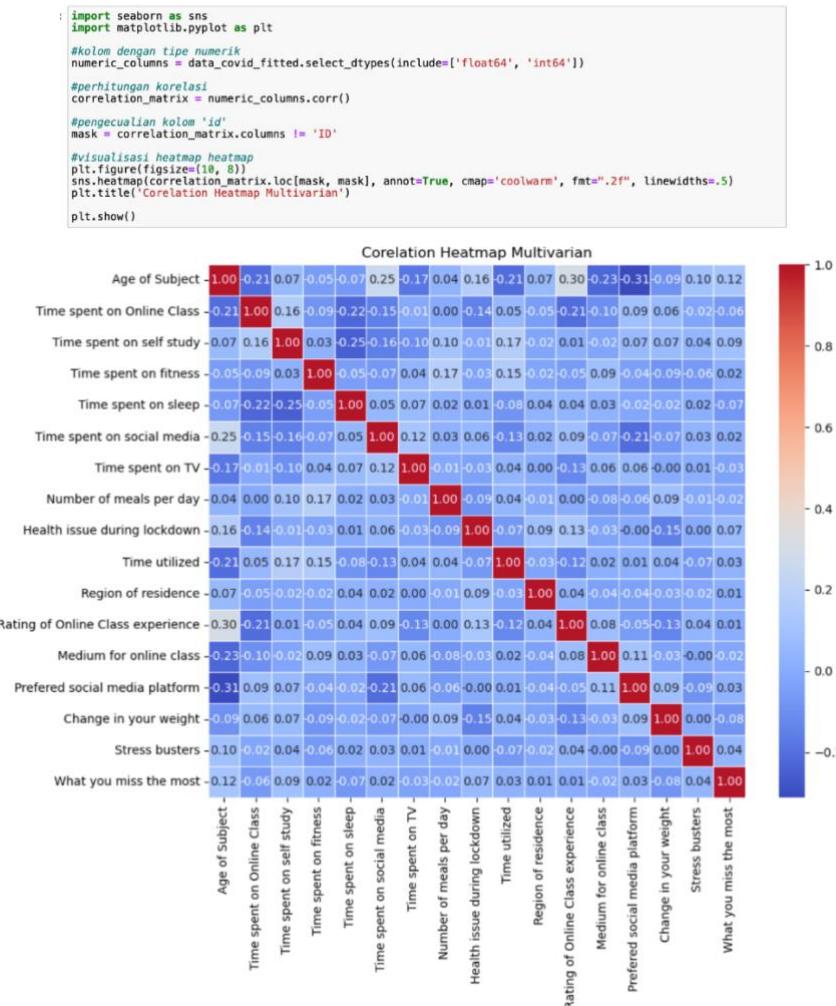
Kode di atas bertujuan untuk mengetahui korelasi pengaruh semua kolom terhadap ‘Health Issue During Lockdown’. Dimana Untuk melakukan analisis korelasi *multivariate* dipergunakan analisis regresi linier sederhana dengan menggunakan *library* statsmodel. Dengan inisiasi variabel x atau variabel independen adalah semua kolom selain variabel y dan variabel y atau variabel dependen adalah ‘Health Issue During Lockdown’. Maka, akan ditampilkan hasil *ordinary least squares* (OLS) dengan nilai p-value masing-masing kolom. Dilakukan penetapan nilai alpha sebesar 0,05 dan dibuat kondisi jika nilai p-value < alpha maka pernyataan H0 ditolak, salah satu variabel independen memiliki pengaruh yang signifikan terhadap variabel dependen demikian sebaliknya.

Interpretasi Hasil :

Output dari kode ini menunjukkan bahwa diperoleh nilai p-value dari kolom ‘Age Of Subject’ sebesar 0,01, ‘Time Spent On Online Class’ sebesar 0,008, ‘Numbers Of Meals Per-Day’ sebesar 0,003, ‘Region Of Residence’ sebesar 0,05, dan ‘Change In Your Weight’ sebesar 0,001’ dimana nilai ini kurang dari nilai

alpha. Maka, dapat ditarik hasil dan kesimpulan bahwa pernyataan H0 ditolak, dan terdapat lebih dari satu variabel yang mempengaruhi variabel y.

4.6.14. Visualisasi Heatmap Untuk Menentukan Hubungan Linier Lebih Dari Dua Kolom



Gambar 4.59. Visualisasi Heatmap Untuk Menentukan Hubungan Linier Lebih Dari Dua Kolom
Penjelasan :

Kode ini menggunakan pustaka *seaborn* dan *matplotlib* untuk membuat *heatmap* korelasi antar kolom. Kemudian, memilih kolom-kolom tertentu untuk pembuatan *dataframe* dan disimpan pada variabel ‘*selected_data*’. Perhitungan korelasi antar kolom menggunakan fungsi bawaan yaitu ‘.corr()’. Lalu, *heatmap* divisualisasikan,

Interpretasi Hasil :

Berdasarkan hasil korelasi *heatmap* dapat diperlihatkan menunjukkan hubungan antara dua variabel atau lebih dengan 13 variabel yang diperhitungkan, diantaranya yakni ‘Age Of Subject’, ‘Time Spent On Online Class’, ‘Time Spent On Self Study’, ‘Time Spent On Fitness’, ‘Time Spent On Sleep’, ‘Time Spent On Soscial Media’, ‘Time Spen On TV’, ‘Number Of Meals Per-day’, ‘Health Issue During Lockdown’, ‘Time Utilized’, ‘Region Of Residence’, ‘Rating Of Online Class Experience’, ‘Medium For Online Class’, ‘Prefreed Social Media Platfrom’, ‘Change In Your Weight’, ‘Stress Busters’, dan ‘What Your Miss The Most’. Dapat ambil hasil bahwa ‘Time Spent On Online Class’ dan ‘Time

'Spent On Self Study' memiliki hubungan yang positif, *'Time Spent On Online Class'* dan *'Time Spent On Social Media'* memiliki hubungan yang negatif. *'Change In Your Weight'* dan *'Time Spent On Fitness'* memiliki hubungan yang positif.

4.7. Pengembangan Model Dan Evaluasi

Pada tahapan ini dilakukan proses pembangunan model dan evaluasi dengan menggunakan K-Means. Metode K-Means merupakan metode yang dipergunakan untuk mengelompokkan ataupun klasterisasi berdasarkan jumlah *cluster* yang telah diinisiasi. Tentunya untuk melakukan penerapan metode K-Means ini diperlukan beberapa langkah-langkah, yakni.

4.7.1. Melakukan Penentuan Jumlah Cluster Yang Optimal

```
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

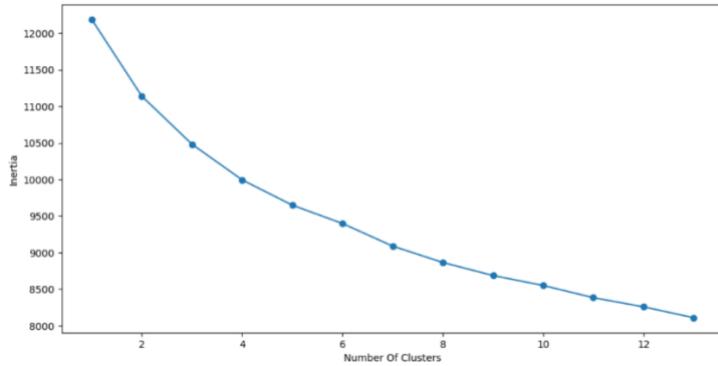
#inisiasi nilai dari data_covid_fitted yang telah dinormalisasi
numeric_columns = data_covid_fitted.select_dtypes(include=['float64', 'int64'])

distance_values = []

#perhitungan visualisasi cluster
for cluster in range(1, 14):
    kmeans = KMeans(n_clusters=cluster, init='k-means++', n_init=10)
    kmeans.fit(numeric_columns)
    distance_values.append(kmeans.inertia_)

cluster_output = pd.DataFrame({'Cluster': range(1, 14),
                               'distance_values': distance_values})

#visualisasi cluster
plt.figure(figsize=(12, 6))
plt.plot(cluster_output['cluster'],
         cluster_output['distance_values'], marker='o')
plt.xlabel('Number Of Clusters')
plt.ylabel("Inertia")
plt.show()
```



Gambar 4.60. Penentuan Jumlah Cluster dengan Metode Elbow

Penjelasan :

Kode di atas bertujuan untuk menentukan jumlah *cluster* menggunakan metode elbow agar memperoleh jumlah *cluster* yang optimal dengan memperhitungkan inersia model K-Means dan menghasilkan visualisasinya. Dipergunakan *library* `sklearn.cluster` dengan *import* `KMeans` dan `matplotlib` untuk visualisasi. Data yang dipergunakan untuk pembentukan K-Means adalah data yang telah dinormalisasi menggunakan nilai *z-score*. Terdapat list *distance_values* yang dipergunakan untuk menghasilkan perhitungan visualisasi dan *cluster* dari rentang 1-13.

Interpretasi Hasil :

Output visualisasi dari kode ini dapat diperlihatkan bahwa grafik inersia (jarak antara data dalam suatu cluster) memiliki siku yang jelas pada titik 2. Sehingga, untuk penentuan jumlah *clustering* perhitungan ini dipergunakan nilai *cluster* sebesar 2. Pemilihan jumlah *cluster* sebesar 2 ini berdasarkan visualisasi yang diperoleh jika dilakukan pemilihan *cluster* lebih besar dari 2 akan ditampilkan bahwa *cluster-cluster* yang dibentuk semakin homogen dan tidak memiliki makna signifikan.

4.7.2. Penerapan Metode K-Means

```
from sklearn.cluster import KMeans
#mengelaskan kolom 'id'
data_covid_fitted = data_covid_fitted.loc[:, data_covid_fitted.columns.difference(['ID'])]

#pembuatan model kmeans
kmeans = KMeans(n_clusters=2, init='k-means++', random_state=1, n_init=10)
kmeans.fit(data_covid_fitted)

label = kmeans.fit_predict(data_covid_fitted)
data_covid_output = data_covid_fitted.copy()
data_covid_output['Cluster'] = label
data_covid_output['Cluster'].value_counts()

Cluster
0    426
1    291
Name: count, dtype: int64
```

Gambar 4.61. Jumlah Masing-Masing Cluster

Penjelasan :

Kode di atas bertujuan untuk memperoleh jumlah dari masing-masing *cluster*. Dengan menggunakan semua data di kolom yang terdapat pada *dataframe* *data_covid_fitted* selain kolom '*ID*'. Lalu dilakukan pembuatan model K-Means dengan jumlah *cluster* sebanyak 2 dan dilakukan perhitungan jumlah *cluster* dan menyimpannya di dalam variabel *data_covid_output*.

Interpretasi Hasil :

Output dari kode ini adalah jumlah *cluster* 0 sebesar 426 dan jumlah *cluster* 1 sebesar 291. Dimana perlu diketahui bahwa disaat melakukan visualisasi data interaktif, *cluster* 0 merupakan *cluster* 1 dan *cluster* 1 merupakan *cluster* 2.

4.8. Visualisasi Data Interaktif

Visualisasi data interaktif merupakan penyajian data atau informasi yang memungkinkan pengguna untuk berinteraksi secara langsung dengan elemen visual tersebut. Pada visualisasi data interaktif ini, kami menggunakan *Dash* untuk memvisualisasi data interaktif menggunakan *scatter plot* dan mengevaluasi korelasi antara dua variabel data yang dipilih dari *dataset*, serta menentukan jumlah *cluster* untuk menganalisis K-means.

4.8.1. Visualisasi Data Interaktif Menggunakan Dash

```
import dash
from dash import dcc, html
from dash.dependencies import Input, Output
import pandas as pd
import plotly.express as px
from scipy.stats import pearsonr
from sklearn.cluster import KMeans

#pembuatan dash
app = dash.Dash(__name__)

#pengaturan tata letak
app.layout = html.Div([
    html.H1("Visualisasi Scatter Plot Interaktif Menggunakan Dash"),
    html.Label("Pilih Kolom X:"), 
    dcc.Dropdown(
        id="dropdown-x",
        options=[{"label": col, "value": col} for col in data_covid_fitted.columns if col != 'id'],
        value='Age of Subject' # Default value
    ),
    html.Label("Pilih Kolom Y:"), 
    dcc.Dropdown(
        id="dropdown-y",
        options=[{"label": col, "value": col} for col in data_covid_fitted.columns if col != 'id'],
        value='Time spent on Online Class' # Default value
    ),
    html.Label("Jumlah Cluster : "), 
    dcc.Dropdown(id="dropdown-class", options=[
        {"label": i, "value": i} for i in range(1, 3)
    ], value=2),
    dcc.Graph(id='scatter-plot'),
    html.Div(id='correlation-output')
])
```

```

#untuk input output
@app.callback(
    Output('scatter-plot', 'figure'),
    Output('correlation-output', 'children'),
    Input('dropdown-x', 'value'),
    Input('dropdown-y', 'value'),
    Input('dropdown-class', 'value')
)
def update_scatter_plot(selected_x, selected_y, selected_class):
    #menggunakan perhitungan k-means yang sudah dilakukan sebelumnya
    kmeans = KMeans(n_clusters=selected_class, init='k-means++', random_state=1, n_init=10)
    label = kmeans.fit_predict(data_covid_fitted)
    data_covid_output = data_covid_fitted.copy()
    data_covid_output['Cluster'] = label

    #perhitungan korelasi dan p-value
    correlation, p_value = pearsonr(data_covid_output[selected_x], data_covid_output[selected_y])

    #pembuatan scatter plot
    fig = px.scatter(data_covid_output, x=selected_x, y=selected_y, color='Cluster',
                      title=f'Visualisasi Scatter Plot Antara {selected_x} Dengan {selected_y}')

    #pembuatan kondisi apakah (x) mempengaruhi (y)
    correlation_output = f"Korelasi Pearson: {correlation:.2f}, p-value: {p_value:.4f}"
    hypothesis_output = ("Hipotesis nol (H0) dapat ditolak. Variabel independen memiliki pengaruh yang signifikan\n"
                         "Jika p-value < 0.05 else\n"
                         "Tidak cukup bukti untuk menolak hipotesis nol (H0). Variabel independen tidak memiliki pengaruh signifikan")
    if p_value < 0.05 else
        "Tidak cukup bukti untuk menolak hipotesis nol (H0). Variabel independen tidak memiliki pengaruh signifikan"

    return fig, html.Div([html.P(correlation_output), html.P(hypothesis_output)])
if __name__ == '__main__':
    app.run_server(debug=True, port=8089)

```

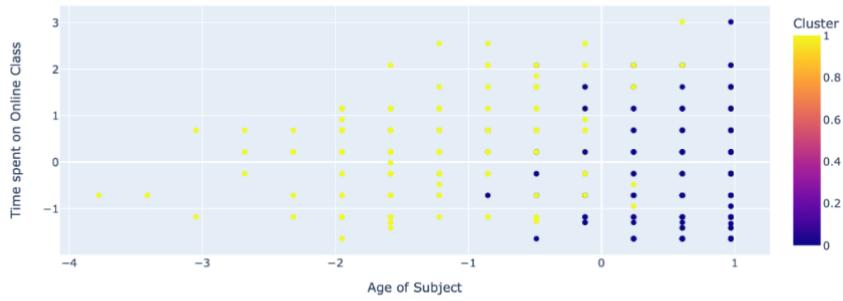
Visualisasi Scatter Plot Interaktif Menggunakan Dash

Pilih Kolumn X:

Pilih Kolumn Y:

Jumlah Cluster :

Visualisasi Scatter Plot Antara Age of Subject Dengan Time spent on Online Class



Korelasi Pearson: -0.21, p-value: 0.0000

Hipotesis nol (H0) dapat ditolak. Variabel independen memiliki pengaruh yang signifikan pada variabel dependen



Gambar 4.62. Visualisasi Data Interaktif

Penjelasan :

Kode tersebut merupakan implementasi aplikasi web interaktif menggunakan *Dash* dengan meng-import pustaka-pustaka yang dibutuhkan, yaitu *Dash* untuk aplikasi web, *plotly.express* untuk visualisasi interaktif, serta pustaka lainnya seperti *pandas*, *dcc*, *html*, *input*, *output*, *scipy.stats.pearsonr*, *sklearn*, *cluster.KMeans*. Selanjutnya, membuat aplikasi *Dash*, kemudian mengatur tata letak aplikasi menggunakan komponen-komponen *HTML*, *Dash Core Components*, dan *dropdown* untuk memilih kolom X dan Y, jumlah klaster, *scatter plot*, serta *output* korelasi. Fungsi *callback* yang digunakan untuk memperbarui *scatter plot* dengan pilihan yang dipilih, serta menghitung korelasi dan *p-value*. Kemudian, menjalankan aplikasi dalam mode debug pada *port 8089*.

Interpretasi Hasil :

Link Dash : <http://127.0.0.1:8089/>

Output dari kode tersebut menunjukkan terdapat dua kolom variabel yaitu X dan Y yang akan dipilih. Setiap kita memilih variabel X dan Y, maka akan muncul hasil dan interpretasi yang berbeda pula, sebagai contoh variabel X yaitu ‘Age of Subject’ dan variabel Y yaitu ‘Time Spent on online class’. Visualisasi cluster ini menunjukkan bahwa

ada hubungan positif antara usia subjek dan rata-rata waktu yang dihabiskan untuk kelas *online*. Hal ini berarti bahwa semakin tua usia subjek, semakin banyak waktu yang dihabiskannya untuk kelas *online* dengan korelasi *Pearson* sebesar -0.21 dan *P-Value* sebesar 0.000 . Kesimpulan berdasarkan *p-value*, hipotesis nol ditolak dan menunjukkan hubungan yang signifikan antara kedua variabel. Selain itu, kesimpulan berdasarkan nilai korelasi menyatakan bahwa terdapat hubungan negatif linear yang signifikan.

4.9. Hasil Analisis Kasus

Berdasarkan analisis yang telah dilakukan, dapat disimpulkan bahwa Pandemi COVID-19 cukup memberikan dampak yang signifikan terhadap kesehatan mental pelajar pada jenjang Sekolah Dasar, Sekolah Menengah Pertama, dan Sekolah Menengah Atas, khususnya untuk pelajar dengan usia $5 - 12$ tahun. Analisis statistika deskriptif, analisis multivariate, dan analisis bivariate yang telah kami lakukan menunjukkan adanya hubungan dengan perubahan-perubahan yang signifikan dalam kehidupan sehari-hari pelajar, mencakup aspek pendidikan, sosial, kesehatan, dan gaya hidup.

4.9.1. Analisis Statistika Deskriptif

Dalam analisis statistika deskriptif, didapatkan jumlah responden dari ketiga jenjang pendidikan adalah 15 responden pelajar SD, 125 responden pelajar SMP, dan 366 responden pelajar SMA. Analisis dilakukan pada beberapa perubahan-perubahan yang terjadi, didapatkan bahwa $53,31\%$ pelajar menggunakan *smartphone* sebagai perangkat utama selama proses pembelajaran *online*, diikuti oleh $40,78\%$ yang menggunakan laptop/*dekstop*. Dengan mayoritas waktu yang dihabiskan untuk belajar secara *online* adalah $2 - 5$ jam didapatkan dari sebagian besar responden dalam data, hal ini juga disertai dengan waktu yang dihabiskan pelajar untuk belajar mandiri yaitu selama $1 - 3$ jam, didapatkan dari modus atau mayoritas jawaban dalam data. Perubahan ini dapat memberikan dampak besar pada kesehatan mental pelajar, terutama dengan perubahan dalam pola pembelajaran selama COVID-19. Waktu belajar yang lebih lama dan penggunaan intensif perangkat digital dapat menyebabkan gangguan pola tidur, meningkatkan risiko stres, kecemasan, dan gejala depresi. Selain itu, pembelajaran *online* juga menciptakan isolasi sosial karena kurangnya interaksi fisik, dapat memicu kesepian dan masalah kesehatan mental lainnya. Penggunaan perangkat digital yang berlebihan juga berkontribusi pada peningkatan tingkat stres pelajar, memicu kelelahan, penurunan motivasi, dan bahkan risiko gangguan mental serius.

Namun, hal ini juga terkait dengan hasil analisis terkait waktu tidur responden. Menurut Unit Pelayanan Kesehatan, Kementerian Kesehatan Republik Indonesia⁶, waktu tidur yang ideal bagi pelajar adalah $7 - 10$ jam sehari. Berdasarkan data yang diperoleh, sebanyak $74,90\%$ responden sudah tidur sesuai dengan rekomendasi tersebut. Adanya proporsi responden yang mematuhi waktu tidur yang direkomendasikan dapat memberikan dampak positif pada kesehatan mental pelajar, mengingat tidur yang cukup terkait erat dengan stabilitas emosional dan kesejahteraan psikologis.

Selain itu, dari hasil penelitian juga dipaparkan bahwa sekitar $66,53\%$ pelajar memanfaatkan waktu selama pandemi COVID-19 untuk melakukan aktivitas *fitness*. Aktivitas fisik seperti ini dapat memiliki dampak positif pada kesehatan mental,

⁶ Kementerian Kesehatan Republik Indonesia. *Lama Waktu Tidur yang Dibutuhkan oleh Tubuh*. Unit Pelayanan Kesehatan, Kementerian Kesehatan Republik Indonesia. <https://upk.kemkes.go.id/new/lama-waktu-tidur-yang-dibutuhkan-oleh-tubuh#:~:text=E.,7%2D8%20jam%20setiap%20hari>.

mengingat olahraga telah terbukti dapat mengurangi tingkat stres, meningkatkan *mood*, dan mendukung kesejahteraan psikologis secara keseluruhan. Dengan demikian, upaya untuk mengintegrasikan aktivitas fisik dalam rutinitas sehari-hari selama pandemi dapat dianggap sebagai strategi yang positif untuk menjaga kesehatan mental pelajar.

Berdasarkan pedoman kesehatan, didapatkan data responden terkait waktu yang dihabiskan untuk mengakses sosial media maksimal dua jam dalam sehari⁷, dan mayoritas dari mereka 67,92% menggunakan sosial media secara wajar. Meskipun demikian, perlu dicatat bahwa 32,08% responden masih menggunakan sosial media secara berlebihan, yang dapat berdampak pada kesehatan psikologis dan fisik. Penggunaan sosial media yang berlebihan telah terkait dengan peningkatan risiko gangguan kesehatan mental, seperti kecanduan internet, dan dapat memperburuk tingkat stres serta kesehatan fisik akibat pola tidur yang terganggu.

Dalam konteks pola makan, mayoritas responden (73,78%), mengonsumsi makanan dalam frekuensi ideal, yaitu 2-3 kali makan utama sehari. Meskipun demikian, sekitar 26,22% responden memiliki frekuensi makan kurang atau lebih dari semestinya (1, 4-7 kali sehari). Didapat juga data yang menggambarkan perubahan berat badan yaitu, 44,91% responden tidak mengalami perubahan berat badan, sebesar 36,54% responden mengalami kenaikan berat badan, dan 18,55% responden mengalami penurunan berat badan. Kebiasaan makan yang tidak teratur dapat berdampak negatif pada kesehatan mental, terutama dalam hal fluktuasi energi dan *mood*. Frekuensi makan yang tidak terkontrol juga dapat menjadi faktor risiko untuk gangguan makan, yang dapat mempengaruhi kesehatan psikologis secara keseluruhan.

Didapatkan juga berdasarkan analisis yang dilakukan, didapatkan bahwa selama COVID-19 terdapat beberapa aktivitas yang dilakukan oleh responden untuk mencegah stres. Diinterpretasikan bahwa aktivitas yang paling sering dilakukan adalah mendengarkan musik dengan persentase sebanyak 25.66%, diikuti dengan bermain *game online* dengan persentase 16.60%. Terdapat pula aktivitas lainnya seperti membaca buku, menonton *web series*, tidur, menggulir media sosial, meditasi, berselancar *online*, berbicara dengan keluarga, menari, memasak, dan banyak lagi, masing-masing dengan persentase yang berbeda.

Selain itu didapatkan juga Sosial Media yang sering diakses oleh responden selama COVID-19, berdasarkan analisis didapatkan bahwa *Instagram* menjadi *platform* media sosial yang paling sering digunakan oleh responden dengan persentase sebesar lebih dari 200 (31,35%), 200 responden (30,21%) menggunakan *YouTube*, 180 responden (27,80%) menggunakan *WhatsApp*, kemudian dilanjutkan dengan persentase penggunaan *Facebook*, *Linkedin*, *Twitter*, *Reddit*, *Snapchat*, *Telegram*, *None*, *Quora*, dan *Talklife* juga tercatat dengan persentase yang lebih rendah, menunjukkan variasi penggunaan sosial media di kalangan responden.

4.9.2. Analisis Bivariate

Dalam analisis *bivariate* yang dilakukan, dapat diperoleh bahwa terdapat hubungan antara usia subjek dan berbagai variabel lainnya. Hubungan tersebut dapat

⁷ Limbong, T. S. (2018). Berapa Lama Waktu Ideal Menggunakan Media Sosial dalam Sehari?.

<https://www.klikdokter.com/psikologi/kesehatan-mental/berapa-lama-waktu-ideal-menggunakan-media-sosial-dalam-sehari>

bersifat positif atau negatif, tergantung pada variabel yang dianalisis. Dilihat dari analisis bivariate, pada variabel ‘Age of Subject’ dengan variabel “Time Spent on Online Class”, ‘Time Spent on Fitness’, ‘Time Spent on sleep’, dan ‘Time Spent on Tv’, terdapat hubungan linear negatif antara 1 variabel dengan variabel lainnya. Hubungan tersebut menunjukkan bahwa semakin tinggi usia subjek, maka semakin sedikit waktu yang dihabiskan untuk aktivitas-aktivitas tersebut, seperti mengikuti kelas online, berolahraga, tidur, dan menonton TV. Terdapat beberapa faktor yang dapat memengaruhi hubungan tersebut adalah faktor waktu, di mana semakin bertambahnya usia, maka semakin banyak tanggung jawab yang dipikul oleh seseorang. Hal ini menyebabkan seseorang memiliki waktu yang lebih sedikit untuk melakukan aktivitas tersebut.

Kemudian, terdapat hubungan antara variabel ‘Age of Subject’ dengan variabel lainnya, yaitu ‘Time Spent on self study’, ‘Time Spent on social media’, ‘Number of meals per day’, dan ‘stress busters’, menunjukkan adanya korelasi positif linear antara 1 variabel dengan variabel lainnya, artinya usia subjek memiliki pengaruh terhadap berbagai aspek kehidupan seseorang. Hubungan tersebut menunjukkan bahwa semakin tinggi usia subjek, maka semakin tinggi pula waktu yang dihabiskan untuk aktivitas tersebut, seperti menghabiskan waktu untuk belajar mandiri, melihat sosial media, pola makan, dan mengelola stres. Faktor yang dapat memengaruhi hubungan tersebut di antaranya, penyebab di mana orang dewasa cenderung memiliki lebih banyak stres dibandingkan remaja, karena faktor pekerjaan, hubungan interpersonal, dan masalah keuangan.

4.9.3. Analisis Multivariate

Dalam analisis *multivariate* yang dilakukan, dapat diperoleh bahwa terdapat hubungan antar dua variabel atau lebih. Hubungan tersebut untuk bisa mengetahui apakah antar lebih dua variabel independen berpengaruh terhadap variabel dependen atau tidak. Dilihat dari analisis *multivariate*, pada variabel ‘Age Of Subject’ dan ‘Time Spent On Online Class’ memiliki pengaruh yang positif (berpengaruh) pada ‘Health Issue During Lockdown’. Kemudian, pada variabel ‘Age Of Subject’ dan ‘Number Of Meals Perday’ memiliki pengaruh yang positif (berpengaruh) pada ‘Change In Your Weight’. Lalu dapat dilakukan analisis *multivariate* untuk mengetahui hubungan semua kolom terhadap kolom ‘Health Issue During Lockdown’ sehingga dalam analisis korelasi *multivariate* ini dapat diketahui lebih dari satu variabel yang mempengaruhi variabel tertentu.

4.9.4. Strategi dan Cara Untuk Mengatasi Dampak Pandemi COVID-19

Untuk mengatasi dampak pandemi COVID-19 pelajar dapat melakukan beberapa strategi maupun cara yang dapat mendukung kesejahteraan fisik dan mental, diantaranya yakni melakukan penerapan cara untuk beradaptasi secara perlahan dengan melakukan kebiasaan yang dimulai dari pembiasaan diri mengikuti kelas *online* menggunakan teknologi yang ada. Selain itu dapat dilakukan mengelola keseimbangan antara tugas akademis maupun kegiatan lainnya, mencari komunitas *virtual* yang membawa pengaruh baik, dan melakukan minat maupun hobi baru.

BAB V

KESIMPULAN

Berdasarkan serangkaian proses yang telah dilakukan untuk melakukan analisis hubungan antara pandemi COVID-19 terhadap kesehatan mental pelajar belum dapat diterapkan dalam *dataset* ini. Hal ini dikarenakan untuk menentukan apakah suatu pelajar memiliki kesehatan mental yang cukup baik atau tidak perlu dilakukan pertimbangan yang kompleks, sedangkan fitur-fitur yang terdapat dalam *dataset* belum cukup memenuhi untuk melakukan penentuan tersebut. Analisis ini lebih menekankan adanya hubungan antara pandemi COVID-19 terhadap beberapa faktor-faktor pendukung yang dapat mempengaruhi kesehatan mental pelajar saat pandemi COVID-19. Selain itu, dengan menggunakan metode K-Means dapat dilakukan analisis pola dan hubungan data dalam suatu kolom dengan kolom lainnya yang dapat menambah pengetahuan mengenai bagaimana data tersebut terbagi berdasarkan *cluster* dan menemukan pola menarik yang berkaitan dengan hubungan antara pandemi COVID-19 dan kesehatan mental dalam konteks tertentu.

Berdasarkan hasil analisis yang telah dilakukan dapat diketahui bahwa pandemi COVID-19 memberi dampak yang signifikan bagi pelajar, tentunya dalam bidang pendidikan karena adanya sekolah *online*, dimana dengan adanya hal ini dapat membuat kebiasaan hidup para pelajar berubah baik dari segi waktu maupun pelaksanaannya. Sekolah *online* ini dapat mengakibatkan timbulnya berbagai tantangan yang dapat menyulitkan pelajar dan mengakibatkan terbentuknya pola-pola baru yang dapat menjadi pendorong maupun penyebab suatu pelajar terkena gangguan mental. Dengan menggunakan metode K-Means dapat dilakukan identifikasi pola-pola dan mengetahui hubungan antar faktor penyebab pada setiap kolom yang tidak dapat ditemukan secara langsung ataupun secara manual.

Seperti contoh dengan menggunakan visualisasi data interaktif dapat diketahui bahwa lamanya '*Time Spent On Online Class*' mempengaruhi hasil survei pada kolom '*Health Issue Duringlockdown*' atau dapat dikatakan jika semakin lama pelajar mengikuti kelas *online* maka pelajar akan cenderung merasa mengalami gangguan kesehatan mental. Meskipun demikian, perlu diketahui bahwa metode K-Means memiliki kekurangan, yakni *sensitive* terhadap penciran (*outliers*) namun setelah dilakukan pengecekan tidak terdapat data yang terindikasi sebagai *outliers* dan telah dilakukan normalisasi sebelumnya.

Untuk mengatasi adanya gangguan kesehatan mental dapat dilakukan beberapa hal, diantaranya yakni menjaga kesehatan fisik, melakukan manajemen waktu yang efisien, melakukan hobi dan minat baru, serta melakukan aktivitas yang menyenangkan bagi tiap individu, seperti mendengarkan musik, membaca, dan berbisnis. Dengan menerapkan hal ini diharapkan para pelajar dapat mengatasi adanya gangguan kesehatan mental.

DAFTAR PUSTAKA

- Chaturvedi Kunal, Singh Nidhi, dkk. (2020). *COVID-19 and its impact on education, social life and mental health of students: A Survey*. Children and Youth Services Review. <http://www.sciencedirect.com/science/article/pii/S019074092032288X>
- Erskine, E. H., Wahdy, E. A., Blum Robert, dkk. (2022). *Laporan Penelitian National Adolescent Mental Health Survei (I-NAMHS), Indonesia*. Universitas Gadjah Mada, The University of Queensland, Johns Hopkins.
- Gloria. (2022). *Hasil Survei I-NAMHS: Satu dari Tiga Remaja Indonesia Memiliki Masalah Kesehatan Mental*. Universitas Gadjah Mada. <https://ugm.ac.id/id/berita/23086-hasil-survei-i-namhs-satu-dari-tiga-remaja-indonesia-memiliki-masalah-kesehatan-mental/>
- Infeksi Semarang. 2020. *Apakah Coronavirus dan COVID-19 itu?*. <https://infeksiemerging.kemkes.go.id/uncategorized/apakah-coronavirus-dan-covid-19-itu>
- Isella Virly, Suarca Kadek, Sari, M. N. (2021). *Kesehatan Mental Anak Selama Pandemi COVID-19. Bagian Ilmu Kesehatan Anak*. Departemen Psikologi RSUD Wangaya, Denpasar, Bali. CDK-198, 48(11).
- Kementerian Kesehatan Republik Indonesia. (2018). *Pengertian Kesehatan Mental*. <https://ayosehat.kemkes.go.id/pengertian-kesehatan-mental>
- Kementerian Kesehatan Republik Indonesia. *Lama Waktu Tidur yang Dibutuhkan oleh Tubuh*. Unit Pelayanan Kesehatan, Kementerian Kesehatan Republik Indonesia. <https://upk.kemkes.go.id/new/lama-waktu-tidur-yang-dibutuhkan-oleh-tubuh#:~:text=E.,7%2D8%20jam%20setiap%20hari>
- Limbong, T. S. (2018). *Berapa Lama Waktu Ideal Menggunakan Media Sosial dalam Sehari?*. <https://www.klikdokter.com/psikologi/kesehatan-mental/berapa-lama-waktu-ideal-menggunakan-media-sosial-dalam-sehari>
- Rahmadani, A., Setianingsih, C., & Dirgantara, F. M. (2022). *Tes Gangguan Depresi, Kecemasan, Dan Stres Pada Mahasiswa Selama Masa Covid-19 Menggunakan Algoritma Naïve Bayes*. eProceedings of Engineering, 9(3).
- Sheikhan, N. Y., Hawke, L. D., Ma, C., Courtney, D., Szatmari, P., Cleverley, K., ... & Henderson, J. (2022). A longitudinal cohort study of youth mental health and substance use before and during the COVID-19 Pandemic in Ontario, Canada: an exploratory analysis. *The Canadian Journal of Psychiatry*, 67(11), 842-854.
- Shindi, I. R., Natasari D. Z., Putri, Q.A. (2022). Analisis Dampak COVID-19 Terhadap Kehidupan Sosial dan Kesehatan Mental Siswa
- UNICEF. Resources for supporting children's emotional well-being during the COVID-19 pandemic [Internet]. (2020). [cited 2020 August 17]. Available from: 17 Agustus 2020. Di dapat dari: <https://www.unicef.org/ukraine/en/stories/resources-supporting-childrens-emotional-well-being-during-covid-19-pandemic>.

LAMPIRAN

Tautan video *YouTube* presentasi kelompok 5,

<https://youtu.be/UtSousuEGqI>

Tautan *drive kelompok 5*,

https://drive.google.com/drive/folders/1gxn8qZ_3bbBeJTzyCMOH-bOJTq9Z9Je0?usp=sharing