

# Analisis Aktivitas Olahraga Masyarakat Indonesia di Masa Pandemi Covid-19



**Alexander Bryan Wiratman**

**00000036371**

**Sistem Informasi 2019**

Universitas Multimedia Nusantara

alexander.bryan@student.umn.ac.id

## **Abstrak**

Tak dapat dipungkiri bahwa tahun 2020 hampir selesai dan diawali dengan kondisi yang kurang baik yaitu pandemi Covid-19 yang melanda seluruh dunia tak terkecuali Indonesia. Berbagai kebijakan yang dibuat oleh pemerintah Indonesia agar grafik pertumbuhan korban terinfeksi Covid-19 tidak bertumbuh lagi. Kebijakan tersebut adalah salah satunya dengan menerapkan kebijakan PSBB (Pembatasan Sosial Berskala Besar). Kebijakan tersebut berdampak pada berbagai sendi kehidupan, baik dalam dunia pendidikan, pekerjaan, dan berbagai aktivitas sehari-hari yang terdampak. Kebijakan yang bertujuan untuk mengurangi kerumunan (*physical distancing*)

pastilah menuai pro dan kontra. Daripada itu, kita dapat mengetahui bahwa kebijakan inilah yang efektif dalam mengurangi penyebaran virus tersebut dengan cepat.

Mudahnya penularan virus Covid-19 yaitu karena virus tersebut berukuran sangat kecil yaitu mencapai 125 nanometer menurut Pusat Informasi Bioteknologi AS, membuat kita perlu waspada dan cermat dalam menyikapi hal tersebut agar diri sendiri dan keluarga tetap sehat dan terlindungi. Tentunya selain menerapkan berbagai protokol, diperlukan daya tahan tubuh yang fit agar tidak mudah terserang virus. Untuk mendukung hal tersebut diperlukanlah olahraga yang cukup. Hal yang menjadi pertanyaan menarik adalah bagaimana aktivitas olahraga di masa pandemik dengan berbagai pembatasan dan protokol kesehatan dapat dilakukan? Hal inilah yang menjadi dasar penelitian ini dilakukan oleh penulis.

## **Kata Kunci**

Indonesia; Pandemi; Covid-19; Olahraga;

# 1. KONSEP BIG DATA DAN PERMASALAHAN

## 1.1 Big Data

Era revolusi industri 4.0 menjadi suatu fenomena yang akan menjadi tantangan bagi pemerintah dan masyarakat dalam suatu negara. Revolusi industri merupakan istilah yang menandakan terjadinya revolusi industri di seluruh dunia dan saat ini sudah mencapai revolusi industri keempat. Revolusi industri 4.0 menerapkan akan mengembangkan konsep otomatisasi dalam berbagai bidang kehidupan. Konsep otomatisasi tersebut bertujuan untuk meningkatkan kecepatan, fleksibilitas, dan meminimalisir berbagai kegagalan/kerugian. Berbagai kemudahan dan pemanfaatan teknologi yang berperan dalam revolusi industri 4.0 dapat dilihat dalam kehidupan sehari-hari, misalnya kemudahan memesan ojek pada jaman sekarang yang hanya melalui aplikasi dalam sebuah *smartphone*. Bahkan jika kita ingin berbelanja, saat ini kita dapat membeli barang yang kita inginkan juga hanya menggunakan aplikasi *e-commerce* melalui *smartphone*. Berbagai kebiasaan baru yang secara cepat atau lambat dapat mengubah pola hidup setiap manusia tersebut muncul seiring perkembangan jaman dengan kemajuan teknologi dan informasi.

Di balik berbagai perkembangan teknologi dalam menghadapi revolusi industri 4.0, terdapat Big Data yang memiliki kontribusi yang besar dalam konsep *Internet of Things* (IoT).

Big data berasal dari bahasa Inggris yaitu “Big” yang artinya besar dan data. Jadi Big Data dapat diartikan sebagai suatu kumpulan data yang memiliki ukuran dan volume yang sangat besar. Big Data tidak sama dengan data tradisional, Big Data dapat terdiri dari berbagai format data seperti *structured*, *unstructured*, dan *semi-structured*. Kaitannya dengan revolusi industri 4.0, Big Data tersebut memerlukan berbagai teknologi dan algoritma yang canggih, bahkan teknologi *Business Intelligence* sederhana tidak dapat terus mengakomodir keperluan Big Data Applications. Konsep Big Data dapat dipahami lebih lanjut dengan mengetahui konsep yang dimilikinya. Konsep Big Data yang paling sering digunakan oleh data scientist di seluruh dunia menurut Furt & Villanustre (2016) adalah *volume*, *velocity*, *variety*. *Volume* memiliki arti bahwa begitu besar jumlah data yang dihasilkan oleh berbagai *device* dan aplikasi, seperti pada *smartphone*, media sosial, dan sebagainya. Berdasarkan data yang diperoleh dari McAfee (2012), 2.5 exabytes dihasilkan

setiap harinya pada tahun 2012 dan tentunya akan terus bertambah besar sampai saat ini. *Velocity* berarti kecepatan. Data haruslah dihasilkan dengan cepat untuk kemudian diproses menjadi suatu informasi yang berguna untuk berbagai bidang. Contohnya adalah supermarket “Walmart” menghasilkan data 2.5 PB (Petabyte) setiap jamnya untuk transaksi yang dilakukan oleh konsumennya. *Variety* memiliki arti keanekaragaman data yang dimiliki oleh Big Data. Contoh variasi data dapat berupa gambar, video, dan teks.

Topik yang dipilih penulis dalam rangka penelitian yang dilakukan adalah aktivitas olahraga yang dilakukan oleh masyarakat Indonesia di masa pandemi. Kaitan Big Data dengan topik yang diambil adalah kaitannya dengan konsep 3V yang dimiliki oleh Big data yaitu *volume*, *velocity*, *variety*. Hubungan topik ini dengan volume adalah banyaknya jumlah responden yang mengisi kuesioner yang disebarkan melalui sosial media dalam rangka memperoleh data. *Velocity* adalah penggunaan media Google Form untuk mendapatkan data yang diperoleh dari responden secara cepat dan mudah dengan disebar ke media sosial. *Variety* adalah data yang diberikan oleh masing-masing responden beragam antara satu sama lain. Konsep Big Data dalam topik

ini juga bertujuan agar dapat diolah menjadi suatu informasi yang berguna. Informasi tersebut dapat dikaitkan dengan berbagai bidang, misalnya dalam bidang bisnis, pemerintahan, dan edukasi. Dalam bidang bisnis, misalnya dengan mengetahui durasi olahraga masyarakat yang kurang, terdapat peluang bisnis untuk sewa pinjam alat olahraga seperti *treadmill*, *barbell*, dan sejenisnya bagi masyarakat, sehingga tidak perlu ke *gym* untuk berolahraga yang dapat berisiko terpapar virus, namun dapat berolahraga di rumah yang lebih aman. Untuk pemerintahan, data yang diperoleh dapat digunakan untuk mengetahui domisili masyarakat yang masih kurang dalam durasi olahraga per harinya, sehingga dapat diketahui bahwa pada domisili tersebut kebugaran tubuh masyarakatnya dinilai kurang. Untuk edukasi, data yang diperoleh dapat digunakan untuk memberikan edukasi bahwa halnya durasi olahraga, konsumsi air putih, dan jam tidur yang dinilai kurang agar dapat menjadi perhatian bagi masyarakat untuk diperbaiki.

## 1.2 Pandemi Covid-19

Virus Covid-19 adalah suatu virus yang pertama kali muncul di Wuhan, China pada akhir Desember 2019. Nama resmi virus ini adalah *Severe Acute Respiratory*

*Syndrome Coronavirus 2* (SARS-CoV2) yang dapat menyebabkan penyakit *Coronavirus Disease-2019* (Covid-19). Virus ini berkembang di paru-paru dan bagi orang yang telah terpapar biasanya menimbulkan ciri-ciri seperti demam, mati rasa, batuk, nyeri di tenggorokan, pusing, dan sesak nafas. Namun saat ini ditemukan juga kasus penderita virus Covid-19 yang tidak memberikan gejala apa pun. Penyebarannya yang begitu cepat dan mudah yaitu ditularkan melalui mulut, hidung, dan telinga membuat virus Covid-19 yang telah dinyatakan oleh WHO sebagai pandemi yang telah menyerang lebih dari 183 negara di seluruh dunia.

Tak terkecuali Indonesia, virus ini juga menyerang dengan angka status positif terinfeksi virus Covid-19 terus bertambah dengan angka yang relatif sangat besar. Dilansir dari covid19.go.id, per 17 Oktober 2020, jumlah positif terinfeksi virus Covid-19 adalah 357.762 orang. Berdasarkan kondisi seperti ini yang dapat membahayakan nyawa setiap orang, maka pemerintah Indonesia memberlakukan PSBB (Pembatasan Sosial Berskala Besar) yang membatasi interaksi satu sama lain di luar rumah dan tetap di rumah saja. Berbagai aktivitas terkena dampaknya seperti, bekerja dilakukan di rumah dengan *Work from Home*

(WFH) dan kegiatan belajar mengajar menjadi Pembelajaran Jarak Jauh (PJJ).

Berbagai kegiatan yang dilakukan di rumah saja secara berulang-ulang berpotensi untuk menimbulkan rasa penat. Hal ini dapat memberikan dampak buruk pada masalah kebugaran fisik dan kesehatan mental. Seharusnya di masa pandemi seperti ini, kebugaran fisik haruslah dijaga agar dapat memiliki imunitas tubuh yang baik dan tidak mudah tertular berbagai virus. Salah satu cara untuk menjaga kebugaran fisik dan meningkatkan imunitas tubuh adalah dengan berolahraga. Aktivitas olahraga diperlukan untuk selalu menjaga kebugaran, meningkatkan imunitas, dan juga dapat mengurangi rasa penat. Atas dasar hal tersebut, penelitian ini bertujuan untuk mengetahui bagaimana aktivitas olahraga masyarakat Indonesia dilakukan di masa pandemi.

## **2. BIG DATA ALGORITMA**

### **2.1 Clustering**

Algoritma yang digunakan dalam mendukung analisis adalah *clustering* dan *decision tree*. Algoritma *clustering* adalah algoritma yang bermanfaat dalam proses *data mining*. *Clustering* bertujuan untuk

mengelompokkan data ke dalam beberapa kelompok yang di dalam kelompok tersebut terdapat kemiripan/kesamaan karakteristik antara anggota-anggotanya. Contohnya adalah untuk menentukan pasar yang dimiliki oleh suatu supermarket berdasarkan barang-barang yang dibeli oleh konsumennya. Pengelompokan ke dalam beberapa kluster dapat dilakukan dengan kemiripan jenis barang yang dibeli. Contohnya adalah *cluster* keluarga yang dominan membeli makanan bayi, gula, susu, dan air. Sedangkan kluster customer kluster dengan *income* tinggi dominan membeli *wine*, *swiss cheese*, dan *Belgian chocolate*. Berdasarkan *clustering* yang telah dilakukan, hasil tersebut dapat digunakan untuk berbagai kepentingan. Misalnya dalam bisnis dapat digunakan untuk menentukan strategi marketing terhadap suatu produk kepada pasar yang tepat. Contohnya adalah dengan mengirimkan email yang berisi katalog *wine* kepada customer dengan *income* yang tinggi.

K-Means clustering adalah salah satu contoh dari algoritma clustering yang populer untuk digunakan. K-Means clustering pertama kali diperkenalkan oleh MacQueen pada tahun 1967. K-Means menarik dalam praktiknya karena algoritma ini cukup sederhana dan dapat menampilkan hasil kluster dengan cepat. K-Means termasuk

kategori *machine learning* yang *unsupervised* yang memiliki arti bahwa algoritma ini tidak memiliki suatu field yang digunakan sebagai target prediksi. Konsep *K-Means* adalah membuat beberapa kluster yang di dalamnya memiliki variasi data yang telah diminimalisir.

Untuk cara kerja algoritma, standar dalam penerapan algoritma *K-Means* pertama kali dibuat oleh Hartigan-Wong pada tahun 1979, dimana total variasi dalam suatu kluster adalah jumlah akar kuadrat jarak *Euclidean* antar anggota kluster dengan pusat kluster. Setiap kluster memiliki titik pusat yang bernama titik *centroid* yang dapat berubah secara adaptif, dimulai dari nilai awal yang bernama *seed-point*. *K-Means* menghitung jarak antara suatu nilai dengan titik *centroid* (jarak *Euclidean*), kemudian nilai tersebut dikelompokkan berdasarkan jarak ke titik *centroid* terdekat.

Langkah-langkah untuk menggunakan algoritma K-Means adalah pertama menentukan jumlah kluster yang diinginkan. Kemudian tentukan titik centroid sebanyak jumlah kluster, titik awal tersebut dinamakan *seed-point*. Kemudian hitung jarak euclidian untuk menentukan anggota dari tiap kluster yang berdekatan dengan titik centroid. Proses tersebut dilakukan sebanyak

beberapa kali iterasi. Biasanya programmer menggunakan 25 iterasi dan iterasi maksimal secara default adalah 10.

Berikut adalah rumus untuk menentukan variasi dalam sebuah klaster:

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2$$

- $x_i$  design a data point belonging to the cluster  $C_k$
- $\mu_k$  is the mean value of the points assigned to the cluster  $C_k$

Untuk menentukan variasi dalam seluruh klaster dapat menggunakan rumus:

$$tot.withinss = \sum_{k=1}^k W(C_k) = \sum_{k=1}^k \sum_{x_i \in C_k} (x_i - \mu_k)^2$$

*K-means* memiliki beberapa kelebihan yaitu:

- Cenderung mudah untuk dilakukan dan diimplementasikan. Hal ini yang menjadikan *K-means* populer dalam dunia analisis data.
- Waktu yang cepat dalam proses hasil sehingga sangat efisien waktu.
- Prinsipnya sederhana sehingga dapat digunakan oleh orang banyak.
- Penggunaan *K-means* dapat diaplikasikan pada berbagai penggunaan yang umum. Relatif sangat banyak hal yang dapat dikelompokkan dengan mudah menggunakan *K-means*.

Meskipun *K-means* sudah sangat populer digunakan dalam analisis data, *K-means*

memiliki beberapa keterbatasan yang menjadi kekurangannya yaitu:

- Jumlah klaster harus ditentukan terlebih dahulu.

Hal tersebut memerlukan estimasi yang tepat dan tidak dapat dilakukan oleh orang secara sembarangan. Ada beberapa metode lagi untuk menentukan jumlah klaster seperti *elbow method*, *silhouette method*, dan *gap statistic*.

- Sensitif terhadap *outliers*.

*Outliers* adalah data yang memiliki nilai jauh berbeda daripada nilai lainnya yang terdapat dalam suatu data. *Outliers* dapat menimbulkan *error* atau menghasilkan sebuah klaster dengan suatu nilai data yang jauh berbeda dengan yang lainnya. Solusinya dapat dengan mempertimbangkan untuk menghapus nilai *outliers* tersebut.

- *K-means* tidak cocok untuk digunakan untuk data dengan jumlah dimensi yang terlalu besar.

Contohnya jika terdapat 20 atribut yang akan menghasilkan 20 dimensi maka akan sulit untuk menentukan

pencarian jarak antar nilai dan titik *centroid*.

## 2.2. Decision Tree

Algoritma *decision tree* merupakan algoritma yang digunakan untuk mengambil suatu keputusan berdasarkan tahapan syarat dan kondisi. Ide dalam *decision tree* adalah untuk memecah beberapa proses pengambilan keputusan yang kompleks untuk menjadi beberapa proses pengambilan keputusan yang lebih sederhana. *Decision tree* menggunakan struktur pohon atau struktur berhierarki. Nama lain dari *decision tree* adalah CART (*Classification and Regression Tree*) yang berarti gabungan dari dua jenis pohon yaitu *classification tree* dan *regression tree*. Selain untuk pengambilan suatu keputusan, *decision tree* juga dapat digunakan untuk mengeksplorasi data seperti mengetahui hubungan antara beberapa variabel. Breiman, et al. (1984) mengatakan bahwa *decision tree* merupakan metode yang sangat populer karena hasil dari model yang terbentuk mudah untuk dipahami.

Struktur *decision tree* dibangun dalam metode *topdown* atau mengalir dari atas ke bawah. *Decision tree* terbangun atas beberapa bagian: *Root node*, *decision node*, dan *leaf node*. *Root node* adalah *node* paling atas yang memiliki beberapa cabang keluar

dan tidak memiliki cabang masuk. *Decision node* adalah penghubung antara *root node* dengan *leaf node* dan *leaf node* dengan *leaf node*. *Decision node* biasanya berhubungan dengan suatu pertanyaan atau pengujian untuk melakukan klasifikasi data. *Leaf node* adalah bagian paling bawah atau akhir dari suatu *decision tree*. Dari *leaf node* tersebut, kita dapat mengetahui keputusan akhir dari *decision tree* tersebut.

Cara kerja *decision tree* adalah sebagai berikut:

- Berawal dari *root node* yang berada di paling atas yang mempresentasikan semua data.
- Setelah *root node* terbentuk, maka proses akan berlanjut ke *decision node* yang akan menentukan ke *selection node* berikutnya/*leaf node* berikutnya.
- Proses pertama dan kedua tersebut akan dilakukan secara rekursif (perulangan) hingga semua data telah sampai ke *leaf node*.

Kelebihan algoritma *decision tree* adalah:

- Dapat menyederhanakan proses pengambilan suatu keputusan yang kompleks menjadi lebih sederhana dan spesifik.

- Mengurangi perhitungan yang tidak diperlukan karena dengan *decision tree*, data hanya diuji berdasarkan kriteria atau kelas tertentu.
- Lebih spesifik dibandingkan dengan proses pengambilan keputusan yang dilakukan oleh manusia. Contohnya adalah dalam *decision node*, algoritma *decision tree* yang merupakan *machine learning* dapat memberikan syarat presisi hingga ke nilai desimal dalam angka, sedangkan manusia cenderung membulatkan suatu nilai desimal.

Algoritma *decision tree* memiliki kekurangan:

- Terjadi *overlap* jika terdapat terlalu banyak *decision node* yang digunakan sebagai kriteria. Hal tersebut dapat meningkatkan waktu dan jumlah memori yang diperlukan.
- Keputusan yang dihasilkan oleh *decision tree* sangat bergantung pada kriteria yang terdapat di dalamnya sehingga keputusan dapat berubah-ubah seiring diubahnya kriteria tersebut yang terdapat dalam *decision node*.
- Tidak mudah untuk menghasilkan desain *decision tree* yang optimal.

### 3. METODOLOGI

#### 3.1 Objek Penelitian

Objek penelitian yang saya lakukan adalah aktivitas olahraga yang dilakukan oleh masyarakat Indonesia selama pandemi Covid-19 ini. Pandemi yang menyerang hampir seluruh negara di dunia termasuk Indonesia telah membuat pemerintah menerbitkan berbagai peraturan dan protokol untuk memutus rantai penyebaran virus Covid-19. Pemerintah memberlakukan PSBB (Pembatasan Sosial Berskala besar) yang membuat berbagai aspek kehidupan terkena dampak. Pekerjaan dan pendidikan misalnya, lebih banyak dilakukan di rumah secara daring untuk menghindari kontak fisik dengan orang banyak. Namun tidaklah mudah bagi aktivitas lainnya untuk juga dilakukan di rumah saja atau secara daring.

Kegiatan pembelajaran jarak jauh (PJJ) dan *work from home* (WFH) adalah masalah yang sering dibahas dan dialami oleh masyarakat dalam masa pandemi ini. Berbagai aktivitas yang membuat tubuh tidak terlalu banyak bergerak rupanya dapat berbahaya pula bagi tubuh. Dilansir dari detik.com, menurut dr. Arie Sutopo, tubuh yang tidak bergerak akan berimbas pada sendi yang akan memberikan dampak yang



tidak sehat pada jantung. Menurutnya, jika tubuh tidak bergerak maka tidak ada nada imunitas tubuh yang baik sehingga perlu bergerak dengan olahraga. Olahraga dapat dilakukan di luar maupun di dalam rumah namun jika dilakukan di luar rumah haruslah mengikuti protokol kesehatan.

Olahraga merupakan kegiatan yang positif bagi kesehatan yaitu untuk menjaga kebugaran tubuh dan dapat meningkatkan imunitas tubuh. Di masa pandemi yang cenderung lebih banyak orang untuk melakukan berbagai aktivitas di rumah, kegiatan olahraga tidak bisa dikesampingkan. Fenomena baru yang harus dihadapi untuk berolahraga menjadi suatu masalah yang menarik untuk dibahas. Bagaimana dengan aktivitas olahraga jika di masa pandemi lebih banyak di rumah saja? Lalu apakah masyarakat Indonesia di masa pandemi ini semakin rutin berolahraga atau meninggalkannya? Berbagai pertanyaan tersebut merupakan hal yang akan menjadi topik dan dibahas dalam penelitian ini.

### **3.2 Metode Pengumpulan Data**

Metode pengumpulan data dilakukan untuk mendapatkan informasi guna mencapai tujuan dalam penelitian. Dalam rangka mengetahui aktivitas olahraga yang dilakukan oleh masyarakat Indonesia di masa

pandemi ini, data diperoleh dengan menyebar kuesioner via Google Form. Kuesioner adalah kumpulan beberapa pertanyaan yang diajukan dengan pertanyaan yang telah didesain sedemikian rupa untuk keperluan penelitian. Google Form merupakan formulir digital yang dapat diakses untuk edit dan diisi oleh responden secara *online* melalui berbagai *device* seperti komputer, *smartphone*, dan tablet. Berdasarkan cara memperoleh data, kuesioner tersebut merupakan data primer yang dikumpulkan dan diolah sendiri oleh peneliti langsung terhadap subjek penelitian yaitu masyarakat Indonesia.

Kuesioner tersebut dilakukan dengan menggunakan Google Form agar sesuai dengan konsep Big Data yaitu volume, *velocity*, dan *variety*. Data yang diperoleh memiliki target volume yang luas karena mencakup masyarakat Indonesia tanpa Batasan umur, jenis kelamin, pendidikan, dan sebagainya. Kuesioner juga disebar dengan menggunakan sosial media untuk mendukung nilai *velocity*. Keberagaman (*variety*) yang ditemukan dalam data yang diperoleh terdapat pada berbagai pertanyaan yang terdapat dalam kuesioner yaitu kategorikal dan numerik.

### **3.3 Metode Penelitian**

Metode penelitian menurut Prof. Dr. Sugiyono (2010) adalah langkah ilmiah untuk mendapatkan data dengan tujuan dan kegunaan tertentu. Metode penelitian yang digunakan adalah deskriptif kuantitatif. Metode deskriptif kuantitatif berarti bahwa penelitian ini bertujuan untuk menggambarkan objek dan subjek dalam kondisi yang apa adanya, dengan tujuan menggambarkan secara fakta dan karakteristik objek yang diteliti secara tepat. Metode ini juga berhubungan dengan korelasi dan asosiatif berdasarkan hubungan antar variabelnya. Dalam arti, penelitian ini mengukur tingkat suatu variabel pada suatu populasi yang diwakilkan dengan *sample*. Korelasi dan asosiatif memiliki arti bahwa penelitian ini menunjukkan hubungan antara variabel- variabel yang terkait.

### **Unit Analisis**

- Subjek penelitian, adalah seluruh masyarakat Indonesia yang saat ini sedang terdampak kondisi pandemi Covid-19 di Indonesia.
- Objek penelitian, aktivitas olahraga di masa pandemi.

### **Populasi dan Sampel**

Populasi yang terdapat dalam penelitian ini adalah seluruh masyarakat Indonesia yang saat ini sedang mengalami

kondisi pandemi Covid-19. Berdasarkan waktu pengumpulannya, kuesioner yang disebar ke masyarakat umum dilakukan pada tanggal 15 September 2020 sampai 30 September 2020 yang menghasilkan 40 responden. Responden tersebut digunakan dalam teknik sampling terhadap masyarakat Indonesia. Teknik sampling teknik sampling yang digunakan adalah *simple random sampling*. Menurut Sugiyono (2001), teknik *simple random sampling* atau yang disebut teknik acak sederhana adalah teknik mengambil *sample* dari suatu anggota populasi yang dilakukan secara acak tanpa memperhatikan strata dalam populasi itu. Dalam kata lain, setiap anggota populasi yang dalam penelitian ini adalah seluruh masyarakat Indonesia, memiliki kesempatan yang sama untuk terpilih sebagai *sample*.

### **Variabel Penelitian**

Variabel yang digunakan dalam penelitian ini adalah variabel bebas dan terikat.

Variabel bebas yang digunakan:

- Umur
- Domisili
- Penting/tidaknya olahraga
- Jumlah konsumsi air putih

Variabel terikat yang digunakan:

- Jenis Olahraga
- Durasi Olahraga
- Berat badan setelah olahraga
- Durasi istirahat selama olahraga

### **Instrumen Penelitian**

Penelitian ini menggunakan Google Form sebagai media untuk menyampaikan kuesioner kepada masyarakat. Kemudian data yang diperoleh diolah dalam Microsoft Excel dan dianalisis dengan Program R dan R Studio.

### **Teknik Analisis Data**

Selanjutnya data yang telah diperoleh dari kuesioner tersebut diolah dalam aplikasi pengolah data yaitu Microsoft Excel untuk kemudian dapat dikelola lebih lanjut. Penelitian selanjutnya diolah dengan menggunakan bahasa pemrograman R dengan program R Studio. R adalah bahasa pemrograman dan perangkat lunak untuk analisis statistika dan grafik yang dibuat oleh Ross Ihaka dan Robert Gentleman dan kemudian dilanjutkan oleh R Development Core Team. R cocok untuk melakukan

pemodelan data baik numerik maupun kategorikal. Data yang diperoleh dari kuesioner mengandung data numerik dan kategorikal. Penggunaan R adalah karena R merupakan salah satu platform yang paling populer untuk digunakan dalam *data science* dan baik untuk melakukan visualisasi data.

Dalam menggunakan R, penelitian ini menggunakan beberapa metode yang disebut dengan algoritma. Algoritma adalah deretan instruksi yang jelas untuk memecahkan suatu masalah yaitu menghasilkan suatu keluaran yang bermanfaat dari masukkan yang diberikan. Algoritma yang diaplikasikan untuk mendukung penelitian ini adalah algoritma *clustering* dan *decision tree*. Penggunaan kedua algoritma tersebut dinilai sesuai untuk mendukung tujuan dari penelitian ini.

## **4. DATA ANALISIS**

### **4.1 Validasi Data**

Validasi data bertujuan untuk memeriksa kebenaran dan kelengkapan data.

```
#Membaca data~
library(readxl)~
data <- read_excel("SurveyKegiatanOlahraga.xlsx")~

#Cek kebenaran dan kelengkapan data~
library(psych)~
describe(data)~
```

	vars <dbl>	n <dbl>	mean <dbl>	sd <dbl>	median <dbl>	trimmed <dbl>	mad <dbl>	min <dbl>	max <dbl>	
Timestamp	1	40	NaN	NA	NA	NaN	NA	Inf	-Inf	
Umur	2	40	18.80	1.07	19	18.72	1.48	17	22	
Domisili*	3	40	1.07	0.27	1	1.00	0.00	1	2	
OlahragaPenting*	4	40	1.95	0.22	2	2.00	0.00	1	2	
DurasiOlahraga	5	40	22.25	36.59	4	13.75	5.93	0	150	
JenisOlahraga*	6	40	4.12	1.62	5	4.22	1.48	1	6	
BBOlahraga	7	40	-1.30	8.84	0	-1.06	0.00	-26	40	
IstOlahraga	8	40	1.94	2.59	1	1.36	1.48	0	12	
MLAirMinum	9	40	623.08	602.37	500	528.84	733.89	0	2000	

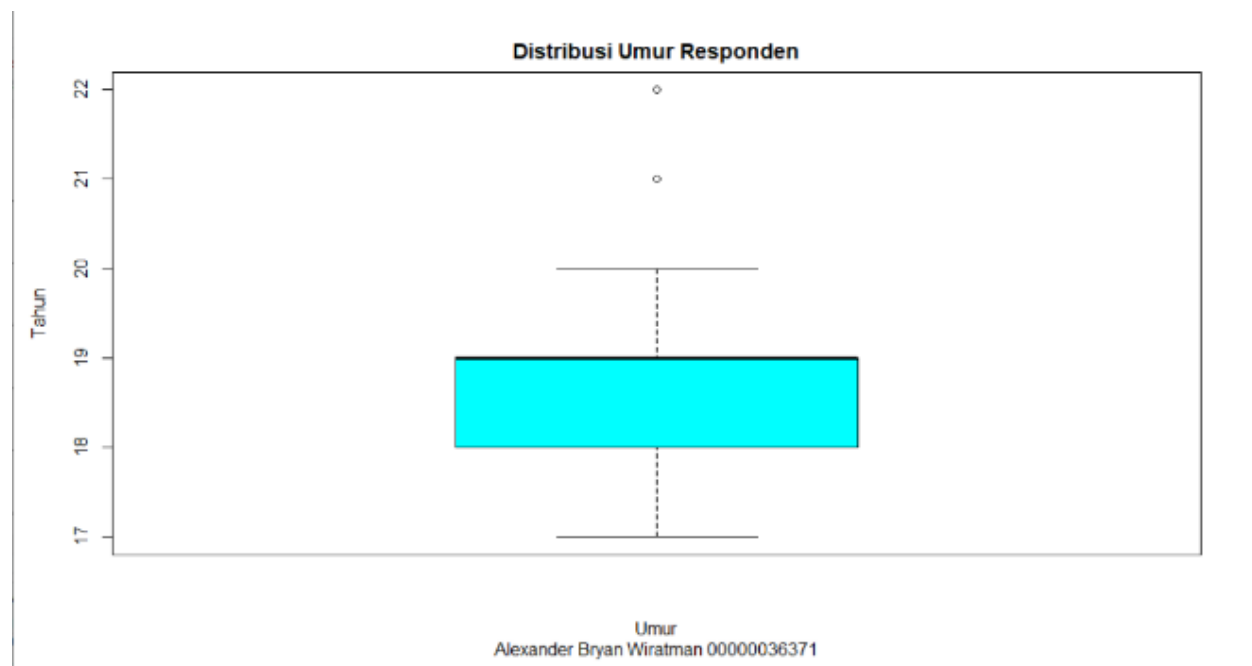
9 rows | 1-10 of 13 columns

**Tabel 4.1.1** Validasi Data

Berdasarkan hasil tersebut, dapat diketahui bahwa hasil penelitian memiliki variabel kategorikal (yang diakhiri tanda \*) dan variabel numerik. Dapat diketahui bahwa seluruh variabel adalah lengkap yang ditandai dengan jumlah tiap variabel adalah

40 sesuai dengan jumlah responden dan mean untuk setiap variabel numerik tidaklah 0 yang menunjukkan bahwa tidak ada data yang kosong.

## 4.2 Visualisasi Data

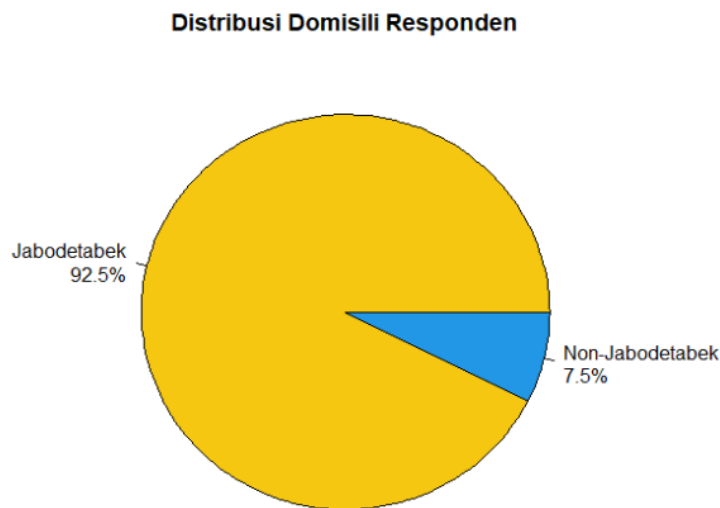


```
#visualisasi data
boxplot(data$umur,
  xlab="Umur",
  ylab="Tahun",
  main="Distribusi Umur Responden",
  sub="Alexander Bryan Wiratman 00000036371",
  col="cyan")
```

**Grafik 4.2.1** Distribusi Umur Responden.

Dapat diketahui dari grafik 4.2.1 bahwa distribusi umur responden memiliki

rata-rata 19 tahun dengan 2 *outliers* yaitu usia 21 tahun dan 22 tahun.



Alexander Bryan Wiratman 00000036371

```
datadomisili <- table(data$Domisili)
lbls <- paste(names(datadomisili), "\n", round(100*datadomisili/sum(datadomisili), 1), "%",
  sep="")
pie(datadomisili,
  labels=lbls,
  main="Distribusi Domisili Responden",
  sub="Alexander Bryan Wiratman 00000036371",
  col=c(7,4))
```

**Grafik 4.2.2** Distribusi Domisili Responden.

Dapat dilihat pada grafik 4.2.2 di atas bahwa 92,5 % responden berdomisili di

daerah Jabodetabek sedangkan terdapat 7,5% responden yang berdomisili di luar Jabodetabek.



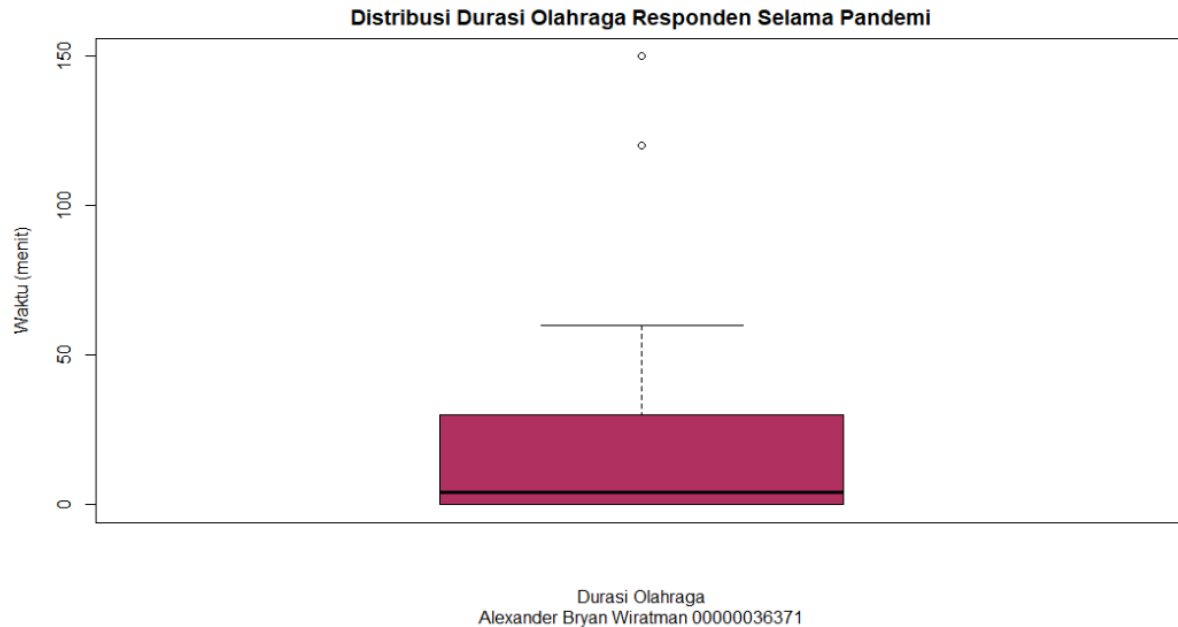
Alexander Bryan Wiratman 00000036371

```
datapenting <- table(data$olahragaPenting)~
lbls <- paste(names(datapenting), "\n", round(100*datapenting/sum(datapenting), 1), "%",
sep="")~
pie(datapenting,~
labels = lbls,~
main = "Anggapan Bahwa Olahraga itu Penting",~
sub = "Alexander Bryan Wiratman 00000036371",~
col = c(6,5))~
```

**Grafik 4.2.3** Anggapan Bahwa Olahraga itu Penting.

Dari grafik 4.2.3, dapat diketahui bahwa 95% responden beranggapan bahwa olahraga itu penting kaitannya dengan Kesehatan sedangkan masih terdapat 5%

responden yang belum menganggap olahraga sebagai suatu hal yang penting untuk dilakukan.

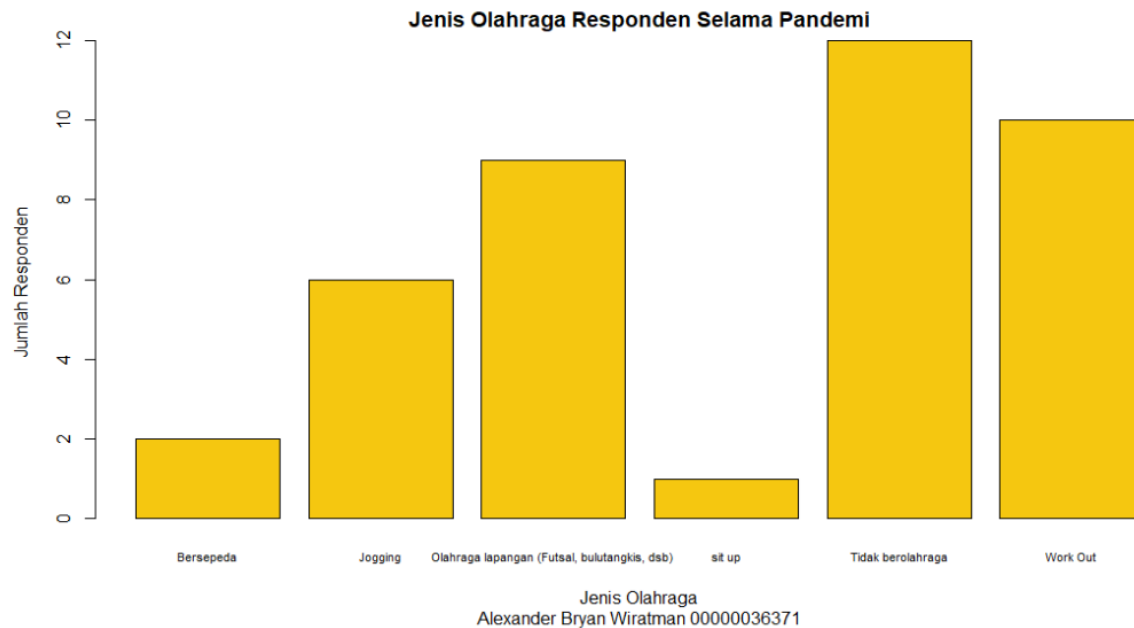


```
boxplot(data$DurasiOlahraga,~
  . . . . . xlab = "Durasi Olahraga",~
  . . . . . ylab = "waktu (menit)",~
  . . . . . main = "Distribusi Durasi Olahraga Responden Selama Pandemi",~
  . . . . . sub = "Alexander Bryan Wiratman 00000036371",~
  . . . . . col = "maroon")~
```

**Grafik 4.2.4** Distribusi Durasi Olahraga Responden Selama Pandemi.

Dari grafik 4.2.4 tersebut, dapat diketahui bahwa durasi olahraga responden memiliki rata-rata yang rendah dan mendekati 0. Hanya terdapat 2 *outliers* yang

berolahraga dengan durasi di atas 100 menit dalam sehari. Dapat ditarik kesimpulan bahwa responden masih sangat sedikit dalam durasi berolahraga sehari-seharinya.



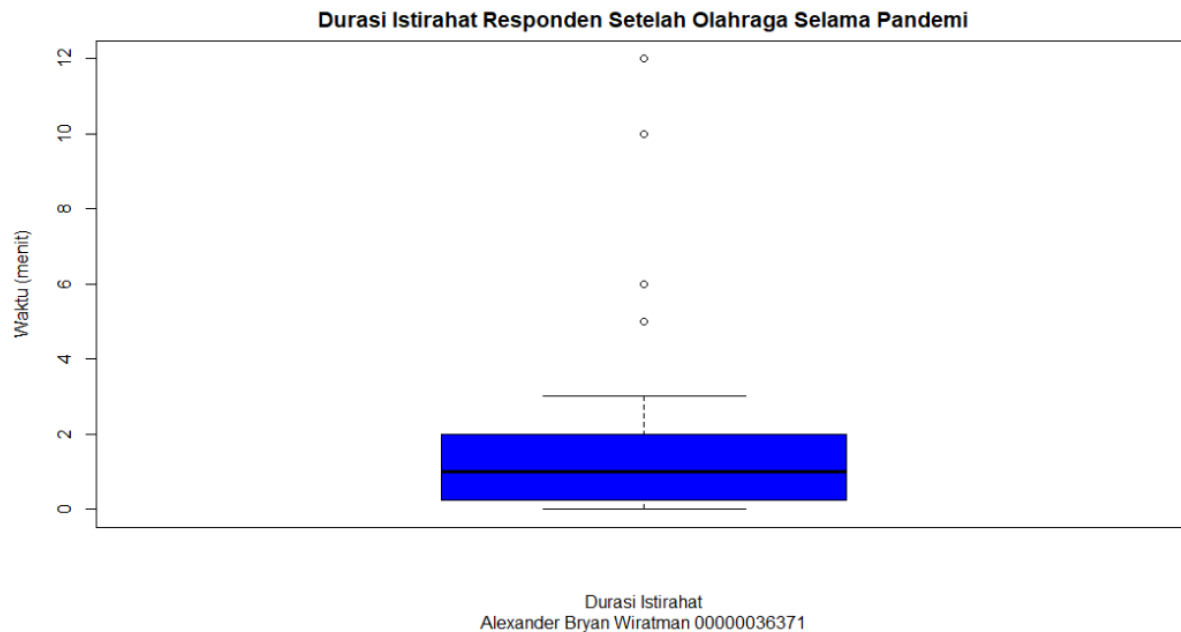
```
datajenisolahraga <- table(data$jenisolahraga)~
view(datajenisolahraga)~
plot(data$jenisolahraga,~
  . . . . . xlab = "Jenis Olahraga",~
  . . . . . ylab = "Jumlah Responden",~
  . . . . . main = "Jenis Olahraga Responden Selama Pandemi",~
  . . . . . sub = "Alexander Bryan Wiratman 00000036371",~
  . . . . . col = 15, cex.names=0.7)~
```

**Grafik 4.2.5** Jenis Olahraga Responden Selama Pandemi.

Berdasarkan grafik 4.2.5, responden yang tidak berolahraga berada di posisi paling atas dalam jumlah, sedangkan olahraga jenis *sit-up* paling sedikit dilakukan

oleh responden. Dari grafik ini juga mendukung kesimpulan bahwa responden masyarakat Indonesia jarang berolahraga.



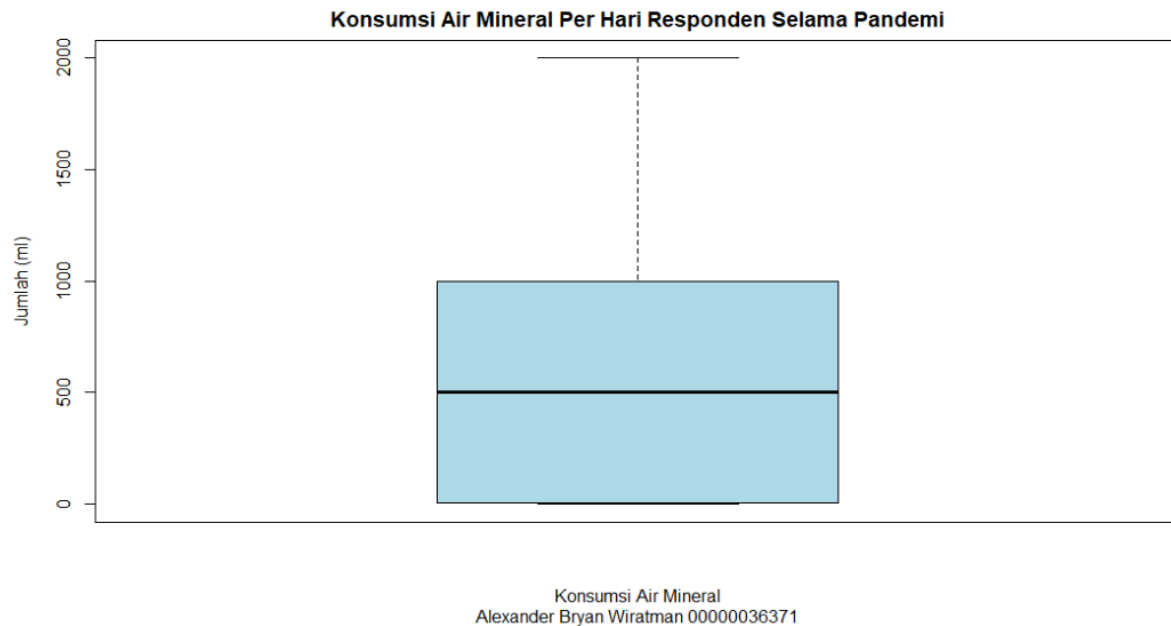


```
boxplot(data$IstOlahraga,~
  . . . . . xlab = "Durasi Istirahat",~
  . . . . . ylab = "waktu (menit)",~
  . . . . . main = "Durasi Istirahat Responden Setelah Olahraga Selama Pandemi",~
  . . . . . sub = "Alexander Bryan Wiratman 00000036371",~
  . . . . . col = "blue")~
```

**Grafik 4.2.6** Durasi Istirahat Responden Setelah Olahraga Selama Pandemi.

Berdasarkan grafik 4.2.6, dapat diketahui bahwa durasi istirahat setelah berolahraga berada pada rata-rata kurang dari

2 menit. Terdapat 4 *outliers* dengan total waktu istirahat setelah berolahraga selama pandemi di atas 4 menit.



```
boxplot(data$MIAirMinum,~
  . . . . . xlab = "Konsumsi Air Mineral",~
  . . . . . ylab = "Jumlah (ml)",~
  . . . . . main = "Konsumsi Air Mineral Per Hari Responden Selama Pandemi",~
  . . . . . sub = "Alexander Bryan Wiratman 00000036371",~
  . . . . . col = "light blue")~
```

**Grafik 4.2.7** Konsumsi Air Mineral Per Hari Responden Selama Pandemi.

Berdasarkan grafik 4.2.7, dapat diketahui bahwa konsumsi air mineral responden berada pada sekitar 500 ml per harinya dengan tidak adanya *outliers*.

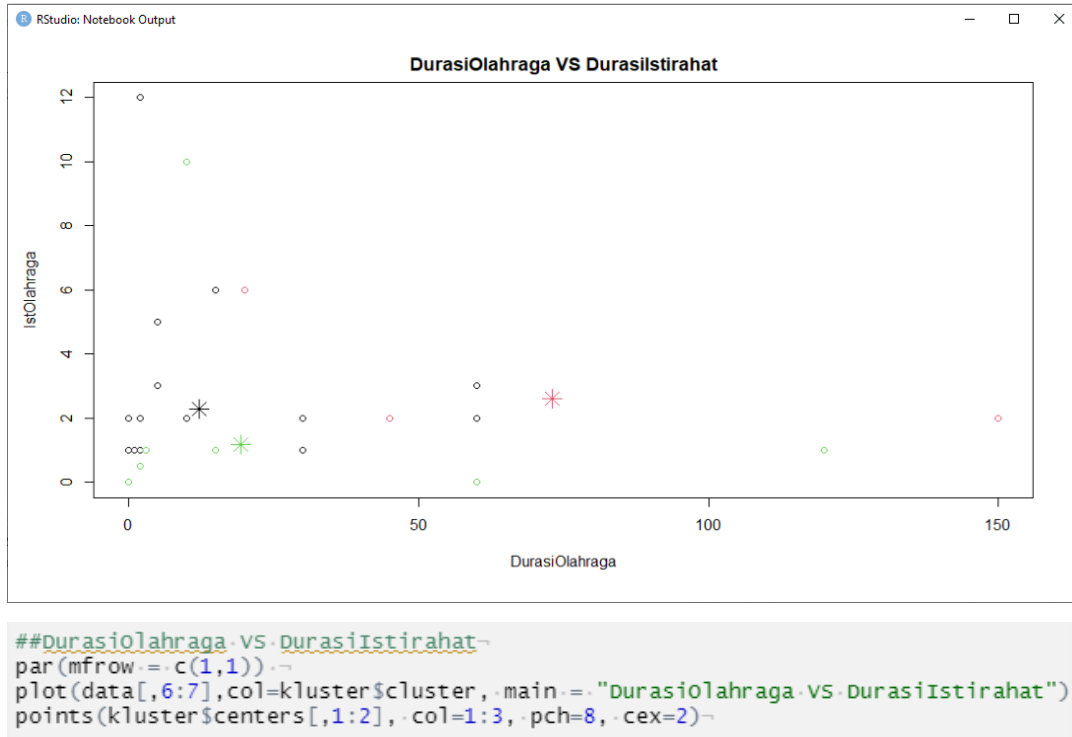
### 4.3 Penerapan Clustering

Penerapan clustering pada data yang berasal dari penelitian yang dilakukan oleh penulis adalah terhadap variabel DurasiOlahraga, BBOlahraga, IstOlahraga, dan MIAirMinum dengan keterangan sebagai berikut:

- DurasiOlahraga  
Menunjukkan durasi lama berolahraga yang dilakukan responden dalam satuan menit.
- IstOlahraga  
Menunjukkan durasi istirahat yang dilakukan responden dalam satuan jam.
- MIAirMinum  
Menunjukkan banyaknya konsumsi air minum (air mineral) dengan satuan ml (mililiter) dalam sehari.

Pertama akan dilakukan pembuatan kluster dengan *k-means*. Kluster dilakukan terhadap variabel DurasiOlahraga dengan

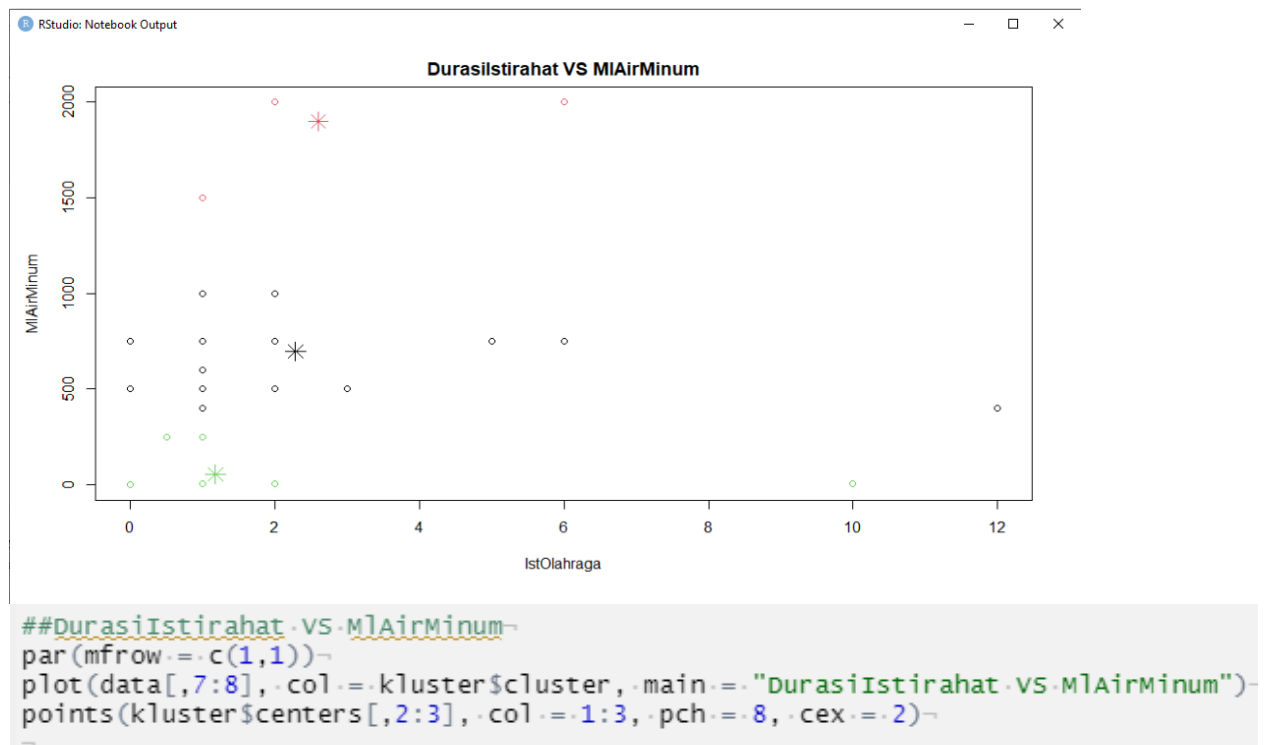
DurasiIstirahat serta variabel DurasiIstirahat VS MIAirMinum.



**Gambar 4.3.1** DurasiOlahraga VS DurasiIstirahat

Dari gambar 4.3.1 di atas, dapat diketahui bahwa persebaran data durasi olahraga dan durasi istirahat tersebar secara

acak dan belum terlihat adanya pengelompokan kluster yang jelas.



**Gambar 4.3.2** DurasiIstirahat VS MIAirMinum

Dari gambar 4.3.2 tersebut, persebaran data durasi istirahat dan jumlah konsumsi air minum tersebar secara acak dan belum terlihat adanya pengelompokan kluster yang jelas.

Terdapat 3 metode populer yang dapat digunakan untuk menentukan jumlah *cluster* yang optimum, yaitu:

- *Elbow method*
- *Silhouette method*
- *Gap statistic*

### **Elbow Method**

Penggunaan *elbow method* dalam menentukan jumlah kluster optimum:

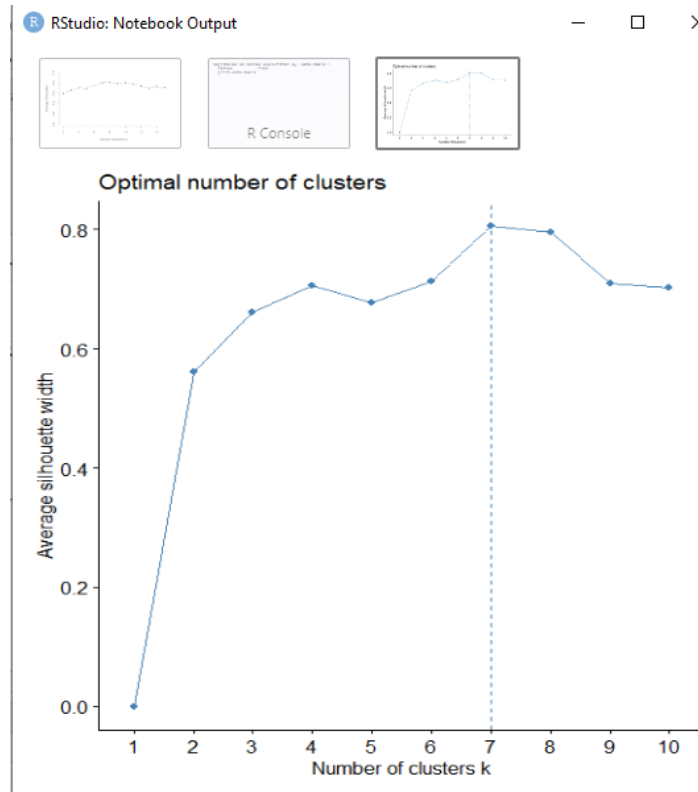


**Gambar 4.3.3 Elbow Plot**

Berdasarkan *elbow plot* pada gambar 4.3.3, dapat dikatakan bahwa jumlah kluster optimum menurut *elbow plot* adalah 3 karena nilai *tot.withinss* sudah semakin melandai.

### Silhouette Plot

Selanjutnya jumlah kluster optimum menurut *silhouette plot*:



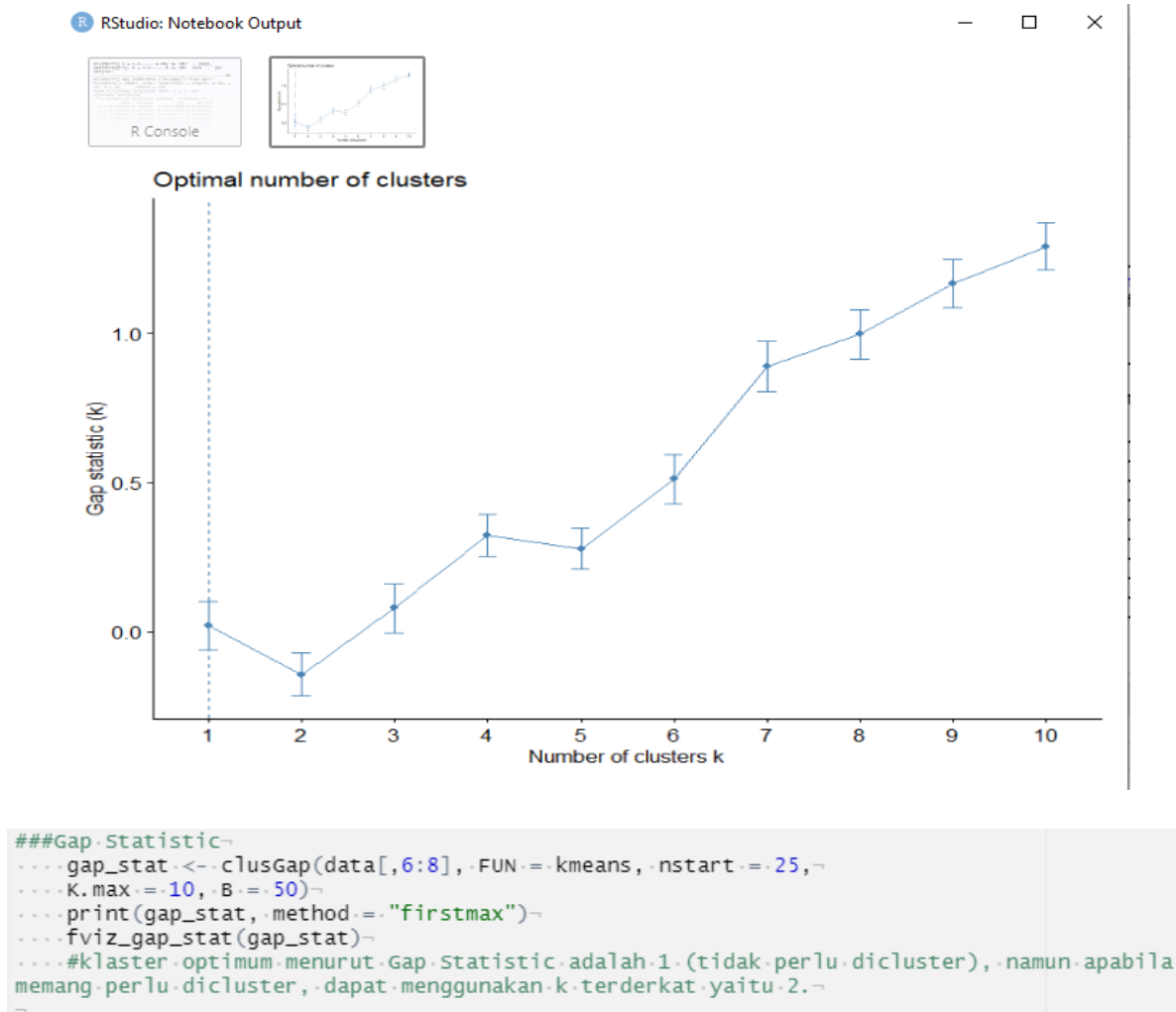
```
###Silhouette Plot
...library(cluster)~
...avg_sil <- function(k) {~
...  km.res <- kmeans(data[,6:8], centers = k, nstart = 25)~
...  ss <- silhouette(km.res$cluster, dist(data[,6:8]))~
...  mean(ss[,3])~
...}~
...k.values <- 2:15~
...avg_sil_values <- purrr::map_dbl(k.values, avg_sil)~
...plot(k.values, avg_sil_values,~
...  type = "b", pch = 19, frame = FALSE,~
...  xlab = "Number of clusters K",~
...  ylab = "Average silhouettes", ylim=c(0,1))~
...factoextra::fviz_nbclust(data[,6:8], kmeans, method = "silhouette")~
...#klaster optimum menurut silhouette plot adalah 7~
```

**Gambar 4.3.4 Silhouette Plot**

Berdasarkan *silhouette plot* pada gambar 4.3.4, didapatkan jumlah kluster optimum yang ditunjukkan oleh garis putus-putus yaitu adalah 7.

### Gap Statistic

Metode yang terakhir akan digunakan adalah *gap statistic* yang menghasilkan:



**Gambar 4.3.5** *Gap Statistic*

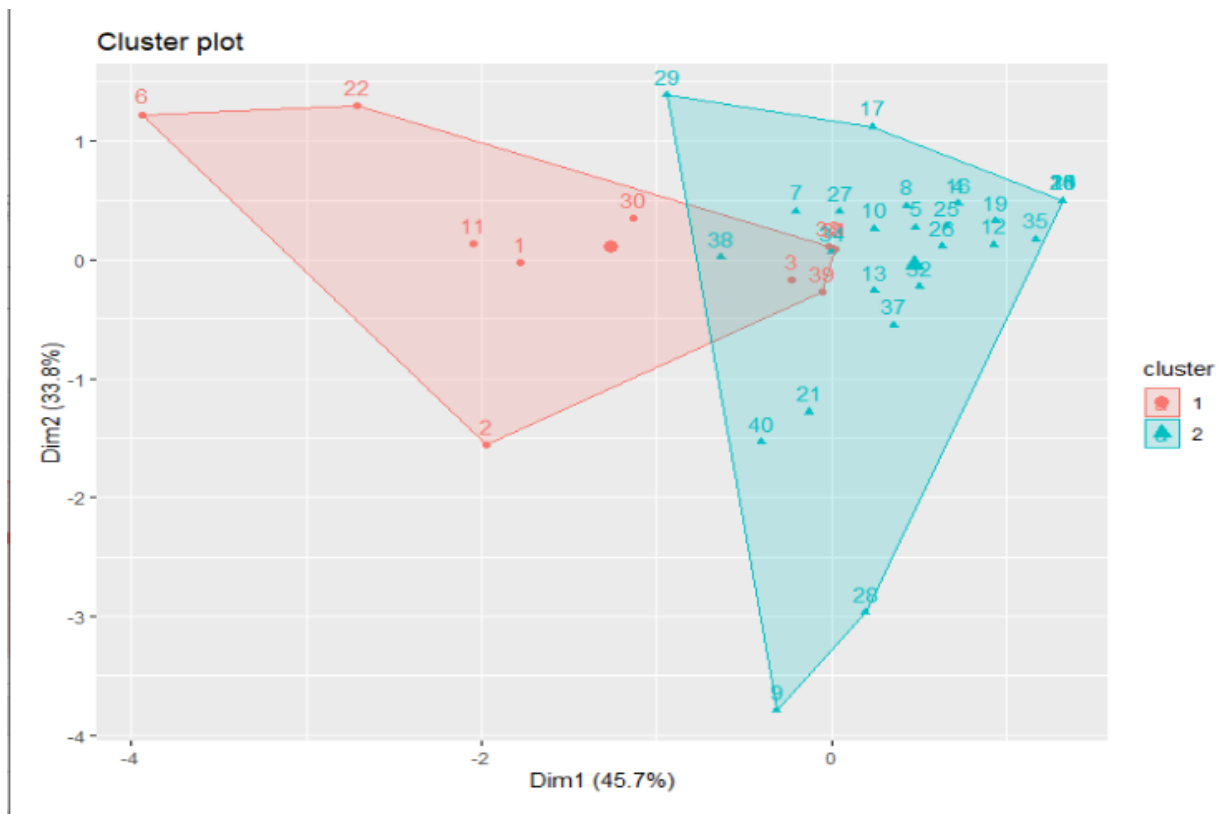
Berdasarkan *gap statistic* pada gambar 4.3.5, jumlah kluster optimum ditunjukkan oleh garis putus-putus yaitu adalah 1 (tanpa kluster), namun jika memang perlu di klaster, dapat menggunakan k terdekat yaitu 2.

### Pilihan Jumlah Kluster

Berdasarkan analisis jumlah kluster optimum dengan metode elbow plot, *silhouette plot*, dan *gap statistics*, dipilihlah jumlah kluster optimum adalah 2.

### Visualisasi Cluster

Visualisasi dengan 2 klaster:



```
#VISUALISASI CLUSTER
final <- kmeans(data[,6:8], 2, nstart = 25)
print(final)
fviz_cluster(final, data = data[,6:8])
library(dplyr)
data[,6:8] %>%
  mutate(cluster = final$cluster) %>%
  group_by(cluster) %>%
  summarise_all("mean")
```

**Gambar 4.3.6** Visualisasi Klaster

Pada gambar 4.3.6, terdapat 2 klaster yaitu klaster 1 dengan komposisi 45,7% untuk kategori kurang sehat, dan klaster 2 dengan komposisi 33,8% untuk kategori cukup sehat.

#### 4.4 Penerapan Decision Tree

*Classification tree* hanya dapat digunakan bila variabel response berupa

variabel kategorikal. Sehingga dibuatlah variabel kategorikal bernama **\*\*sehat\*\*** dengan 2 value: Cukup dan Kurang. Cukup bila DurasiOlahraga  $\geq 40$  menit dan Kurang bila DurasiOlahraga  $< 40$  menit. Variabel ini disimpan dengan tipe factor ke dalam data frame.

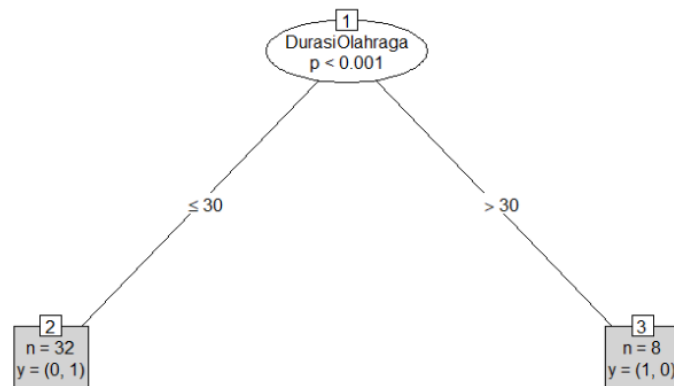


```
#Membuat Variabel Response -
#Kategori cukup jika durasi olahraga >= 40 menit dalam sehari -
data$sehat <- ifelse(data$DurasiOlahraga >= 40, "Cukup", "Kurang")
data$sehat <- as.factor(data$sehat)
```

**Gambar 4.4.1** Variabel Response

Setelah variable response ditentukan, dapat langsung menerapkan algoritma decision tree pada data.

RStudio: Notebook Output



```
library("party") -
ctree <- ctree(sehat~DurasiOlahraga, data=data) -
plot(ctree, type="simple") -
```

**Gambar 4.4.2** Decision Tree

Kesimpulan yang dapat diperoleh dari *decision tree* tersebut adalah terdapat 6 responden dengan kategori durasi olahraga

cukup dan 32 responden dengan kategori olahraga kurang.

## 5. SIMPULAN DAN SARAN

Kesimpulan yang dapat diperoleh dari analisis data yang dilakukan oleh penulis adalah dapat diketahui bahwa Sebagian besar masyarakat Indonesia kurang dalam menerapkan aktivitas olahraga selama pandemi. Meskipun hal tersebut dapat dikatakan wajar karena berbagai keterbatasan dan protocol yang membuat tidak mudah melakukan olahraga terutama di luar rumah seperti biasanya. Namun, tak dapat dipungkiri bahwa olahraga dapat meningkatkan Kesehatan dan kebugaran tubuh yang mana sangat berdampak positif

bagi tubuh dalam menghadapi pandemi virus Covid-19.

Saran dari penulis untuk seluruh masyarakat Indonesia adalah tetap ingat dan peduli akan Kesehatan karena protocol Kesehatan yang baik belum lengkap jika tubuh tidak fit. Oleh karena itu, perlunya olahraga setiap hari selama 40 menit dapat menjaga tubuh agar tetap bugar dan daya tahan tubuh kuat terlebih dalam menghadapi masa pandemi Covid-19 seperti ini.

## DAFTAR PUSTAKA

- Agus, C. S. (2015). Penerapan Teknik Data Mining Untuk Menentukan Hasil Seleksi Masuk SMAN 1 Gibeber Untuk Mahasiswa Baru Menggunakan Decision Tree. TEDC, 9 (1).
- Dong, dkk (2020). An Interactive Web-based Dashboard to Track COVID-19 in Real Time. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1).
- Guha, et al (2000). ROCK: A Robust Clustering Algorithm for Categorical Attributes. Information Systems Vol. 25, No.5, pp. 346-366. [https://doi.org/10.1016/S0306-4379\(00\)00022-3](https://doi.org/10.1016/S0306-4379(00)00022-3).
- Kasih, Patmi (2019). Pemodelan Data Mining Decision Tree dengan Classification Error untuk Seleksi Calon Anggota Tim Paduan Suara. Inovation in Research of Informatics, 1 (2), 63-69.
- Kaushik, M. & Mathur, B. (2014). Comparative Study of K-Means and Hierarchical Clustering Techniques.

- International Journals of Software & Hardware Research in Engineering, volume 2 (6 June 2014).
- Maryanto, Budi (2017). Big Data dan Pemanfaatannya dalam Berbagai Sektor. Media Informatika Vol.16 No. 2.  
<https://doi.org/10.30596/jih.v1i2.5035>.
- Oussous, et al. (2017). Big Data Technologies: A Survey  
<http://dx.doi.org/10.1016/j.jksuci.2017.06.001>.
- Prasetyo, Hoedi & Sutopo Wahyudi (2018). INDUSTRI 4.0: TELAAH KLASIFIKASI ASPEK DAN ARAH PERKEMBANGAN RISET. DOI:10.14710/jati.13.1.17-16.
- R, Gusbakti & Boy, Elman (2020). Edukasi Olah Raga di Rumah Saja Sebagai Upaya Memelihara Kebugaran Fisik dan Kesehatan Mental di Masa Pandemi Covid-19 bagi Mahasiswa Fakultas Kedokteran.  
<https://doi.org/10.30596/jih.v1i2.5035>.
- Rizman, Krista (2008). An Efficient k'-means Clustering Algorithm. Pattern Recognition letters 29 (2008) 1385-1391.  
<https://doi.org/10.1016/j.patrec.2008.02.014>.
- Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. IEEE Transactions on Systems, Man, and Cybernetics, 21(3), 660–674.  
doi:10.1109/21.97458
- Syafrida & Hartati, R (2020). Bersama Melawan Virus Covid-19 di Indonesia. DOI: 10.15408/sjsbs.v7i6.15325.
- Wang, H. & Song, M. Ckmeans.1d.dp: Optimal k-means Clustering in One Dimension by Dynamic Programming. R.J. Author manuscript available in PMC 2016 Dec