

Analisis Survey e-Channel dan e-Lifestyle di Indonesia

Aditya Anandita Dharma Putra (13219403)

Teknik Elektro, Institut Teknologi Bandung

Bandung, Indonesia

aditsareski23@gmail.com

Abstract— Peserta diharapkan mampu untuk memahami cara pengumpulan data, pembersihan data, menganalisis data yang didapat, uji statistik, serta visualisasi statistika melalui hasil survey mengenai e-lifestyle, banking, Fintech dll. Pada tugas ini terdapat empat tahap yang dilalui yakni data screening, data cleaning, data visualisation, dan uji data statistik berupa pencarian confidence interval (Selang Kepercayaan). Beberapa kesimpulan penting yang dapat diambil ialah tingginya jumlah responden dari kalangan 15-20 tahun yang dapat membuat hasil statistik bias dan mayoritas responden yang sudah terbiasa menggunakan aplikasi/media untuk WFH. Telah didapat pula berbagai kesimpulan atas preferensi responden dalam pendataan kegiatan online selama 3 bulan terakhir.

Index Terms— *e-Lifestyle, Channel, Statistika, Confidence Interval, Visualisasi*

----- ◆ -----

1. INTRODUCTION

Dalam menghadapi wabah COVID-19, tentunya diperlukan adaptasi baru terhadap kebijakan yang telah ditetapkan oleh pemerintah. Akibatnya banyak kegiatan-kegiatan yang dialihkan dari luring menjadi daring. Adaptasi ini tentunya akan menciptakan tren baru di masyarakat baik itu mencoba hal-hal baru yang bersifat online atau lebih intensif lagi dalam melaksanakan kegiatan online yang juga telah dilakukan sebelumnya.

Hal ini tentu saja berkaitan erat dengan pengambilan keputusan. Keterbatasan manusia dalam memahami keseluruhan faktor yang mempengaruhi keputusan terkadang membuat keputusan tersebut tidak efektif. Hal ini menjadi sesuatu yang bisa berakibat fatal seperti salah satunya bagi perusahaan dan pengusaha, karena kesalahan pengambilan keputusan akan arah usaha dapat menyebabkan kerugian investasi, modal yang tidak kembali, dan bahkan menyebabkan kebangkrutan.

Dengan keterbatasan ini, diperlukan suatu analisis data yang dapat menerjemahkan kerumitan dan banyaknya data dari populasi yang terjadi menjadi lebih ringkas dan mudah dipahami. Kemudian, menerjemahkan data tersebut menjadi suatu kesimpulan untuk membantu dalam pengambilan keputusan yang lebih baik. Ilmu probabilitas dan statistika menantang kepasrahan manusia terhadap luasnya kejadian yang mungkin terjadi di dunia ini

Tugas besar kali ini difokuskan ke arah statistika yang bertujuan untuk menganalisis data berdasarkan survey yang terdiri dari 68 buah pertanyaan termasuk biodata responden, terbagi ke 5 buah bagian yaitu eLifestyle, eChannel & Digital Banking, eMoney/ePayment, eCom- merce, dan Fintech. Tugas ini memperkenalkan mahasiswa dalam mengumpulkan, membersihkan, menganalisis, menguji, dan melakukan visualisasi data yang telah didapat dalam suatu bentuk yang lebih mudah dipahami pembaca, kemudian menyimpulkan hasil dan menentukan manfaat

hasil statistik tersebut.

2. STUDI PUSTAKA

2.1 Ilmu Statistik

Ilmu statistik adalah ilmu (teknik) untuk mengumpulkan data, menganalisis data, mengambil kesimpulan dan menyajikan berdasarkan data yang berhasil dihimpun untuk menjelaskan dan mendukung keputusan yang dicapai. Ilmu statistik disebut juga statistika yaitu ilmu data dimana melibatkan kegiatan mengumpulkan, mengklasifikasikan, meringkas, mengorganisir, menganalisis, dan menafsirkan informasi numerik.

2.2 Confidence Interval

Salah satu bagian penting dari statistika ialah tingkat kepercayaan terhadap kesimpulan, yang biasa diwujudkan dalam suatu interval kepercayaan. Interval ini menyatakan berapa persen kita mengkehendaki sampel yang diambil benar-benar mewakili keadaan populasi. Confidence interval kerap kali dinyatakan dalam persen, antara 1 – 100% namun yang umumnya dipakai ialah 95% dan 99%.

Dalam percobaan ini, confidence interval yang akan diambil ialah dari segi proporsi, yang berarti akan diambil nilai p dan q , dengan p ialah “peluang” atau proporsi munculnya nilai X (p = jumlah kemunculan X / n (jumlah data) dan q ialah $1 - p$. Adapun rumus confidence interval yang digunakan ialah:

$$p - Z \sqrt{\frac{pq}{n}} < P < p + Z \sqrt{\frac{pq}{n}}$$

Ket:

p = proporsi kemunculan nilai X

P = proporsi populasi sesungguhnya

Z = nilai yang diambil dari tabel distribusi normal

Untuk interval selang kepercayaan 95%, maka digunakan $Z_{0,025} = 1,95996$. Adapun nilai di atas melambangkan batas bawah dan batas atas dari interval tersebut.

2.3 Uji Hipotesis

Bagian penting lainnya dari ilmu

statistika yaitu uji hipotesis dari data yang telah dianalisis. Hipotesis terdapat 2 macam, yaitu hipotesis nol yang berisi tentang dugaan awal kita berupa hasil pasti dan hipotesis alternatif yang berisi tentang perkiraan lain berupa daerah kritis dengan batas yang diambil dari Z tabel.

Dalam percobaan ini, uji hipotesis yang digunakan yaitu uji untuk satu proporsi (sampel besar). Pengujian ini menggunakan Z tabel untuk menentukan daerah kritis dan nilai statistik uji dengan rincian seperti berikut.

Penentuan daerah kritis:

- $Z < -z_{\alpha}$, untuk $H_1: p < p_0$
- $Z > z_{\alpha}$, untuk $H_1: p > p_0$
- $Z < -z_{\frac{\alpha}{2}}$, $Z > z_{\frac{\alpha}{2}}$ untuk $H_1: p \neq p_0$

Rumus Z-score dengan jumlah data yang sama:

$$Z = \frac{(\bar{x} - p_0)n}{\sqrt{np(1-p)}}$$

Ket:

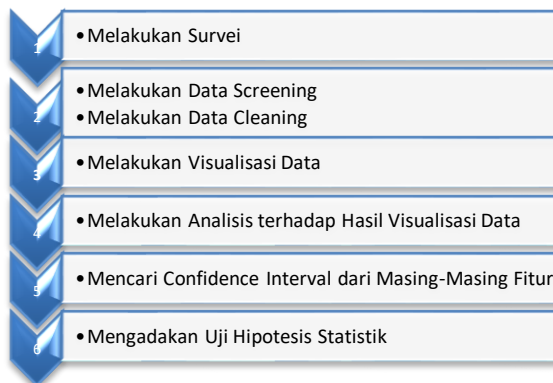
\bar{x} = rata – rata data yang asli

p_0 = hipotesis nol proporsi data

n = jumlah data

3. METODOLOGI

Berikut ialah cara kerja dari tugas besar ini.



4. INITIAL DATA ANALYSIS

4.1 Data Screening

Tahap pertama yang harus dilakukan yaitu data screening. Proses ini dilakukan untuk merapikan, mendefinisikan isi dataframe dan menggabungkan kolom agar data siap untuk di-*cleaning* dan dianalisis lebih lanjut, serta divisualisasikan. Berikut beberapa langkah yang dilakukan untuk

melakukan data screening.

- Penggabungan kolom untuk pilihan jawaban tunggal

Pada survey terdapat pertanyaan dimana responden hanya bisa memilih satu jawaban seperti jenis kelamin, usia, penghasilan, pekerjaan dll. Maka dibutuhkan penggabungan kolom untuk dapat menganalisisnya. Berikut contoh program yang digunakan untuk menggabungkan kolom

```
df["Jenis kelamin ?"] = df["Jenis
kelamin ?"] +
df["Unnamed
: 6"]
```

Perbandingan sebelum dan setelah digabung kolom sebagai berikut.

➤ Sebelum

	Jenis kelamin ?	Unnamed: 6
0	Laki-laki	0
1		1
2		0
3		0
4		0
5		0
6		0
7		0
8		1
9		0

➤ Setelah

	Jenis kelamin ?
1	Laki-laki
2	Perempuan
3	Perempuan
4	Perempuan
5	Perempuan
6	Perempuan
7	Perempuan
8	Laki-laki
9	Perempuan

- Penghapusan kolom yang telah digabung atau yang tidak perlu dianalisis lagi

Setelah kolom digabung, maka kolom yang lainnya akan dihapus karena tidak diperlukan lagi. Pada proses ini akan digunakan fungsi `del` untuk menghapus kolomnya, seperti

```
del df["Unnamed: 6"]
```

- Mendefinisikan angka pada dataframe

Pada hasil survey ditemukan angka 1, 2, 3 pada dataframe yang mana tidak memiliki arti. Oleh karena itu, digunakan fungsi `df.loc` untuk mendefinisikan angka pada tiap dataframe dengan contoh program seperti berikut.

```
df.loc[(df["Jenis kelamin"]==1), "Jenis kelamin"] = "Laki-laki"
df.loc[(df["Jenis kelamin"]==2), "Jenis kelamin"] = "Laki-laki"
```

- Mengganti nama kolom

Mengingat pada saat visualisasi dibutuhkan keterangan yang jelas, maka untuk beberapa dataframe akan digunakan fungsi `df.rename` dengan contoh program seperti berikut.

```
df.rename(columns={"Unnamed: 59": "Pengeluaran belanja kebutuhan internet menggunakan broadband/wifi per bulan"}, inplace=True)
```

4.2 Data Cleaning

Setelah semua kolom digabung dan diperjelas pada data screening, maka dilanjutkan dengan data cleaning. Data cleaning bertujuan untuk membuat data menjadi unik dan tidak ada data yang terduplikasi.

Dalam melakukan survey, data yang didapatkan biasanya belum tentu valid secara keseluruhan sehingga dibutuhkan yang namanya data cleaning. Data cleaning adalah proses dimana data yang didapatkan dari hasil survey dilakukan pengecekan terhadap duplikasi data, kevalidan data dan data kosong. Untuk membersihkan data secara umum, digunakan program python seperti berikut.

```
df = df.drop_duplicates()
```

Setelah itu, dilakukan pembersihan data terhadap Nama Surveyor, Nama Responden, Usia, Pendidikan Terakhir, Pekerjaan, dan Penghasilan per bulan yang dimiliki oleh responden.

```
df = df.drop_duplicates(subset=["Nama Surveyor (Mahasiswa)", "Nama Responden", "Usia", "Pendidikan Terakhir", "Pekerjaan", "Penghasilan per bulan"])
```

```
Responden", "Usia :",
"Pendidikan Terakhir?",
"Pekerjaan?",
"Penghasilan per bulan?"])
```

Kemudian, lakukan pembuangan baris dengan index [0].

```
df = df.drop(df.index[0])
```

Data cleaning kemudian dilakukan dengan membuang semua baris data yang tidak menjawab pertanyaan wajib, seperti: nama responden, jenis kelamin, usia, pendidikan terakhir, penghasilan per bulan, dan domisili. Semua data yang NaN dan 0 akan dihapus barisnya dengan program seperti berikut.

```
df = df[(df["Nama Surveyor (Mahasiswa)"] != na) &
(df["Nama Surveyor (Mahasiswa)"] != 0)]
```

Berikut adalah perbandingan jumlah data setelah di-cleaning.

- Sebelum

```
In [200]: len(df)
```

```
Out[200]: 2021
```

- Sesudah

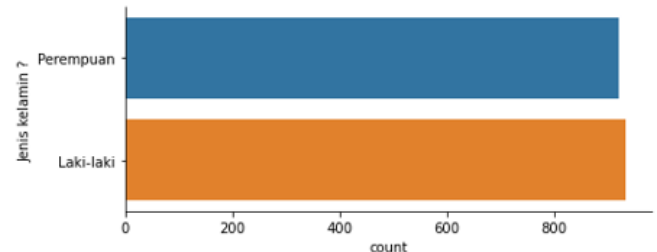
```
In [33]: len(df)
```

```
Out[33]: 1860
```

5. EXPLORATORY DATA ANALYSIS

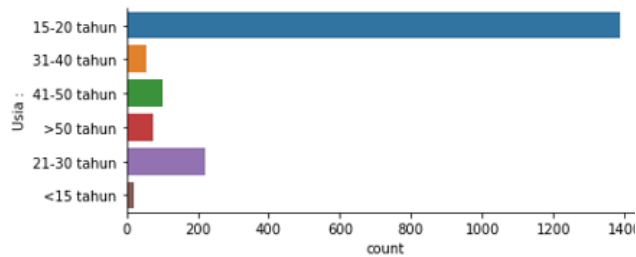
5.1 Visualisasi Data 1 Variabel

5.1.1 Jenis Kelamin



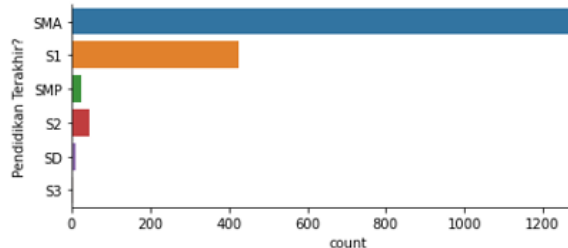
Dari tabel di atas dapat dilihat bahwa responden terdiri dari pria dan wanita yang jumlahnya kurang lebih sama dengan jumlah responden laki-laki lebih banyak sedikit dibandingkan responden perempuan.

5.1.2 Usia



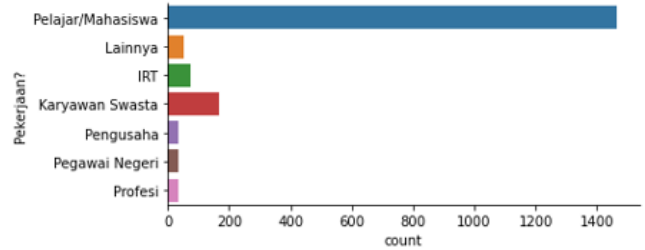
Dari tabel di atas, dapat dilihat bahwa sebagian besar responden adalah remaja yang berusia sekitar 15-20 tahun. Jika dikaitkan dengan topik yang dilampirkan pada survey, tentu saja responden dengan umur 15-20 tahun adalah sasaran utama dalam tugas ini karena variabel yang akan dihasilkan tentunya lebih beragam dibanding responden yang berumur lebih dari 30 tahun.

5.1.3 Pendidikan Terakhir



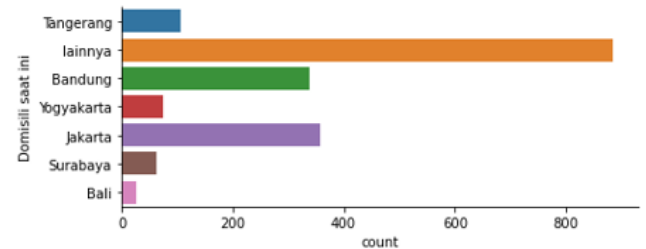
Berkaitan dengan pada tabel sebelumnya (usia responden), berdasarkan tabel ini juga dapat dilihat bahwa sebagian besar (lebih dari 1200 dari total 1738 atau 70%) responden adalah lulusan SMA yang sebagian besar berusia usia 15-20 tahun dan jumlah responden lulusan S1 kurang dari 1/3 nya. Sementara responden lulusan S2, SMP, SMA, dan S3 adalah kelompok responden berdasarkan pendidikan terakhir yang minoritas.

5.1.4 Pekerjaan



Pada tabel ini dapat dilihat bahwa sebagian besar responden (lebih dari $\frac{1400}{1860} \times 100\% = 75\%$) adalah pelajar atau mahasiswa. Jika dilihat dari 2 tabel sebelumnya, hal ini sesuai karena mengingat bahwa kebanyakan responden adalah dari rentang usia 15-20 tahun, dan lulusan SMA.

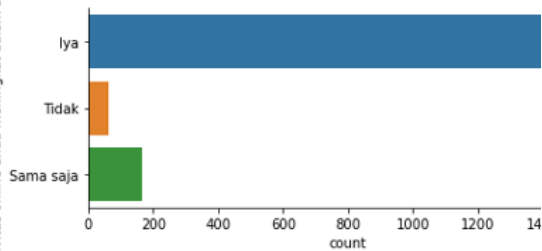
5.1.5 Domisili



Berdasarkan diagram di atas dapat kita simpulkan bahwa sebagian besar responden berasal dari luar kota Tangerang, Bandung, Yogyakarta, Jakarta, Surabaya dan Bali. Namun, kurang efektifnya data domisili lainnya yaitu kita tidak dapat memastikan secara pasti domisili manakah yang paling dominan pada survey.

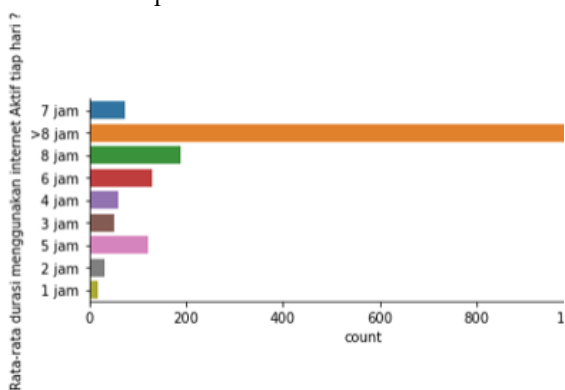
5.1.6 Apakah aktivitas online meningkat dalam 3 bulan terakhir

Apakah aktivitas online anda meningkat dalam 3 bulan terakhir?



Berdasarkan gambar di atas, dapat disimpulkan bahwa wabah pandemi COVID-19 memaksa kita untuk beradaptasi dengan kondisi sehingga menyebabkan lebih dari 75% ($\frac{1400}{1860} \times 100\%$) responden menjawab aktivitas online mereka meningkat karena kuliah online, WFH, dan sebagainya. Di sisi lain, juga terdapat responden yang aktivitasnya tidak terdampak COVID alias aktivitas onlinenya sama saja, bahkan ada juga yang menurun.

5.1.7 Rata-rata durasi menggunakan internet Aktif tiap hari



Berkaitan dengan poin sebelumnya, akibat dari wabah COVID-19 adalah kenaikan durasi rata-rata penggunaan internet aktif yang hasilnya didominasi lebih dari 8 jam

5.1.8 Akses atau provider internet yang digunakan selama ini

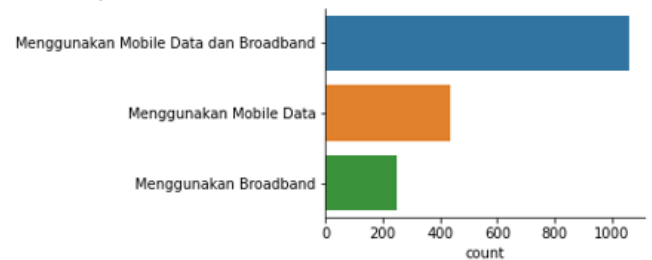
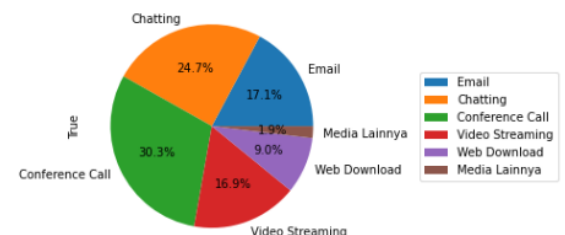


Diagram diatas menunjukkan akses atau provider internet yang digunakan responden. Sebagian besar responden menggunakan mobile data sekaligus broadband. Jika dikaitkan dengan kehidupan serba online sekarang, tentunya hal ini relatif relevan mengingat lebih murah menggunakan broadband karena tidak ada batasan kuota dan bayaran tetap tiap bulan, kemudian dikombinasikan dengan mobile data jika sedang berada di luar jangkauan wifi/broadband.

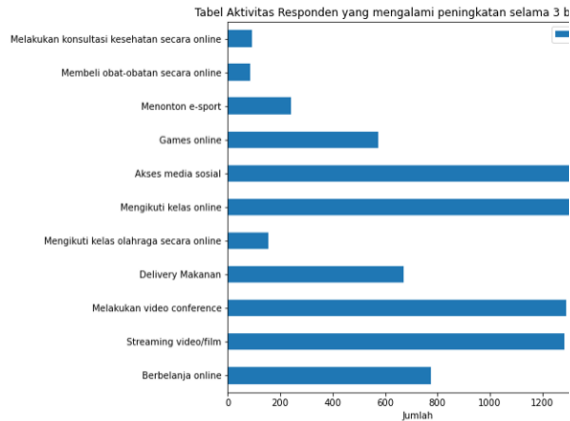
5.1.9 Media Komunikasi yang Digunakan Responden untuk Work From Home

Proporsi Media Komunikasi yang Digunakan Responden untuk Work From Home



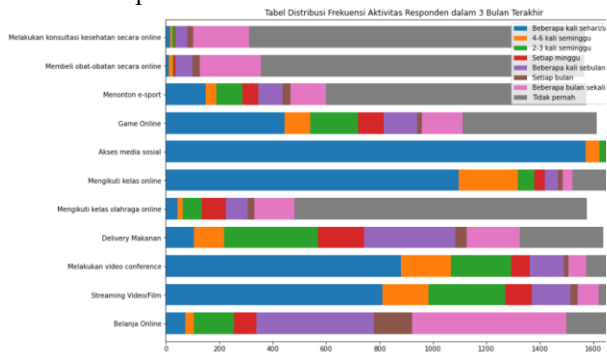
Berdasarkan diagram, didapatkan bahwa media komunikasi yang paling banyak digunakan oleh responden selama WFH/SFH adalah Conference call dan diurutkan kedua yaitu chatting. Sedangkan media komunikasi lainnya sangat sedikit digunakan yaitu hanya sekitar 1.9%

5.1.10 Aktivitas responden yang mengalami peningkatan selama 3 bulan terakhir



Berdasarkan tabel di atas, dapat disimpulkan bahwa kegiatan yang meningkat selama online yang paling banyak dialami oleh responden adalah mengikuti kelas online, kemudian diikuti oleh akses media sosial, streaming video/film dan melakukan video conference. Sedangkan, kegiatan yang paling sedikit dialami peningkatannya oleh responden adalah membeli obat-obatan secara online

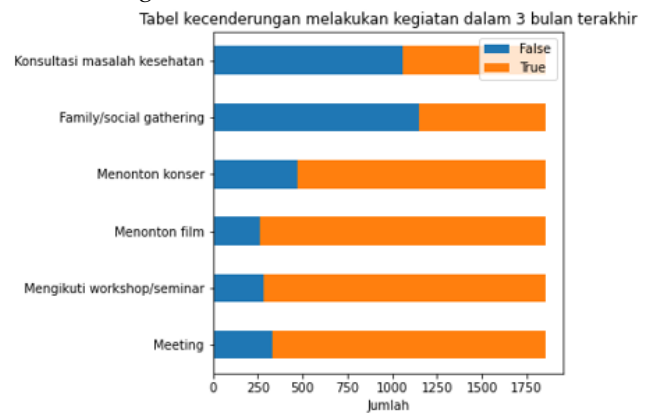
5.1.11 Distribusi frekuensi aktivitas responden dalam 3 bulan terakhir



Grafik di atas menunjukkan frekuensi responden melakukan aktivitas-aktivitas sebelumnya. Sekilas langsung terlihat bahwa mengakses media sosial adalah hal yang dilakukan hampir semua responden setiap hari. Kemudian diikuti dengan kegiatan mengikuti kelas online, video

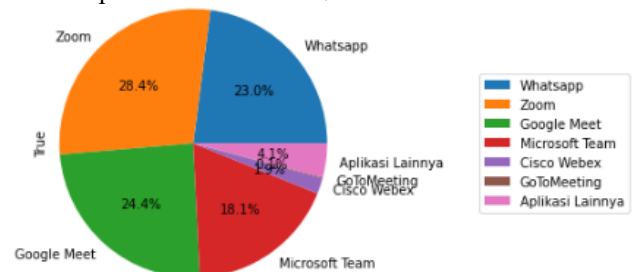
conference, dan streaming video/film juga tampaknya hampir setiap hari dilakukan responden. Hal ini tentu relevan dengan kebiasaan sehari-hari mayoritas responden yaitu pelajar/mahasiswa yang melakukan pembelajarannya secara online. Disamping itu, juga didapatkan informasi bahwa kegiatan-kegiatan seperti konsultasi kesehatan online, membeli obat-obatan online, menonton e-sports, dan mengikuti kelas olahraga online hampir tidak pernah dilakukan oleh semua responden.

5.1.12 Kecenderungan melakukan kegiatan online dalam 3 bulan terakhir



Berdasarkan diagram diatas dapat kita simpulkan bahwa responden cenderung melakukan kegiatan seperti belajar/bekerja, meeting, mengikuti workshop, dan menonton konser secara online. Sedangkan konsultasi kesehatan dan family/social gathering cenderung dilakukan secara offline.

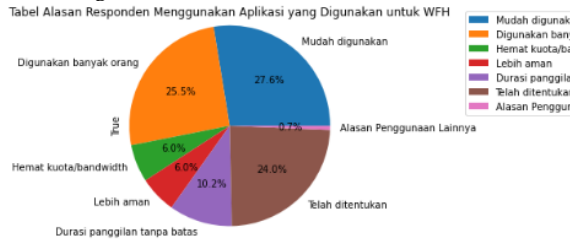
5.1.13 Aplikasi yang Digunakan oleh Responden untuk WFH/SFH



Berdasarkan grafik di atas, dapat disimpulkan bahwa responden

cenderung menggunakan aplikasi Zoom dan Google Meet. Sedangkan yang paling jarang untuk digunakan yaitu Cisco Webex dan GoToMeeting.

5.1.14 Alasan Responden Menggunakan Aplikasi yang Digunakan untuk WFH/SFH



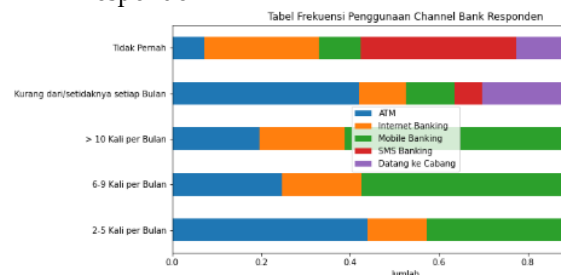
Data di atas berkaitan erat dengan visualisasi sebelumnya, dimana alasan terbesar responden menggunakan suatu aplikasi karena mudah digunakan dan digunakan oleh banyak orang.

5.1.15 Layanan-Layanan Bank yang Digunakan Responden



Berdasarkan data di atas, dapat disimpulkan bahwa layanan bank yang paling sering digunakan oleh responden adalah ATM dan Mobile Banking. Sedangkan layanan yang paling sedikit digunakan yaitu Kartu Kredit dan SMS Banking.

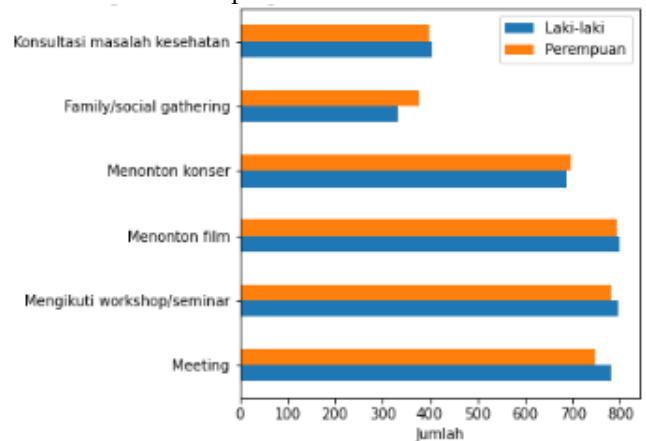
5.1.16 Frekuensi Penggunaan Channel Bank Responden



Berdasarkan diagram di atas, didapatkan informasi bahwa mobile banking adalah layanan yang paling sering digunakan dalam sebulan. Sedangkan SMS Banking adalah layanan yang paling sedikit digunakan bahkan beberapa responden tidak pernah menggunakannya.

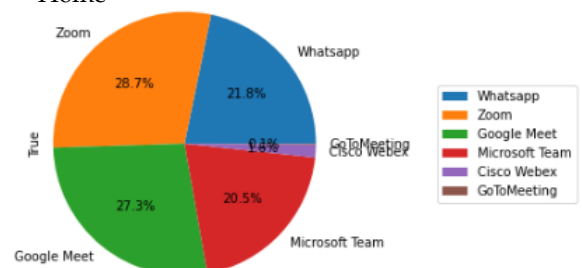
5.2 Visualisasi Data 2 Variabel

5.2.1 Kecenderungan Kegiatan Online dalam 3 Bulan Terakhir Berdasarkan Jenis Kelamin Responden



Grafik di atas menjelaskan bahwa dalam kegiatan online seperti menonton film, mengikuti seminar/workshop, meeting dan konsultasi masalah kesehatan lebih banyak dilakukan oleh responden laki-laki sedangkan untuk sosial gathering dan menonton konser lebih sering dilakukan oleh perempuan.

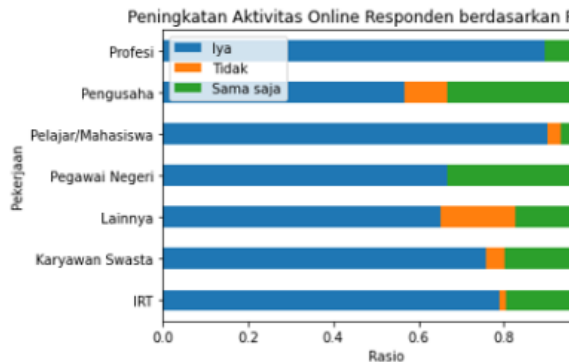
5.2.2 Aplikasi yang Digunakan Pelajar/Mahasiswa untuk School From Home



Dari grafik di atas, didapatkan informasi bahwa pelajar/mahasiswa cenderung menggunakan aplikasi Zoom dan Google Meet. Namun, Aplikasi

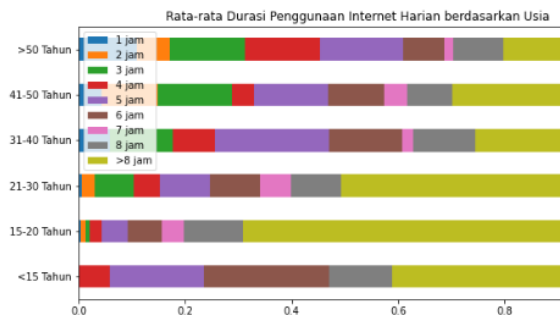
Whatsapp dan Microsoft Team juga sering digunakan walaupun tidak sebanyak Zoom dan Google Meet. Disamping itu, didapatkan juga informasi bahwa untuk School From Home sangat jarang yang menggunakan Cisco Webex ataupun GoToMeeting.

5.2.3 Peningkatan Aktivitas Online Responden Berdasarkan Pekerjaan



Grafik di atas menjelaskan bahwa hampir setiap aspek pekerjaan selama 3 bulan terakhir mengalami peningkatan dalam hal aktivitas secara online. Namun, juga terdapat beberapa yang tidak mengalami peningkatan seperti pengusaha dan pekerjaan lainnya.

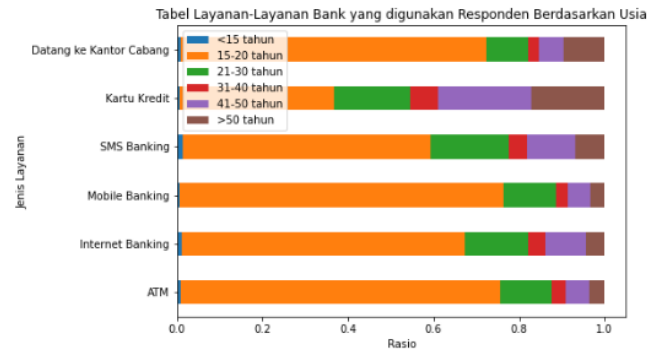
5.2.4 Rata-rata Durasi Penggunaan Internet Harian Berdasarkan Usia



Berdasarkan grafik di atas, dapat disimpulkan bahwa responden yang usia berapapun cenderung menggunakan internet lebih dari 8 jam dalam sehari. Untuk responden yang berusia 15-20 tahun merupakan kelompok usia yang paling banyak menggunakan internet lebih dari 8 jam dalam sehari, sedangkan untuk

penggunaan yang relatif menyebar yaitu responden yang berusia lebih dari 50 tahun.

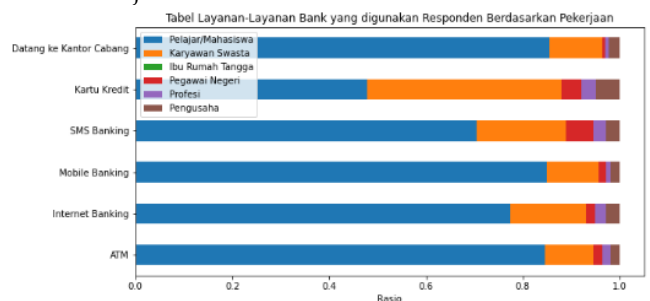
5.2.5 Layanan-Layanan Bank yang Digunakan Responden Berdasarkan Usia



Berdasarkan grafik di atas, didapatkan informasi seperti berikut.

- Responden yang berusia <15 tahun cenderung menggunakan layanan SMS banking dan Internet Banking.
- Responden yang berusia 15-20 tahun cenderung menggunakan layanan mobile banking dan ATM
- Responden yang berusia 21-30 tahun cenderung menggunakan layanan SMS Banking dan Kartu Kredit
- Responden yang berusia lebih dari 30 tahun cenderung menggunakan layanan Kartu Kredit

5.2.6 Layanan-Layanan Bank yang Digunakan Responden Berdasarkan Pekerjaan



Berdasarkan grafik di atas, didapatkan informasi seperti berikut.

- Responden yang bekerja sebagai pelajar/mahasiswa cenderung menggunakan layanan bank yang langsung datang ke kantor cabang,

- mobile banking dan ATM
- Responden yang bekerja sebagai karyawan swasta cenderung menggunakan layanan bank kartu kredit
- Responden yang bekerja sebagai pegawai negeri cenderung menggunakan layanan bank SMS banking dan kartu kredit
- Responden yang bekerja sebagai pengusaha cenderung menggunakan layanan bank kartu kredit
- Responden yang bekerja sebagai profesi cenderung menggunakan layanan bank internet banking, sms banking, dan kartu kredit.
- Responden IRT cenderung tidak menggunakan layanan bank apapun

6 UJI STATISTIK DAN HIPOTESIS

Pengujian statistik pada tugas besar kali ini menggunakan $\alpha = 0.05$ atau *confidence interval*-nya yaitu 95%. Berikut adalah beberapa pengujian statistik terhadap variabel yang diuji.

6.1 Jenis Kelamin Responden

Berikut adalah *confidence interval* 95%-nya dan Uji Hipotesis dengan
 H_0 : Proporsi responden laki – laki = 50%
 H_1 : Proporsi responden laki – laki \neq 50%

```
CI95("Jenis kelamin ?", "Laki-laki")
print()
UH3("Jenis kelamin ?", "Laki-laki", 0.5)
```

Confidence Interval 95% untuk variabel Laki-laki pada Jenis kelamin:
 n = 30 :
 0.3548123 < Proporsi Laki-laki < 0.7118543
 n = 100 :
 0.4832645 < Proporsi Laki-laki < 0.6767355
 n = 1858 :
 0.4799564 < Proporsi Laki-laki < 0.5254257

Untuk $H_1 \neq 0.5$, maka:
 Z score untuk 30 data random:
 0.36514837167011066
 Untuk 30 data random, H_0 gagal ditolak
 Z score untuk 100 data random:
 1.5999999999999992
 Untuk 100 data random, H_0 gagal ditolak
 Z score untuk semua data:
 0.2319942094967968
 Untuk semua data, H_0 gagal ditolak

Dikarenakan H_0 gagal ditolak, maka dapat kita simpulkan bahwa proporsi responden laki-laki pada survey adalah 50%. Artinya proporsi

responden laki-laki sama dengan perempuan

6.2 Usia

Berikut adalah *confidence interval* 95%-nya dan uji hipotesis dengan
 H_0 : Proporsi responden berusia < 15 tahun = 70%
 H_1 : Proporsi responden berusia < 15 tahun < 70%

```
CI95("Usia :", "15-20 tahun")
print()
UH1("Usia :", "15-20 tahun", 0.70)
```

Confidence Interval 95% untuk variabel 15-20 tahun pada Usia : adalah:
 n = 30 :
 0.5360180 < Proporsi 15-20 tahun < 0.8639820
 n = 100 :
 0.6762934 < Proporsi 15-20 tahun < 0.8437066
 n = 1858 :
 0.7256175 < Proporsi 15-20 tahun < 0.7652329

Untuk $H_1 < 0.7$, maka:
 Z score untuk 30 data random:
 0.0
 Untuk 30 data random, H_0 gagal ditolak
 Z score untuk 100 data random:
 1.3093073414159553
 Untuk 100 data random, H_0 gagal ditolak
 Z score untuk semua data:
 4.272774218019384
 Untuk semua data, H_0 gagal ditolak

Dikarenakan H_0 gagal ditolak, maka dengan selang kepercayaan 95% dapat kita simpulkan bahwa responden yang berusia 15-20 tahun yaitu 70% dari total responden.

6.3 Pendidikan Terakhir

Berikut selang kepercayaan 95% dan uji hipotesisnya dengan
 H_0 : Pendidikan Terakhir responden SMA = 70%
 H_1 : Pendidikan Terakhir responden SMA < 70%

```
CI95("Pendidikan Terakhir?", "SMA")
print()
UH1("Pendidikan Terakhir?", "SMA", 0.7)
```

Confidence Interval 95% untuk variabel SMA pada Pendidikan Terakhir? adalah:
 n = 30 :
 0.3210809 < Proporsi SMA < 0.6789191
 n = 100 :
 0.5671550 < Proporsi SMA < 0.7528450
 n = 1858 :
 0.7013669 < Proporsi SMA < 0.7421207

Untuk $H_1 < 0.7$, maka:
 Z score untuk 30 data random:
 -2.3904572186687867
 Untuk 30 data random, H_0 ditolak
 Z score untuk 100 data random:
 -0.8728715609439678
 Untuk 100 data random, H_0 gagal ditolak
 Z score untuk semua data:
 2.04526159251165
 Untuk semua data, H_0 gagal ditolak

Dikarenakan H_0 gagal ditolak untuk semua data, maka dengan selang

kepercayaan 95% dapat kita simpulkan bahwa responden yang berusia 15-20 tahun yaitu 70% dari total responden. Sedangkan untuk 30 data random, H_0 ditolak maka dengan selang kepercayaan 95% dapat kita simpulkan bahwa responden yang berusia 15-20 tahun tidak mencapai 70% dari total responden

6.4 Penghasilan per Bulan

Berikut selang kepercayaan 95% dan uji hipotesisnya dengan
 H_0 : Penghasilan per bulan kecil dari Rp2jt = 75%
 H_1 : Penghasilan per bulan kecil dari Rp2jt < 75%

```
CI95("Penghasilan per bulan?","<Rp 2 Juta")
print()
UH1("Penghasilan per bulan?","<Rp 2 Juta", 0.75)
```

Confidence Interval 95% untuk variabel <Rp 2 Juta pada Penghasilan? adalah:
n = 30 :
0.5360180 < Proporsi <Rp 2 Juta < 0.8639820
n = 100 :
0.5885725 < Proporsi <Rp 2 Juta < 0.7714275
n = 1858 :
0.6997161 < Proporsi <Rp 2 Juta < 0.7405422

Untuk $H_1 < 0.75$, maka:
Z score untuk 30 data random:
-0.632455320336764
Untuk 30 data random, H_0 gagal ditolak
Z score untuk 100 data random:
-1.6165807537309507
Untuk 100 data random, H_0 gagal ditolak
Z score untuk semua data:
-2.9735106085356873
Untuk semua data, H_0 ditolak

Dikarenakan H_0 ditolak untuk semua data, maka dengan selang kepercayaan 95% dapat disimpulkan bahwa responden dengan penghasilan per bulan < 2 juta tidak mencapai 75% dari total responden.

6.5 Pekerjaan

Berikut selang kepercayaan 95% dan uji hipotesis dengan
 H_0 : Responden Pelajar/Mahasiswa = 50%
 H_1 : Responden Pelajar/Mahasiswa > 50%

```
CI95("Pekerjaan?","Pelajar/Mahasiswa")
print()
UH2("Pekerjaan?","Pelajar/Mahasiswa", 0.5)
```

Confidence Interval 95% untuk variabel Pelajar/Mahasiswa pada Pekerjaan? adalah:
n = 30 :
0.6153181 < Proporsi Pelajar/Mahasiswa < 0.9180152
n = 100 :
0.6875185 < Proporsi Pelajar/Mahasiswa < 0.8524815
n = 1858 :
0.7693575 < Proporsi Pelajar/Mahasiswa < 0.8065305

Untuk $H_1 > 0.5$, maka:
Z score untuk 30 data random:
2.9211869733608866
Untuk 30 data random, H_0 ditolak
Z score untuk 100 data random:
5.4
Untuk 100 data random, H_0 ditolak
Z score untuk semua data:
24.823380416157125
Untuk semua data, H_0 ditolak

Dikarenakan H_0 ditolak, maka dengan selang kepercayaan 95% dapat disimpulkan bahwa responden yang bekerja sebagai pelajar/mahasiswa lebih dari 50% total responden.

6.6 Peningkatan Aktivitas Secara Online dalam 3 Bulan Terakhir

Berikut selang kepercayaan 95% dan uji hipotesisnya dengan

H_0 : Domisili responden dari Jakarta = 50%

H_1 : Domisili responden dari Jakarta \neq 50%

```
CI95("Domisili saat ini","Jakarta")
print()
UH3("Domisili saat ini","Jakarta",0.5)
```

Confidence Interval 95% untuk variabel Jakarta pada Domisili saat ini adalah:
n = 30 :
0.0333083 < Proporsi Jakarta < 0.3000251
n = 100 :
0.1047008 < Proporsi Jakarta < 0.2552992
n = 1858 :
0.1742277 < Proporsi Jakarta < 0.2100565

Untuk $H_1 \neq 0.5$, maka:
Z score untuk 30 data random:
-3.651483716701108
Untuk 30 data random, H_0 ditolak
Z score untuk 100 data random:
-6.4
Untuk 100 data random, H_0 ditolak
Z score untuk semua data:
-26.54013756643341
Untuk semua data, H_0 ditolak

Dikarenakan H_0 ditolak, maka dengan selang kepercayaan 95% dapat disimpulkan bahwa responden yang berdomisili dari Jakarta tidak sama dengan 50% dari total responden.

6.7 Aktivitas Online dalam 3 Bulan Terakhir

Berikut selang kepercayaan 95% dan uji hipotesis dengan

H_0 : Responden mengalami peningkatan aktivitas online = 80%

H_1 : Responden mengalami

peningkatan aktivitas online < 80%

```
CF95("Apakah aktivitas online anda meningkat dalam 3 bulan terakhir?", "Iya")
print()
UH1("Apakah aktivitas online anda meningkat dalam 3 bulan terakhir?", "Iya", 0.8)

Confidence Interval 95% untuk variabel Iya pada Apakah aktivitas online anda meningkat dalam 3
terakhir? adalah:
n = 30 :
0.6153181 < Proporsi Iya < 0.9180152
n = 100 :
0.6762934 < Proporsi Iya < 0.8437066
n = 1858 :
0.8011012 < Proporsi Iya < 0.8361431

Untuk H1 < 0.8 , maka:
Z score untuk 30 data random:
-0.45643546458763834
Untuk 30 data random, H0 gagal ditolak
Z score untuk 100 data random:
-1.0000000000000001
Untuk 100 data random, H0 gagal ditolak
Z score untuk semua data:
2.0067499121472756
Untuk semua data, H0 gagal ditolak
```

Dikarenakan H_0 gagal ditolak, maka dengan selang 95% dapat disimpulkan bahwa responden yang mengalami peningkatan aktivitas online sama dengan 80% dari total responden.

6.8 Akses atau Provider Internet yang Digunakan Selama Ini

Berikut selang kepercayaan 95% dan uji hipotesis dengan

H_0 : Responden menggunakan mobile data dan broadband = 78%

H_1 : Responden menggunakan mobile data dan broadband > 78%

```
Akses atau provider internet yang digunakan selama ini?", "Menggunakan Mobile Data dan Broadbar
)
kses atau provider internet yang digunakan selama ini?", "Menggunakan Mobile Data dan Broadbar

Confidence Interval 95% untuk variabel Menggunakan Mobile Data dan Broadband pada Akses atau
internet yang digunakan selama ini? adalah:
n = 30 :
0.4608932 < Proporsi Menggunakan Mobile Data dan Broadband < 0.8057735
n = 100 :
0.5353723 < Proporsi Menggunakan Mobile Data dan Broadband < 0.7246277
n = 1859 :
0.5471539 < Proporsi Menggunakan Mobile Data dan Broadband < 0.5921683

Untuk H1 > 0.78 , maka:
Z score untuk 30 data random:
-1.9392482462757996
Untuk 30 data random, H0 gagal ditolak
Z score untuk 100 data random:
-3.6210341218895072
Untuk 100 data random, H0 gagal ditolak
Z score untuk semua data:
-21.89276597932349
Untuk semua data, H0 gagal ditolak
```

Dikarenakan H_0 gagal ditolak, maka dengan selang kepercayaan 95% dapat disimpulkan bahwa responden yang menggunakan mobile data dan juga menggunakan broadband.

7. KESIMPULAN

1. Dari proses data screening berupa penggabungan kolom dan pendefinisian dataframe, didapatkan dataframe yang lebih rapi dan siap untuk di-cleaning
2. Berdasarkan proses data cleaning berupa penghapusan data responden yang tidak menjawab pertanyaan wajib, dan mengantisipasi terjadinya duplikasi,

didapatkan panjang *row* yaitu 1860 responden.

3. Melalui data analisis tiap fitur didapatkan beberapa kesimpulan, seperti responden yang dominan berusia 15-20 tahun sehingga menimbulkan bias, mayoritas responden yang mengalami kenaikan aktivitas online, dan berbagai macam kegiatannya. Hal ini relevan dengan kebijakan setiap sekolah yang melaksanakan kegiatan belajar mengajar secara online (*School From Home*). Visualisasi data telah dijelaskan pada Bab 5 yaitu *Exploratory Data Analysis*, lebih tepatnya pada 5.1.7., 5.2.3, dan 5.2.4.
4. Melalui analisis dan visualisasi antar fitur, dapat disimpulkan bahwa wabah COVID-19 mengakibatkan berbagai kalangan usia mengalami peningkatan dalam durasi rata-rata penggunaan internet harian. Hal ini tentu saja relevan dengan adaptasi yang dilakukan yaitu menjadikan aktivitas harian dari yang awalnya luring menjadi daring. Untuk memperkuat argumen, dapat diperhatikan pada gambar 5.2.3 yang menjelaskan peningkatan aktivitas online responden berdasarkan pekerjaan mereka.
5. Melalui pencarian confidence interval, didapat bahwa perbesaran nilai n akan mengecilkan jarak antara nilai batas confidence interval, sehingga mempertinggi keakuratan hasil statistik.
6. Melalui uji statistik, didapat bahwa proporsi populasi yang aktivitas onlinenya mengalami peningkatan dalam 3 bulan terakhir yaitu sekitar 80% dengan taraf signifikansi uji 5%.

ACKNOWLEDGMENT

Penulis ingin mengucapkan terima kasih kepada bapak Dimitri Mahayana selaku dosen mata kuliah Probabilitas dan Statistika, dan kepada kak Abbas dan kak Aldo selaku asisten yang membimbing selama pengerjaan tugas besar ini.

REFERENCES

1. DataScience Made Simple, "Reindex or Rearrange Rows in Python Pandas", Pandas Python, <https://www.datasciencemadesimple.com/reindex-python-change-order-column-rows-pandas-dataframe.2020>. Diakses 21 Desember 2020.
2. "Pengertian Ilmu Statistik dan Tahapan Kegiatannya | Lentera Kecil".lenterakecil.com, <https://lenterakecil.com/pengertian-ilmu-statistik-dan-tahapan-kegiatan/>. Diakses 23 Desember 2020.
3. Kanoki, "Pandas dataframe filter with Multiple conditions," Pandas Python, <https://kanoki.org/2020/01/21/pandas-dataframe-filter-with-multiple-conditions.2020>. Diakses 22 Desember 2020.
4. Mahayana, Dimitri. "Estimasi Statistik," Catatan Kuliah Probabilitas dan Statistik STEI ITB. Diakses 22 Desember 2020.
5. Mahayana, Dimitri. "Uji Hipotesis," Catatan Kuliah Probabilitas dan Statistik STEI ITB. Diakses 23 Desember 2020.
6. Pandas Development Team, "Pandas Data Frame User Guide," Sphinx, <https://pandas.pydata.org/pandas-docs/stable/reference/frame.html.2008>. Diakses 22 Desember 2020
7. W3Schools, "Python Matplotlib Tutorial," Matplotlib, https://www.w3schools.com/python/matplotlib_intro.asp.2012. Diakses 22 Desember 2020.

BIOGRAPHY

Aditya Anandita Dharma Putra, lahir di Padang, 23 Desember 2000. Penulis merupakan anak tunggal di keluarganya. Penulis memulai pendidikan formal sejak berusia 7 tahun di SD No.23 Painan Utara di Kab. Pesisir Selatan. Kemudian, melanjutkan pendidikannya di SMPN 1 Painan dan di SMAN Agam Cendekia, Sumatra Barat. Pada tahun 2019 penulis lulus SMA dan melanjutkan studinya di Institut Teknologi Bandung melalui jalur SBMPTN pada tahun yang sama. Sampai saat laporan ini dibuat, penulis sedang menekuni tahun keduanya di Sekolah Teknik Elektro dan Informatika pada program studi Teknik Elektro. Kesan penulis selama mengikuti dan mempelajari mata kuliah Probabilitas dan Statistika (EL2004) sangat menyenangkan dan menarik. Di tengah padatnya jadwal akademik, baik itu lomba, ujian semester dan tugas lainnya, penulis berusaha memberikan hasil terbaik yang ia mampu untuk menyelesaikan tugas besar ini. Penulis merasa tugas ini dapat bermanfaat baik bagi penyedia layanan internet, dan e-lifestyle lainnya maupun menjadi referensi bagi generasi selanjutnya.