

# Tugas Besar Probstat 18220053

November 29, 2021

## 1 Identitas dan Deskripsi

Nama : I Putu Andika Bagas Jiwanta

Nim : 18220053

Pada tugas ini saya menganalisis data survey e-money

## 2 Analisis Data

### 2.1 Imports

```
[1]: import urllib.request as urlreq
import pandas as pd
import os
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import norm
```

```
[2]: # Dibutuhkan karena mengerjakan di google colab
url_emoney = "https://raw.githubusercontent.com/bagasjiwanta/Tugas-Besar_II2111/
↳master/emoney.csv"
os.makedirs(name="data", exist_ok=True)
path_emoney = os.path.join("data", "emoney.csv")
urlreq.urlretrieve(url=url_emoney, filename=path_emoney)
```

```
[2]: ('data/emoney.csv', <http.client.HTTPMessage at 0x7f73e05ed590>)
```

### 2.2 Melihat data

```
[3]: df = pd.read_csv(path_emoney)
df.head(2)
```

```
[3]: responden jenis_kelamin ... keluhan_2 keluhan_3
0          1    Laki - laki ...      NaN      NaN
1          2    Laki - laki ...      NaN      NaN
```

[2 rows x 29 columns]

```
[4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2095 entries, 0 to 2094
Data columns (total 29 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   responden                            2095 non-null   int64
1   jenis_kelamin                        2095 non-null   object
2   usia                                 2095 non-null   object
3   pendidikan                           2095 non-null   object
4   penghasilan                           2095 non-null   object
5   aktivitas_meningkat_1                999 non-null    object
6   aktivitas_meningkat_2                1445 non-null   object
7   aktivitas_meningkat_3                1426 non-null   object
8   aktivitas_meningkat_4                1510 non-null   object
9   aktivitas_meningkat_5                100 non-null    object
10  menggunakan_emoney?                 2095 non-null   object
11  emoney_pilihan_1                     1664 non-null   object
12  emoney_pilihan_2                     1396 non-null   object
13  emoney_pilihan_3                     354 non-null    object
14  emoney_pilihan_4                     855 non-null    object
15  emoney_pilihan_5                     1225 non-null   object
16  penggunaan_1                         1238 non-null   object
17  penggunaan_2                         726 non-null    object
18  penggunaan_3                         1254 non-null   object
19  penggunaan_4                         887 non-null    object
20  penggunaan_5                         443 non-null    object
21  metode_topup_1                       498 non-null    object
22  metode_topup_2                       1551 non-null   object
23  metode_topup_3                       196 non-null    object
24  metode_topup_4                       147 non-null    object
25  metode_topup_5                       534 non-null    object
26  keluhan_1                           951 non-null    object
27  keluhan_2                           102 non-null    object
28  keluhan_3                           249 non-null    object
dtypes: int64(1), object(28)
memory usage: 474.8+ KB
```

```
[5]: n = df.shape[0]
      print("Jumlah data:", n)
```

```
Jumlah data: 2095
```

## 2.3 Analisis Proporsi Demografi

```
[6]: for x in list(df.columns)[1:5]:
      print('tabel frekuensi {}'.format(x))
      dfx = pd.DataFrame(df[x].value_counts()).reset_index()
      dfx['proporsi'] = dfx[x]/n
      dfx.columns = [x, 'frekuensi', 'proporsi']
      print(dfx)
      plt.pie(dfx['proporsi'], labels = list(dfx[x]), autopct = '%.0f%%')
      plt.title('proporsi {} responden (n={})'.format(x,n))
      plt.show()
      print('')
```

tabel frekuensi jenis\_kelamin

	jenis_kelamin	frekuensi	proporsi
0	Laki - laki	1049	0.500716
1	Perempuan	1031	0.492124
2	tidak menjawab	15	0.007160

proporsi jenis\_kelamin responden (n=2095)

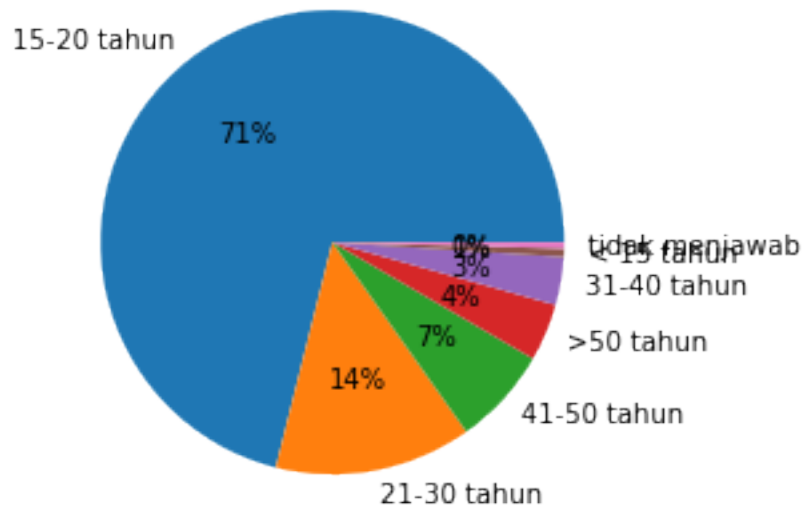


tabel frekuensi usia

	usia	frekuensi	proporsi
0	15-20 tahun	1489	0.710740
1	21-30 tahun	288	0.137470
2	41-50 tahun	143	0.068258
3	>50 tahun	84	0.040095
4	31-40 tahun	70	0.033413

5	< 15 tahun	12	0.005728
6	tidak menjawab	9	0.004296

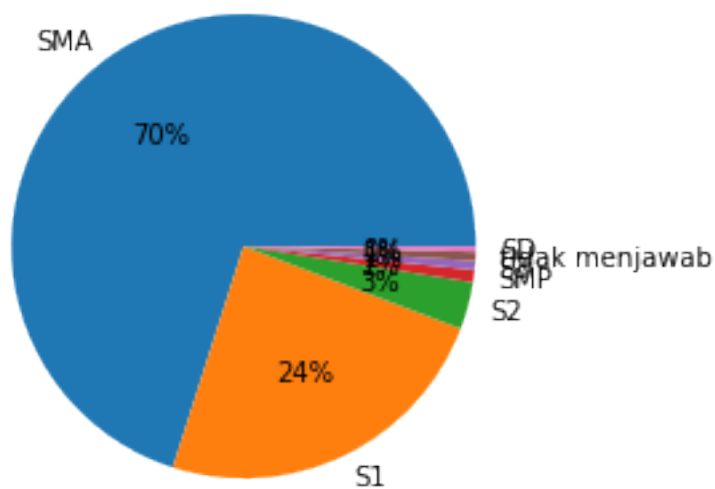
proporsi usia responden (n=2095)



tabel frekuensi pendidikan

	pendidikan	frekuensi	proporsi
0	SMA	1468	0.700716
1	S1	506	0.241527
2	S2	69	0.032936
3	SMP	18	0.008592
4	S3	13	0.006205
5	tidak menjawab	13	0.006205
6	SD	8	0.003819

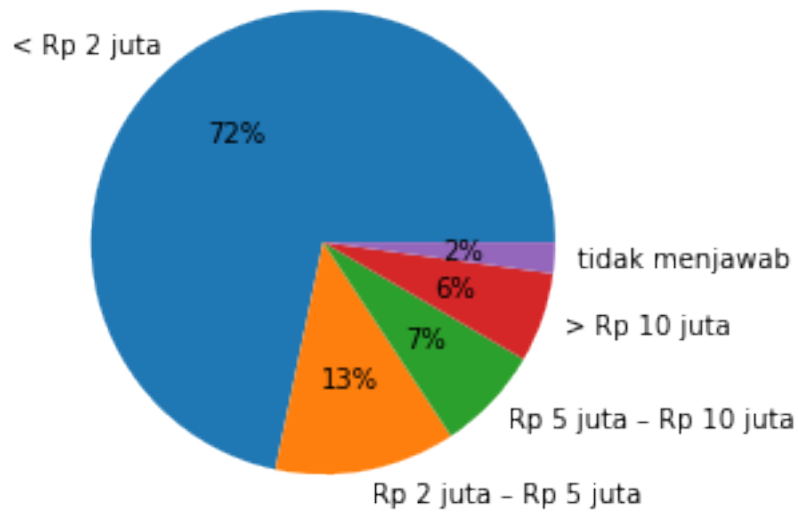
proporsi pendidikan responden (n=2095)



tabel frekuensi penghasilan

	penghasilan	frekuensi	proporsi
0	< Rp 2 juta	1502	0.716945
1	Rp 2 juta - Rp 5 juta	264	0.126014
2	Rp 5 juta - Rp 10 juta	153	0.073031
3	> Rp 10 juta	131	0.062530
4	tidak menjawab	45	0.021480

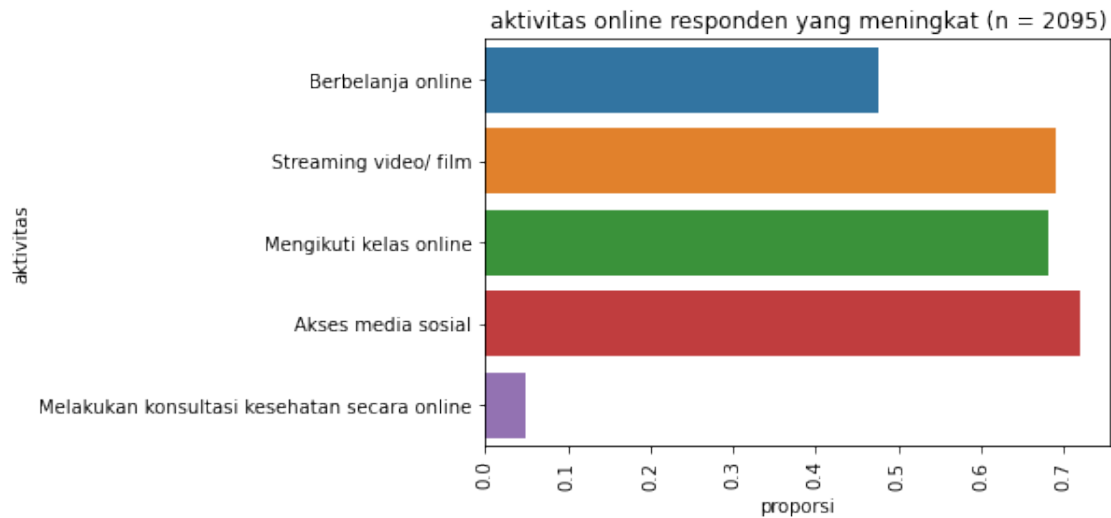
### proporsi penghasilan responden (n=2095)



## 2.4 Proporsi peningkatan aktivitas online

```
[7]: aktivitas = pd.DataFrame()
for x in list(df.columns)[5:10]:
    dfx = pd.DataFrame(df[x].value_counts()).reset_index()
    dfx['proporsi'] = dfx[x]/n
    dfx.columns = ['aktivitas', 'frekuensi', 'proporsi']
    aktivitas = pd.concat([aktivitas, dfx], axis = 0)
aktivitas = aktivitas.reset_index(drop = True)
print(aktivitas)
sns.barplot(data= aktivitas, y='aktivitas', x='proporsi')
plt.xticks(rotation = 90)
plt.title('aktivitas online responden yang meningkat (n = {})'.format(n))
plt.show()
```

	aktivitas	frekuensi	proporsi
0	Berbelanja online	999	0.476850
1	Streaming video/ film	1445	0.689737
2	Mengikuti kelas online	1426	0.680668
3	Akses media sosial	1510	0.720764
4	Melakukan konsultasi kesehatan secara online	100	0.047733



## 2.5 Confidence interval 95% untuk proporsi peningkatan aktivitas

```
[8]: alpha = 1 - 0.95
      aktivitas['z(alpha/2)'] = norm.ppf(1 - alpha/2)
      aktivitas['error'] = aktivitas['z(alpha/
      →2)']*(aktivitas['proporsi']*(1-aktivitas['proporsi'])/n)**(1/2)
      aktivitas['batas bawah'] = aktivitas['proporsi'] - aktivitas['error']
      aktivitas['batas atas'] = aktivitas['proporsi'] + aktivitas['error']
      aktivitas
```

```
[8]:
```

	aktivitas	...	batas atas
0	Berbelanja online	...	0.498237
1	Streaming video/ film	...	0.709546
2	Mengikuti kelas online	...	0.700632
3	Akses media sosial	...	0.739974
4	Melakukan konsultasi kesehatan secara online	...	0.056862

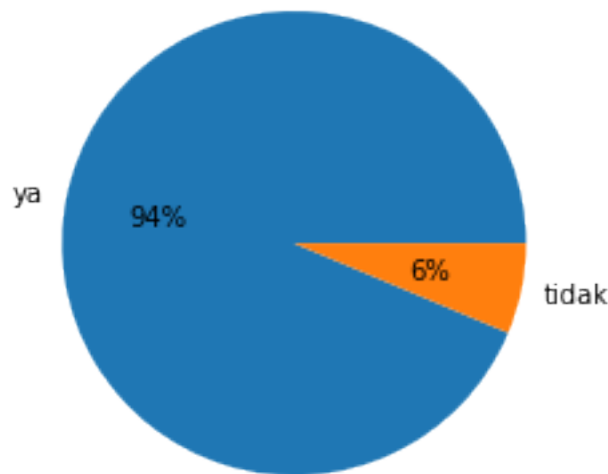
[5 rows x 7 columns]

## 2.6 Proporsi kepemilikan akun/kartu e-money

```
[9]: temp = pd.DataFrame(df.iloc[:,10].value_counts())
temp = temp.reset_index()
temp.columns = ['punya emoney?', 'frekuensi']
temp['proporsi'] = temp['frekuensi']/n
print(temp)
plt.pie(temp['proporsi'], labels = list(temp['punya emoney?']), autopct = '%.0f%%')
plt.title('proporsi responden yang memiliki akun/kartu e-money (n={})'.format(n))
plt.show()
```

	punya emoney?	frekuensi	proporsi
0	ya	1963	0.936993
1	tidak	132	0.063007

proporsi responden yang memiliki akun/kartu e-money (n=2095)

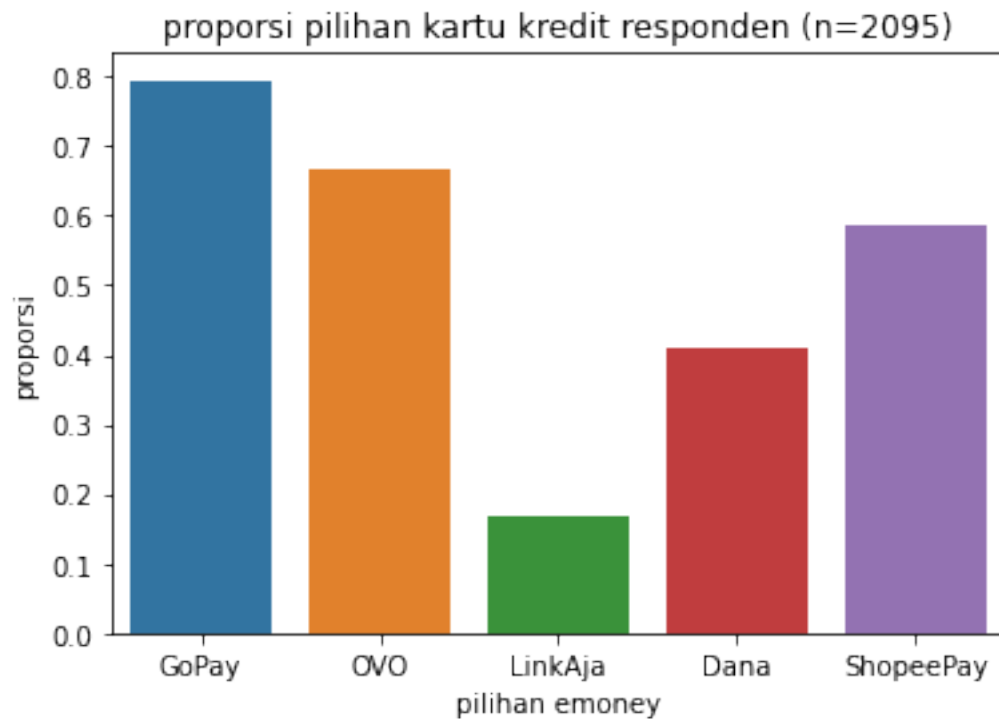


## 2.7 Proporsi pilihan e-money

```
[10]: pilihan_emoney = pd.DataFrame()
for x in list(df.columns)[11:16]:
    dfx = pd.DataFrame(df[x].value_counts()).reset_index()
    dfx['proporsi'] = dfx[x]/n
    dfx.columns = ['pilihan emoney', 'frekuensi', 'proporsi']
    pilihan_emoney = pd.concat([pilihan_emoney, dfx], axis=0)
    pilihan_emoney = pilihan_emoney.reset_index(drop = True)
print(pilihan_emoney)
sns.barplot(data = pilihan_emoney, x = 'pilihan emoney', y = 'proporsi')
plt.title('proporsi pilihan kartu kredit responden (n={})'.format(n))
plt.show()
```



	pilihan emoney	frekuensi	proporsi
0	GoPay	1664	0.794272
1	OVO	1396	0.666348
2	LinkAja	354	0.168974
3	Dana	855	0.408115
4	ShopeePay	1225	0.584726

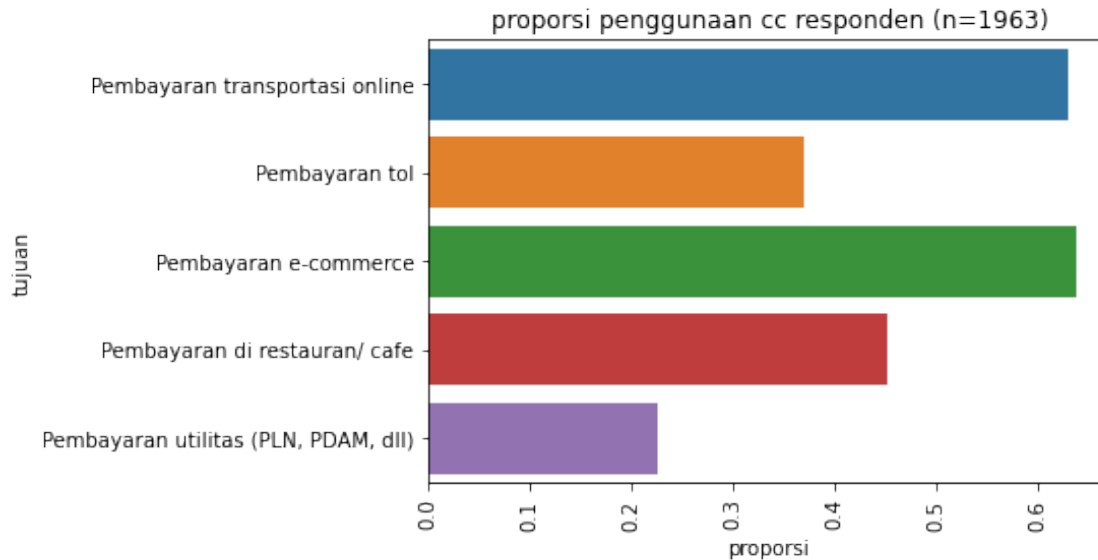


## 2.8 Proporsi tujuan kepemilikan emoney

```
[11]: n2 = (df['menggunakan_emoney?'] == 'ya').sum()
tujuan_cc = pd.DataFrame()
for x in list(df.columns)[16:21]:
    dfx = pd.DataFrame(df[x].value_counts()).reset_index()
    dfx['proporsi'] = dfx[x]/n2
    dfx.columns = ['tujuan', 'frekuensi', 'proporsi']
    tujuan_cc = pd.concat([tujuan_cc, dfx], axis=0)
    tujuan_cc = tujuan_cc.reset_index(drop = True)
print(tujuan_cc)
sns.barplot(data = tujuan_cc, y = 'tujuan', x = 'proporsi')
plt.title('proporsi penggunaan cc responden (n={})'.format(n2))
plt.xticks(rotation = 90)
plt.show()
```

tujuan	frekuensi	proporsi
--------	-----------	----------

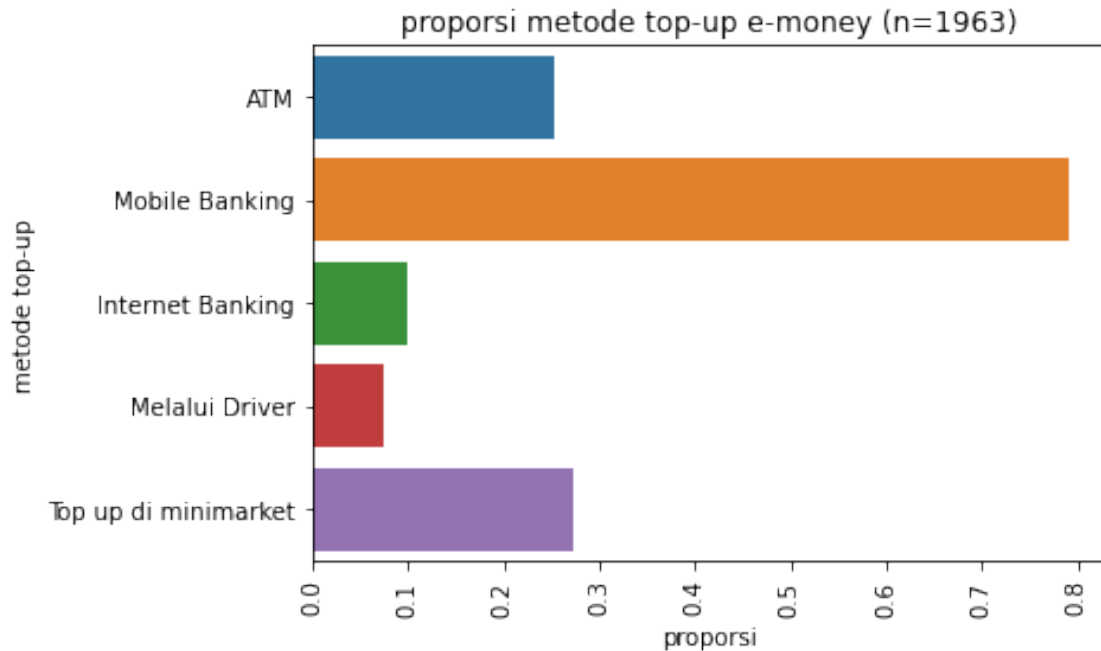
0	Pembayaran transportasi online	1238	0.630667
1	Pembayaran tol	726	0.369842
2	Pembayaran e-commerce	1254	0.638818
3	Pembayaran di restoran/ cafe	887	0.451859
4	Pembayaran utilitas (PLN, PDAM, dll)	443	0.225675



## 2.9 Proporsi metode top-up e-money

```
[12]: metode_topup = pd.DataFrame()
for x in list(df.columns)[21:26]:
    dfx = pd.DataFrame(df[x].value_counts()).reset_index()
    dfx['proporsi'] = dfx[x]/n2
    dfx.columns = ['metode top-up', 'frekuensi', 'proporsi']
    metode_topup = pd.concat([metode_topup, dfx], axis=0)
    metode_topup = metode_topup.reset_index(drop = True)
print(metode_topup)
sns.barplot(data = metode_topup, y = 'metode top-up', x = 'proporsi')
plt.title('proporsi metode top-up e-money (n={})'.format(n2))
plt.xticks(rotation = 90)
plt.show()
```

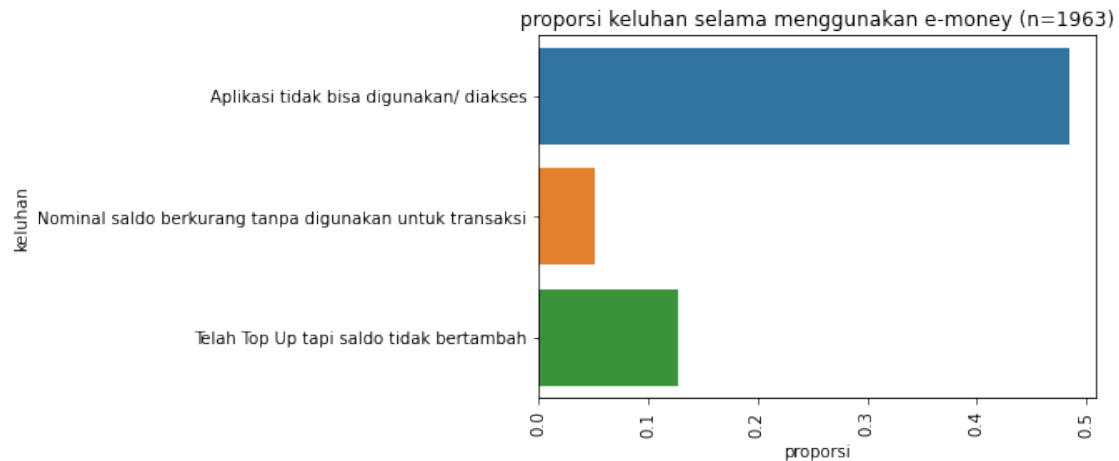
	metode top-up	frekuensi	proporsi
0	ATM	498	0.253693
1	Mobile Banking	1551	0.790117
2	Internet Banking	196	0.099847
3	Melalui Driver	147	0.074885
4	Top up di minimarket	534	0.272033



## 2.10 Proporsi keluhan menggunakan e-money

```
[13]: metode_topup = pd.DataFrame()
for x in list(df.columns)[26:]:
    dfx = pd.DataFrame(df[x].value_counts()).reset_index()
    dfx['proporsi'] = dfx[x]/n2
    dfx.columns = ['keluhan', 'frekuensi', 'proporsi']
    metode_topup = pd.concat([metode_topup, dfx], axis=0)
    metode_topup = metode_topup.reset_index(drop = True)
print(metode_topup)
sns.barplot(data = metode_topup, y = 'keluhan', x = 'proporsi')
plt.title('proporsi keluhan selama menggunakan e-money (n={})'.format(n2))
plt.xticks(rotation = 90)
plt.show()
```

	keluhan	frekuensi	proporsi
0	Aplikasi tidak bisa digunakan/ diakses	951	0.484463
1	Nominal saldo berkurang tanpa digunakan untuk ...	102	0.051961
2	Telah Top Up tapi saldo tidak bertambah	249	0.126847



## 2.11 Analisis confidence interval 95% selisih proporsi

Rumus :

$$(\hat{p}_1 - \hat{p}_2) - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

```
[14]: alpha = 0.05
      zalpha_2 = norm.ppf(1-alpha/2)
```

```
[15]: gopay = df.iloc[:, 11].count()
      ovo = df.iloc[:, 12].count()

      n1 = df.shape[0]
      n2 = n1

      p1 = gopay / n1
      q1 = 1 - p1

      p2 = ovo / n2
      q2 = 1 - p2

      base = p1 - p2
      error = zalpha_2 * (((p1*q1/n1)+(p2*q2/n2)) ** 0.5)
      lower = base - error
      upper = base + error

      print("Pengguna gopay - pengguna ovo")
      print("%.6f < p1 - p2 < %.6f" %(lower, upper))
```

```
Pengguna gopay - pengguna ovo
0.101329 < p1 - p2 < 0.154519
```

```
[16]: n1 = (df['menggunakan_emoney?'] == 'ya').sum()
n2 = n1

transport = df.iloc[:, 16].count()
restoran = df.iloc[:, 19].count()

p1 = transport / n1
q1 = 1 - p1

p2 = restoran / n2
q2 = 1 - p2

base = p1 - p2
error = zalpha_2 * (((p1*q1/n1)+(p2*q2/n2)) ** 0.5)
lower = base - error
upper = base + error

print("Pembayaran transportasi online - pembayaran restoran")
print("%.6f < p1 - p2 < %.6f" %(lower, upper))
```

```
Pembayaran transportasi online - pembayaran restoran
0.148140 < p1 - p2 < 0.209476
```

### 3 Kesimpulan

#### 3.1 Data dan bentuk data

Data dapat dilihat [disini](#). Data adalah hasil survey terhadap 2095 responden. Dalam data, terdapat informasi informasi terhadap responden : 1. Identitas (jenis kelamin, usia, pendidikan terakhir, penghasilan) 2. Aktivitas apa saja yang meningkat selama online (pandemi) 3. Apakah menggunakan emoney dan emoney pilihannya 4. Penggunaan emoney 5. Keluhan selama menggunakan emoney

#### 3.2 Demografi

1. Jumlah pria dan wanita yang menjawab survey kurang lebih sama
2. Usia yang paling banyak dijumpai adalah 15-20 tahun sebanyak 71%, diikuti dengan 20-25 tahun sebanyak 14%, dan mengecil seterusnya
3. Sekitar 70% responden adalah lulusan SMA, 21% lulusan S1, dan diikuti dengan sisanya.
4. Sekitar 72% responden memiliki penghasilan dibawah 2 juta, 13% diantara 2-5 juta dan sisanya diatas 5 juta

#### 3.3 Analisis Proporsi

1. Aktivitas streaming, kelas online, dan akses media sosial menjadi kegiatan populer dalam pandemi, hal ini selaras dengan mayoritas usia responden yang berada di rentang 15-20 tahun yang merupakan mahasiswa / siswa SMA.
2. Hampir semua responden memiliki emoney, dengan persentase sebesar 94%

3. Emoney yang paling laris dari terlaris ke terjangkau secara berturut-turut : Gopay, OVO, Shopeepay, Dana, Link aja.
4. Pembayaran transportasi online (63%) dan ecommerce (64%) menjadi dua alasan utama responden menggunakan emoney. Naiknya penggunaan transportasi online dapat disebabkan oleh maraknya jasa ojek online di Indonesia. Sementara penggunaan ecommerce juga meningkat akibat meningkatnya jumlah online shop dan pembayaran non-tunai selain kartu.
5. Metode topup dengan mobile banking menjadi sangat populer, dengan proporsi sebesar 79%, diikuti dengan melalui minimarket, dengan proporsi sebesar 27%. Dapat disimpulkan bahwa para kecepatan dan kemudahan menjadi salah satu faktor krusial. Metode top up lain seperti bank, driver, dan ATM kalah jauh.
6. Keluhan yang paling sering disampaikan responden adalah aplikasi yang tidak bisa digunakan/diakses.

### **3.4 Analisis confidence interval 95 persen untuk selisih proporsi**

1. Pengguna gopay - ovo. Dalam analisis, didapatkan bahwa selisih proporsi pengguna gopay dikurangi proporsi pengguna ovo berada di rentang 10.13% sampai 15.19%. Artinya, untuk populasi masyarakat Indonesia yang menggunakan emoney, gopay akan selalu lebih populer minimal 10.13% dan maksimal 15.19% daripada ovo. Kemungkinan besar ini terjadi karena gopay adalah metode pembayaran utama (selain cash) untuk layanan gojek (ojek online). Hal ini juga dapat dilihat dari Analisis Proporsi pada poin 3.3 poin ke-4, sebesar 63% responden yang menggunakan emoney mengalami peningkatan dalam penggunaan emoney untuk pembayaran transportasi online selama online.
2. Penggunaan emoney untuk transportasi online - restoran. Dalam analisis, didapatkan bahwa selisih proporsi pengguna gopay dikurangi proporsi pengguna ovo berada di rentang 14.81% sampai 20.95%. Artinya, untuk populasi masyarakat Indonesia, jauh lebih banyak yang menggunakan emoney untuk transportasi online ketimbang restoran. Hal ini mungkin disebabkan oleh pilihan pembayaran di restoran masih lebih beragam seperti kartu, dan lain-lain. Sementara untuk transportasi online, cenderung menggunakan pembayaran yang lebih gampang (emoney / cash).