

```
pip install faker
```

```
Collecting faker
  Downloading Faker-19.10.0-py3-none-any.whl (1.7 MB)
    1.7/1.7 MB 18.0 MB/s eta 0:00:00
Requirement already satisfied: python-dateutil>=2.4 in /usr/local/lib/python3.10/dist-packages (from faker) (2.8.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.4->faker) (1.16.0)
Installing collected packages: faker
Successfully installed faker-19.10.0
```

```
from faker import Faker

fake = Faker()

jumlah_data = 20

dummy_data = []

for _ in range(jumlah_data):
    nama = fake.name()
    nomor_telepon = fake.phone_number()
    alamat = fake.address()
    email = fake.email()

    data = f>Nama: {nama}\nNomor Telepon: {nomor_telepon}\nAlamat: {alamat}\nEmail: {email}\n"
    dummy_data.append(data)

file_name = "dummy_data.txt"
with open(file_name, "w") as file:
    for data in dummy_data:
        file.write(data + "\n")

print(f"{jumlah_data} data dummy telah disimpan dalam {file_name}.")

20 data dummy telah disimpan dalam dummy_data.txt.
```

```
import re

pattern = r"[a-zA-Z0-9_+-.]+@[a-zA-Z0-9-]+\.[a-zA-Z.]+"
```

```
with open("dummy_data.txt", "r") as f:
    data = f.readlines()

emails = []

for line in data:
    match = re.search(pattern, line)
    if match:
        emails.append(match.group())
print(emails)
```

```
['jacob06@example.com', 'susan21@example.net', 'matthewedwards@example.org', 'knoxwillie@example.com', 'katie45@example.com', 'karihammond@example.com', 'reynoldsrachel@example.com', 'jenniferhuc
```

```
with open("dummy_data.txt", "r") as f:
    teks = f.readlines()
    text = "".join(teks)

pola_nomor_telepon = r'Nomor Telepon:\s+([\d+()-]+)'\
nomor_telepon = re.findall(pola_nomor_telepon, text)

nums = []
# Mengambil nomor telepon yang cocok
for num in nomor_telepon:
    # Substituting multiple spaces with single space
    num = re.sub(r'\s+', ' ', num, flags=re.I)

    # Menghilangkan karakter spesial
    num = re.sub(r'\W', '', num)
    nums.append(num)

# Output data Preprocessing
nums
```

```
['9552359284',
 '0019236416728',
 '5697836687',
 '16454154463',
 '0012274212083',
 '7677227330',
 '3064109090',
 '4335576335',
 '2174126853',
 '298',
 '310',
 '4885646426',
 '17968081654',
 '9857671885',
 '17713453817',
 '0019478736936',
 '3039872471',
 '6473274598',
 '4285214025',
 '9717384150']
```

```
import pandas as pd
pd.DataFrame({"email": emails, "nomor_telepon": nums})
```

	email	nomor_telepon
0	jacob06@example.com	9552359284
1	susan21@example.net	0019236416728
2	matthewedwards@example.org	5697836687
3	knoxwillie@example.com	16454154463
4	katie45@example.com	0012274212083
5	karihammond@example.com	7677227330
6	reynoldsrachel@example.com	3064109090
7	jenniferhudson@example.org	4335576335
8	castillocaroline@example.org	2174126853
9	eduardo65@example.org	298
10	raymondvang@example.org	310
11	jonathan46@example.net	4885646426
12	brandon31@example.com	17968081654
13	sthompson@example.com	9857671885
14	lopezmichael@example.org	17713453817
15	jspencer@example.org	0019478736936
16	sean20@example.com	3039872471
17	elizabeththompson@example.com	6473274598
18	jwilliams@example.org	4285214025
19	steventorres@example.net	9717384150