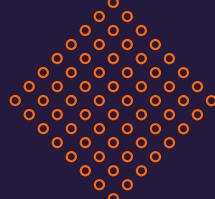


# FINAL PROJECT EDA

Scraping Berita

# SHOPEE



Kelas A



# Meet The Team!



SAFIRA RAMADHANI

2206026656



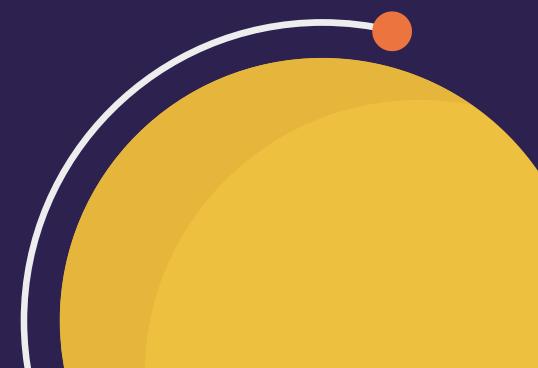
GIAN SINAR KATULISTIWA

2206031675



AMIRA SHOHIFA

2206829130



# DAFTAR ISI



Pendahuluan



Preprocessing



Pengolahan dan  
Analisis Data



Kesimpulan

# PENDAHULUAN

## ○ Latar Belakang



Sejak tahun 2015, shopee memulai operasionalnya di Indonesia. Hingga saat ini, shopee tumbuh menjadi salah satu platform e-commerce yang paling populer di Indonesia. Shopee telah memperluas cakupan operasionalnya dan terus menghadirkan inovasi dalam aspek teknologi dan pelayanan. Pada tahun 2023, Shopee juga terus meningkatkan fokus pada sektor-sektor kunci seperti fashion, makanan dan minuman, serta elektronik.

# PENDAHULUAN

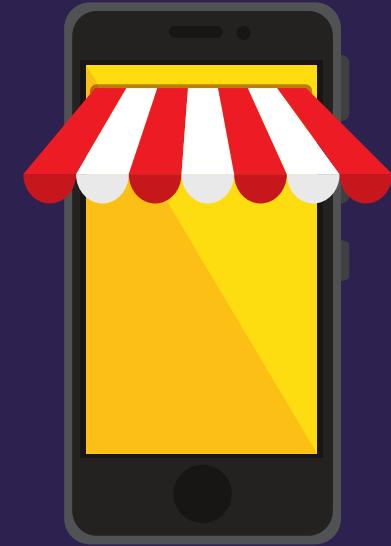
## ○ Latar Belakang

Saat ini, shopee menjadi e-commerce yang menjadi pilihan utama bagi para penjual dan pembeli di Indonesia untuk bertransaksi online dengan berbagai pilihan produk yang tersedia.

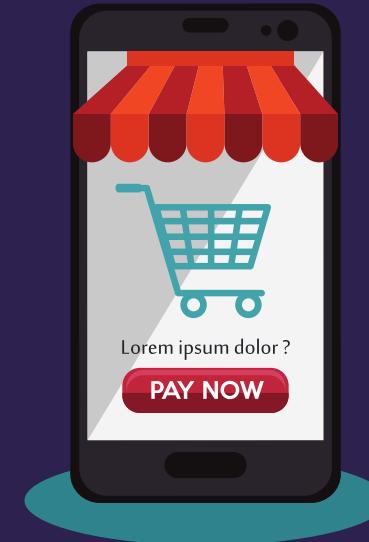
Oleh karena itu, kami memilih topik scrapping “Shopee” karena ingin mengetahui bagaimana analisis sentimen Shopee pada masyarakat. Selain itu juga untuk mengetahui apa saja keterkaitan antara shopee dengan topik lainnya.



# Apa saja langkah yang dilakukan?



Scraping Berita



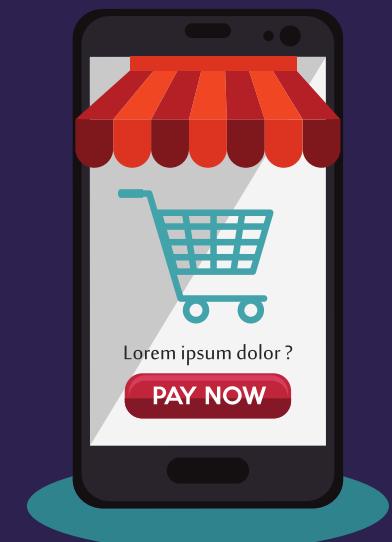
Preprocessing



Statistika Deskriptif



Data Wrangling



Visualisasi Data



Social Network Analysis

# SCRAPPING

Hasil pencarian "shopee", 5165 hasil ditemukan

detikcom

Semua Tanggal

Cari Berita

Melakukan impor library dan modul yang diperlukan

Kami memilih topik "Shopee" untuk dilakukan scrapping, url yang digunakan berasal dari detik.com

```
#Mengimpor modul yang akan digunakan
import requests as req
from bs4 import BeautifulSoup as bs
from datetime import datetime
import csv
import pandas as pd
import numpy as np
from datetime import datetime,timedelta
import pytz
import re
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
from nltk.sentiment import SentimentIntensityAnalyzer
import ast
import string
from wordcloud import WordCloud
!pip install PySastrawi
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import itertools
import matplotlib.pyplot as plt
import seaborn as sns
!pip install googletrans==4.0.0-rc1
from googletrans import Translator

hades = {'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/111.0.0.0 Safari/537.36'}
```

# PREPROCESSING

1. Mendownload data yang diperlukan oleh modul Natural Language Toolkit (NLTK) dari server NLTK
2. Menghapus baris-baris duplikat dari DataFrame dan mengatur ulang indeks dari DataFrame
3. Menghapus kolom yang tidak diperlukan dari DataFrame df
4. Membersihkan teks dari karakter, tanda baca, stop words, dan mengembalikan teks yang telah diproses secara bersih
5. Melakukan proses pemfilteran kata
6. Menghitung sentimen dari setiap kata dalam kolom 'Headline' dari DataFrame df

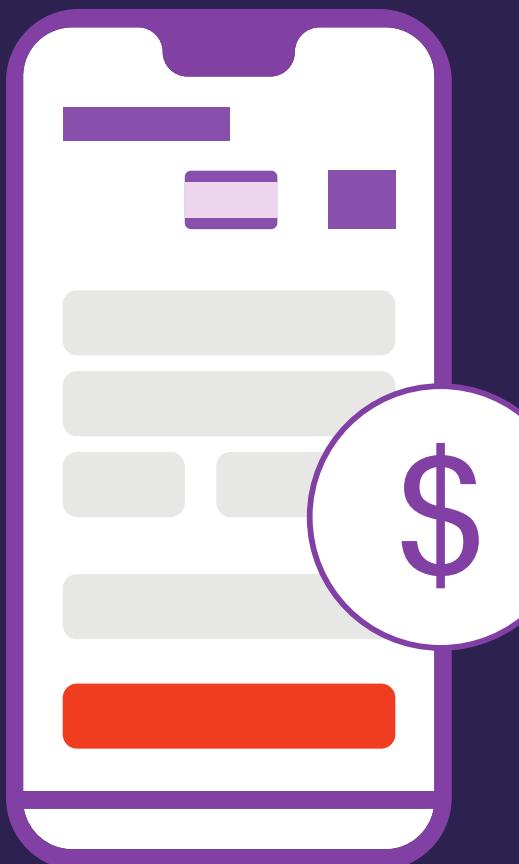


Dibawah ini merupakan hasil output dari data yang telah dilakukan preprocessing (penghapusan baris duplikat, pengaturan ulang indeks, penghapusan kolom yang tidak diperlukan, membersihkan teks dari karakter, tanda baca, stop words, pemfilteran kata, dan menghitung sentimen.

```
[ ] df
```

	Headline	Date	Content
0	bluetti indonesia hadirkan kemudahan mengakses...	16 Jun 2023 13:21	Perusahaan penyedia portable power stations da...
1	blueprintada solusi integrasi terbaik bisnis r...	16 Jun 2023 13:06	Sejak pandemi COVID-19, Indonesia menjadi salah...
2	kisah niniek beromzet rp juta bikin abon samba...	16 Jun 2023 06:30	Sambal menjadi salah satu favorit masyarakat I...
3	eksklusif interview ceo makuku inovasi produk ...	16 Jun 2023 00:00	Perkembangan industri popok di Indonesia mence...
4	tips investasi emas pemula	15 Jun 2023 10:05	Beberapa waktu lalu dunia investasi dikejutkan...
...	...	...	...
128	kartu sih informasi lengkap prefix daerah	27 Jan 2023 18:48	Mungkin kalian pernah mendapatkan telepon dari...
129	badai phk melanda perusahaan indonesia	20 Nov 2022 12:44	Badai pemutusan hubungan kerja (PHK) masih ter...
130	uang kertas rp gambar kapal pinisi dijual mahal	29 Okt 2022 15:20	Banyak orang mencari uang lama yang sudah tida...
131	dalam badai phk landa perusahaan	24 Okt 2022 16:17	Kinerja sejumlah perusahaan memburuk. Mereka p...
132	peristiwa populer pemukulan petugas shopee gin...	16 Okt 2022 11:30	Ada 5 peristiwa menggegerkan yang terjadi dala...

133 rows × 3 columns



# Statistika Deskriptif

Pada proses ini, dilakukan running code untuk menghitung mean, median, modus, kuartil, variansi, dan standar deviasi. Namun, output yang bisa menghasilkan informasi hanya modus saja. Hal ini dikarenakan keseluruhan data yang digunakan pada analisis ini merupakan data bertipe string, sehingga tidak bisa dilakukan operasi numerik.

```
# Menampilkan ringkasan statistika deskriptif
df.describe()



|        | Headline                                         | Date              | Content                                           |
|--------|--------------------------------------------------|-------------------|---------------------------------------------------|
| count  | 133                                              | 133               | 133                                               |
| unique | 128                                              | 128               | 128                                               |
| top    | pilih oli terbaik tingkatkan kinerja mesin mobil | 22 Apr 2021 00:00 | Secara umum, oli berfungsi melumasi bagian-bag... |
| freq   | 2                                                | 2                 | 2                                                 |


```

# Data Wrangling

Pada proses data wrangling dilakukan identifikasi missing value pada dataset yang dimiliki. Namun, karena tidak ditemukan missing value, maka tidak perlu dilakukan langkah selanjutnya berupa *handle missing value*

## ▼ Data Wrangling

Langkah yang dilakukan pada data wrangling adalah dengan menghitung jumlah missing value pada setiap kolom dan mengatasinya.

```
[ ] # Menghitung jumlah missing value pada setiap kolom
df.isnull().sum()

# Menghitung jumlah missing value dalam keseluruhan DataFrame
df.isnull().sum().sum()
```

0

# VISUALISASI DATA

## 1) Visualisasi Cirrus

Pada visualisasi cirrus di samping, dapat terlihat bahwa ukuran kata menunjukkan frekuensi kemunculan kata tersebut dalam data yang sedang dianalisis. Selain itu, jarak antara kata-kata juga berhubungan dengan keterkaitan kata tersebut dalam data yang sedang dianalisis.

# VISUALISASI DATA

## 2) Visualisasi Term

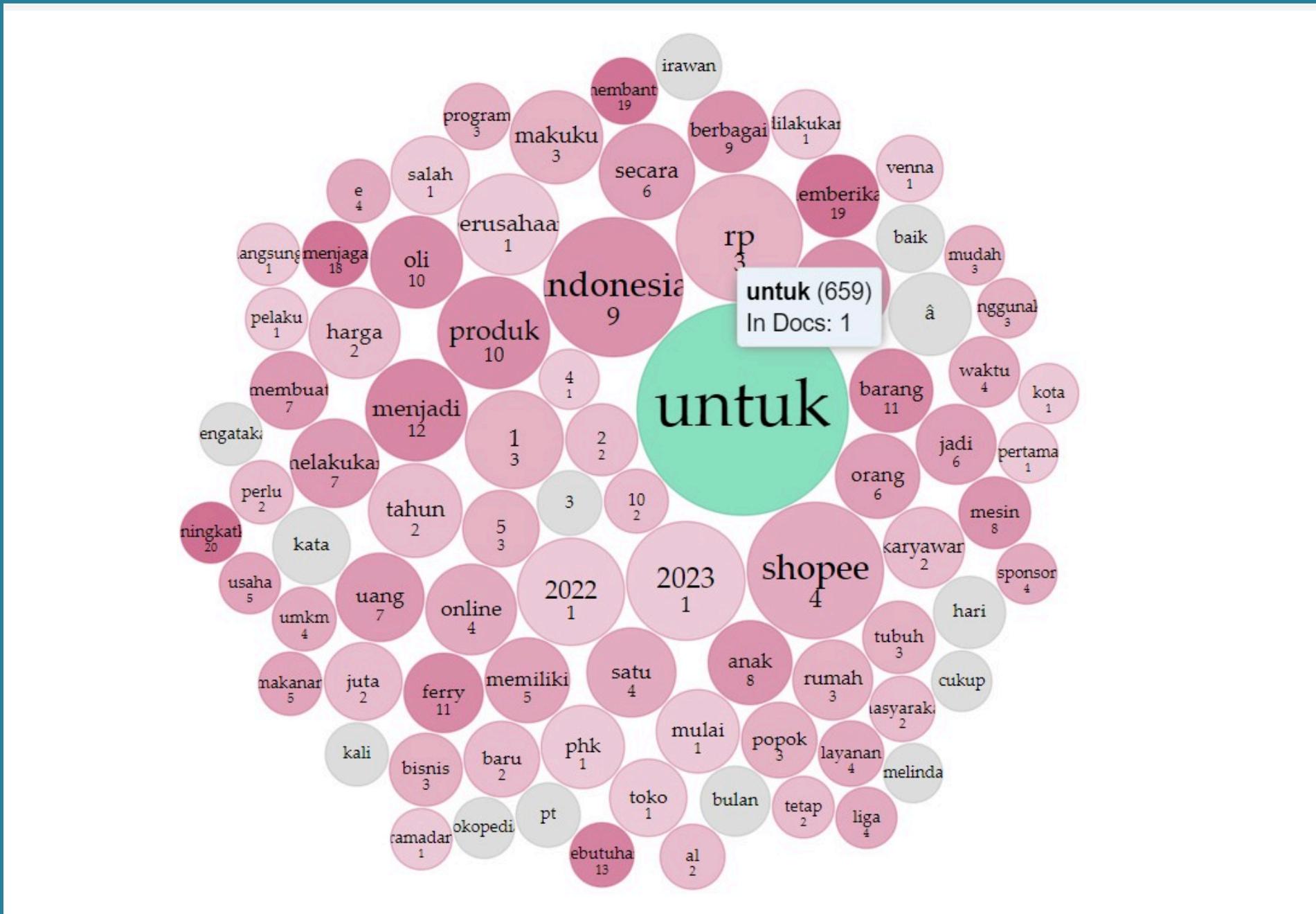
		Term	Count	Trend
1	untuk		659	
2	indonesia		285	
3	shopee		274	
4	rp		235	
5	2023		207	
6	produk		185	
7	2022		169	
8	menjadi		152	
9	perusahaan		150	
10	1		141	
11	mobil		137	
12	secara		135	
13	tahun		129	
14	makuku		127	
15	oli		125	
16	harga		122	
17	online		121	
18	satu		117	
19	uang		115	

Visualisasi Terms digunakan untuk melihat jumlah munculnya sebuah kata dan mengurutkan dari yang terbanyak hingga yang paling sedikit.

Pada visualisasi tersebut terlihat kata yang paling banyak muncul adalah "untuk" yaitu sebanyak 659 kali.

# VISUALISASI DATA

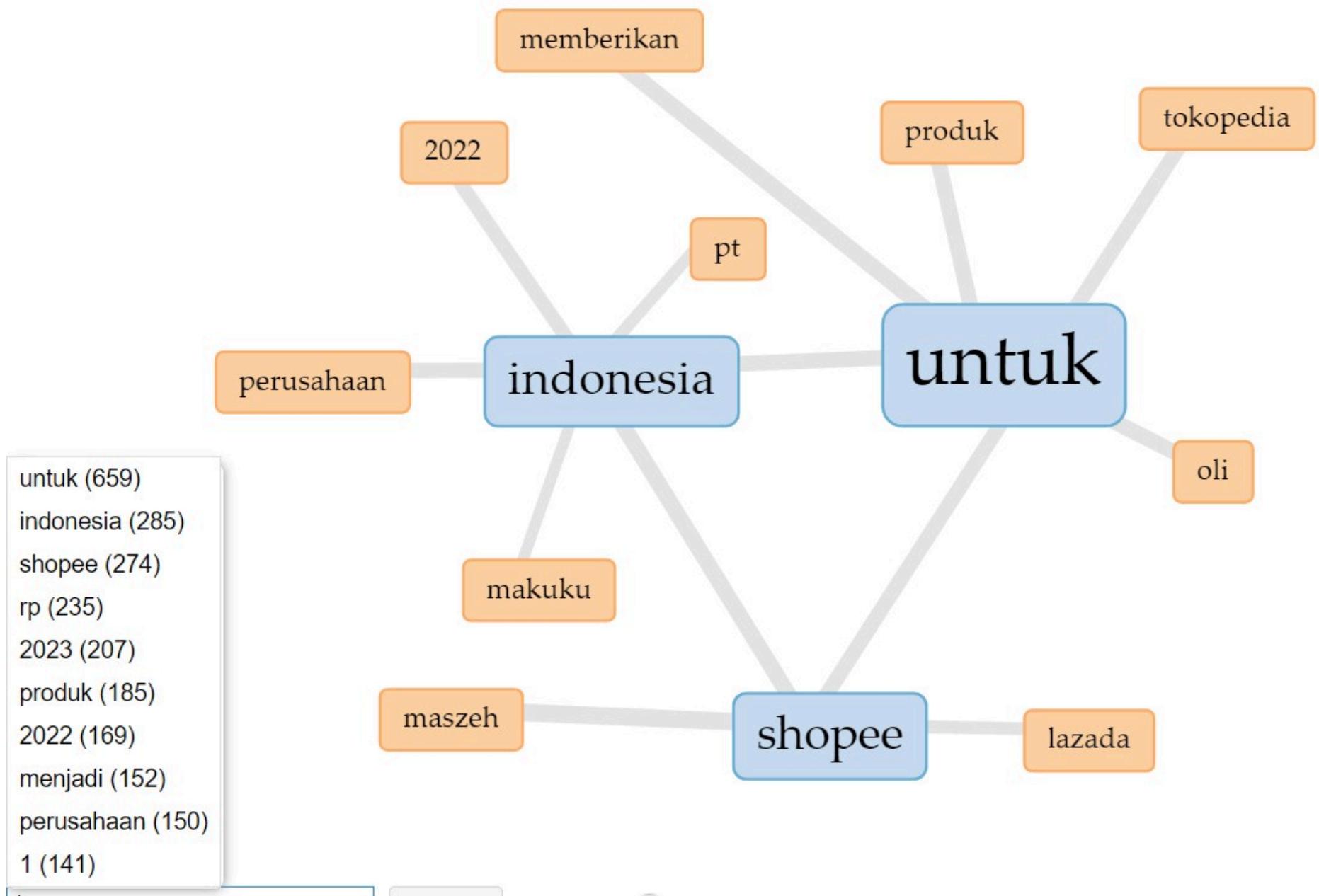
### 3) Visualisasi Terms Berry



Pada visualisasi terms berry di samping, dapat terlihat bahwa ukuran berry yang berbeda menunjukkan frekuensi kemunculan kata tersebut. Selain itu, jarak antara kata-kata juga berhubungan dengan keterkaitan kata tersebut dalam data yang sedang dianalisis.

# VISUALISASI DATA

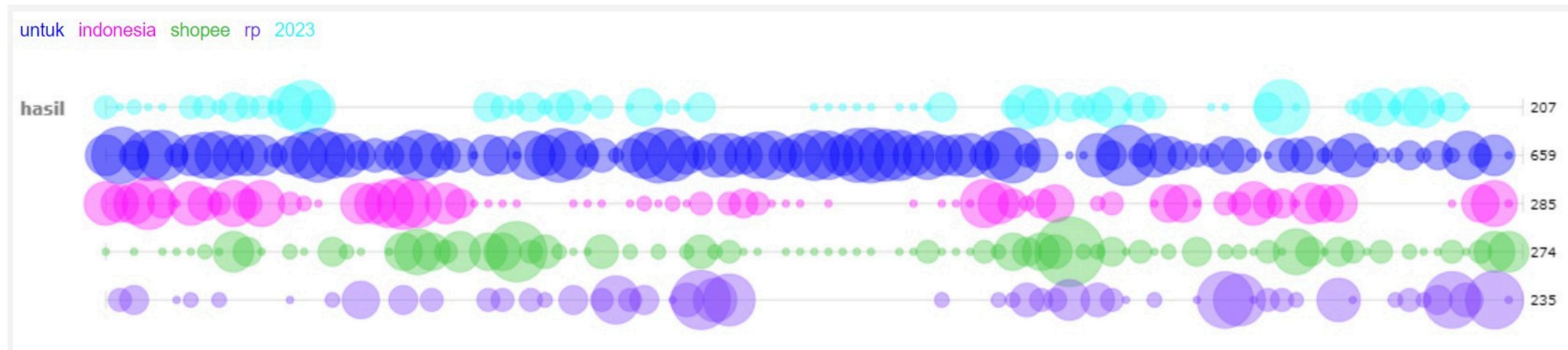
## 4) Visualisasi Links



Visualisasi Links digunakan untuk melihat hubungan atau koneksi .antar kata. Pada visualisasi links disamping terlihat bahwa kata yang paling banyak berhubungan dengan kata lainnya adalah "untuk", "indonesia" dan "shopee".

# VISUALISASI DATA

## 5) Visualisasi Bubblelines



Pada visualisasi bubble lines di atas, dapat dilihat bahwa ukuran gelembung menunjukkan ukuran dari data tersebut.

# VISUALISASI DATA

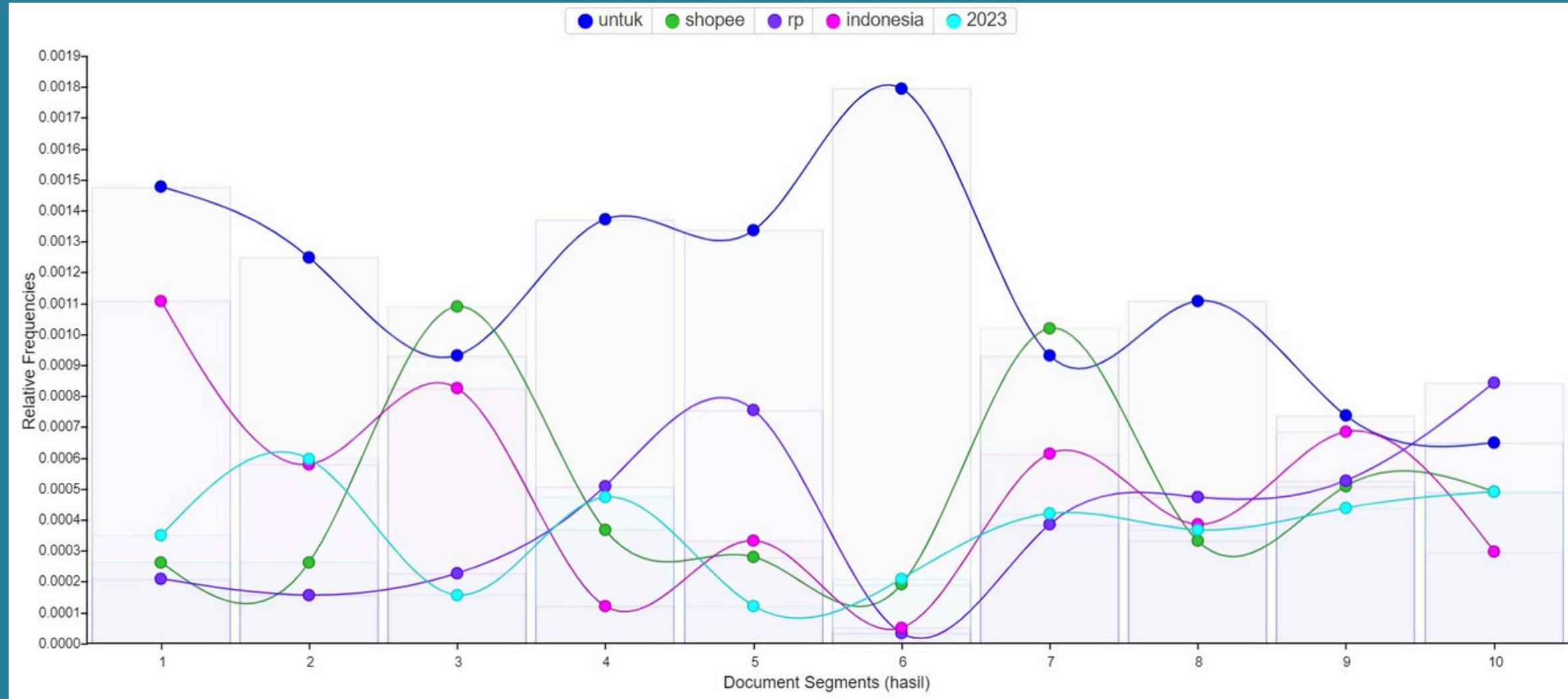
## 6) Visualisasi Dreamscape



Berdasarkan  
pemahaman kami  
dalam melihat  
visualisasi ini, kami  
menyimpulkan bahwa  
visualisasi ini  
menggambarkan  
wilayah yang  
disebutkan dalam data  
berita yang kami  
ambil.

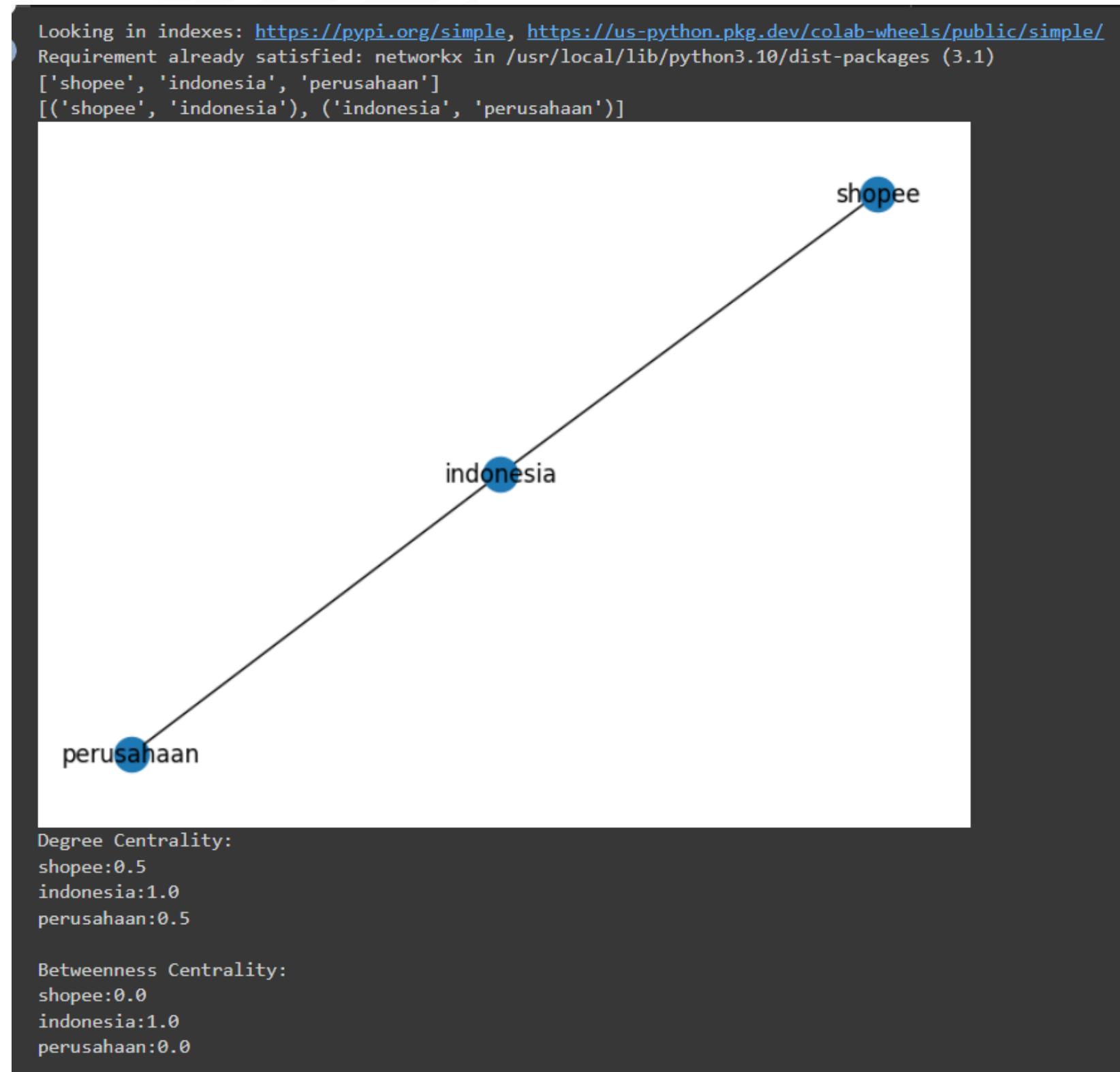
# VISUALISASI DATA

## 7) Visualisasi Trend



Pada visualisasi trend di samping, dapat dilihat bahwa tren kata dari waktu ke waktu mengalami perubahan seiring perubahan waktu.

# Social Network Analysis



Langkah-langkah dalam melakukan social network analysis adalah sebagai berikut.

1. Mengimpor modul yang diperlukan
2. Menggambar grafik
3. Melakukan social network analysis

# KESIMPULAN

This corpus has 1 document with 56,860 total words and 8,397 unique word forms. Created about 32 minutes ago.

**Vocabulary Density:** 0.148

**Readability Index:** 18.196

**Average Words Per Sentence:** 24.0

Most frequent words in the corpus:

- untuk (659); indonesia (285); shopee (274); rp (235); 2023 (207); produk (185); 2022 (169); menjadi (152); perusahaan (150); 1 (141); mobil (137); secara (135); tahun (129); makuku (127); oli (125); harga (122); online (121); satu (117); uang (115); melakukan (115); memiliki (106); orang (105); anak (104); phk (102); mulai (102); barang (102); â (101); memberikan (100); berbagai (97); karyawan (96); rumah (94); jadi (94); ferry (93); salah (90); membuat (89); kata (89); juta (89); toko (88); 5 (86); popok (84); baru (84); bisnis (78); tubuh (77); hari (77); mesin (76); 2 (76); waktu (73); bulan (72); dilakukan (70); baik (68); venna (65); program (65); membantu (65); menjaga (64); irawan (64); masyarakat (63); layanan (63); al (62); 3 (62)

Dari hasil scrapping berita dengan topik "Shopee" yang bersumber dari detik.com, kami memproses dan menganalisis suatu data dengan total kata 56.860 dan 8.397 kata unik. Didapatkan kesimpulan sebagai berikut.

1. Hingga saat ini, Shopee masih menjadi salah satu e-commerce yang populer di kalangan masyarakat, dibuktikan dengan masih banyaknya jumlah berita yang beredar di situs berita.
2. Dari hasil analisis, didapatkan kata kunci yang paling sering muncul adalah kata 'Shopee', 'Untuk', dan 'Indonesia'
3. Dari hasil analisis sentimen, didapatkan sebagian besar masyarakat menunjukkan sentimen yang positif terhadap Shopee sebagai salah satu e-commerce di Indonesia.

# Terima Kasih!

