

# **Deteksi Komunitas pada Pengguna Aplikasi Kencan Berdasarkan Kebiasaan dan Perilaku Menggunakan Metode Louvain**

Kelompok 8 : Tarisma Dwi Putri Alifianti

Aisyah Mufidah Najwa

Ikfan Putra Maesru Dwi Pradana

# Pendahuluan

Pertumbuhan dunia digital yang pesat membawa tantangan baru dalam melakukan interaksi sosial. salah satunya adalah munculnya aplikasi kencan yang memungkinkan individu untuk berinteraksi lintas batas geografis. Interaksi digital ini menciptakan jaringan sosial yang kompleks dan dinamis, sehingga diperlukan pendekatan analisis untuk memahami struktur komunitas yang terbentuk secara implisit dari perilaku dan preferensi pengguna. Penelitian ini menggunakan pendekatan cosine similarity untuk mengukur kemiripan antar pengguna dan algoritma Louvain untuk mengidentifikasi komunitas dalam jaringan, guna memberikan wawasan terhadap perilaku pengguna serta pengembangan fitur rekomendasi yang lebih personal.

# Tujuan Penelitian

1

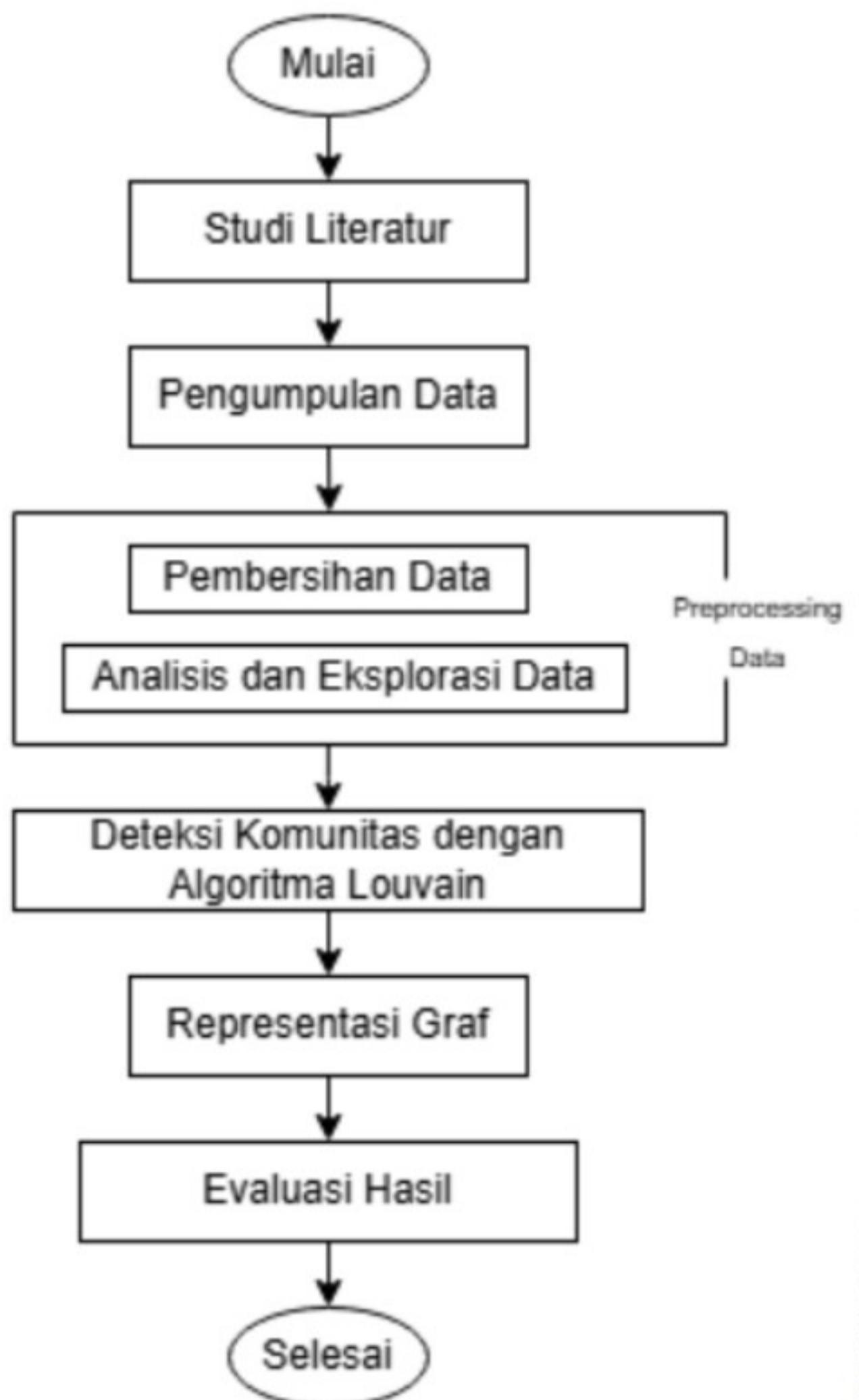
Mengidentifikasi  
komunitas pengguna  
aplikasi kencan

2

Menganalisis pola  
keterhubungan  
berdasarkan kemiripan  
perilaku.

menggunakan pendekatan cosine similarity dan  
algoritma deteksi komunitas Louvain.

# Alur Penelitian



# Studi Literatur

M. Newman, “Networks: An Introduction,” *Networks: An Introduction*, pp. 1–784, Sep. 2010,  
doi:  
[10.1093/ACPROF:OSO/9780199206650.001.0001](https://doi.org/10.1093/ACPROF:OSO/9780199206650.001.0001).

D. H. Felmlee and D. A. Kreager, “The Invisible Contours of Online Dating Communities: A Social Network Perspective,” *Journal of Social Structure*, vol. 18, no. 1, pp. 1–28, Jan. 2017, doi: [10.21307/joss-2018-004](https://doi.org/10.21307/joss-2018-004).

# Data Understanding

Dataset yang digunakan terdiri dari 59.946 profil publik pengguna aktif aplikasi OkCupid, dengan 31 atribut yang diambil dari website kaggle. Data ini mencakup informasi demografis seperti usia, jenis kelamin, orientasi seksual, dan etnis, serta gaya hidup seperti pola makan, kebiasaan minum, dan merokok. Selain itu, terdapat data semi-terstruktur berupa deskripsi profil yang menggambarkan preferensi dan nilai hidup pengguna. Dataset ini dipilih berdasarkan kriteria lokasi, aktivitas akun, dan kelengkapan profil.

## DATASET



### OkCupid Profiles

Explore dating app profiles in search for love

[kaggle.com](https://www.kaggle.com)

# Pre-Processing Data

Pemilihan kolom dilakukan berdasarkan relevansi terhadap karakteristik sosial, tingkat kelengkapan data, serta hasil korelasi awal terhadap pembentukan komunitas. Setelah melalui proses seleksi, dipilih 12 kolom yang dianggap paling informatif, yaitu: education, job, religion, orientation, drinks, location, status, diet, offspring, smokes, drugs, dan age.

Data kemudian dibersihkan dari nilai kosong (missing values) pada kolom-kolom terpilih. Seluruh baris yang mengandung nilai kosong pada salah satu kolom tersebut dihapus, sehingga jumlah data menyusut dari 59.946 menjadi 9.922 baris.

Selain itu, dilakukan penyaringan pada kolom usia (age) karena terdapat outlier dengan nilai hingga 110 tahun. Berdasarkan analisis interkuartil (IQR), batas atas ditentukan pada usia 53 tahun. Oleh karena itu, hanya pengguna berusia antara 18 hingga 53 tahun yang disertakan dalam analisis.

# Pre-Processing Data

Usia kemudian dikelompokkan ke dalam enam rentang interval 6 tahun, yaitu 18–23, 24–29, 30–35, 36–41, 42–47, dan 48–53. Setiap pengguna diberikan label kelompok usia (age\_group) yang sesuai.

informasi location pengguna yang semula dalam format "city, state" dipisahkan, dan hanya bagian negara bagian (state) yang diambil untuk penyederhanaan.

Untuk mengurangi beban komputasi dan memudahkan visualisasi data dalam tahap eksplorasi, dilakukan pengambilan sampel sebanyak 1.000 data secara acak dari total 9.128 baris hasil pembersihan. Pengambilan sampel ini menggunakan metode random sampling.

# Hasil

Fokus analisis ditujukan pada individu-individu yang berasal dari California karena setelah ditelusuri, terdapat 995 individu yang berasal dari negara bagian tersebut. Sementara itu, sisanya, hanya lima individu, berasal dari lokasi yang beragam seperti Switzerland, New York, Arizona, Netherlands, dan Ohio. Keberadaan lima individu ini justru membentuk dua komunitas outlier yang muncul karena lokasi mereka sangat berbeda dari mayoritas data.

Setelah dilakukan penyaringan data, jumlah individu yang dianalisis menjadi 995, dan jumlah komunitas yang awalnya 12 menyusut menjadi 10 komunitas utama yang lebih relevan untuk dianalisis lebih lanjut. Digunakan pendekatan cosine similarity antar pengguna, dengan hanya menghubungkan pasangan individu yang memiliki nilai kemiripan di atas 0,7. Proses deteksi komunitas dilakukan menggunakan algoritma Louvain.

Struktur komunitas yang terbentuk menunjukkan adanya pola keterhubungan yang cukup kuat di dalam tiap kelompok, dengan nilai modularity sebesar 0,686.

# Implementasi Cosine similarity dan Louvain

```
import pandas as pd
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.metrics.pairwise import cosine_similarity
import networkx as nx
import community as community_louvain # pip install python-louvain

# 1. Filter umur 18-53 dari df3
df_filtered = df3[(df3['age'] >= 18) & (df3['age'] <= 53)].copy()

# 2. Ambil 1000 data random dari hasil filter (jika data kurang dari 1000, ambil semua)
n_sample = 1000
if len(df_filtered) < n_sample:
    n_sample = len(df_filtered)
df_sample = df_filtered.sample(n=n_sample, random_state=42).copy()

# 3. Buat kelompok umur (age_group)
Tabnine | Edit | Test | Explain | Document
def age_group(age):
    if 18 <= age <= 23:
        return '18-23'
    elif 24 <= age <= 29:
        return '24-29'
    elif 30 <= age <= 35:
        return '30-35'
    elif 36 <= age <= 41:
        return '36-41'
    elif 42 <= age <= 47:
        return '42-47'
    elif 48 <= age <= 53:
        return '48-53'
    else:
        return 'other'

df_sample['age_group'] = df_sample['age'].apply(age_group)
```

```
# 4. Kolom kategorikal yang akan diencode
cat_cols = ['education', 'job', 'religion', 'orientation', 'drinks',
            'state', 'status', 'diet', 'offspring', 'smokes', 'drugs', 'age_group']

df_sample = df_sample[df_sample['state'] == 'california'].copy()

df_encoded = df_sample.copy()

# 6. Encoding
le_dict = {}
for col in cat_cols:
    le = LabelEncoder()
    df_encoded[col] = le.fit_transform(df_encoded[col])
    le_dict[col] = le

# 7. Scaling fitur
scaler = StandardScaler()
df_scaled = scaler.fit_transform(df_encoded[cat_cols])

# 8. Hitung similarity
sim_matrix = cosine_similarity(df_scaled)

# 9. Buat graph dari similarity
threshold = 0.7
G = nx.Graph()
n = len(df_scaled)

for i in range(n):
    G.add_node(i)

for i in range(n):
    for j in range(i + 1, n):
        if sim_matrix[i, j] > threshold:
            G.add_edge(i, j, weight=sim_matrix[i, j])

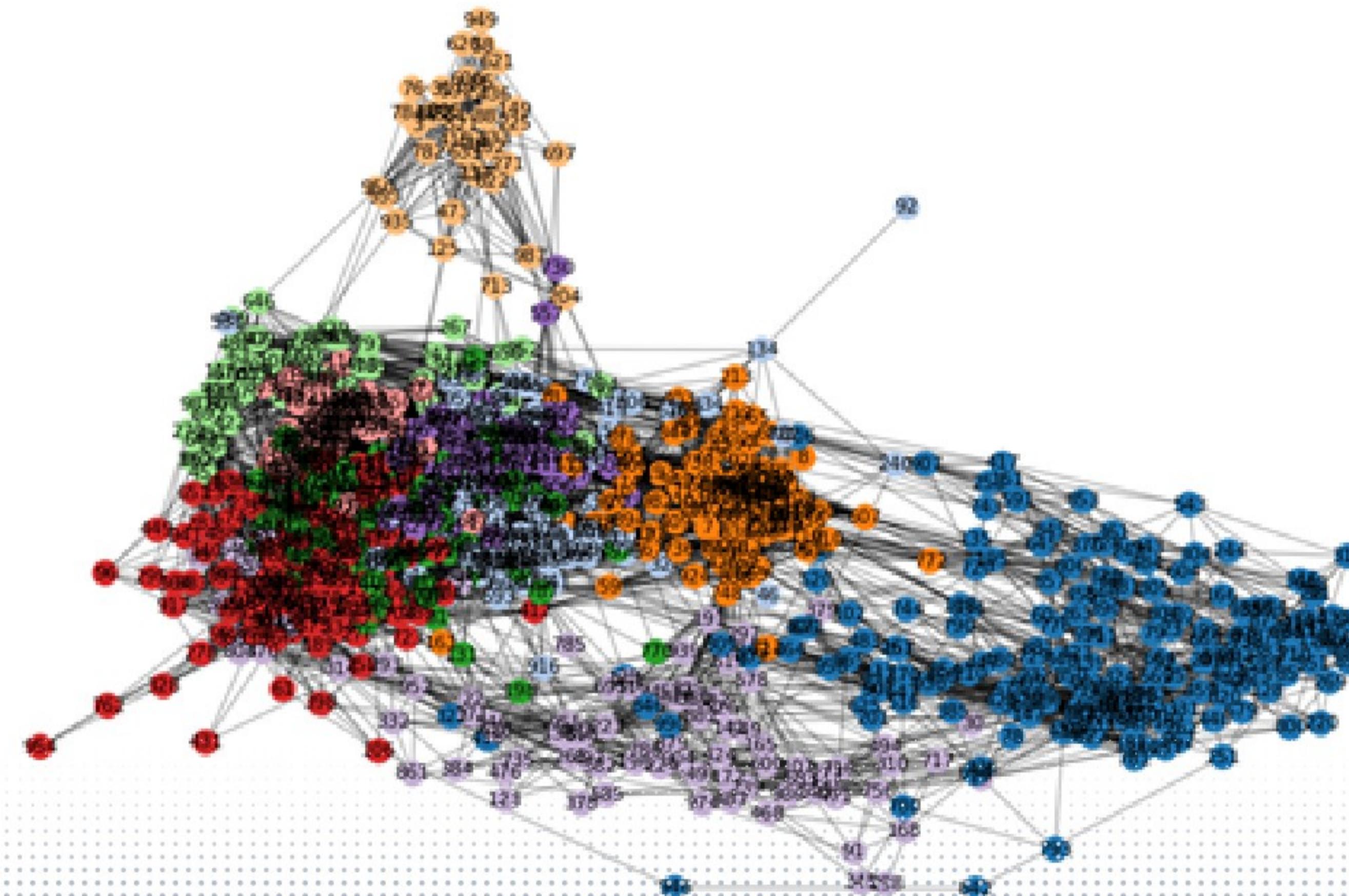
# 10. Louvain community detection
partition = community_louvain.best_partition(G, weight='weight')

# 11. Tambahkan hasil ke df_sample
df_sample['community'] = df_sample.index.map(partition)

# 11. Tampilkan hasil partition (node ke komunitas)
print(partition)
```

# Visualisasi graph

Community Detection Louvain on df3 Similarity Graph



# Jumlah Komunitas

Komunitas	Jumlah Anggota
0	195
1	90
2	117
3	40
4	96
5	59
6	145
7	81
8	85
9	87

# Pembahasan

## **Komunitas 0:** Generasi Muda Liberal Urban (195 anggota)

Komunitas terbesar yang didominasi anak muda usia 18-29 tahun di California dengan pendidikan tinggi dan profesi beragam. Mayoritas ateis/agnostik, heteroseksual, dan single dengan pola konsumsi alkohol dan narkoba yang tinggi (138 orang kadang narkoba, 126 minum sosial). Karakteristik utama: kelompok muda urban yang sangat terbuka dan aktif secara sosial dengan gaya hidup eksploratif.

## **Komunitas 1:** Profesional Terdidik (90 anggota)

Mayoritas usia 24-35 tahun dengan pendidikan S1/S2 dan profesi di marketing/teknologi. Religius campuran (ateis, agnostik, Kristen), hampir semua heteroseksual dan single dengan gaya hidup sehat (87 orang tidak narkoba, mayoritas tidak merokok). Karakteristik utama: profesional muda yang rasional dan cenderung menghindari risiko dalam gaya hidup.

## **Komunitas 2: Mahasiswa Fokus (117 anggota)**

Komunitas mahasiswa dengan mayoritas usia 18-29 tahun yang sedang kuliah S1/S2/PhD dengan profesi dominan mahasiswa. Religius beragam, hampir semua heteroseksual dan single dengan disiplin tinggi (115 tidak narkoba, 109 tidak merokok). Karakteristik utama: mahasiswa dengan fokus studi tinggi dan gaya hidup sangat terkontrol.

## **Komunitas 3: Dewasa Kreatif Beragam (40 anggota)**

Mayoritas usia 36-41 tahun dengan profesi di seni, akademik, dan kesehatan. Mayoritas ateis/agnostic dengan orientasi seksual beragam (27 heteroseksual, 7 biseksual, 6 gay) dan status mayoritas "available". Karakteristik utama: kelompok dewasa yang terbuka terhadap keragaman dan ekspresif dalam pilihan hidup.

## **Komunitas 4: Profesional Urban Terdidik (96 anggota)**

Komunitas berpendidikan tinggi mayoritas usia 24-47 tahun dengan profesi beragam dan religius mayoritas "other"/Yahudi. Hampir semua heteroseksual dan single dengan gaya hidup sangat sehat (96 tidak narkoba, 92 tidak merokok). Karakteristik utama: elit profesional dengan nilai inklusif dan komitmen tinggi terhadap kesehatan.

### **Komunitas 5: Profesional Konservatif (59 anggota)**

Mayoritas usia 24-47 tahun dengan pendidikan tinggi dan religius mayoritas Kristen/agnostik. Pola hidup konservatif dengan 35 orang tidak minum alkohol sama sekali dan 59 tidak pernah narkoba. Karakteristik utama: kelompok yang menjunjung tinggi nilai-nilai tradisional dan gaya hidup teratur.

### **Komunitas 6: Kelas Menengah Mapan (145 anggota)**

Mayoritas dewasa usia 36-53 tahun dengan profesi manajemen/kesehatan dan 57 orang memiliki anak. Religius beragam, mayoritas single meski punya anak, dengan gaya hidup sehat (142 tidak narkoba, 135 tidak merokok). Karakteristik utama: kelompok mapan dengan pengalaman hidup luas dan tanggung jawab keluarga.

### **Komunitas 7: Profesional Teknologi Muda (81 anggota)**

Mayoritas usia 24-35 tahun dengan profesi dominan teknologi/finansial. Seluruh anggota heteroseksual, single, dan sangat disiplin (81 orang tidak pernah narkoba, 78 tidak merokok). Karakteristik utama: profesional teknologi dengan kedisiplinan tinggi dan fokus karier yang kuat.

### **Komunitas 8: Profesional Sains Religius (85 anggota)**

Mayoritas usia 24-35 tahun dengan profesi sains/pemasaran dan religius mayoritas Katolik/Kristen. Seluruh anggota heteroseksual dengan gaya hidup sangat sehat (85 tidak narkoba, 83 tidak merokok). Karakteristik utama: perpaduan unik antara latar belakang sains dan nilai-nilai religius yang kuat.

### **Komunitas 9: Komunitas Kreatif LGBTQ+ (87 anggota)**

Komunitas unik dengan seluruh anggota non-heteroseksual (54 biseksual, 33 gay) dengan mayoritas usia 18-35 tahun. Profesi dominan mahasiswa/kreatif dengan pola konsumsi lebih bervariasi (32 kadang narkoba, 12 kadang merokok). Karakteristik utama: komunitas LGBTQ+ yang ekspresif dan terbuka terhadap eksplorasi identitas.

# Perbedaan Antar Komunitas

Analisis terhadap sepuluh komunitas yang teridentifikasi menunjukkan adanya variasi signifikan dalam karakteristik demografis, gaya hidup, dan preferensi personal. Setiap komunitas yang terbentuk merepresentasikan kelompok pengguna dengan kesamaan tertentu, sehingga dapat digunakan sebagai dasar dalam sistem rekomendasi, baik untuk pertemanan maupun pencarian pasangan. Misalnya, pengguna dalam komunitas yang sama cenderung memiliki kecocokan dalam minat dan gaya hidup, sehingga rekomendasi interaksi atau pertemanan dapat difokuskan pada individu dari komunitas tersebut.

# Perbedaan Antar Komunitas

## a. Perbedaan Berdasarkan Demografi Usia

- Komunitas 0, 2, dan 9 didominasi oleh generasi muda berusia 18-29 tahun, yang umumnya merupakan mahasiswa atau profesional pemula. Kelompok ini menunjukkan karakteristik eksplorasi identitas dan gaya hidup yang lebih eksperimental.
- Komunitas 6 merepresentasikan kelompok dewasa mapan berusia 36-53 tahun yang telah memiliki tanggung jawab keluarga dan stabilitas karier.
- Komunitas 1, 4, 7, dan 8 berada di kategori menengah (24-35 tahun), mencerminkan fase transisi dari eksplorasi menuju stabilitas.

# Perbedaan Antar Komunitas

## a. Variasi Orientasi Seksual dan Identitas

Perbedaan paling mencolok terlihat pada Komunitas 9 yang secara eksklusif terdiri dari individu dengan orientasi non-heteroseksual (biseksual dan gay). Komunitas ini menunjukkan karakteristik unik dalam hal ekspresivitas, keterbukaan terhadap eksperimen, dan keterlibatan dalam bidang artistik. Sebaliknya, sembilan komunitas lainnya didominasi oleh individu heteroseksual dengan variasi yang minimal dalam orientasi seksual.

## a. Spektrum Gaya Hidup (pola konsumsi alkohol, narkoba, dan kebiasaan merokok)

Komunitas 0 menempati posisi paling liberal dengan tingkat konsumsi alkohol dan narkoba yang tinggi, mencerminkan gaya hidup urban yang terbuka dan eksperimental. Komunitas 5, 7, dan 8 berada di ujung konservatif dengan tingkat abstinensi yang tinggi terhadap narkoba dan kebiasaan merokok. Komunitas 5 bahkan menunjukkan proporsi signifikan individu yang sama sekali tidak mengonsumsi alkohol, mengindikasikan pengaruh nilai-nilai religius atau budaya yang lebih ketat.

# Perbedaan Antar Komunitas

## a. Pola Keagamaan dan Spiritualitas

Distribusi afiliasi keagamaan menunjukkan polarisasi yang menarik. Komunitas 0 dan beberapa komunitas muda lainnya didominasi oleh ateis dan agnostik, mencerminkan tren sekularisasi di kalangan generasi muda urban. Komunitas 4 menunjukkan keunikan dengan dominasi agama Yahudi dan kategori "lainnya", mengindikasikan diversitas religius yang lebih tinggi. Sebaliknya, Komunitas 6, 8, dan 9 menunjukkan prevalensi yang lebih tinggi untuk agama-agama tradisional seperti Katolik dan Kristen, meskipun dengan interpretasi dan praktik yang mungkin beragam.

# Kesimpulan

Penelitian ini berhasil mengidentifikasi komunitas pengguna aplikasi kencan OkCupid dengan menerapkan algoritma Louvain pada jaringan sosial yang dibangun berdasarkan cosine similarity. Dari hasil analisis terhadap 995 pengguna, ditemukan 10 komunitas utama dengan nilai modularitas sebesar 0,686, menunjukkan keterhubungan internal yang tinggi dalam tiap kelompok. Setiap komunitas memiliki karakteristik unik yang mencerminkan segmentasi sosial berdasarkan usia, orientasi seksual, pekerjaan, pola hidup, dan nilai-nilai pribadi.

Temuan ini menunjukkan bahwa pengguna cenderung membentuk relasi dalam kelompok yang serupa secara sosial dan perilaku, bukan secara acak. Hal ini membuka peluang penerapan sistem rekomendasi berbasis komunitas untuk meningkatkan relevansi dan efektivitas interaksi dalam aplikasi. Selain itu, penelitian ini juga memberikan kontribusi dalam memahami dinamika jejaring sosial dalam konteks digital yang semakin kompleks dan personal.

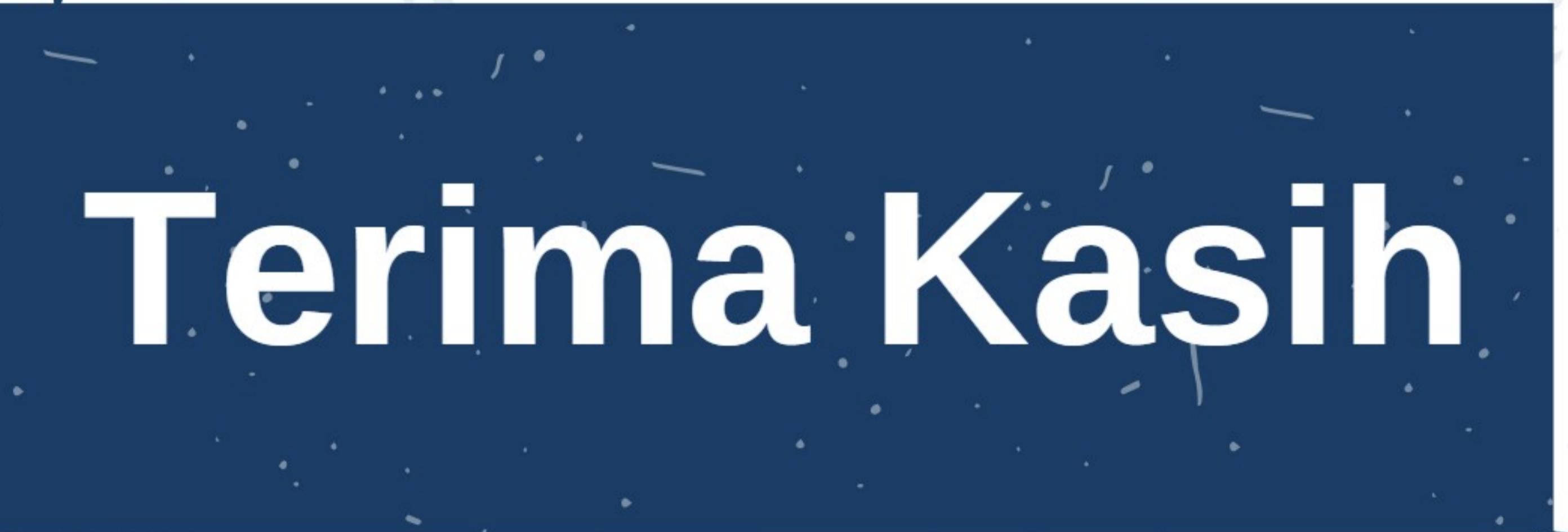


Google Colab  
[google.com](https://google.com)



colab

**CODE**



**Terima Kasih**