

Import Libraries

```
In [2]: import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
from sklearn.preprocessing import normalize
```

Import beberapa library yang dibutuhkan.

Loading Data

```
In [3]: df = pd.read_csv('/content/drive/MyDrive/nlp/responsi-nlp/dataset_mentalhealth.csv')
df.head()
```

```
Out[3]:
```

	Question_ID	Questions	Jawaban
0	1590140	Apa yang dimaksud dengan penyakit mental?	Penyakit mental adalah kondisi kesehatan yang ...
1	2110618	Siapa yang terpengaruh oleh penyakit mental?	Diperkirakan bahwa penyakit mental mempengaruh...
2	6361820	Apa penyebab penyakit mental?	Diperkirakan bahwa penyakit mental mempengaruh...
3	9434130	Apa sajakah tanda-tanda peringatan penyakit me...	Gejala gangguan kesehatan mental bervariasi te...
4	7657263	Apakah penderita penyakit jiwa bisa sembuh?	Ketika penyembuhan dari penyakit mental, ident...

Memuat data QnA tentang mental health.

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

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 98 entries, 0 to 97
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype  
---  -
 0   Question_ID 98 non-null    int64  
 1   Questions   98 non-null    object  
 2   Jawaban      98 non-null    object  
dtypes: int64(1), object(2)
memory usage: 2.4+ KB
```

Dataset ini berisi 3 kolom yaitu Question_ID, Questions, dan Jawaban.

Preprocessing

```
In [5]: def remove_unnecessary_char(text):
        text = re.sub(r'^a-zA-Z0-9\s', '', text)
        text = re.sub(r'\s+', ' ', text).strip()
        return text

        def lowercase(text):
            return text.lower()

        def preprocess(text):
            text = remove_unnecessary_char(text)
            text = lowercase(text)
            return text

        df['Questions'] = df['Questions'].apply(preprocess)
```

Langkah-langkah *preprocessing* data yang dilakukan yang pertama adalah pembersihan karakter yang tidak dibutuhkan seperti tanda baca dan spasi tambahan, lowercase seluruh karakter.

```
In [6]: df.head()
```

```
Out[6]:
```

	Question_ID	Questions	Jawaban
0	1590140	apa yang dimaksud dengan penyakit mental	Penyakit mental adalah kondisi kesehatan yang ...
1	2110618	siapa yang terpengaruh oleh penyakit mental	Diperkirakan bahwa penyakit mental mempengaruhi...
2	6361820	apa penyebab penyakit mental	Diperkirakan bahwa penyakit mental mempengaruhi...
3	9434130	apa sajakah tandatanda peringatan penyakit mental	Gejala gangguan kesehatan mental bervariasi te...
4	7657263	apakah penderita penyakit jiwa bisa sembuh	Ketika penyembuhan dari penyakit mental, ident...

Tampilan data setelah dibersihkan.

Vectorization

TF-IDF

```
In [7]: question = df['Questions']
        answers = df['Jawaban']
```

```
In [8]: vectorizer = TfidfVectorizer()
        tfidf_matrix = vectorizer.fit_transform(question)
```

```
In [9]: tfidf_matrix_normalized = normalize(tfidf_matrix)
```

Vektorisasi pada program ini menggunakan teknik TF-IDF sederhana.

Cosine Similarity

```
In [10]: def get_response(user_query, answers):  
    user_vector = vectorizer.transform([user_query])  
    user_vector_normalized = normalize(user_vector)  
  
    cosine_similarities = cosine_similarity(user_vector_normalized, tfidf_ma  
  
    most_similar_index = np.argmax(cosine_similarities)  
  
    response = answers[most_similar_index]  
    return response
```

Cosine similarity digunakan untuk mengukur kesamaan antara dua vektor yang berbeda. Cosine similarity mengukur kedekatan antara dua vektor dengan mengukur kosinus sudut antara mereka.

Prompt Input

```
In [12]: print("Halo! Saya adalah bot yang siap membantu Anda dengan pertanyaan seputar kesehatan mental. Tanyakan sesuatu atau ketik 'exit' untuk keluar.")

while True:
    user_query = input("Anda: ")

    if user_query.lower() in ['exit', 'quit', 'stop']:
        print("Terima kasih! Sampai jumpa.")
        break

    response = get_response(user_query, answers)

    if response:
        print("Bot:", response)
    else:
        print("Bot: Maaf, saya tidak mengerti pertanyaan Anda.")

    print("=" * 40)

print("Bot: Sampai jumpa! Semoga harimu menyenangkan.")
```

Halo! Saya adalah bot yang siap membantu Anda dengan pertanyaan seputar kesehatan mental. Tanyakan sesuatu atau ketik 'exit' untuk keluar.

Anda: Apa yang dimaksud dengan penyakit mental?

Bot: Penyakit mental adalah kondisi kesehatan yang mengganggu pikiran, emosi, hubungan, dan fungsi seseorang. Mereka dikaitkan dengan kesusahan dan berkurangnya kapasitas untuk terlibat dalam kegiatan kehidupan sehari-hari yang biasa.

Penyakit mental jatuh di sepanjang rangkaian keparahan: beberapa cukup ringan dan hanya mengganggu beberapa aspek kehidupan, seperti fobia tertentu. Di ujung lain spektrum terletak penyakit mental yang serius, yang mengakibatkan gangguan fungsional utama dan gangguan dengan kehidupan sehari-hari. Ini termasuk gangguan seperti depresi berat, skizofrenia, dan gangguan bipolar, dan mungkin mengharuskan orang tersebut menerima perawatan di rumah sakit.

Penting untuk mengetahui bahwa penyakit mental adalah kondisi medis yang tidak ada hubungannya dengan karakter, kecerdasan, atau kemauan seseorang. Sama seperti diabetes adalah kelainan pankreas, penyakit mental adalah kondisi medis karena biologi otak.

Demikian pula dengan bagaimana seseorang akan mengobati diabetes dengan

Conclusion

Keseluruhan kode diatas merupakan perancangan chatbot sederhana menggunakan dataset mental health menggunakan teknik vektorisasi TF-IDF dan Cosine similarity untuk memberikan response terhadap pertanyaan yang diberikan. Preprocessing data tidak dilakukan secara mendalam karena data yang dianalisis sudah cukup bersih dari awal.

Pengujian dilakukan dengan 10 pertanyaan berikut:

1. Apa yang dimaksud dengan penyakit mental?
2. Siapa yang terpengaruh oleh penyakit mental?
3. Apa penyebab penyakit mental?
4. Apa sajakah tanda-tanda peringatan penyakit mental?
5. Apakah penderita penyakit jiwa bisa sembuh?
6. dimana saya bisa mencari terapi
7. gimana saya bisa bertemu dengan psikolog
8. perbedaan antara psikiater dan psikolog
9. apa yang dimaksud gangguan skizoid
10. obat untuk masalah kesehatan mental

Pertanyaan nomor 1-5 merupakan pertanyaan yang memiliki kesamaan 100% dengan data yang ada pada dataset. Sedangkan pertanyaan 6-10 merupakan pertanyaan yang