

ANALISIS SIMPTOM KEMARAHAN MELAMPAU DALAM ISU KESIHATAN MENTAL MENGGUNAKAN TEKNIK PEMODELAN TOPIK

NUR FARZANA SYUHADAH BINTI KASHFULLAH

DR. LAILATUL QADRI BINTI ZAKARIA

Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia

ABSTRAK

Kemarahan adalah emosi biasa yang ada dalam diri seseorang. Perasaan marah itu sangat bagus untuk mengelakkan kita daripada dipijak dan menghadapi tekanan kehidupan seharian. Namun begitu, setiap individu akan bertindak balas terhadap perasaan marah yang berbeza dalam situasi yang sama. Sebagai contoh, seseorang mungkin berasa sangat marah sehingga sukar untuk menangani sesuatu situasi, manakala yang lain mungkin dapat mengendalikan perasaan marah dengan baik. Pemodelan topik ialah teknik yang sering digunakan untuk mengenal pasti topik atau tema sesuatu dokumen. Projek ini akan menganalisis masalah kemarahan yang melampau dengan menggunakan teknik pemodelan topik. Kajian ini bertujuan untuk mengenal pasti topik yang biasa dikaitkan dengan kemarahan melampau dan melakukan kategori bagi setiap topik yang dikeluarkan. Kesedaran tentang faktor faktor yang mencetuskan kemarahan boleh membantu seseorang memahami dirinya dengan lebih baik, membangunkan kesedaran diri dan mengurus emosi dengan lebih baik. Seseorang yang boleh mengelakkan konflik atau situasi yang boleh mencetuskan emosi ke arah negatif. Dengan mengetahui faktor-faktor yang membawa kepada masalah kemarahan yang melampau, seseorang boleh membina hubungan yang lebih baik dan meningkatkan kesejahteraan mental dan fizikal.

Kata kunci: kemarahan, teknik pemodelan topik, faktor kemarahan

PENGENALAN

Kesihatan mental telah wujud sejak zaman dahulu lagi, namun pada zaman moden ini telah berlaku banyak pendedahan, yang menyebabkan ramai tertumpu kepada perkara ini. Gaya hidup yang semakin berubah dengan peredaran zaman juga merupakan salah satu faktor yang membawa kepada masalah kesihatan mental. Apabila berbicara tentang emosi, ia merupakan punca utama dalam kesihatan mental. Emosi, satu subjektif yang melibatkan perasaan tertentu dalam apa jua keadaan. Ia melingkupi tindak balas mental dan fizikal terhadap sesuatu situasi atau peristiwa tertentu. Komponen utama emosi adalah tindakan dan perilaku di mana emosi sering kali mengarah kepada tindakan atau perilaku tertentu seperti rasa takut yang menarik picu respons lawan atau lari, rasa gembira yang memberi motivasi

positif, rasa marah yang mengundang kepada jeritan atau tengkingan, rasa sedih yang membawa kepada kesihatan mental. Kesihatan mental mengandungi masalah tertekan, kesedihan yang berpanjangan dan perubahan emosi secara kerap, kehilangan minat dalam perkara yang disukai dalam tempoh masa yang lama.

Bagaimana terbitnya masalah ini? Masalah kesihatan mental terdiri daripada pelbagai faktor, dan salah satunya adalah masalah kemarahan melampau. Masalah kemarahan melampau atau dikenali sebagai 'anger issues' adalah satu masalah di mana seseorang itu hilang pertimbangan dalam mengawal emosi marah, yang boleh menjadi penyebab kepada depresi dan masalah mental yang lain. Badan seseorang individu tersebut akan berubah kepada peningkatan denyut jantung, tekanan darah serta tingkatan adrenalin apabila berasa marah yang tak terkata. Emosi negatif yang terbit daripada masalah kemarahan ini adalah seseorang itu berada dalam keadaan sedih, takut, rasa dikhianati, dan lain - lain lagi. Kemarahan yang melampau mempunyai impak yang signifikan dari pelbagai masalah kesihatan mental seperti depresi, tekanan, bipolar atau PTSD. Faktor - faktor yang mempengaruhi masalah kemarahan dalam simptom kesihatan mental dapat dikenal-pasti dalam diri seseorang individu.

Terdapat banyak kajian yang telah dijalankan oleh penyelidik yang terdahulu berkaitan dengan emosi tapi masih kurang kajian dijalankan berkaitan dengan isu-isu yang berkaitan dengan emosi terutamanya emosi marah. Oleh itu, projek ini akan menganalisis topik-topik yang berkaitan simptom kemarahan melampau untuk mengesan simptom ini daripada awal. Emosi marah yang mempunyai banyak faktor akan dikenal pasti dalam teks menggunakan kaedah pemodelan topik. Di dalam dataset emosi yang sedia ada, emosi marah dan penerangannya akan dikeluarkan dan daripada situ, analisis akan dilakukan daripada penerangan yang diberikan.

Untuk mengatasi masalah seperti yang telah dibincangkan, projek ini akan menghasilkan satu model pengesan kategori yang menyebabkan kemarahan yang membolehkan para pengguna untuk membantu mengenalpasti simptom-simptom marah. Model pengesan penyebab kemarahan akan menganalisis dan mengekstrak simptom-simptom kemarahan daripada set data yang berkaitan dengan kesihatan mental. Model ini diadakan bertujuan untuk mengenal pasti simptom yang berkaitan dengan marah dan memberikan maklumat berkaitan penyebab kemarahan. Selain itu, projek ini akan menambah baik dalam membahagikan kategori penyebab kemarahan yang akan digunakan dalam model. Dengan adanya model yang dapat menganalisis dan mengekstrak simptom-simptom kemarahan dan kategori penyebab kemarahan, individu boleh menambah maklumat dan pengetahuan mereka berkenaan dengan simptom-simptom kemarahan melampau yang membawa kepada masalah kesihatan mental.

Usulan yang dilakukan adalah untuk mengenal pasti faktor yang berkaitan emosi marah dalam teks dengan menggunakan teknik pemodelan topik. Dengan menggunakan teknik pemodelan topik seperti Latent Dirichlet Allocation (LDA) dan juga BERTopic, topik-topik yang berkaitan dengan emosi marah dapat dikenal pasti dari kumpulan teks yang

dianalisis. Ini membolehkan penyelidik untuk memahami aspek-aspek tertentu yang mempengaruhi kemarahan. Pembangunan aplikasi sistem pengesanan kategori emosi marah untuk membantu pengguna memahami isu marah melampau juga dilakukan.

METODOLOGI KAJIAN

Metodologi yang digunakan dalam pembangunan projek ini ialah CRISP-DM (*Cross-Industry Standard Process for Data Mining*), satu metodologi yang biasa digunakan dalam analisis data dan data mining. Metodologi ini adalah amat sesuai dengan projek ini yang memerlukan pendekatan pembangunan sistem BERTopic yang lebih fleksibel dan berorientasikan kepada pengguna. Dengan menggunakan metodologi CRISP-DM untuk projek ini, produk akhir yang berkualiti boleh dihasilkan.

Fasa pemahaman bisnes

Fasa ini mengenal pasti tujuan analisis ini dilakukan, untuk mengesan masalah kemarahan dalam diri seseorang individu. Fasa ini juga dijalankan untuk memastikan kategori penyebab kemarahan yang dibangunkan mencapai objektif yang telah ditetapkan.

Fasa pemahaman data

Fasa pemahaman data adalah fasa mengenal pasti sumber data yang digunakan. Sorotan susastera juga dijalankan ke arah kemarahan melampau untuk meningkatkan pemahaman mengenai faktor-faktor kemarahan.

B	C
joy	On days when I feel close to my partner and other friends.
fear	Every time I imagine that someone I love or I could contact a
anger	When I had been obviously unjustly treated and had no possibility
sadness	When I think about the short time that we live and relate it to
disgust	At a gathering I found myself involuntarily sitting next to two
shame	When I realized that I was directing the feelings of discontent
guilt	I feel guilty when when I realize that I consider material things
joy	After my girlfriend had taken her exam we went to her parent's
fear	When, for the first time I realized the meaning of death.
anger	When a car is overtaking another and I am forced to drive off the
sadness	When I recently thought about the hard work it takes to study, and
disgust	When I found a bristle in the liver paste tube.
shame	When I was tired and unmotivated, I shouted at my girlfriend and
guilt	When I think that I do not study enough. After the weekend I
joy	When I pass an examination which I did not think I did well.
fear	When one has arranged to meet someone and that person arrives
anger	When one is unjustly accused of something one has not done.
sadness	When one's studies seem hopelessly difficult and uninteresting.
disgust	When one finds out that someone you know is not at all like one
shame	When one has been unjust, stupid towards someone else.
guilt	When one has neglected or been unjust to a good friend.
joy	Passing an exam I did not expect to pass.
fear	When I climbed on a tree to pick apples. The angle of the ladder

Rajah 1 Dataset

Dataset yang digunakan mempunyai 7514 data emosi. Di dalam dataset tersebut, terdapat beberapa emosi iaitu joy (gembira), fear (takut), anger (marah), sadness (sedih), disgust (jijik) dan shame (malu). Dataset yang sedia ada mempunyai pelbagai penerangan dan juga emosi yang dikategorikan mengikut penerangan individu. Jadi, daripada dataset ini akan mengesan penerangan yang membabitkan emosi marah dan analisis kemarahan yang melampau dalam isu kesihatan mental menggunakan pemodelan topik.

Fasa pra-pemrosesan data

Fasa ini melakukan pembersihan data dan melakukan transformasi data untuk mendapatkan data yang bersih.

```

      CLEAN_DESCRIPTION
2  obviously unjustly treated possibility elucidate...
9  car overtaking another forced drive road
16 one unjustly accused something one done
26 partner attacked lost three teeth
33 unjust accusation directed way acting someone ...
...
7483 someone spread rumour
7498 saw bed hostel mess guessed someone else used ...
7497 surname brother different often grinned home p...
7504 september 1984 forced live someone like first ...
7506 roommate liked listen meaningless song melody ...

[1079 rows x 1 columns]
<ipython-input-9-046c8f4643be>:28: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html
anger_df['CLEAN_DESCRIPTION'] = anger_df['DESCRIPTION'].apply(preprocess_text)

```

Rajah 2 Pengeluaran pra-pemrosesan data

Pra-pemrosesan yang dilakukan adalah seperti berikut:

- i) Penyingkiran kata henti - menyingkirkan kata-kata umum yang tidak bermakna dan juga penambahan kata-kata umum untuk memastikan data lebih bersih.
- ii) Tokenisasi - memecahkan teks ke dalam kata-kata individu.
- iii) Tukar kepada huruf kecil - menyeragamkan teks supaya huruf besar dan huruf kecil dianggap sama dalam analisis teks untuk mengurangkan bunyi di dalam data.
- iv) Lematisasi - mengurangkan kata-kata ke bentuk dasar dan mengurangkan variasi perkataan yang sama.

Kajian ini akan menggunakan proses pra-pemrosesan yang sama bagi kedua-dua algoritma pemodelan topik LDA dan juga BERTopic. Hal ini untuk memastikan pengeluaran untuk LDA dan juga BERTopic menggunakan data bersih yang sama dan seragam kerana untuk mendapatkan ketekalan dalam keputusan yang dihasilkan oleh kedua-dua model.

Fasa pemodelan

Fasa ini adalah yang paling lama kerana ia melibatkan semua kerja berat seperti pengaturcaraan pemodelan topik dan sebagainya. Dalam fasa ini, semua keperluan yang dikumpulkan dalam fasa analisis dibangunkan dan diusahakan. Keperluan ini pula akan digabungkan untuk mencapai target kategori penyebab dalam kemarahan. Melalui fasa ini juga, kelemahan kategori itu dapat ditentukan selepas pelaksanaan berjaya dilakukan.

```

from sklearn.feature_extraction.text import CountVectorizer
from sklearn.decomposition import LatentDirichletAllocation

# tokenize the cleaned descriptions
tokenized_descriptions = cleaned_df['CLEAN_DESCRIPTION'].apply(word_tokenize)

# Join tokens back into a string for tokens in tokenized_descriptions
cleaned_text = [' '.join(tokens) for tokens in tokenized_descriptions]

# Create a CountVectorizer to create a bag-of-words representation
vectorizer = CountVectorizer(max_df=0.95, min_df=2, stop_words='english')
bow_matrix = vectorizer.fit_transform(cleaned_text)

# Initialize and fit LDA model
num_topics = 10 # Adjust the number of topics as per your requirement
lda_model = LatentDirichletAllocation(n_components=num_topics, random_state=42)
lda_model.fit(bow_matrix)

# Print the topics
def print_top_words(model, feature_names, n_top_words):
    for topic_idx, topic in enumerate(model.components_):
        message = f"Topic #{topic_idx + 1}:"
        message += " ".join([feature_names[i] for i in topic.argsort()[::-1-n_top_words : -1]])
        print(message)
    print()

n_top_words = 11 # Adjust the number of top words per topic
print_top_words(lda_model, vectorizer.get_feature_names_out(), n_top_words)

def categorize_lda_topics(lda_model, feature_names):
    topic_categories = {}

    # Define subcategories of anger
    anger_categories = {
        'friendship': ['friend', 'trust', 'betrayal', 'disloyal', 'loyal', 'confidant'],
        'relationship': ['partner', 'love', 'cheat', 'breakup', 'romantic', 'family'],
        'family': ['mother', 'father', 'sibling', 'home', 'children', 'marriage'],
        'work': ['work', 'stress', 'frustration', 'challenge', 'exam', 'colleague']
    }

    # Define subcategories of friendship
    friendship_subcategories = {
        'conflict': ['discussion', 'wrong', 'argue', 'confrontation'],
        'betrayal': ['betrayal', 'disloyal', 'backstabbing', 'untrustworthy', 'broken promise'],
        'fight': ['angry', 'anger', 'treated', 'insulted', 'situation']
    }

    # Iterate through each topic
    for topic_idx, topic in enumerate(lda_model.components_):
        top_words = [feature_names[i] for i in topic.argsort()[::-1-n_top_words : -1]] # Get top 10 words per topic

        # Determine the category based on keyword matches
        for category, keywords in anger_categories.items():
            if any(word in top_words for word in keywords):
                if category == 'friendship':
                    for subcategory, sub_keywords in friendship_subcategories.items():
                        if any(word in top_words for word in sub_keywords):
                            topic_categories[topic_idx] = f"Friendship ({subcategory})"
                            break
                else:
                    topic_categories[topic_idx] = "Friendship (other)"
            else:
                topic_categories[topic_idx] = category
            break

    else:
        topic_categories[topic_idx] = 'other' # Assign as 'other' if no clear match

    return topic_categories

# Get feature names from the CountVectorizer
feature_names = vectorizer.get_feature_names_out()

# Categorize LDA topics
topic_categories = categorize_lda_topics(lda_model, feature_names)

# Print categorized topics
print("Categorized Topics in LDA model:")
for topic_idx, category in topic_categories.items():
    print(f"Topic #{topic_idx + 1}: Category - {category}")

```

Rajah 3 Program aturcara pembahagian topik kategori dalam LDA

```

import numpy as np
import matplotlib.pyplot as plt
from sklearn.manifold import MDS
from sklearn.metrics.pairwise import euclidean_distances

# Create a matrix of topic-word distributions
topic_word_distributions = lda_model.components_

# Compute the pairwise euclidean distance between topics
distances = euclidean_distances(topic_word_distributions)

# Apply MDS to reduce dimensionality
mds = MDS(n_components=2, dissimilarity="precomputed", random_state=1)
topic_coords = mds.fit_transform(distances)

# Plotting the topics
plt.figure(figsize=(12, 8))
plt.scatter(topic_coords[:, 0], topic_coords[:, 1], c='blue', marker='o')

# Annotating topics with their category
for i, category in topic_categories.items():
    plt.annotate(category, (topic_coords[i, 0], topic_coords[i, 1]), textcoords="offset points", xytext=(0, 30), ha="center")

plt.title("Intertopic Distance Map via MDS")
plt.xlabel("MDS dimension 1")
plt.ylabel("MDS dimension 2")
plt.grid(True)
plt.show()

```

Rajah 4 Program aturcara Peta Jarak Intertopik melalui Scaling Berbilang Dimensi

Untuk pembangunan algoritma LDA, kategori utama adalah persahabatan, perhubungan, keluarga, kerja. Namun, kategori persahabatan adalah terlalu umum, jadi kategori persahabatan pecah menjadi subkategori selanjutnya seperti konflik, pengkhianatan dan pergaduhan. Fungsi ini berulang melalui setiap topik, mengekstrak perkataan teratas dan memadankannya dengan kata kunci dalam kategori dan subkategori yang ditentukan. Rajah 4 menunjukkan algoritma peta jarak intertopik ini memetakan topik-topik ke dalam ruang dua dimensi berdasarkan jarak atau kesamaan mereka, sehingga memudahkan untuk melihat bagaimana topik-topik tersebut terkait satu sama lain.

```

# Assign categories based on the cause of anger
def assign_category(description):
    if 'mother' in description.lower() or 'father' in description.lower():
        return 'family'
    elif 'boyfriend' in description.lower() or 'girlfriend' in description.lower():
        return 'relationship'
    elif 'angry' in description.lower() or 'argument' in description.lower():
        return 'friendship (Conflict)'
    elif 'fight' in description.lower():
        return 'friendship (Fight)'
    elif 'betrayed' in description.lower():
        return 'friendship (Betrayal)'
    elif 'team' in description.lower() or 'work' in description.lower():
        return 'work'
    else:
        return 'other'

# Fit BERTopic model
topics, _ = bertopic_model.fit_transform(anger_df['CLEAN_DESCRIPTION'])

# Add categories to topics
anger_df['TOPIC'] = topics
anger_df['CATEGORY'] = anger_df['CLEAN_DESCRIPTION'].apply(assign_category)

# Print the topics with categories
topics_with_categories = anger_df[['TOPIC', 'CATEGORY']].drop_duplicates()
print(topics_with_categories)

# Get topic information
topic_info = bertopic_model.get_topic_info()

```

Rajah 5 Program aturcara pembahagian topik kategori dalam BERTopic

```

from bertopic import BERTopic

topic_model = BERTopic()
topics, _ = bertopic_model.fit_transform(anger_df['CLEAN_DESCRIPTION'])
bertopic_model.visualize_topics()

```

Rajah 6 Program aturcara Peta Jarak Intertopik

Sama seperti LDA, BERTopic juga akan mengeluarkan kategori yang terlibat untuk setiap topik. Dalam BERTopic, peta jarak intertopik adalah visualisasi yang menunjukkan hubungan antara topik-topik yang dihasilkan oleh model.

Fasa penilaian

Fasa penilaian merupakan salah satu fasa yang memfokuskan kepada penilaian model teknikal, dimana model yang lebih sesuai dengan projek yang dilakukan. Di dalam projek ini membangunkan dua pemodelan topik iaitu LDA dan juga BERTopic. Jadi di fasa ini adalah untuk mengenal pasti model manakah yang memenuhi kriteria untuk projek.

<pre> Topics in LDA model: Topic #1: team match football angry time leader class wanted went university played lost Topic #2: angry friend got come time thing came boyfriend father said class promised Topic #3: mother angry home friend parent sister time felt stole father year quarrel Topic #4: brother angry mother bus money mistake thing saw went child student south Topic #5: friend angry girl asked school boyfriend thing saw close work car lost Topic #6: friend angry exam time got felt called like brother money tried talking Topic #7: friend car angry accused father girl school felt year away teacher time Topic #8: anger time let felt wanted asked feel girlfriend refused thing father colleague Topic #9: friend room boy girlfriend mother started certain late anger way good felt </pre>	<pre> Categorized Topics in LDA model: Topic #1: Category - other Topic #2: Category - friendship (fight) Topic #3: Category - friendship (fight) Topic #4: Category - family Topic #5: Category - friendship (fight) Topic #6: Category - friendship (fight) Topic #7: Category - friendship (fight) Topic #8: Category - family Topic #9: Category - friendship (fight) </pre>
--	--

Rajah 7 Pengeluaran topik dan kategori topik LDA



Rajah 8 Visualisasi Peta Jarak Intertopik melalui Scaling Berbilang Dimensi sembilan topik

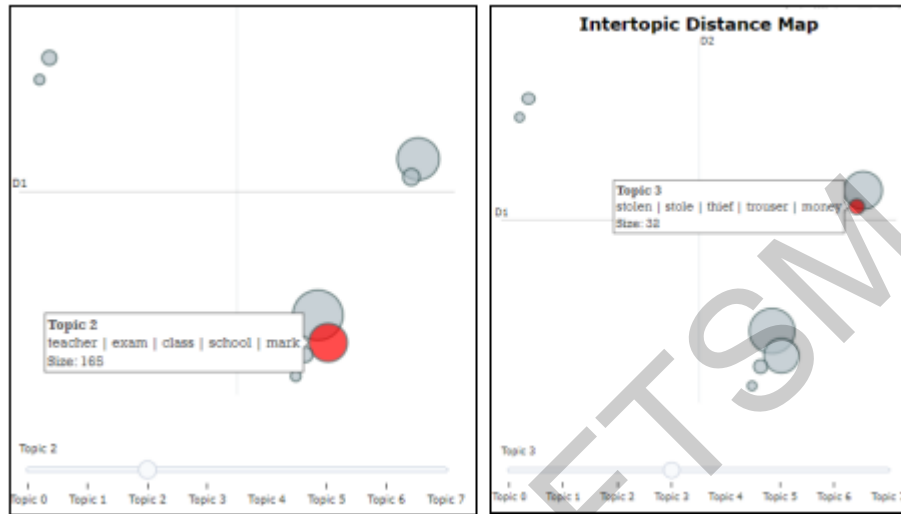
Topik #1 mempunyai perkataan atas seperti team, match dan juga football. Topik #1 juga berada dalam kategori lain-lain iaitu ianya mungkin tiada kaitan dengan senarai kategori yang ada. Dalam Topik #3 mempunyai perkataan atas seperti angry, mother, home namun berada dalam kategori persahabatan dan bukannya kategori keluarga walaupun perkataan atas topik tersebut berkenaan dengan kekeluargaan. Variasi dalam had limit topik ini boleh mempengaruhi model pengeluaran dan juga visualisasi. Titik-titik tersebut mempunyai salah satu kiraan MDS Dimension 1 atau 2 yang hampir sama membuatkan titik-titik tersebut seiringan dengan titik yang lain.



Rajah 9 Pengeluaran topik dan kata atas topik bagi BERTopic

Setiap topik dicirikan oleh satu set perkataan yang mewakili dokumen yang ada. Jika dilihat pengeluaran kepada Topik 2 di index 3, ([teacher, exam, class, mark, school, ...], kategori kemarahan untuk Topik 1 adalah disebabkan oleh kerja. Namun, apabila dilihat kepada kategori yang disenaraikan kepada semua Topik 2, terdapat kategori keluarga, kerja dan juga lain-lain, sama seperti pengeluaran tiada had diatas. Faktor keduanya adalah liputan kata

kunci yang terhad. Pendekatan pengkategorian berasaskan kata kunci mungkin tidak merangkumi semua kemungkinan variasi atau nuansa punca kemarahan yang terdapat dalam huraian, yang membawa kepada pengkategorian yang tidak konsisten.



Rajah 10 Visualisasi Peta Jarak Intertopik

Rajah 10 menunjukkan buih Topik 2 dan Topik 3 berjauhan antara satu sama lain. Ini menunjukkan kedua-dua topik mempunyai perbezaan perkataan yang ketara. Pengeluaran perkataan untuk Topik 2 adalah *teacher*, *exam*, *class*, *school* dan juga *mark* manakala untuk Topik 3 adalah *stolen*, *stole*, *thief*, *trouser* dan juga *money*. Topik 0, Topik 2, Topik 4 dan juga Topik 6 berdekatan antara satu sama lain tetapi berjauhan dengan Topik 1 dan juga Topik 3. Topik 5 dan juga Topik 7 juga berjauhan dengan semua topik memandangkan saiz buih bagi Topik 5 dan juga Topik 7 kecil berbanding dengan semua topik.

LDA menghasilkan topik dengan perkataan atas yang kadang-kadang bertindan antara topik yang berbeza, yang boleh mengurangkan kejelasan beberapa topik, manakal BERTopic menghasilkan perkataan atas yang lebih unik untuk setiap topik, mengurangkan risiko perkataan bertindan dan meningkatkan keseragaman.

Fasa implementasi / pengujian

Fasa pengujian merupakan salah satu fasa yang penting dalam pembangunan sistem BERTopic. Hal ini kerana pengujian dijalankan untuk mencari kecacatan, ralat dan kelemahan dalam sistem ini. Apabila ralat dapat dikenalpasti, pembaikan dan pembedahan dapat dilakukan untuk memastikan kategori penyebab kemarahan yang lancar dan memastikan pengalaman yang terbaik bagi pengguna. Sistem beroperasi daripada pembangunan algoritma di *Visual Studio Code*, lalu diintegrasikan ke *Streamlit* menggunakan *GitHub*.


```

app.py 1 x
C:\Users\User\Desktop> app.py ...
1 import streamlit as st
2 import numpy as np
3 import pandas as pd
4 from bertopic import BERTopic
5 import joblib
6
7 # Load your data
8 @st.cache_data
9 def load_data(uploaded_file):
10     if uploaded_file is not None:
11         data = pd.read_csv(uploaded_file)
12         return data
13     return None
14
15 # Perform BERTopic modeling
16 def create_model(texts, num_topics, min_topic_size, nr_topics, n_gram_range=(1,3)):
17     model = BERTopic(n_gram_range=n_gram_range, calculate_probabilities=True, nr_topics=num_topics, min_topic_size=min_topic_size)
18     topics, probs = model.fit_transform(texts)
19
20     topic_info = model.get_topic_info()
21
22     st.write(topic_info.head(nr_topics)) # Display top topics based on nr_topics
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45 # Function to assign categories based on the cause of anger
46 def assign_category(description):
47     description = description.lower()
48     if 'mother' in description or 'father' in description or 'sister' in description or 'brother' in description:
49         return 'Family'
50     elif 'boyfriend' in description or 'girlfriend' in description or 'couple' in description:
51         return 'Relationship'
52     elif 'angry' in description or 'argument' in description or 'angry' in description:
53         return 'Friendship (Conflict)'
54     elif 'fight' in description or 'punch' in description or 'friend' in description or 'hurt' in description:
55         return 'Friendship (Fight)'
56     elif 'betrayed' in description or 'anger' in description:
57         return 'Friendship (Betrayal)'
58     elif 'exam' in description or 'work' in description:
59         return 'Work'
60     else:
61         return 'Other'
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90

```

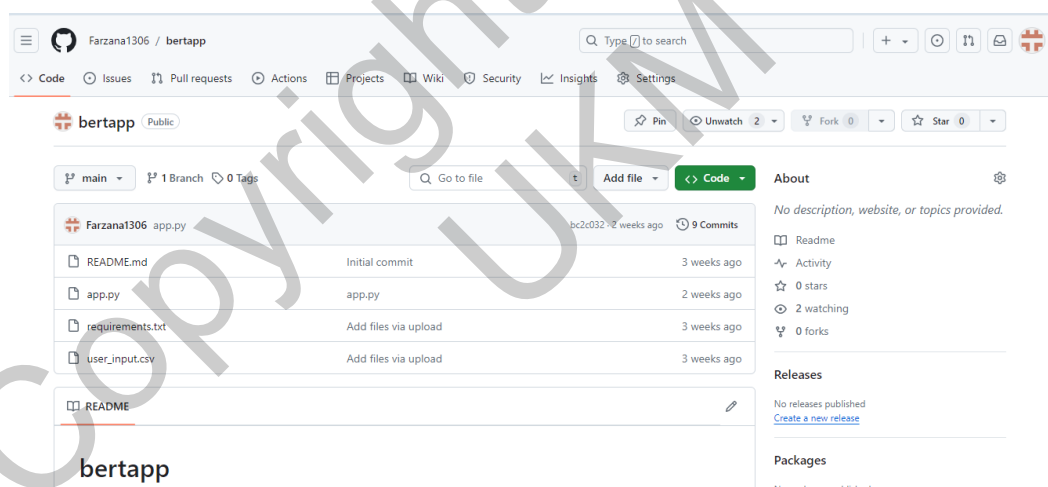
```

92 # Assign categories based on the cause of anger
93 data['category'] = data['text'].apply(assign_category)
94
95 topics, topic_info = generate_topics(texts, num_topics, min_topic_size, nr_topics)
96 data['topic'] = topics
97
98 # Calculate predominant category for each topic
99 predominant_categories = data.groupby('topic')['category'].agg(lambda x: x.value_counts().idxmax())
100 topic_info['category'] = topic_info['Topic'].map(predominant_categories)
101
102 # Merge the topic information with the main dataframe
103 topic_info = topic_info.rename(columns={"Name": "topic_name"})
104 data_with_info = data.merge(topic_info[['Topic', 'topic_name', 'category']], left_on='topic', right_on='Topic', how='left')
105
106 # Drop the redundant 'Topic' column
107 data_with_info = data_with_info.drop(columns=['Topic'])
108
109 st.write("Data with topics and categories:")
110 st.write(data_with_info)
111
112 st.write("Topic Info with Categories:")
113 st.write(topic_info)

```

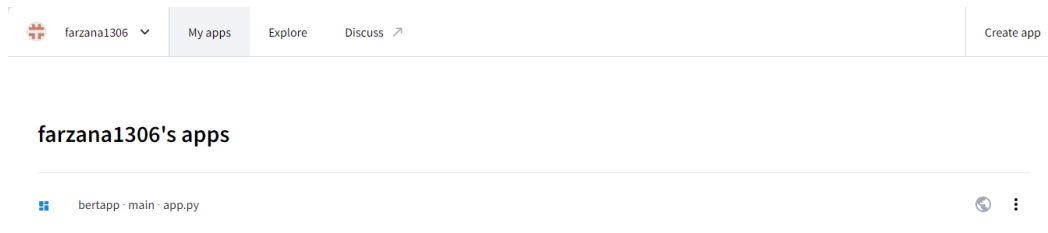
Rajah 11 Pembangunan algoritma sistem BERTopic di *Visual Studio Code*

Untuk membangunkan sistem BERTopic, pembangunan algoritma dilakukan melalui Visual Studio Code. Rajah 11 menunjukkan pembangunan algoritma dilakukan menggunakan Streamlit untuk antara muka. Pembangunan algoritma untuk menentukan perkataan atas, topik, kategori adalah hampir sama dengan pembangunan algoritma di *Google Colab*. Streamlit IU menyediakan kawasan teks untuk pengguna memasukkan sampel data yang diberikan. Selepas data dimasukkan, ia akan memproses teks dan menentukan kategori topik menggunakan model BERTopic. Keputusan, termasuk data dengan topik dan kategori, visualisasi data serta maklumat topik akan dipaparkan dalam aplikasi.



Rajah 12 Implementasi algoritma sistem BERTopic di GitHub

Rajah 12 menunjukkan laman web repositori di GitHub, di mana pembangunan algoritma di Visual Studio Code akan diimplementasikan di sini. Fail *requirements.txt* adalah fail yang mengandungi perpustakaan dan fungsi yang digunakan. Fail ini penting untuk memastikan algoritma tidak mempunyai sebarang ralat.



Rajah 13 Sistem beroperasi Streamlit

Melalui GitHub, sistem BERTopic akan dicipta menggunakan laman web Streamlit. Streamlit akan memanggil repositori daripada GitHub untuk mencipta sistem BERTopic.

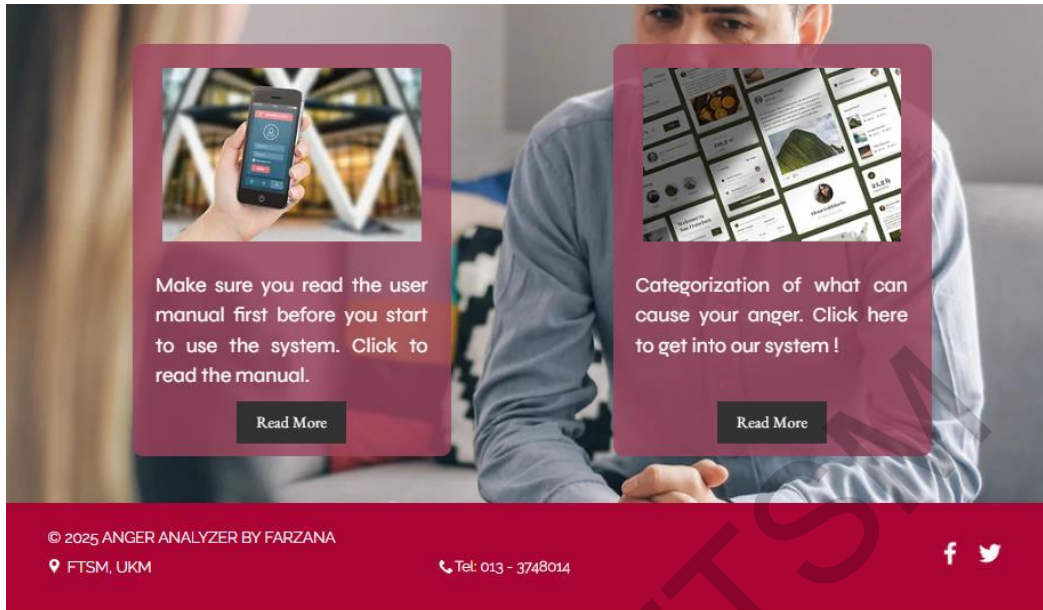
KEPUTUSAN DAN PERBINCANGAN

Analisis Simptom Kemarahan Melampau Dalam Isu Kesihatan Mental Menggunakan Teknik Pemodelan Topik telah berjaya dibangunkan dan semua dokumentasinya telah dilengkapi. Semasa proses pembangunan, sistem kategori penyebab kemarahan dibangunkan menggunakan Streamlit dengan bahasa pengaturcaraannya yang dipanggil Visual Studio Code. Sistem ini akan dipanggil melalui laman web utama Anger Analyzer.



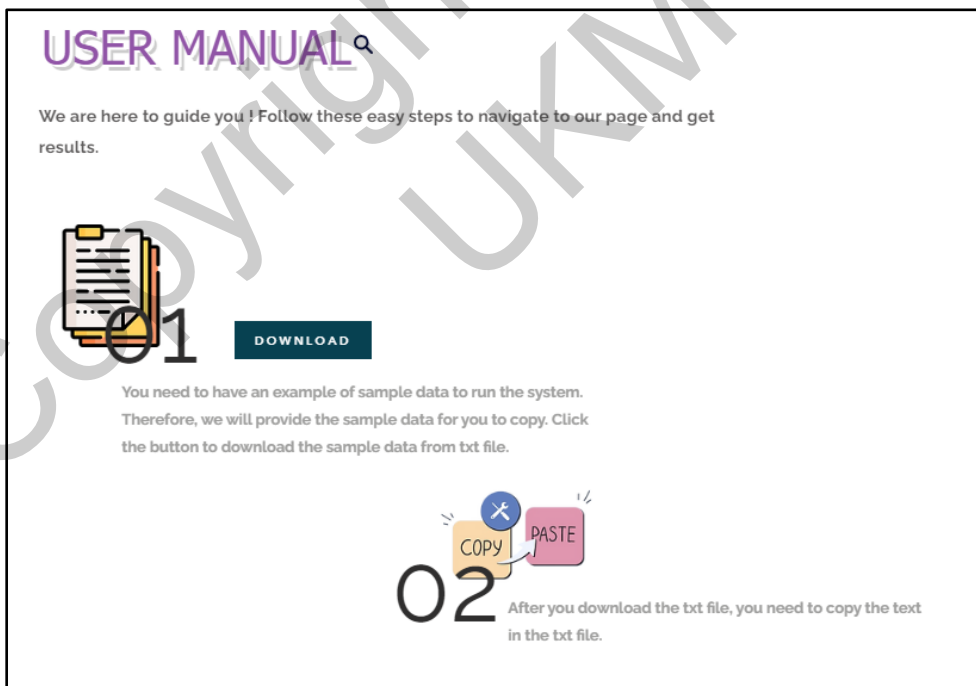
Rajah 1 Antara Muka Deskripsi

Apabila memasuki laman web Anger Analyzer, pengguna akan disambut dengan skrin utama dimana terdapat deskripsi mengenai teknik BERTopic dan juga kemarahan.

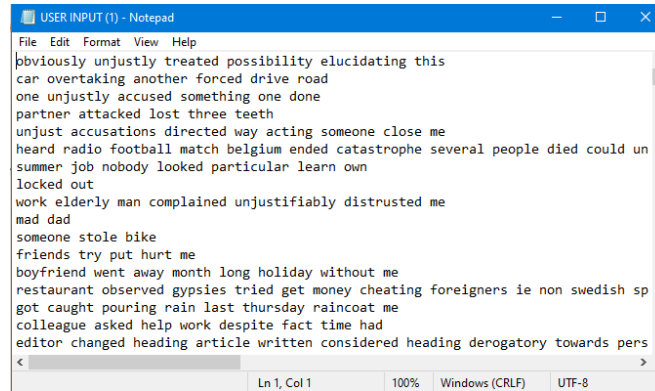


Rajah 2 Antara Muka Pemilihan Muka Surat

Apabila pengguna melihat kepada bahagian bawah, terdapat arahan untuk pengguna menggunakan sistem. Kepada pengguna yang tidak mempunyai kemahiran dalam bidang seperti ini, pengguna boleh menggunakan manual pengguna untuk mengetahui cara untuk menggunakan sistem. Semua pengguna haruslah membaca manual pengguna dahulu sebelum menggunakan sistem.



Rajah 3 Antara Muka Laman Manual Pengguna



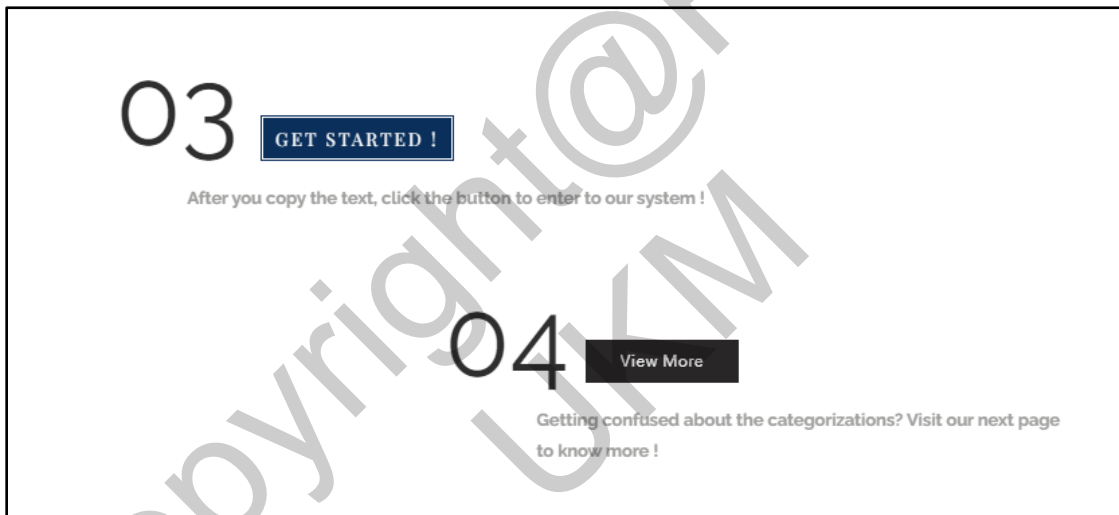
```

File Edit Format View Help
bbviously unjustly treated possibility elucidating this
car overtaking another forced drive road
one unjustly accused something one done
partner attacked lost three teeth
unjust accusations directed way acting someone close me
heard radio football match belgium ended catastrophe several people died could un
summer job nobody looked particular learn own
locked out
work elderly man complained unjustifiably distrusted me
mad dad
someone stole bike
friends try put hurt me
boyfriend went away month long holiday without me
restaurant observed gypsies tried get money cheating foreigners ie non swedish sp
got caught pouring rain last thursday raincoat me
colleague asked help work despite fact time had
editor changed heading article written considered heading derogatory towards pers

```

Rajah 4 Sampel data penerangan emosi marah

Di dalam laman web manual pengguna, terdapat langkah-langkah yang harus diikuti sebelum memasuki sistem. Disebabkan sistem yang dibangunkan haruslah menggunakan dataset yang banyak, jadi projek ini akan menyediakan sampel data untuk pengguna ikuti. Jadi, pengguna boleh memuat naik data yang disediakan dan menyalin semua dataset yang diberikan untuk digunakan di dalam sistem. Rajah 4 menunjukkan sampel data yang diberikan.



Rajah 5 Antara Muka Laman Manual Pengguna

Selepas pengguna telah 'copy' dataset yang diberikan, pengguna boleh terus ke laman sistem projek ini. Langkah yang keempat adalah laman web penerangan mengenai kategori penyebab kemarahan untuk pengguna lebih memahami berkenaan kategori yang dilakukan.

WHAT CAN CAUSE ANGER ?



Anger can arise in various contexts, each with its unique triggers and manifestations. There are some categories of anger specifically in the contexts of family, friendship, relationships, and work. Understanding the context and triggers of anger is crucial for addressing and managing it effectively in different areas of life.

FAMILY ?



Family conflicts about roles and responsibilities, misunderstanding, and unfulfilled expectations are all common causes of anger. Family members may hold one another to certain standards for behaviour, assistance, or responsibilities, and when these standards are not reached, it can cause resentment and irritation. Misunderstandings or a closed discourse are examples of communication problems that might exacerbate rage. Tension can also arise from disagreements on parental duties, financial obligations, or housework tasks. Frequent disputes, emotional retreat, or, in severe situations, physical violence or outbursts are common signs of this kind of rage.

RELATIONSHIP ?



In romantic relationships, anger can be triggered by trust issues, poor communication, unmet expectations, and power struggles. Trust issues, such as suspicions of infidelity or actual infidelity, are common sources of anger. Poor communication, where partners frequently misunderstand each other or fail to express their needs and concerns effectively, can lead to frustration. Power struggles and controlling behaviors can also contribute to conflict. Anger in romantic relationships often results in heated arguments, long-term resentment, emotional withdrawal, or, in some cases, physical allegations.

FRIENDSHIP ?



Anger in friendships frequently results from miscommunication, envy, emotions of abandonment, and betrayal. Friendships are built on trust, and any betrayal—real or imagined—can cause profound sorrow and resentment. When one buddy is envious of another's connections, accomplishments, or attention, jealousy may result. Resentment can also be sparked by neglect, where a buddy feels neglected or underappreciated. Conflicts and rage can result from misunderstandings and presumptions that are not clearly communicated. Passive-aggressive behaviour, physical or emotional distance from the buddy, or outright confrontations or heated arguments are some ways in which this kind of rage can be exhibited.

WORK ?



Anger at work is frequently caused by a heavy workload, a lack of appreciation, disagreements with coworkers or superiors, and worries about job security. Workers may believe that duties are assigned unjustly or feel overburdened by an excessive workload. Anger and a sense of being underappreciated might result from not receiving credit or acknowledgment for their work. Emotional rivalry, personality conflicts, or disagreements with superiors can all lead to conflicts with coworkers or managers and a toxic work environment. Anger and tension can be increased by worries about one's job security or by actions that one feels are unjust. Anger can show itself as high stress levels, irritation, arguments with coworkers or management, a drop in motivation or job output, and an increase in tardiness or absenteeism.

Rajah 6 Antara Muka Laman Web Penerangan Kategori Penyebab Kemarahan

Rajah 6 menunjukkan penerangan yang dilakukan dalam muka surat kategori penyebab kemarahan. Penerangan ini dilakukan adalah untuk memberikan penjelasan yang lebih mendalam terhadap pengguna tentang faktor-faktor kemarahan melampau.

BERTOPIC TOPIC MODELING AND CATEGORIZATION TOPICS

Please provide BERTopic parameters as follows:

Note:

i USER INPUT : You can adjust the parameters such as the number of topics, minimum topic size, and number of top words to fine-tune the BERTopic model according to your specific data and analysis needs. Note that the default parameters are 9, 5 and 12.

Number of Topics

-
+

Minimum Topic Size

-
+

Number of Top Words

-
+

Rajah 7 Parameter BERTopic Streamlit

Berdasarkan rajah di atas, terdapat parameter yang boleh ditentukan untuk pengguna mengubah suai mengikut bilangan data dan analisis yang dimahukan. Jumlah topik adalah untuk menentukan berapa banyak topik yang akan dikenal pasti oleh model. Jumlah topik yang optimal dapat bervariasi tergantung pada kompleksiti data dan tujuan analisis. Pengguna boleh memilih jumlah topik yang dianggap paling sesuai untuk menggambarkan variasi topik dalam data mereka. Topik yang paling berkesan untuk projek ini adalah 9 atau 10 topik. Ukuran minimum topik menentukan jumlah minimum dokumen atau teks yang harus ada dalam suatu topik agar dianggap signifikan. Hal ini membantu dalam memfilter topik-topik kecil yang mungkin tidak relevan atau representatif secara statistik.

Dan jumlah kata kunci atasan merupakan jumlah kata kunci teratas yang akan ditampilkan untuk setiap topik. Kata-kata ini mewakili kata-kata yang paling muncul dalam topik tersebut, membantu dalam memahami setiap topik dengan lebih baik. Kata kunci atas yang paling ideal adalah 10 kata kunci atau 12 kata kunci. Dengan menyesuaikan parameter-parameter ini, pengguna dapat meningkatkan ketepatan dan relevansi model BERTopic terhadap data yang disediakan dalam projek ini, serta meningkatkan interpretabiliti hasil analisis topik yang dihasilkan.

Note:

i USER INPUT : Please copy all the text in the txt file that we have provided, and paste it in textbox below. The text should be more than 100 text otherwise it cannot proceed to process the topics.

Please input your text (one per line):

Input text

felt much anger perhaps feel till now girlfriends sister incited betray needed much think anything simply passed everything without employers knowledge even without communication bank accounts changed personal accounts closed exboyfriend tried prejudice fianct family saying things let family worried postpone wedding already using wedding rings honest feel anger felt hate brother went new shirt adored returned dirty grease grandmother highly authoritative person lives alone independent know live without commanding others hate this serious argument dear person discussion mother way educate children felt anger mother said absurd things false things filled ears inflated head seemed caught carcass heart wanted explode arm wanted knock world

Press Ctrl+Enter to apply

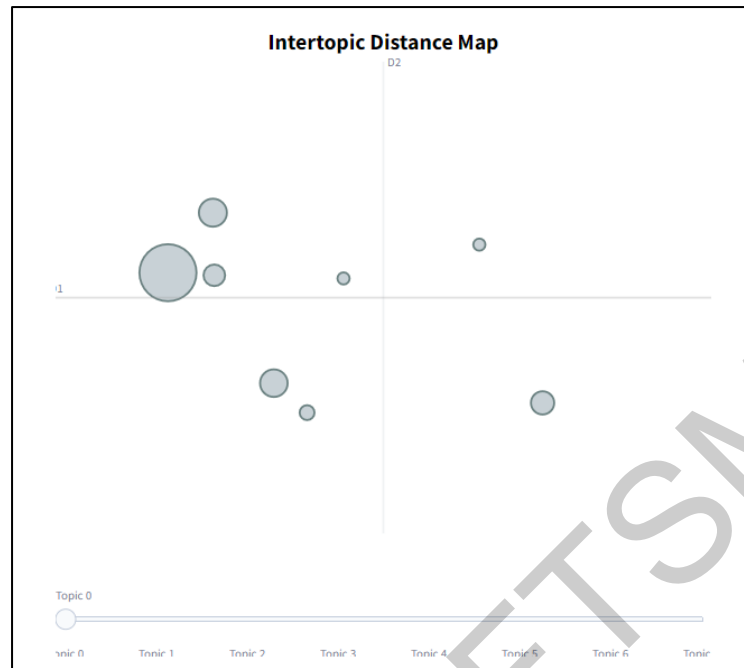
Analyze

Rajah 8 Papan Input Data

Rajah 8 menunjukkan input teks data yang dimasukkan, haruslah melebihi 100 deskripsi untuk menghasilkan analisa ini, jika tidak sistem tidak boleh membaca dan proses untuk mengeluarkan topik dan pola yang efektif. Hal ini kerana dataset yang kecil mempunyai risiko overfitting, di mana model tidak sesuai dengan detail-detail kecil atau anomali yang tidak mewakili umum dari data. Dalam analisis topik seperti yang dilakukan dengan BERTopic, bilangan dokumen atau teks yang mencukupi diperlukan untuk mendapatkan perwakilan statistik yang mencukupi bagi pelbagai topik yang terdapat dalam data. Jika jumlah teks terlalu kecil, topik yang terhasil mungkin kurang mewakili atau berbeza dengan ketara. Dengan mempunyai set data yang lebih besar, hasil analisis cenderung menjadi lebih stabil dan konsisten dari semasa ke semasa. Ini kerana variasi dalam data kurang mempengaruhi pengenalpastian dan tafsiran topik.

Topic	Count	Name	Representation
0	-1	-1_friend_me_angry_person	friend me angry person wanted gc
1	0	0_mother_brother_anger_felt	mother brother anger felt angry pe
2	1	1_teacher_school_examination_marks	teacher school examination marks m
3	2	2_promised_boyfriend_room_flat	promised boyfriend room flat friend
4	3	3_car_traffic_driving_towed_away	car traffic driving towed away towed
5	4	4_tennis_football_people_match	tennis football people match goal
6	5	5_16_roll_cash_less	16 roll cash less money anything
7	6	6_tv_news_saw_south_africa	tv news saw south africa hainburg
8	7	7_girl_girlfriend_me_work	girl girlfriend me work me angry prob

Rajah 9 Pengeluaran Topik



Rajah 10 Pengeluaran Topik Peta Jarak Intertopik

Topic Info with Categories:

		Representative_Docs	category
0	without r	friend called told wanted come see me lived village 50 km distance fom home town	Other
1	something	high school guys sometimes provoked friend mine called pig arunted irritated girls	Family
2	course	studied almost one week physicsexamination difficulty passed examn angry teache	Other
3	Jay com	friend shares room me opened window morning left room whole day came home ev	Relationship
4	nother fi	told boy friend park certain place car would get towed away listen towed away stag	Other
5	onship	heard radio football match belgium ended catastrophe several people died could ur	Other
6	t one	eaten anything 2 days able pay fine fianct stayed another town spent one thousand	Other
7	thi angr	tv news heard nicaragua complained increasing threat war usa saw report tv new	Other
8	d me gir	girlfriend guys like made fuss girl girlfriend understand me angry problems sepe	Relationship

Rajah 11 Pengeluaran Topik Beserta Kategori Penyebab Kemarahan

Berdasarkan pengeluaran di bawah Perwakilan (Representation), di situ kita boleh membahagikan kategori berdasarkan perkataan atas yang dikeluarkan. Walaupun perkataan atas di topik mempunyai kata kunci yang boleh dikategorikan sebagai keluarga, perhubungan, persahabatan ataupun kerja, namun model tidak dapat menangkap dan mengeluarkan topik kategori tersebut sebagai others. Hal ini adalah disebabkan kepada beberapa faktor. Yang pertama, model yang digunakan untuk mengelaskan dokumen mungkin tidak dilatih dengan cukup baik atau tidak mempunyai data yang mencukupi untuk membuat perbezaan yang jelas antara kategori.

Ini mungkin menyebabkan banyak dokumen dimasukkan ke dalam kategori lain-lain. Yang kedua adalah teks mungkin mempunyai konteks yang tidak jelas atau terlalu umum, menjadikannya sukar bagi model untuk mengelaskan topik dengan betul. Contohnya, jika teks tidak memberikan maklumat kontekstual yang mencukupi, model mungkin memilih kategori

lain-lain sebagai pilihan yang selamat.

Pengujian Kebolehgunaan

Pengujian kebolehgunaan ialah satu proses yang melibatkan pengujian akhir yang dilaksanakan oleh wakil pengguna untuk memastikan sistem yang dibangunkan mampu menyediakan fungsi yang diperlukan sebelum ia dikeluarkan kepada umum. Tujuan pengujian kebolehgunaan adalah untuk menilai kebolehgunaan sistem, mengumpul data kuantitatif, dan menilai kepuasan pengguna. Kaedah untuk mengumpul data atau mendapatkan kepuasan pengguna ialah melalui soal selidik terhadap 18 orang responden. Terdapat 10 item telah dibangunkan yang merupakan gabungan rubrik/kajian yang dijalankan oleh pakar lain melalui kertas penyelidikan mereka. Ia merangkumi soalan skala 1 sampai 5.

Soalan dimuat naik ke Google Form dan pautannya telah diberikan kepada pengguna. Sasaran responden bagi soal selidik ini dipilih berdasarkan pengguna remaja zaman sekarang yang berkemungkinan mempunyai masalah kemarahan. Selepas mengedarkan pautan soal selidik, seramai 18 orang responden telah menjawab soal selidik ini. Data yang diterima daripada penilaian kebolehgunaan dianalisis melalui kaedah analisis data yang bernama statistik deskriptif dengan menggunakan skala bagi setiap aspek. Jadual 1 menunjukkan Tafsiran Skala.

Jadual 1 Tafsiran Skala

Skala	Tafsiran
1.00	Sangat Tidak Setuju
2.00	Tidak Setuju
3.00	Neutral
4.00	Setuju
5.00	Sangat Setuju

Jadual 2 menunjukkan skala yang diterima daripada setiap item aspek Bahagian A (Laman Web Anger Analyzer) dan juga Bahagian B (Sistem BERTopic). Majoriti pengguna memberikan skala 4 dan 5, dan ebanyakkan item mencapai skala 5. Terdapat beberapa soalan yang mempunyai skala 3 namun hanya dalam satu atau dua orang responden yang memberikan skala 3 (Neutral).

Jadual 2 Majoriti Skala

No	Item	3	4	5
1	The website is easy to navigate and use.		7 (38.6%)	11 (61.6%)
2	It is easy to upload data from the Anger Analyzer homepage to the BERTopic Topic Modeling interface.	2 (11.1%)	10 (55.6%)	6 (33.3%)
3	The steps given to the 'BERTopic Topic Modeling' system are easy to understand and follow	1 (5.6%)	9 (50%)	8 (44.4%)
4	The brief information on the website makes it easier for me to understand anger.		6 (33.3%)	12 (66.7%)
5	I love looking at the design of this website and am satisfied using the website.		6 (33.3%)	12 (66.7%)

6	The topics generated in the category of causes of anger made me better understand the factors of anger.	1 (5.6%)	6 (33.3%)	11 (11.61%)
7	The given parameters work well and perfectly with the topic production.		11 (11.61%)	7 (38.9%)
8	I can analyze the causes of anger when using this system with the data examples provided.	1 (5.6%)	11 (11.61%)	6 (33.3%)
9	The production of a visual representation of the displayed data along with the categories works well and is in line with the specified parameters.		11 (11.61%)	7 (38.9%)
10	Overall, I am satisfied with this system.		6 (33.3%)	12 (66.7%)

Berdasarkan jawapan responden dan analisis yang dibuat, dapat disimpulkan bahawa kebolegunaan laman web Anger Analyzer dan sistem BERTopic ini adalah pada skala positif. Kesemua skala soalan yang diterima dianggap tinggi kerana melebihi 10 responden yang memberikan majoriti skala 4 dan skala 5 di setiap item. Dapat juga disimpulkan bahawa objektif penilaian projek ini tercapai.

Cadangan Penambahbaikan

Selepas menjalankan kajian yang menyeluruh, penambahbaikan projek analisis simptom kemarahan melampau dan kategori penyebab kemarahan menggunakan pemodelan topik LDA dan BERTopic boleh dilihat dalam beberapa aspek. Yang pertama adalah berdasarkan pemilihan data yang berkualiti. Penggunaan data yang relevan dan berkualiti tinggi daripada pelbagai sumber seperti forum, media sosial, artikel dan kajian kes yang menangani kemarahan yang melampau adalah lebih berkesan berbanding dataset yang didapati di laman web. Untuk visualisasi data, penggunaan Streamlit boleh digunakan untuk membuat dashboard yang interaktif yang lebih menarik minat pengguna dalam meneroka topik dalam data kemarahan. Maklum balas pengguna yang dikumpulkan melalui Google Form juga boleh diperbaiki dengan menghantar borang kepada pengguna dalam masa yang panjang. Hal ini menyebabkan kutipan pengguna hanyalah sebanyak 18 orang sahaja.

KESIMPULAN

Pembangunan algoritma LDA dan BERTopic menyediakan pendekatan yang berbeza untuk menjana topik daripada data teks dan boleh dipilih berdasarkan keperluan projek dan pilihan pengguna. LDA lebih mudah untuk dilaksanakan dan difahami, manakala BERTopic menawarkan prestasi yang lebih baik dan pemahaman topik yang lebih mendalam dengan menggunakan model BERT yang kompleks. Secara keseluruhannya, sistem BERTopic ini telah berjaya dibangunkan dengan menggunakan data yang telah dikaji dan diperolehi. Objektif kajian dan keperluan yang telah ditetapkan sebelum ini telah berjaya dicapai. Walaupun terdapat beberapa halangan, ia berjaya diatasi menggunakan pelbagai cara. Diharapkan sistem ini dijadikan titik kajian untuk kajian lain pada masa hadapan.

Kekuatan Sistem

Kedua-dua algoritma ini berjaya dalam mengesan topik yang terdapat dalam teks yang diberikan, lalu menjana kategori penyebab kemarahan tersebut. Sumbangan projek ini juga adalah menghasilkan perbandingan di antara teknik pemodelan topik seperti LDA dan juga BERTopic. Manakala pembangun sistem menggunakan Visual Studio Code dan juga Streamlit adalah salah satu kekuatan dalam projek ini. Apabila sorotan susastera dilakukan, tiada sistem pemodelan topik menggunakan teknik LDA dan BERTopic yang dijumpai. Jadi, projek ini telah membina templat untuk pemodelan topik. Pengguna yang lain boleh menggunakan system ini untuk menganalisis bukan sahaja mengenai topik kemarahan, bahkan menganalisis topik yang lain seperti dalam kategori industri makanan atau kesihatan.

Kelemahan Sistem

Kekangan projek ini adalah algoritma LDA dan juga BERTopic agak mencabar kerana kekurangan info dan penerangan yang tepat akan cara untuk membina kedua-dua algoritma ini. Pengeluaran yang dikeluarkan dalam sistem juga mempunyai perbezaan setiap kali menjalankan kod terutamanya pengeluaran BERTopic. Hal ini mungkin kerana terdapat beberapa topik yang tersembunyi dalam satu topik yang menggabungkan banyak perkataan rawak. Pemilihan parameter yang tepat juga memberikan kekangan yang penting dalam projek ini. Larasan parameter untuk LDA seperti bilangan topik, alfa dan beta untuk mendapatkan topik yang lebih koheren. Untuk BERTopic, pemilihan model yang bersesuaian dengan domain juga penting untuk memastikan hasil pengeluaran berkualiti dan senang untuk difahami oleh pengguna. Visualisasi yang dikeluarkan oleh BERTopic juga adalah kurang menepati cita rasa yang sesuai dalam konteks reka bentuk.

PENGHARGAAN

Saya ingin ucapkan setinggi-tinggi penghargaan dan jutaan terima kasih kepada Dr. Lailatul Qadri Binti Zakaria, penyelia penulis kajian ini yang telah memberi tunjuk ajar serta bimbingan untuk menyiapkan projek ini dengan jayanya. Dr.Laila adalah sumber inspirasi yang tak terganti, membimbing kami untuk mencapai potensi terbaik. Semua ilmu yang diterima akan menjadi bekal berharga dalam perjalanan akademik dan profesional diri sendiri dan juga rakan seperjuangan.

Saya juga ingin mengucapkan terima kasih kepada semua pihak yang membantu secara langsung mahupun tidak langsung dalam menyempurnakan projek ini. Segala bantuan yang telah dihulurkan amatlah dihargai kerana tanpa bantuan mereka, projek ini tidak dapat dilaksanakan dengan baik. Semoga tuhan merahmati dan memberikan balasan yang terbaik.

RUJUKAN

Alteryx Community. Topic Modeling Intertopic Distance Map - what is t. from <https://community.alteryx.com/t5/Alteryx-Designer-Desktop-Discussions/Topic-ModelingIntertopic-Distance-Map-what-is-the-meaning-of/td-p/1092545> [6 March 2023]

Analytics Vidhya. Unveiling the Future of Text Analysis: Trendy Topic Modeling with BERT. from <https://www.analyticsvidhya.com/blog/2023/07/text-analysis-topic-modeling-with-bert/> [28 July 2023]

David, D. 2021. NLP tutorial: Topic modeling in python with bertopic. Medium. <https://medium.com/geekculture/nlp-tutorial-topic-modeling-in-python-with-bertopicda760e1d03aa> [2 Oktober 2021]

Egger, R., & Yu, J. 2022. A topic modeling comparison between LDA, NMF, Top2Vec, and Bertopic to demystify twitter posts. Frontiers in sociology. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9120935/> [6 May 2022]

Fadhil, M. 2023 .3.10 penyediaan Spesifikasi Keperluan Sistem [f2.6]. MySQA Portal. <https://sqa.mampu.gov.my/index.php/en/3-10-penyediaan-spesifikasi-keperluan-sistem-f2-6>

Singh, S. 2023. power of topic modelling : A guide to text analysis with Python and Machine Learning. Medium. <https://medium.com/@simranjeetsingh1497/power-of-topicmodelling-a-guide-to-text-analysis-with-python-and-machine-learning-abd0c858f472> [4 Februari 2023]

Nur Farzana Syuhadah Binti Kashfullah (A187107)

Dr. Lailatul Qadri Binti Zakaria

Fakulti Teknologi & Sains Maklumat

Universiti Kebangsaan Malaysia